Image denoising beyond additive Gaussian noise - Patch-based estimators and their application to SAR imagery

Charles-Alban Deledalle

École Doctorale
d'Informatique,
Télécommunications
et Électronique de Paris

Thèse

présentée pour obtenir le grade de docteur

de l'Ecole Nationale Supérieure des Télécommunications

Spécialité : Signal et Images

Charles-Alban DELEDALLE

Débruitage d'images au-delà du bruit additif gaussien
Estimateurs à patchs et leur application à l'imagerie SAR

Image denoising beyond additive Gaussian noise
Patch-based estimators and their application to SAR imagery

Soutenue le 15 novembre 2011 devant le jury composé de

Jose	BIOUCAS DIAS	Instituto de Telecomunicações	Rapporteur
Laure	BLANC-FÉRAUD	CNRS – Sophia Antipolis	Président
Jean-François	GIOVANNELLI	Univiversité Bordeaux 1	Examinateur
Jean-Michel	MOREL	ENS Cachan	Rapporteur
Philippe	RÉFRÉGIER	Ecole centrale de Marseille	Rapporteur
Andreas	REIGBER	German Aerospace Center (DLR)	Examinateur
Loïc	DENIS	Télécom Saint-Etienne	Co-directeur
Florence	TUPIN	Télécom ParisTech	Directeur

For to strange sores strangely they strain the cure.

Much ado about nothing,
William Shakespeare

A des maux étranges on applique d'étranges remèdes.

Beaucoup de bruit pour rien,
William Shakespeare

Abstract

Noise in images often limits visual and automatic interpretation of the scene. *Speckle* in synthetic aperture radar (SAR) imagery and *shot* noise in photon-limited imagery are two examples of strong corruptions that require the use of denoising techniques. Patches are small image parts that capture both textures and local structures. Though being crude low-level features (compared to higher level descriptors), they have led to very powerful image processing approaches by exploiting the natural redundancy of images. Patch-based methods achieve *state-of-the-art* denoising performance.

The classical patch-based denoising technique – non-local (NL) means – is designed for images corrupted by an additive Gaussian noise (i.e., fluctuations being symmetrical, signal-independent without outliers). NL means cannot be applied directly on images corrupted by a non-Gaussian process especially with non-symmetrical distribution, signal-dependence and heavy-tail such as *speckle* and *shot* noise.

The goal of this thesis is to bridge the gap between patch-based denoising methods restricted to Gaussian noise and techniques dedicated to SAR despeckling. After reviewing image denoising techniques for Gaussian noise and for non-Gaussian noise, we propose an extension of the NL means that adapts to a given noise distribution.

Besides the problem of image denoising, we study the problem of patch comparison under non-Gaussian conditions. Many tasks in computer vision require matching image parts. We introduce a similarity criterion grounded on the generalized likelihood ratio test and illustrate its effectiveness on different applications including detection, stereo-vision and motion-tracking.

This criterion is at the heart of the proposed patch-based estimator. An iterative scheme is proposed to deal with strong noise corruptions and we develop an unsupervised method for parameter setting. Our approach leads to *state-of-the-art* denoising results in SAR imagery for amplitude images, as well as interferometric or polarimetric data. The proposed technique is applied successfully to one of the latest aerial SAR sensor: F-SAR from the German Aerospace Center (DLR).

Images with strong contrasts suffer from denoising artefacts known as *noise halo* due to the absence of similar patches in the vicinity of some structures. This residual noise can be reduced by considering patches with shapes of various scales and orientations. Local selection of relevant shapes leads to an improved denoising quality, especially close to edges.

Contents

Notations

General notations

a	a normal letter denotes a scalar
$\boldsymbol{a}$	a bold letter denotes a vector
$\boldsymbol{M}$	a bold capital letter denotes a matrix
$\mathcal{S}$	a caligraphic letter denotes a set
$\|\boldsymbol{a}\|_p$	ℓ_p norm of $\boldsymbol{a}$
$\|\boldsymbol{A}\|_{p,q}$	$\ell_{p,q}$ norm of $\boldsymbol{A}$
$< \boldsymbol{a} \| \boldsymbol{b} >$	dot product between $\boldsymbol{a}$ and $\boldsymbol{b}$
z^*	denotes the complex conjugate of z
$\boldsymbol{M}^t$	denotes the matrix transpose of M
$\boldsymbol{M}^\dagger$	denotes the matrix hermitian transpose of M
$\boldsymbol{M}^{-1}$	denotes the inverse matrix of M (when it exists)
$\boldsymbol{M}^+$	denotes the pseudo-inverse matrix of M
$\|\boldsymbol{M}\|$	determinant of $\boldsymbol{M}$
$\mathrm{tr}(\boldsymbol{M})$	trace of $\boldsymbol{M}$
∇	gradient
$\star$	convolution
$\{.\}$	a sequence
$[\![\ldots]\!]$	a discrete interval
$\#\mathcal{S}$	the cardinality of the set $\mathcal{S}$

Notations related to statistical models

v	a normal letter denotes a scalar observation
$\boldsymbol{v}$	a bold letter denotes a vectorial observation
V	a capital letter denotes a random variable modeling observations v
$\boldsymbol{V}$	a bold capital letter denotes a random vector modeling observations $\boldsymbol{v}$
$V \| V' = v'$	a conditional random variable V given the event $V' = v'$
$V \| u$	a random variable parameterized by u

$\mathbb{P}$	probability
p_V	pdf or pmf associated to the random variable V
p	a shorter version of p_V used when their is no ambiguity
$\mathbb{E}$	expectation
Var	variance
R	risk
$\hat{u}$	an estimate of u
$\overline{u}$	a pre-estimate of u used to derive a new estimate $\hat{u}$
$\hat{U}$	a random variable modeling the variability in the estimation of u
$\overline{U}$	a random variable modeling the variability in the pre-estimation of u

Notations related to image models

Ω	image spatial domain
x	a pixel position in Ω
x_k	pixel position associated to the pixel of index k
u	a noise-free / true / underlying value
v	a noisy observed value
$\hat{u}$	an estimated / filtered value
$v(x)$	the value observed at the pixel position x In the same way, $\boldsymbol{v}(x)$, $V(x)$ and $\boldsymbol{V}(x)$ are the (random) variable or vector at pixel position x
v_k	The pixel value observed at pixel index k, i.e., $v(x_k)$ In the same way, $\boldsymbol{v}_k$, V_k and $\boldsymbol{V}_k$ are the (random) variable or vector at pixel index k

Notations related to patch models

P_x	a patch, i.e., a crude low-level feature composed of neighbors of x (usually located in a square window)
W_x	a search window, i.e., a neighborhood of pixel candidates (usually located in a square window)
$\lvert\mathrm{P}\rvert$	the size of the patches
$\lvert\mathrm{W}\rvert$	the size of the search windows
$\boldsymbol{v}(x)$	the patch extracted at pixel position x, i.e., $\boldsymbol{v}(x) = v(\mathrm{P}_x)$
$\boldsymbol{v}_k$	the patch extracted at pixel index x_k, i.e., $\boldsymbol{v}_k = v(\mathrm{P}_{x_k})$
$d(\boldsymbol{v}_1, \boldsymbol{v}_2)$	a dissimilarity measure between $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$
$\mathscr{C}(\boldsymbol{v}_1, \boldsymbol{v}_2)$	a similarity measure between $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$

$w(x, x')$	the weight between x and x'
$\varphi(t)$	the kernel function $\mathbb{R}^+ \mapsto [0, 1]$
S	The shape of the patches

Notations related to specific distributions

Gaussian or homoscedastic distributions

σ	the standard-deviation (when it exists)
Σ	the covariance matrix (when it exists)

Cauchy distribution

γ	the shape parameter

Distribution for speckle

L	the shape parameter (also referred to as number of looks)

Impulse and salt-and-pepper distributions

L	the number of level
P	the rate of corruption

Notations related to SAR imagery

z	the complex value of the measured echo
A	the amplitude of a measured echo
I	the intensity of a measured echo
φ	the phase of a measured echo
ϕ	the phase difference between two echos
d	the empirical coherence between two echos
$\boldsymbol{k}$	the observed scattering complex vector
$\boldsymbol{C}$	the empirical complex covariance matrix
R	the reflectivity
D	the true coherence
β	the true phase difference
$\boldsymbol{\Sigma}$	the true complex covariance matrix
L	the number of looks

Chapter 1

Introduction

In the 80s–90s, most image denoising techniques were based either on regularization/variational methods or sparse wavelet decompositions. In 2005, Buades et al. introduced the non-local means filter which marked a turning point in the story of image denoising. The basic idea – to search for similar patterns and combine their pixel values – has led to impressive results. All recent techniques in image denoising rely on the concepts of non-locality and patches. Among them, the K-SVD (Aharon et al., 2006), the BM3D (Dabov et al., 2007) and the NLSM (Mairal et al., 2009) are generally considered as the current *state-of-the-art* filters.

Concurrently, synthetic aperture radar (SAR) imagery entered into a new golden age. Since 2006, several new sensors have appeared with polarimetric and/or interferometric configurations and very high range resolutions:

- 2006 – ALOS-2 (a Japanese L-band polarimetric spaceborne sensor),
- 2007 – TerraSAR-X (a German X-band spaceborne sensor),
- 2007 – Cosmo-SkyMed (an Italian constellation of four X-band spaceborne sensors),
- 2007 – RadarSAT-2 (a Canadian polarimetric spaceborne sensor),
- 2008 – Sethi (a French airborne polarimetric and interferometric sensor),
- 2010 – TanDEM-X (a replica of TerraSAR-X used jointly in interferometric mode),
- 2011 – F-SAR (a German polarimetric airborne sensor with multi-frequencies).

Due to the multiplication of sensors, the quantity of polarimetric and interferometric data is increasing quickly. In spite of the widest availability of SAR images, their analysis remains a difficult task that requires taking into account the geometry, height, roughness and moisture of the objects. Automatic techniques are then required to process these wide data-sets for tasks such as target tracking, change detection, classification, 3D reconstruction and estimation of the forest biomass evolution.

In this context, the initial motivation of this thesis was to adapt the idea of patch-based non-local filtering to reduce *speckle* that corrupts SAR images. *Speckle* is an undesirable grain that appears in coherent imaging systems (including SAR, sonar, ultrasound and laser imagery). Concurrently, optical images also suffer from undesirable fluctuations ascribed to the *thermal* noise and the *shot* noise. Our curiosity led us to study the more general restoration problem of noisy images or how to *heal* the images from noise. Beyond SAR image processing, the topic of this thesis became more generally the problem of image denoising.

Image denoising is an important problem in image processing since noise often limits visual and automatic interpretation of the scene. It is often necessary to pre-process images with

a suitable method of noise reduction before analyzing them. Image denoising is a key pre-processing step in many cases, e.g., low-light or high-speed imaging, low-cost sensor usage in embedded systems, and also ultrasound, sonar and radar coherent imagery.

1.1 Approach followed in this thesis

1.1.1 Sources and models of degradation

Images are multi-dimensional signals that can represent different kinds of information such as pictures, drawings, logos, *etc.* According to the specificity of the scene, the sensor and/or the digital storage of an image, different types of degradation can occur:

Inherent degradations: There are many types of degradation which are inherent to the nature of the signal emitted in the direction of the sensor. The received signal is a projection of a three dimensional scene to two dimensions. This leads to geometric deformations, masking effects and mixtures of sources. Moreover, if objects are moving during the exposure, this leads to motion blur. A shorter exposure time should then be used, providing only a limited number of photons in low light conditions. The signal then has high fluctuations known as *shot* noise (see Fig. 1.1). The atmosphere can also introduce blurring and geometric perturbations. In coherent imagery, the interferences of many elementary scatterers produce high fluctuations referred to as *speckle* (see Fig. 1.1).

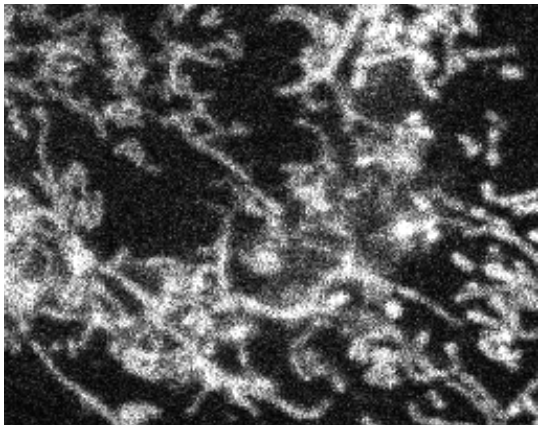
Degradations assigned to the sensor: The stability, sensibility and exposure time are other factors that play an simportant role in the resulting image quality. *Shot* noise and motion blur are characterized by the parameters of the sensor. Undesired blurring can also arise from the internal impulse response of the system characterized by its point spread function. Moreover, sensors generally suffer from internal fluctuations referred to as *thermal* noise. Color cameras measure the signal on a mosaic of red, blue and green sensors (Bayer matrix). The reconstruction of a color image from this mosaic is called the *demosaicing* problem. This process can introduce false colors, chromatic aliasing, zippering and purple fringing. Moreover, damages on this mosaic can lead to missing pixels.

Degradations ascribed to digitizing and storage: The analog signal is next sampled and quantized. The sampling of a continuous signal to a discrete signal can lead to aliasing effects (undersampling with respect to the sampling rate given by Niquist-Shanon theorem). The discretization of continuous values into a finite number of levels leads to quantization distortions, also called the quantization noise. Due to the large size of images, their storage often requires resorting to lossy compression, hence, leading to compression artifacts (e.g., JPEG artifacts). Sometimes the transmission of this digital information towards a storage unit can lead to a loss of information resulting, for instance, in missing or aberrant pixels (i.e., impulse or *salt-and-pepper* noise).

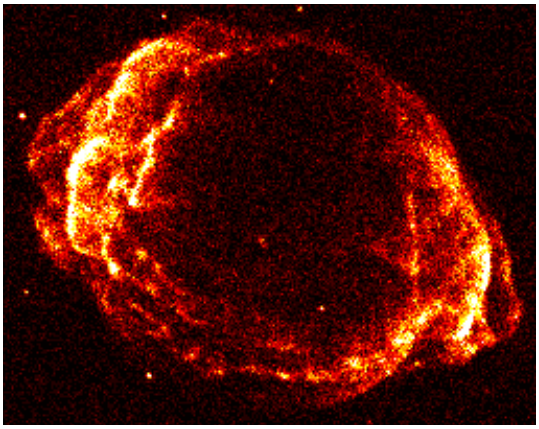
In this thesis we only focus on the problem of denoising. However, deblurring, demosaicing, in-painting, compression, inverse half-toning and super-resolution are examples of other tasks that are closely related to our problem.

1.1.2 Different manifestations of noise in images

Most denoising techniques are designed for additive white Gaussian noise (AWGN). The success of this model is due to its efficiency to represent noise fluctuations in many situations as well as its simplicity. However, in many situations, the AWGN model is known to be limited and non-realistic to properly describe noise fluctuations. Figure 1.1 gives an illustration of different manifestation of noise in images. Many solutions have been designed for particular applications



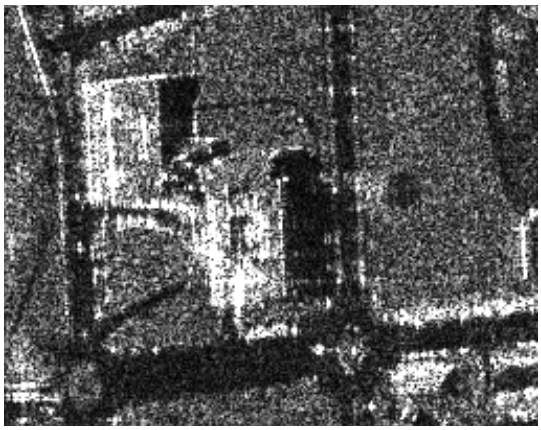
(a) Mitochondrion in confocal microscopy



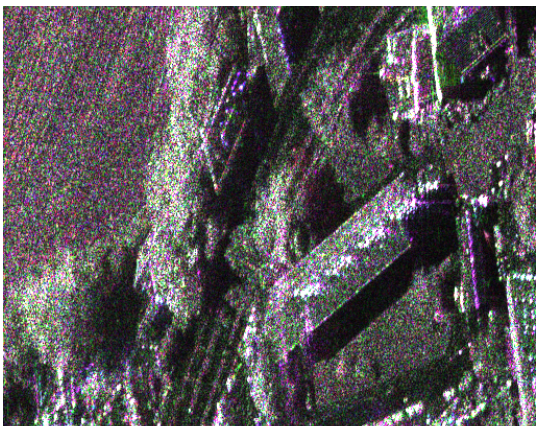
(b) Supernova in X-ray imagery

(c) Fetus using ultrasound imagery

(d) Plane wreckage in SONAR imagery



(e) Urban area using SAR imagery



(f) Urban area using polarimetric SAR imagery

Figure 1.1: (a) A mitochondrion observed in fluorescence confocal microscopy (image courtesy of Y. Tourneur) and (b) a supernova observed with X-ray imagery (image courtesy of Chandra X-ray Observatory – data identifier: `ADS/ Sa.CXO#Contrib/ ChandraDeepField`). Images (a) and (b) suffer from *shot* noise. (c) A fetus obtained by ultrasound imagery (image courtesy of Elise Nicolas), (d) a plane wreckages in SONAR imagery (acquired by Shadows system, image courtesy of IXsea), (c) a single building observed by SAR imagery (acquired by RAMSES ©ONERA CNES) and (f) an urban area observed by polarimetric SAR imagery (acquired by F-SAR ©DLR). Images (c-f) suffer from *speckle*.

under non-Gaussian noise (often the *speckle* or the *shot* noise). They provide a first approximation of the real (more difficult) noise structure. Few provide a general methodology that can be applied to different kinds of noise models.

In several situations, noise is correlated. Noise correlation may have different causes. When noise is signal-dependant, as is the case of Poisson or Gamma noise due to shot noise or speckle, noise is correlated to the noise-free signal which is itself spatially correlated. By accounting for the (non-Gaussian) distribution of observed values, such correlations will be correctly handled by the denoising approach described in this thesis. Due to a transformation of the noisy data such as upsampling or local averaging, noise may exhibit a stationary covariance. A whitening process can then be applied[1] to decorrelate noise. In the most general case, noise is non Gaussian and the correlations are non-stationary, such as "blocky" JPEG compression artefacts. Statistical modeling of such noises is highly challenging and dependant on the underlying noise-free signal. The robustness of the denoising technique designed for uncorrelated noise can in some cases extend its applicability to such noises.

This thesis essentially focuses on extending denoising approaches to non-Gaussian noise when noise is modeled by a random process that is independent from one pixel to another. *Speckle* and *shot* noise will particularly retain our attention with several examples and illustrations given throughout the different chapters. The application of the proposed methodology to synthetic aperture radar images will be the topic of a dedicated chapter.

1.1.3 Why image denoising?

Noise is usually considered as undesired fluctuations corrupting a signal or an image. Noise is often non-informative and visually unpleasant: it is a pathology of images which then need to be *healed*. However, noise is not always non-informonative or ascribed to errors. In coherent imagery, the fluctuations of *speckle* are fully determined by the random organization or disposition of punctual scatterers and are then inherent to the imaged scene. Among others, these fluctuations characterize the surface roughness or the scattering material (Sprague, 1972; McKinney et al., 2000). The painting presented in Fig. 1.2 also presents high fluctuations placed purposely by the artist (a technique known as pointillism) and are then part of the work. It is legitimate to ask why we should denoise such images.

Image interpretation and pipeline of image processing

Let us consider the two previous examples. *Speckle* can be informative, however, one might not be interested on this overload of information. For instance, synthetic aperture radar imagery can be used to observe a street in an urban area. *Speckle* provides information on the random disposition of scatterers inside each pixel of the road (more precisely inside the resolution cells). Why should we be interested in this information? In practice, we would rather be interested in a summary of this information: the local average rate of fluctuations reflecting the roughness of the street, i.e., a *speckle*-free image. In remote sensing, automatic classification into streets, buildings, vegetation or water areas are usually based on such a summary of information instead of *speckle* directly.

In the second example, from an artistic point of view, one would not want to modify the work of the artist. However, imagine that you want to automatically extract the features such as: the number of people, the presence of a river or the colors of the umbrellas. This could be of interest for indexing the works of a museum and then be able to retrieve a painting with simple queries. In this case, the suppression of the fluctuations is a necessary pre-processing step in order to use standard image processing algorithms to extract such high level features.

[1] this inversion procedure requires special care due to ill-conditioning issues that may arise

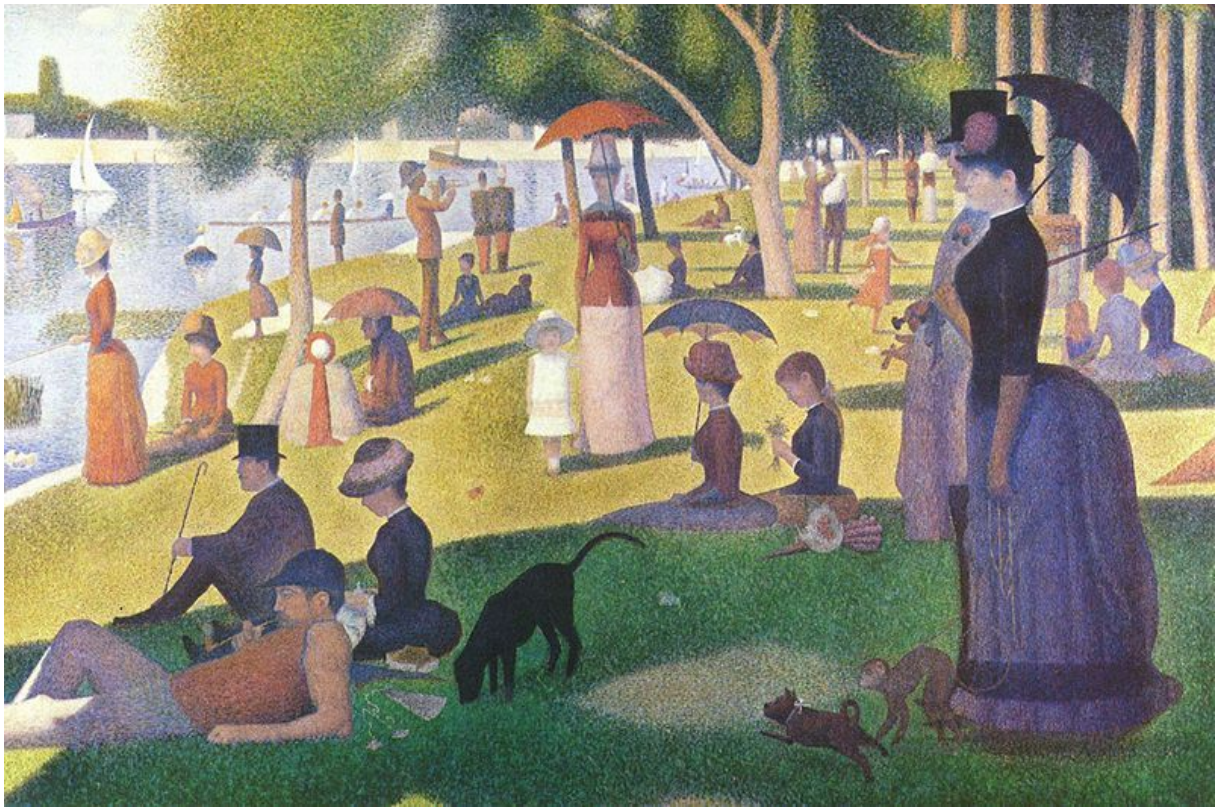


Figure 1.2: *Un dimanche après-midi sur l'île de la Grande Jatte* (Georges Seurat)

In both examples, the deterministic fluctuations will still be considered as noise since they would defeat standard algorithms of image processing. The estimation of a noise-free image from a noisy image is paramount to scene interpretation or low-level processing tasks such as segmentation, classification and 3D reconstruction.

An end or a methodological enrichment?

If denoising can be used as a pre-processing step for higher level interpretations of natural images, it is also possible to design high-level techniques that are robust and directly adapted to the noisy data. For instance, the segmentation model of Mumford and Shah (1989) considers that the measurement of an image "always produces a corrupted, noisy approximation of the true image". It is actually sometimes easier to take into account the noisy data rather than a denoised version since the statistics of the noisy data are usually well-known while the statistics of the processed data are generally not available when non-linear processing are involved. If such models take noise into account, it is no longer required to denoise the image in a pre-processing step. The resulting pipeline is then simpler with fewer parameters and, as a consequence, with better control on the production of the final results.

Of course, it is not always straightforward to take noise into account in high-level techniques. For instance, to the best of our knowledge, there is no powerful extension of the famous descriptors of Lowe (2004), known as shift-invariant feature transform (SIFT), to deal with different statistics of noise.

Denoising does not aim only to improve the robustness of such tools. Indeed, as stated by Mumford and Shah (1989) themselves, their segmentation model is closely related to the functional proposed by Geman and Geman (1984) for image restoration purposes. Hence, the

elaboration of denoising algorithms is not necessarily the final objective but rather it allows us to find means, methodologies or even concepts in order to improve solutions for other tasks. This is an important point of this thesis, since the methodology that we propose for denoising in Chap. 5 has led to the definition of similarity criteria (see Chap. 4) relevant for other applications such as detection, stereo-vision or object-tracking.

Link to image restoration: similarities, differences and competition

As we have just mentioned, solutions proposed for different tasks can lead to mutual enrichments. According to the multiple sources of degradations, a multiplicity of problems emerge with more or less satisfying solutions. Denoising is a problem often linked to the more general problem of image restoration. Image restoration aims to retrieve a signal of interest from measurements suffering from deformations (e.g., blur, sub-sampling or missing-pixels) and errors (i.e., noise). It usually involves the inversion of an operator. However, noise affecting the measurements leads to strong aberrations when directly inverting the operator from the noisy observations. The inversion leads to solve an *ill*-conditioned system of equations explaining the poor stability of such approaches. More generally, inverse problems are said to be *ill*-posed. To solve such problems, it is then required to add extra assumptions or constraints, for instance, on the regularity of the solution. Such assumptions form the *prior* model of the solution.

Unlike denoising, the difficulty of image restoration is the inversion of a possibly unknown operator that may be spatially varying. However, as we will see, like image restoration, image denoising requires having a *prior* model of the solution (i.e., the underlying noise-free image). The choice of this image model is the key to the success of image restoration and image denoising methods. Several models have then been proposed for both tasks, and, not surprisingly, better models for image restoration lead to better models for image denoising and vice-versa. Among the approaches many were originally proposed for other purposes, such as, deblurring, demosaicing, in-painting, compression, inverse half-toning and super-resolution.

1.2 Issues considered and main contributions

1.2.1 The initial problem: accounting for the statistics of SAR images

SAR intensity images suffer from multiplicative fluctuations called *speckle*. *Speckle* is usually reduced by spatially averaging neighbor pixels at the expense of a loss of resolution. Inspired by the work of Buades et al. (2005), our starting point was to use a non-local approach to efficiently reduce *speckle* in the image and prevent this loss of resolution. To extend the non-local means (NL means) to SAR data, a similarity criterion had to be designed to find resembling *patches* in a SAR image. In the NL means the *patch* similarity is expressed as the square difference between pixel values. For SAR images, this square difference does not adapt to the signal-dependent fluctuations of multiplicative noise. Since noise is multiplicative, a naive idea could be to consider a criterion based on the ratio of the intensities rather than a criterion based on the differences.

Interferometric and polarimetric SAR images have a multi-dimensional content. The pixels contain a complex vector formed by stacking different co-registered SAR images. The phase difference between each channel appears as a crucial quantity that suffers from a *decorrelation* noise. Our extension of the NL means had to jointly reduce the *speckle* in the intensity and the *decorrelation* noise in the phase differences by considering the similarity between *patches* of complex vectors. The square difference cannot take into account the wrapping of the phase differences. Simple heuristics could have been used to take into account this wrapping. A naive joint criterion with a weighted combination of the ratio of intensities and a heuristics criterion

between the phase differences could have been proposed. We considered instead a statistical framework based on the joint distribution of the intensities and phase differences.

Contribution 1 (NL-SAR): We designed a similarity criterion that takes into account the joint statistical model of *speckle* and *decorrelation* noise (e.g., the Gamma, the circular complex Gaussian or the Wishart distribution). This idea has been fruitful leading to the non-local SAR filter (in short NL-SAR) or also denoted as the *probabilistic patch-based* (PPB) filter. It has been the topic of two journal papers: in the IEEE Transactions on Image Processing for *speckle* reduction on SAR amplitude images (Deledalle et al., 2009b) and in the IEEE Transactions on Geoscience and Remote Sensing for the estimation of SAR interferometric parameters (Deledalle et al., 2011a). At the same time, NL-SAR has been presented at GRETSI'2009 and IGARSS'2010 (Deledalle et al., 2009a, 2010b,d). This work has inspired other teams for designing new non-local filters for SAR images (Parrilli et al., 2010; Teuber and Lang, 2011; Feng et al., 2011; Zhong et al., 2011), for the extraction of high quality information from the fusion of optical and radar data (Palubinskas et al., 2011) and for the reduction of disturbing effects like layover and shadowing in InSAR data (Schmitt et al., 2011).

Speckle reduction is a critical task since it strongly influences the global performances of SAR processing pipelines. To evaluate the performance of NL-SAR, we needed the evaluation by experts on such aforementioned processing pipelines, from the elaboration of sensors to the interpretation of high-level features. I spent one month at the German Aerospace Center (DLR) under the supervision of Andreas Reigber and Marc Jäger to test and validate the applicability of NL-SAR to high-resolution polarimetric SAR images sensed by their latest aerial SAR sensor: F-SAR. The work done during my stay confirmed the usefulness of our SAR denoising method on real data. NL-SAR is on the road to be integrated into the processing pipeline of DLR and be used routinely to ease the interpretation of polarimetric and interferometric SAR images. The source code of NL-SAR has been released in open source as described in Appendix A.

Contribution 2 (Remote sensing applications): Under our supervision, Sami Benzid (from URISA, SUPCOM, Tunis) successfully used the NL-SAR for change detection in multi-temporal series of SAR images (Benzid et al., 2010). It has also led to joint works with Sofiène Hachicha (from URISA, SUPCOM, Tunis) and Fang Cao (from Telecom ParisTech, Paris) on multi-temporal SAR classification and polarimetric SAR classification, respectively. These two joint works have been respectively presented at IGARSS'2010 (Hachicha et al., 2011) and IGARSS'2011 (Cao et al., 2011). I have also collaborated with Antoine Lucas (from Caltech, California) using NL-SAR for the analysis of SAR images of Titan. This joint work was presented at the Titan science meeting (Lucas et al., 2011a) and the Titan surface workshop (Lucas et al., 2011b).

1.2.2 Towards a general methodology: a variety of applications

Our extension of the NL means to SAR data has led to a general methodology that can be applied to data of arbitrary nature when a statistical model of the undesirable fluctuations is provided. To assess the validity of this general methodology, we applied it also on images of other natures with non-Gaussian fluctuations.

Contribution 3 (Poisson NL means): Our framework has been used for photon-limited images that appear under low-light conditions, for instance, in fluorescence microscopy or astronomical imagery. These images suffer from a *shot* noise that can be modeled by a Poisson distribution. We proposed an unsupervised setting of the denoising parameters driven by the Stein unbiased risk estimate. It led to the Poisson NL means filter that we presented at ICIP'2010 (Deledalle et al., 2010c). For this paper, I received the ICIP'2010 best student paper award. Poisson NL means recently inspired the work of Lee et al. (2011).

Contribution 4 (GLR-based criterion): The design of this general methodology led us to study several similarity criteria under non-Gaussian noise conditions. We considered several definitions from a dense literature on this topic extracted from the communities of image processing, detection theory and machine learning. We enumerated the basic properties that they should fulfill and evaluated their performance on different tasks. From this study, we recommended using the generalized likelihood ratio test (GLR) to define *patch* similarity. The performance of GLR has been studied in a task of estimation of the glacier velocity with a pair of SAR images. This was a joint project with Renaud Fallourd (from Université de Savoie, Chambéry) that was presented in IGARSS'2010 (Deledalle et al., 2010a). We recently presented a deeper comparative study of similarity criteria under non-Gaussian noise at ICIP'2011 (Deledalle et al., 2011e) and we submitted a journal version currently under review (Deledalle et al., 2012).

1.2.3 Signal-adaptivity in non-local filtering

While our general methodology has proven effective for SAR images and photon-limited images, we were confronted by an inherent problem of the NL means: the *rare patch effect*. Some features are (almost) unique (i.e., not found elsewhere inside the image). Due to the lack of redundancy for such features, noise cannot be reduced and the resulting image would present a persistent residual noise. We noticed that this *rare patch effect* is all the more important when the data are vectorial such as in interferometric or polarimetric SAR data. Concurrently, Vincent Duval (from Telecom ParisTech) and Joseph Salmon (from Université Paris Diderot) have also been focussing on spatial adaptation in non-local filtering.

Contribution 5 (NLM-SAP): Together, we proposed a solution to solve the problem of the *rare patch effect* that arises around edges with high contrast. Such a phenomenon essentially appears in imagery with high dynamic ranges and produces a residual noise that we refer to as a *noise halo*. In this joint work, we proposed to use *patches* with locally adaptive shapes and sizes in order to favor the redundancy of "rare" features. This yielded an anisotropic version of the NL means that we coined NLM-SAP for non-local method with spatial adaptive *patches*. It was published in Journal of Mathematical Imaging and Vision (Deledalle et al., 2011c) and presented at the Journées des Statistiques (Salmon et al., 2011) and SSVM'2011 (Deledalle et al., 2011b).

1.3 Organization of the manuscript

Chapter 2 introduces the problem of image denoising and reviews the main approaches used under the AWGN model. After highlighting the necessity to adapt denoising techniques when noise departs from the AWGN model, we review in Chap. 3 different approaches. Among them, we focus on selection-based filters which include the NL means. We motivate the need of a proper definition of *patch* similarity and we introduce a similarity criterion based on the generalized likelihood ratio test. We draw up in Chap. 4 a comparative study of similarity criteria under non-Gaussian noise. Theoretical properties and task-based evaluations are in favor of our proposed criterion. Based on this criterion, we then present a general methodology to extend the NL means to images damaged by non-Gaussian noise in Chap. 5. An iterative scheme is also proposed to deal with high noise corruptions and we develop an unsupervised setting of the parameters. The application of the proposed method to SAR imagery is given in Chap. 6. The efficiency of the proposed technique is, among others, illustrated in recent airborne F-SAR data. The application of non-local approaches to SAR data emphasizes the problem of the *rare patch effect*. The spatial adaptation of *patches* is presented in Chap. 7. Conclusions and perspectives are discussed in Chap. 8.

Chapter 2

The problem of image denoising

A large number of denoising methods have been proposed for image denoising. Their differences lie in the assumptions made on the property of the underlying scene or signal (e.g., the regularity, sharpness or repetitions) and the nature of the noise (e.g., additive or multiplicative). The quality of denoising depends on the adequacy of these assumptions on the processed images. As a consequence, the best denoising methods are based on flexible models applying efficiently to a large variety of images.

The most powerful methods are currently based on the concept of non-locality and the decomposition of images as a collection of *patches* (Katkovnik et al., 2010; Chatterjee and Milanfar, 2010). We speak about *patch* decomposition, *patch* model, *patch* representation or filtering in the *patch* space. The first authors to propose such an approach were Awate and Whitaker (2005, 2006) and Buades et al. (2005) with, respectively, the unsupervised information-theoretic adaptive (UINTA) filter and the non-local (NL) means filter. State-of-the art denoising techniques all rely on patches, either for dictionary learning (Elad and Aharon, 2006), collaborative denoising of blocks of similar patches (Dabov et al., 2007) or non-local sparse models (Mairal et al., 2009). Regularization with non-local patch-based weights have shown to improve on classical regularization involving only local neighborhoods (Gilboa and Osher, 2007; Peyré et al., 2008; Zhang et al., 2010b).

Organization of the chapter— In this chapter, we first introduce the framework and notations that we will use to describe the problem of denoising in Sec. 2.1. In particular we will see that denoising should adapt to the nature of the underlying signal and we will propose criteria to evaluate denoising performance. Section 2.2 describes filters designed for smooth regular images, Sec. 2.3 for piece-wise constant images, Sec. 2.4 for images with sparse decomposition on a dictionary and Sec. 2.5 for self-similar images. Since such models are generally valid only locally, we will focus in Sec. 2.6 on aggregation filters that combine estimates resulting from different assumptions on the signal. Besides the problem of modeling the underlying image, we will see in Sec. 2.7 the influence of the noise model on the denoising performance.

2.1 Introduction to image denoising

Denoising amounts to estimating the underlying spatial information, or noise-free image $\boldsymbol{u}$, that a noisy image $\boldsymbol{v}$ contains. Numerous denoising techniques have been proposed in the image processing literature. The majority of them consider a Gaussian noise model. The goal of denoising is to find a powerful method for noise reduction and the preservation of structures of interest such as edges or textures without introducing undesired artifacts.

2.1.1 Statistical framework and notations

The methodology presented in this thesis generalizes well to multi-dimensional data (3D, video, ...). In the following, for notational simplicity, we consider the case of images that are a collection of N observations placed on a two dimensional regular grid $\Omega \subset \mathbb{Z}^2$ (with $N = |\Omega|$). Each element of the grid (or pixel) is identified with a unique index $k \in [1, N]$, its position is denoted as $x_k \in \Omega$ and its associated value v_k. At each pixel, the observation v_k may be $D-$dimensional, e.g., $D = 1$ for gray-level images, $D = 3$ for RGB color images or $D = 256$ for hyper-spectral images. In this thesis, we are not only considering gray-level images or color images, but a variety of different modalities where v_k can be a matrix or even a tensor with real, integer or complex values. An image will be modeled as the function v, defined from Ω to the observation space, such that for all pixel index k:

$$v_k = v(x_k). \tag{2.1}$$

Since there is a bijection between the pixel index k and the position x_k, an image can also be viewed as the vector of the observations such that $\boldsymbol{v} = (v_k)_{k\in[1,N]}$ is an N-dimensional vector obtained by stacking the observations of each pixel of the image. The notations v and $\boldsymbol{v}$ will be used respectively to denote an image either in its functional representation or in its vectorial representation.

As we have mentioned previous chapter, the observed image $\boldsymbol{v}$ suffers from many sources of degradation. We are not so interested in the image $\boldsymbol{v}$ itself but in the underlying spatially varying information that it contains. This underlying information, free of degradations, will be modeled with an image of M pixels and denoted by $\boldsymbol{u}$. This underlying image $\boldsymbol{u}$ contains the set of information allowing to explain or understand the observation $\boldsymbol{v}$. A classical image model is to link the noisy observation $\boldsymbol{v} \in \mathbb{R}^N$ and $\boldsymbol{u} \in \mathbb{R}^M$ with the following relation:

$$\boldsymbol{v} = \Phi\boldsymbol{u} + \sigma\boldsymbol{\varepsilon} \tag{2.2}$$

where $\boldsymbol{\varepsilon} \in \mathbb{R}^N$ models measuring errors or noise fluctuations, $\sigma > 0$ controls/amplifies the noise level and $\Phi \in \mathcal{M}_{N,M}(\mathbb{R})$ is a linear operator modeling, for instance, blur, sub-sampling or missing pixels. To retrieve $\boldsymbol{u}$ from $\boldsymbol{v}$ is a particular case of an image restoration problem which requires the inversion of an operator through noisy perturbations. In this thesis, we focus only on noisy data, i.e., Φ is the identity matrix. Furthermore, we consider non-additive noise and the underlying values u_k can be of a different nature from the observed values v_k (several examples are given in the next chapter). While $\boldsymbol{v}$ lies in the observation space, we say that $\boldsymbol{u}$ takes its values from a parameter space, since u_k acts as parameters of a generative process producing the observations v_k. Here, other non-spatially varying parameters Θ of this generative process which are not of main interest (i.e., the nuisance parameters) are assumed to be known (e.g., the factor σ in Eq. (2.2)).

In this thesis, we focus only on the case of noisy images such that the observed image corresponds to the underlying noise-free image viewed through noise fluctuations. Hence, the noisy image and the noise-free image share the same definition domain Ω (hence, $M = N$) and the noisy values are directly statistically linked to the noise-free values. Noise can generally be well modeled with parametric distributions, either grounded on physical or empirical considerations. We assume that the noisy image is modeled by a given distribution so that $\boldsymbol{v}$ is a realization of an N-dimensional random variable $\boldsymbol{V}$ described by the probability density function (pdf) $p_{\boldsymbol{V}}(\boldsymbol{v} \mid \boldsymbol{u}, \Theta)$ (or a probability mass function for discrete observations) such that:

$$\mathbb{P}(\boldsymbol{V} \in \mathcal{A} \mid \boldsymbol{u}, \Theta) = \int_{\mathcal{A}} p_{\boldsymbol{V}}(\boldsymbol{v} \mid \boldsymbol{u}, \Theta)\, \mathrm{d}\boldsymbol{v} \tag{2.3}$$

for any subset $\mathcal{A}$ of the observation space. Since nuisance parameters Θ are assumed to be known and for the sake of notational simplicity, this pdf will be written as $p(\boldsymbol{v}|\boldsymbol{u})$.

For example, an image $\boldsymbol{v}$ damaged by additive white Gaussian noise (in short AWGN or Gaussian noise) with standard deviation σ can be modeled by:

$$\boldsymbol{v} = \boldsymbol{u} + \sigma\boldsymbol{\varepsilon} \tag{2.4}$$

where $\boldsymbol{u}$ is the noise-free image and $\boldsymbol{\varepsilon}$ is the realization of a normalized zero-mean Gaussian random vector. It is straightforward to see that $\boldsymbol{V}|\boldsymbol{u}$ follows a Gaussian distribution with mean $\boldsymbol{u}$ and standard deviation σ. While such decompositions also exist for other distributions (e.g., the gamma distribution involves a multiplicative decomposition), there is not necessarily a decomposition of $\boldsymbol{v}$ in terms of $\boldsymbol{u}$ and an independent noise component (for example, in the case of Poisson noise). In general, when noise departs from additive Gaussian noise, the link between $\boldsymbol{v}$ and $\boldsymbol{u}$ is described by its likelihood function $p(\boldsymbol{v}|\boldsymbol{u})$.

2.1.2 Denoising as a bias-variance trade-off

One would want to search for an estimator h such that $\hat{\boldsymbol{u}} = h(\boldsymbol{v})$ minimizes the square error with the noise-free image $\boldsymbol{u}$:

$$\|\hat{\boldsymbol{u}} - \boldsymbol{u}\|_2^2 = \|h(\boldsymbol{v}) - \boldsymbol{u}\|_2^2 \tag{2.5}$$

where $\hat{\boldsymbol{u}}$ results from the application of a given estimator h on the noisy image $\boldsymbol{v}$. However, since the noise-free image $\boldsymbol{u}$ is unknown, direct minimization of Eq. (2.5) is not possible. In fact, the optimal solution canceling the square error is the estimator h producing the noise-free image itself: $h(\boldsymbol{v}) = \boldsymbol{u}$. The design of such an estimator for any noise-free image $\boldsymbol{u}$ is unfeasible since:

1. two different noisy images can share the same noise-free image, and
2. two identical noisy images can arise from two different noise-free images.

This optimal estimator is therefore not a deterministic function associating the same entry to the same estimate. The elaboration of the optimal function h seems impossible.

An alternative is to model the noisy image $\boldsymbol{v}$ as a realization of a random vector $\boldsymbol{V}$. Since there is usually a bijection between the random vector $\boldsymbol{V}$ (understand its pdf) and the parameter of interest $\boldsymbol{u}$, hopefully one could design a robust estimator solving the denoising problem. Instead of minimizing Eq. (2.5), denoising techniques are usually expressed as the research of an estimator h such that $\hat{\boldsymbol{U}} = h(\boldsymbol{V})$ is a random vector lying in a narrow neighborhood of $\boldsymbol{u}$. This leads to the minimization of the mean square error (MSE):

$$\mathbb{E}\left[\|\hat{\boldsymbol{U}} - \boldsymbol{u}\|_2^2\right] = \int \|h(\boldsymbol{v}) - \boldsymbol{u}\|_2^2 \, p(\boldsymbol{v}|\boldsymbol{u}) \, \mathrm{d}\boldsymbol{v} \, . \tag{2.6}$$

where the expectation is taken over the random vector $\boldsymbol{V}$. Note that this expectation is mathematically the conditional expectation of $\|\hat{\boldsymbol{U}} - \boldsymbol{u}\|_2^2$ knowing $\boldsymbol{u}$. For the sake of notations and because $\boldsymbol{u}$ is considered as deterministic, we simply refer to the expectation when there is no ambiguity. The minimization of the MSE rewrites as the research of an optimal bias-variance trade-off:

$$\mathbb{E}\left[\|\hat{\boldsymbol{U}} - \boldsymbol{u}\|_2^2\right] = \underbrace{\left(\mathbb{E}\left[h(\boldsymbol{V})\right] - \boldsymbol{u}\right)^2}_{\text{bias}} + \underbrace{\operatorname{Var}\left[h(\boldsymbol{V})\right]}_{\text{variance}} \, . \tag{2.7}$$

where the bias term penalizes the non-preservation of structures or the introduction of systematic artifacts and the variance term penalizes the residual fluctuations of the estimator.

If the image $\boldsymbol{u}$ is also modeled as the realization of a random vector $\boldsymbol{U}$, the minimization of the MSE is known as the Bayesian least square (BLS) problem, whose solution is given by the posterior expectation or conditional mean:

$$h(\boldsymbol{v}) = \int \boldsymbol{u}\, p(\boldsymbol{u}|\boldsymbol{v})\, \mathrm{d}\boldsymbol{u} = \frac{\int \boldsymbol{u}\, p(\boldsymbol{v}|\boldsymbol{u})p(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u}}{\int p(\boldsymbol{v}|\boldsymbol{u})p(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u}}\,. \tag{2.8}$$

In this Bayesian formulation, the distribution $p(\boldsymbol{u})$ models in a statistical way the *a priori* configuration that can occur. The term $p(\boldsymbol{v}|\boldsymbol{u})$ is the likelihood of $\boldsymbol{u}$ given $\boldsymbol{V} = \boldsymbol{v}$, which is fully described by the given noise distribution model. With perfect knowledge of *prior* pdf $p(\boldsymbol{u})$, Eq. (2.8) leads to optimal performance. Despite its theoretical performance, this approach requires the computation of an integral over a huge state space which, depending on the distributions, may not be known in closed-form and therefore is time-consuming to evaluate numerically. An alternative solution to avoid integration issues is to search for the noise-free image $\boldsymbol{u}$ that maximizes the *a posteriori* (MAP) pdf:

$$h(\boldsymbol{v}) = \arg\max_{\boldsymbol{u}} p(\boldsymbol{u}|\boldsymbol{v}) = \arg\max_{\boldsymbol{u}} p(\boldsymbol{v}|\boldsymbol{u})p(\boldsymbol{u}) \tag{2.9}$$

since the evidence $p(\boldsymbol{v})$ is constant with respect to $\boldsymbol{u}$. Instead of minimizing the mean square error, MAP is known to minimize the probability of errors in the estimation. The optimization problem (2.9) should be solved by algorithms robust to local extrema since the posterior distribution $p(\boldsymbol{u}|\boldsymbol{v})$ is usually multi-modal. Both approaches require the knowledge of the *prior* pdf. The *prior* is usually a (potentially difficult) choice left to the statistician/practitioner or it can sometimes be estimated, but is often unknown. Moreover, the Bayesian approach is usually criticized since it models noise-free data as random variables while they are fully deterministic. However, both estimators indicate that to perform efficient image denoising, it is necessary to introduce assumptions on both the nature of the noise-free image and the statistics of the noise.

The accuracy of models on noise-free images acts on the bias-variance trade-off of the solution. An overly restrictive model, non-adapted to the majority of processed images, will reduce the variance significantly while introducing a bias in the estimation. At the opposite, an overly flexible model will usually leave too much noise in the solution and have a small bias. Let us illustrate this influence by referring to two toy examples, using MAP estimation, where $\boldsymbol{v}$ is a real image damaged by Gaussian noise:

The case of a constant noise-free image: Consider a *prior* modeling a constant image $\boldsymbol{u}$, such that, for all pixel position x, $u(x) = a$ where a is an unknown real value. Under Gaussian noise assumption, denoising, i.e., estimation of MAP value a, is obtained thanks to the sample mean of noisy realizations:

$$\hat{a} = \frac{1}{N}\sum_{l=1}^{N} v_l\,. \tag{2.10}$$

Suppose that our image model is wrong and $\boldsymbol{u}$ can instead take any possible real value. In this case it is straightforward to show that the variance at each position index k is small:

$$\mathrm{Var}[\hat{U}_k] = \frac{\sigma^2}{N} \tag{2.11}$$

while the bias can be arbitrarily large:

$$\mathbb{E}[\hat{U}_k] - u_k = \frac{1}{N} \sum_{l \neq k} (u_l - u_k) \ . \tag{2.12}$$

The case of the *prior*-less model: Without any *prior* model on the noise-free image $\boldsymbol{u}$, the MAP estimation boils down to the maximum likelihood (ML) estimation. Under Gaussian assumption, ML estimation provides at each pixel k the estimated value given by:

$$\hat{u}_k = \arg\max_u p(v_k|u) = v_k \ . \tag{2.13}$$

In this case it is straightforward to show that the variance at each position index k is maximal:

$$\mathrm{Var}[\hat{U}_k] = \sigma^2 \tag{2.14}$$

while the bias is null:

$$\mathbb{E}[\hat{U}_k] - u_k = 0 \ . \tag{2.15}$$

The parameterization of noise-free models acts as the parameterization of machine learning algorithms used, for instance, for statistical inference. In the first example, we restricted the model to one degree of freedom leading to a prediction with poor quality but with low variance. In the second example, we let the model be excessively complex leading to an over-fitting of the solution on the noisy input data.

Figures 2.1 and 2.2 show, on two different images, the influence of two image models that we will present in the following. When the assumption fits the processed image well, the performance is good, whereas when the assumption is violated, the results are of poor quality.

2.1.3 Evaluation of denoising techniques

The most common approach to measure the performance of a given estimator h is based on the evaluation of the square error (2.5) or of an estimation of the MSE (2.6) using Monte-Carlo simulations on a large data set of noise-free images $\boldsymbol{u}$. Optimal performance is reached for a MSE of zero and the smaller the MSE, the better the estimator. Due to the law of large numbers and the large size of images, one or few simulations are usually enough to approach the MSE.

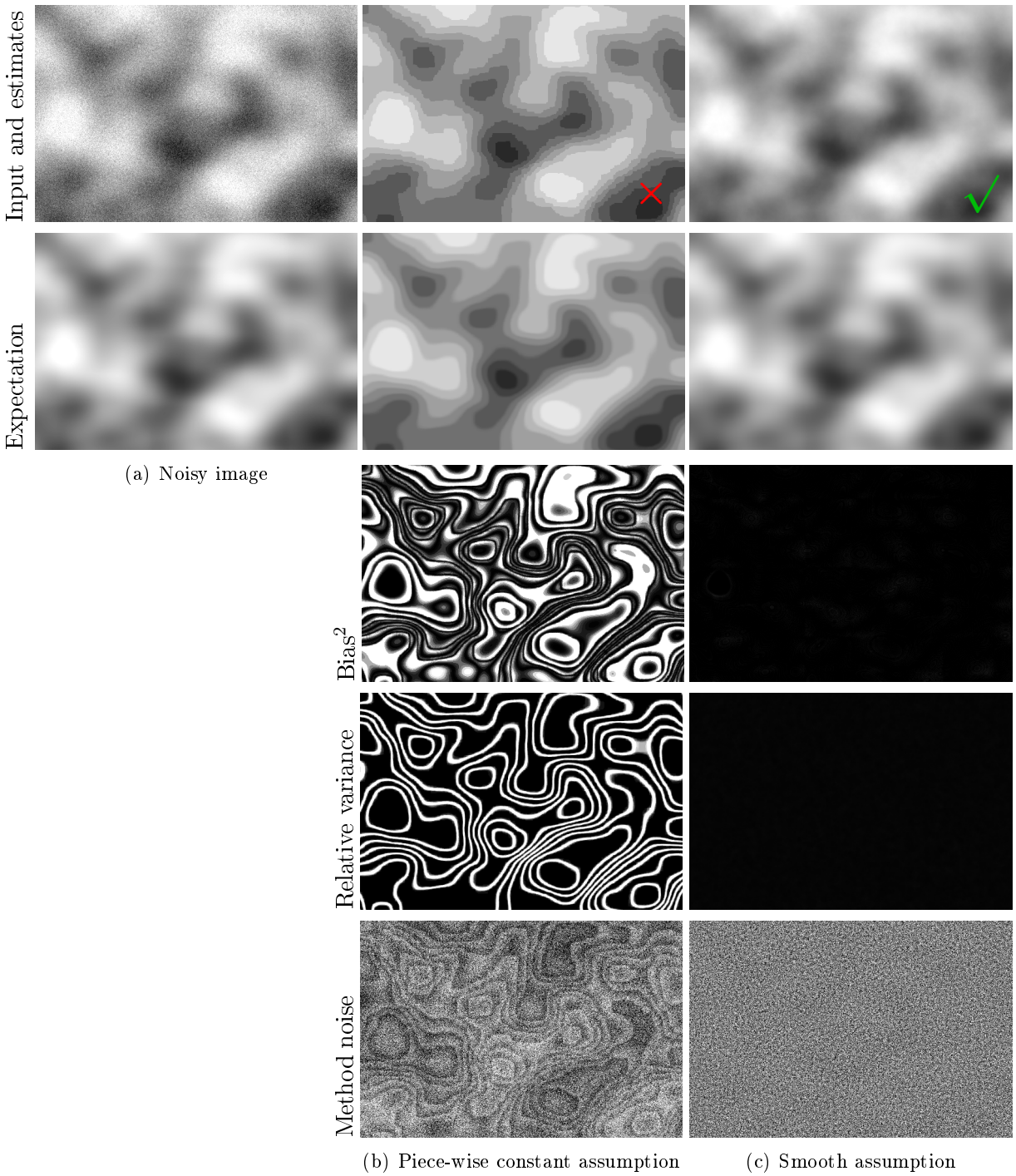
The MSE is, by definition, a quadratic criterion which varies on a large range of values. For better quality assessment, the MSE is usually mapped on a logarithmic decibel scale. This leads for instance to the peak signal-to-noise ratio (PSNR) criterion which has been introduced to measure denoising quality of 8-bit coded images:

$$\mathrm{PSNR}(\hat{\boldsymbol{u}}, \boldsymbol{u}) = 10 \log_{10} \frac{255^2}{\frac{1}{N} \|\hat{\boldsymbol{u}} - \boldsymbol{u}\|_2^2} \ . \tag{2.16}$$

The higher the PSNR, the better the estimator. Estimates of good quality have a PSNR bigger than 30 dB. For images with unbounded values or with a high dynamic range, the signal-to-noise ratio (SNR) criterion can be used instead:

$$\mathrm{SNR}(\hat{\boldsymbol{u}}, \boldsymbol{u}) = 10 \log_{10} \frac{\mathrm{Var}(\boldsymbol{u})}{\frac{1}{N} \|\hat{\boldsymbol{u}} - \boldsymbol{u}\|_2^2} \tag{2.17}$$

where $\mathrm{Var}(\boldsymbol{u})$ denotes the empirical variance of the set of variables $\{u_k\}$. Typical values of SNR for images of good quality are around 20 dB or higher.

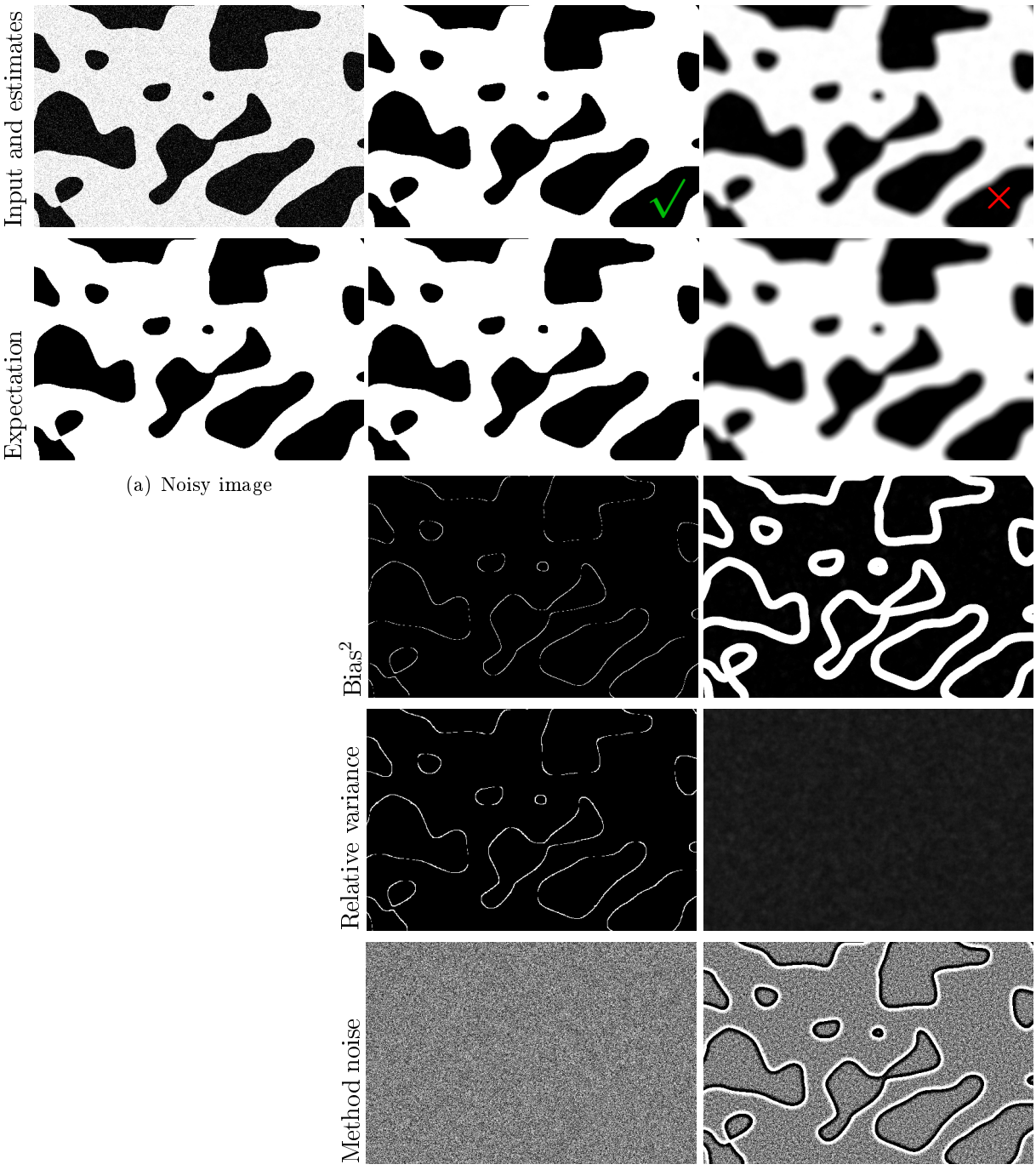


(a) Noisy image

(b) Piece-wise constant assumption

(c) Smooth assumption

Figure 2.1: First row: (a) the noisy input image and (b-c) two estimates resulting from the respective assumptions of a piece-wise constant image (obtained by the minimization of the total-variation) and the assumption of a smooth image (obtained by a moving average filter). Second row: expectations estimated over 100 noise realizations. Third to fifth rows: visual criteria using the local bias image, the local *relative variance* image and the method noise. Only the estimate lying on the good assumption provides a satisfying result. The corresponding PSNR are (b) 28.42 and (c) 43.74. The corresponding SSIM are (b) 0.790 and (c) 0.991.



(a) Noisy image

(b) Piece-wise constant assumption

(c) Smooth assumption

Figure 2.2: First row: (a) the noisy input image and (b-c) two estimates resulting from the respective assumptions of a piece-wise constant image (obtained by the minimization of the total-variation) and the assumption of a smooth image (obtained by a moving average filter). Second row: expectations estimated over 100 noise realizations. Third to fifth rows: visual criteria using the local bias image, the local *relative variance* image and the method noise. Only the estimate lying on the good assumption provides a satisfying result. The corresponding PSNR are (b) 46.07 and (c) 17.52. The corresponding SSIM are (b) 0.999 and (c) 0.746.

MSE-based criteria are only based on the difference of pixel values between the noisy image and the noise-free image. The preservation of high-level structures is not taken into account by such criteria. Wang et al. (2004) introduce the structural similarity (SSIM) to measure the visual quality of compressed images. The SSIM is based on the average of local statistics performed on a $N \times N$ sliding window. It provides a number between 0 and 1 measuring the mean distortion of the estimation compared to the true underlying image. Optimal quality is reached at 1. The SSIM has proven to be efficient for discriminating typical artifacts arising from compression such as the block artifacts inherent in JPEG-like compressions. Denoising methods lead instead to punctual artifacts or localized persistence of noise which are then not reflected well by the SSIM.

Such numerical criteria measure the average quality of an estimator. Due to this average, it does not provide relevant information on the restoration of small rare features or the introduction of small artifacts. For instance, an estimator can leave a residual variance everywhere in the image without introducing undesired artifacts while another estimator can reduce significantly the noise variance but it can introduce a single large undesired artifact in the image. As a consequence, our visual system as well as vision algorithms can prefer one estimator to another while having a lower PSNR.

Even if the relative performance between two estimators can be evaluated on a dataset of noise-free images, it may not be stated that one of them will be more efficient in any given situation. What is important is to choose the estimator for which we know and we can localize the features preserved, the features lost and the artifacts introduced.

When the noise-free image is not available, denoising quality can be evaluated by studying the residues $v(x) - \hat{u}(x)$ referred to as the *method noise* by Buades et al. (2008). Method noise is a visual criterion, designed for AWGN, and evaluating the information removed by the estimator. An ideal denoising procedure would give a method noise consisting of pure noise without any structure (i.e., uncorrelated), and following the noise statistics. If object structures are present in the method noise, that means that the related objects are not well restored in the denoised image. When the noise-free image is available, we suggest instead using Monte-Carlo simulations at the cost of a time-consuming simulation.

Proposed criteria: We suggest evaluating the preservation or introduction of information by comparing the expectation of the estimator $\mathbb{E}[\hat{\boldsymbol{U}}]$ with the true image $\boldsymbol{u}$ using the image of square bias, i.e., $(\mathbb{E}[\hat{U}(x)] - u(x))^2$. When this image is zero the estimator is said to be *unbiased*. Of course the estimator should also reduce the noise level. This reduction can be locally measured by studying the *relative (residual) variance*, i.e., the ratio $\mathrm{Var}[\hat{U}(x)]/\mathrm{Var}[V(x)]$. When this ratio is constant everywhere in the image the estimator is said to have a *stationary relative variance* (the same smoothing effect appears everywhere in the image). This study can be performed directly by measuring the local square bias and the local *relative variance* using Monte-Carlo simulations where the expectations are obtained by averaging over the different simulations. Usually, the noise variance $\mathrm{Var}[V(x)]$ is either constant and equal to σ^2 or *signal-dependent*, i.e., connected to the underlying value $u(x)$. Note that when the noise variance is constant, minimizing jointly the square bias and the variance is equivalent to minimize the MSE. This is no longer relevant when noise variance is signal-dependent.

Such statistics evaluate the bias-variance trade-off at each position of a given image. The residual noise and the variations of the introduced artifacts can be directly observed and localized thanks to the image of local variance. The elimination of structures and the introduction of systematic artifacts can then rather be observed and localized thanks to the image of local bias. While the method noise has the advantage of note requiring knowledge of the noise-free image, it is however limited to measuring correctly small bias or errors under strong noise level. These estimated local errors can provide better clues. Since statistics are local, the limitation of this

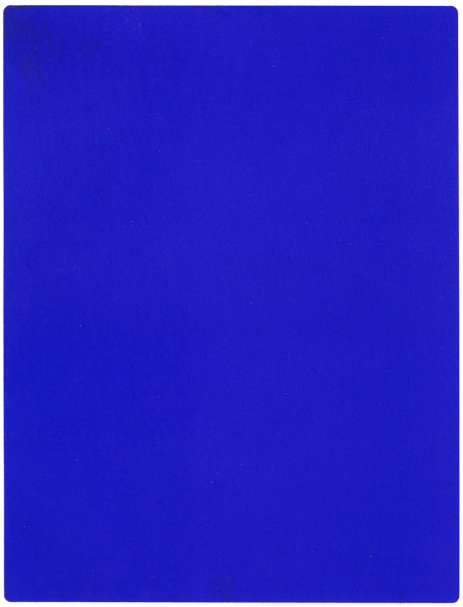


Figure 2.3: Example of an extremely smooth image: *The Blue Epoch* (Yves Klein, 1962)

approach is the necessity to perform Monte-Carlo simulations on many samples, which could potentially be time consuming.

Figures 2.1 and 2.2 illustrate the aforementioned evaluation principle with visual and numerical criteria on two different images with two different algorithms that we will present in the following. The performance of the two estimators are well reflected by all criteria on these toy examples.

2.2 Filtering of smooth regular images

The basic denoising approaches model noise-free images as smooth and regular functions. Typically, they assume that we can control the variations of the function, for instance, using the Hölder condition (Tsybakov, 2008):

$$\exists C > 0, \alpha > 0 \quad \forall x, x', \quad |u(x) - u(x')| < C\|x - x'\|^{\alpha} \, . \tag{2.18}$$

Figure 2.3 presents an example of an extremely (almost constant) smooth regular image. A less extreme case was given in Fig. 2.1 where the noise-free signal evolves slowly in space without abrupt discontinuities.

2.2.1 Linear-based filtering, moving average or boxcar filters

The crudest denoising approach is the direct application of the spatial coherence principle or ergodicity assumption. It considers noisy samples in a neighborhood on a given pixel as all following the distribution of that pixel. The ergodicity assumption leads to the moving average (boxcar filter). The estimated image is constructed by replacing each noisy value by a local average computed with a sliding window. It thus results from the convolution of the noisy image with a given kernel function modeling the size and the shape of the window. Moving average filters are then linear filters:

$$\hat{\boldsymbol{u}} = A\boldsymbol{v} \tag{2.19}$$

where A is a $N \times N$ matrix. This method lies on the following observation: the average of K independent random variables with variance σ^2 has a residual variance of σ^2/K. Thus, the moving average filter decreases the noise variance with a factor proportional to the number of pixels present inside the window. While the sliding window is usually chosen as being square, the isotropy of the filter can be improved by considering a weighted average with a circular symmetric kernel such as a Gaussian kernel of given bandwidth h:

$$\hat{u}(x) = \frac{\sum_{x'} w(x,x')v(x')}{\sum_{x'} w(x,x')} \quad \text{using} \quad w(x,x') = \exp\left(-\frac{\|x'-x\|^2}{2h^2}\right) . \tag{2.20}$$

This filter is coined the linear Gaussian filter.

Note that linear filters minimize the Bayesian risk, i.e., they are the solution of the following optimization problem:

$$\hat{u}(x) = \arg\min_{u} \sum_{x'} w(x,x')(u - v(x'))^2 \tag{2.21}$$

where $(u - v(x'))^2$ is called the quadratic cost. Compared to Eq. (2.6), it appears that linear filters decrease the mean square error as soon as the weights model the likelihood $p(\boldsymbol{v}|\boldsymbol{u})$ with unknown parameter $\boldsymbol{u}$.

Since their implementations are easy and the statistical results of these models are well known, such filters are commonly used by taking care to limit the size of the window to reach satisfying bias-variance trade-off: problem of local adaptivity to variable smoothness discussed in (Kervrann and Boulanger, 2008).

2.2.2 Heat equation based models

Smooth and regular solutions $\boldsymbol{u}$ can be expressed as the solution of the heat equation given by the following partial differential equation (PDE):

$$\frac{\partial u(x)}{\partial t} = \operatorname{div}\left(\nabla u(x)\right) . \tag{2.22}$$

where $\nabla u(x)$ is the gradient vector of $\boldsymbol{u}$ at location x. The numerical solution of (2.22) can be obtained by iterative scheme using finite difference methods. One step of this scheme realizes an isotropic diffusion of the noisy pixel values: it corresponds to the linear Gaussian filter given in Eq. (2.20).

2.2.3 Local polynomial approximation

As we have seen, linear filters can be interpreted as the minimizers of Bayesian risk (2.21). Such an optimization problem leads to locally approximating the image by a zero-order polynomial function. In regression theory, this estimator is known as the Nadaraya-Watson estimator which estimates the conditional expectation $\mathbb{E}[V(x)|u(x)]$. Under the assumption of smooth regular images, Katkovnik et al. (2006) suggest rather searching for a local polynomial approximation (LPA) with higher order (idea introduced for density estimation in Silverman, 1986; Scott, 1992; Fan and Gijbels, 1996). For each pixel position x, the best polynomial approximation $\hat{Q}_x$ is obtained by solving the following optimization problem:

$$\hat{Q}_x = \arg\min_{Q_x} \sum_{x'} w(x,x')(Q_x(x') - v(x'))^2 . \tag{2.23}$$

Figure 2.4: Example of a cartoon image: *the Flinstones*

Once the coefficients of the polynomial $\hat{Q}_x$ are obtained, the estimated noise-free value at position x is given by $\hat{u}(x) = \hat{Q}_x(x)$. This idea linked with spatially-adaptive windows (see next section) is at the heart of recent phase unwrapping methods dealing with noisy data (Katkovnik et al., 2008; Bioucas-Dias et al., 2008).

2.2.4 Common limitation: the loss of resolution

If an image presents sharp discontinuities, violating the regularity assumption, the resulting image will suffer from a resolution loss, since the same smoothing effect will be applied equally to homogeneous regions and to edges. Linear filters are low-pass filters eliminating high frequencies which encode, among others, the edges of the image. Edge preserving filtering aims at limiting this problem.

2.3 Edge preserving filtering of piece-wise constant images

Edge preserving filters try to model the noise-free image $\boldsymbol{u}$ as a regular image with discontinuities, i.e, as a piece-wise constant image. As shown on Fig. 2.4, cartoon-like images are good examples of piece-wise constant images sometimes referred to as the cartoon model. Unlike for smooth regular images, the same smoothing cannot be performed equally everywhere in the image but should rather adapt to the local content of the image: we speak about adaptive filtering. They assume that the regularity assumption only holds true in certain direction or anisotropic neighborhood.

2.3.1 Anisotropic diffusion: Perona and Malik model

Linear filters realize an isotropic diffusion where the noise values of each side of an edge are mixed together leading to a blur. To avoid such an undesired effect, the information should instead be diffused separately on each side of the edges without mixing different populations. This is the principle of anisotropic diffusion. Perona and Malik (1990) adopted the heat equation with

spatially varying coefficients:

$$\begin{cases} \frac{\partial u(x,t)}{\partial t} = \operatorname{div}\left(g(|\nabla u(x,t)|)\nabla u(x,t)\right) \\ u(x,0) = v(x) \end{cases} . \tag{2.24}$$

where g is a decreasing function, for instance, $g(.) = \exp(-.^2/\kappa^2)$ or $g(.) = \frac{1}{1+.^2/\kappa^2}$. In this latter case, the parameter κ controls the anisotropy of the diffusion (the larger κ, the more isotropic the diffusion). This function adapts the diffusion as a function of the image content. In the direction where the gradient is large, e.g., in the orthogonal direction of an edge, the diffusion is null. When the gradient is low in all directions, e.g., in homogeneous areas, this filter acts as the isotropic diffusion.

2.3.2 Filtering with adaptive windows

In the same vein as the anisotropic diffusion, local adaptivity can be reached by considering shape-adaptive windows instead of a predetermined neighborhood. Refined Lee's filter (Lee, 1981) selects at each pixel one among eight oriented windows (Fig. 2.5(a)). Here windows are of fixed size and only their orientations are spatially varying. In homogeneous areas it can be preferable to use larger windows while around fine details smaller windows would be selected to preserve the resolution. Adaptive window sizes or spatially variable bandwidth selection can be used to reach this goal (Park et al., 1999; Takeda et al., 2007) (Fig. 2.5(b)). Of course better performance can be obtained by using shape and size adaptive windows. Katkovnik et al. (2002) suggest independently adapting the size of the four quadrants around the pixel of interest (Fig. 2.5(c)). This idea has been extended in (Katkovnik et al., 2004) where the scale of many directional windows is adapted. Vasile et al. (2006) proposed to build an intensity driven adaptive neighborhood (IDAN) thanks to a two-step region growing algorithm (Fig. 2.5(d)).

These approaches can require a high computation load depending on the model complexity: selection of a window among a given set of windows usually leads to faster algorithms than techniques defining locally windows of arbitrary shape. All these methods are based on local analysis of the image content, such as, the gradient orientation, the statistics inside the selected window and/or the analysis of the confidence or the risk of the estimation. Such approaches are then linked to aggregation-based filtering since in general multiple estimates are provided by the different possible neighborhoods (see Sec. 2.6).

2.3.3 Total-variation minimization and Markov random fields

Total-variation (TV) is an example of a *prior* that enforces smoothness while preserving edges. Rudin et al. (1992) introduce a variational formulation acting as a compromise between the regularity of the solution and the fidelity of the solution with the noisy data. The Rudin-Osher-Fatemi (ROF) model describes the noise-free image $\boldsymbol{u}$ as the solution of the following optimization problem:

$$\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}} \|\boldsymbol{v} - \boldsymbol{u}\|_2^2 + \lambda \sum_x |\nabla u(x)| \tag{2.25}$$

where λ is a Lagrangian multiplier controlling the regularity of the solution. When $\lambda = 0$, the solution is the noisy image itself and when $\lambda \to \infty$, the solution is a constant image. The ℓ_2 norm measures the consistency of the solution with the observation $\boldsymbol{v}$. The total variation, i.e., $\sum_x |\nabla u(x)|$, penalizes the variations or transitions present in the solution (it sums the size of each transition). In practice, the first term is directly evaluated by the pair of pixel value difference

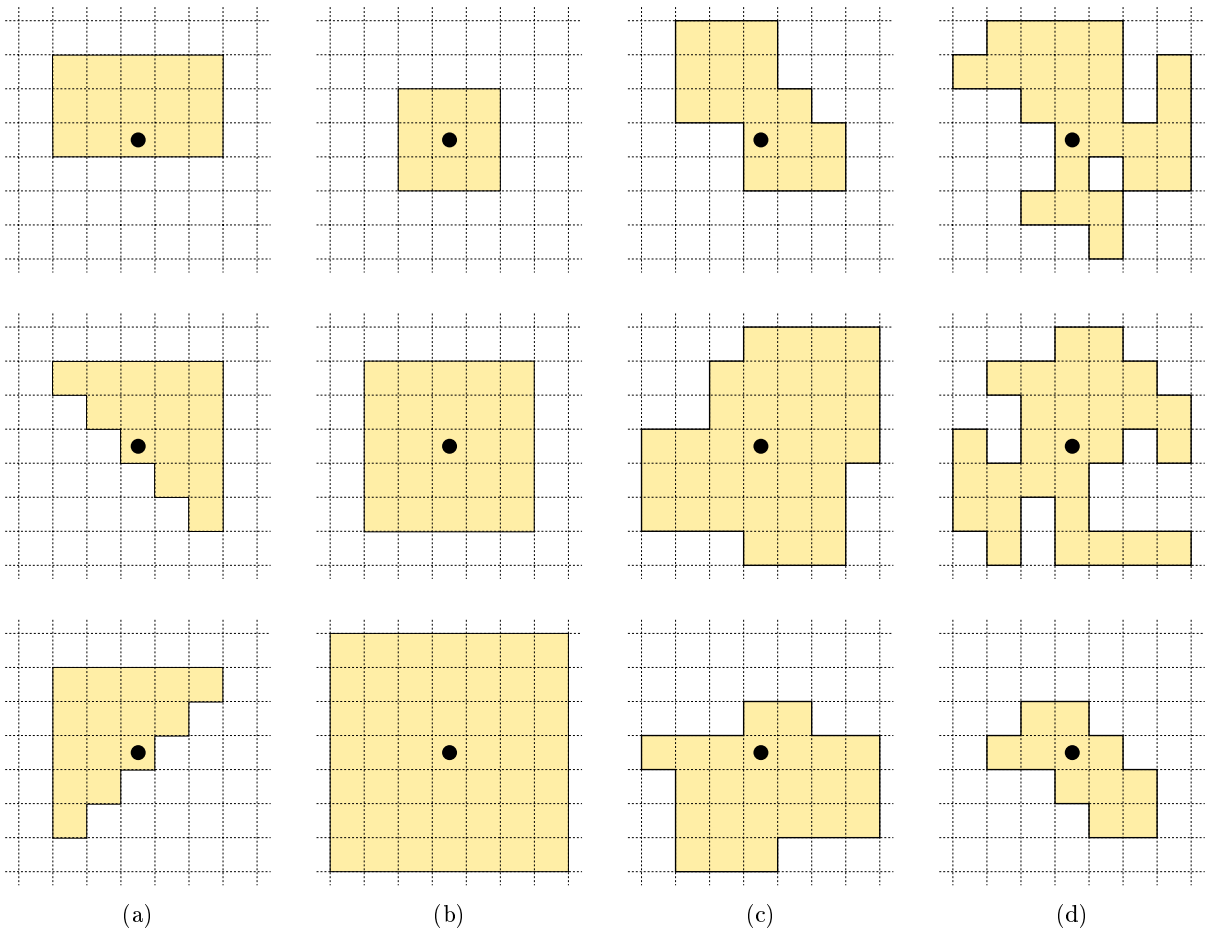


Figure 2.5: Examples of selected local windows by adapting (a) the orientation, (b) the size (c) the sizes of the four quadrants and (d) an adaptive neighborhood obtained by a region growing algorithm.

while the second term is evaluated using finite difference on a graph formed from the grid Ω and a local connectivity system such as the 4 or 8 connectivity. The ROF model provides good results on cartoon images and more generally on regular images in the sense of Besov spaces. However, it leads to stronger attenuation of several small disconnected regions than that of a single region (Strong and Chan, 2003). When used on non-cartoon images, TV minimization leads to the *stair-casing* effect (Nikolova, 2000) (see the effect of TV minimization on a smooth image in Fig. 2.1).

As we will see in Chap. 3, the ROF model can be interpreted in a Bayesian setting as the maximization of the *a posteriori* probability, i.e., $\hat{\boldsymbol{u}} = \arg\max_{\boldsymbol{u}} p(\boldsymbol{u}|\boldsymbol{v})$, under the Gaussian noise assumption on a Markov random field.

Note that other fidelity terms can be used. In particular the use of the ℓ_1 norm leads to the so-called TV-L1 model first introduced by (Alliney, 1992):

$$\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}} \|\boldsymbol{v} - \boldsymbol{u}\|_1 + \lambda \sum_x |\nabla u(x)| \; . \tag{2.26}$$

Nikolova (2003) shows that non-smooth fidelity terms are more robust to outliers than smooth fidelity terms. Compared to the ROF model, the TV-L1 model does not include loss of contrast in the solution but remove small objects instead (Chan and Esedoḡlu, 2005; Duval et al., 2009).

Note that the relation between the ROF and the TV-L1 model is the same as the relation between the average or the median filter, since, the average minimizes the ℓ_2 norm while the median minimizes the ℓ_1 norm.

Other norms on the *prior* term can also be used to model piece-wise constant images. For instance, the Potts model penalizes transitions whatever the size of the step (Wu, 1982). It corresponds to the ℓ_0 pseudo-norm of the gradient counting the number of non-null transitions. The Potts model sums the length of the perimeter of each object in the image. Variational problems using the Potts model lead to solve non-polynomial (NP) hard problems. Tikhonov regularization $\sum_x |\nabla u(x)|^2$ is a convex *prior* leading to regular solutions unadapted to model sharp transitions but easy to be solved numerically.

All these approaches are influenced by one or more filtering parameters that control the amount of discontinuities present in the image. The limitation of edge preserving filters is that they cannot restore large homogeneous regions as well as punctual targets and high frequency patterns or textures. Dictionary-based filtering aims at restoring these features thanks to a dictionary or a *code-book* encoding punctual targets, edges, textures or homogeneous areas.

2.4 Dictionary-based filtering: sparse decompositions

The above methods define denoising directly in the image domain. It has been shown by Donoho and Johnstone (1994) that spatial adaptation of the smoothing can instead be obtained by using sparse decompositions of images on a well-chosen alternative space described by a collection of *atoms* or *code-book*. This collection is referred as the dictionary. The dictionary is all the more relevant that the image can be represented with few *words*.

Dictionary-based filters model an image as a linear combination of a family of K atoms $\boldsymbol{d}_k$ of dimension N. The $N \times K$ matrix $\boldsymbol{D}$, whose columns are the atoms $\boldsymbol{d}_k$ is called the dictionary. In such approaches, a noise-free image is assumed to be sparse meaning that it can be decomposed as the product $\boldsymbol{D\alpha}$ where the K-dimensional vector $\boldsymbol{\alpha}$ is said to be sparse (i.e., only a few of its coefficients are non-null). Dictionary-based filters look for a sparse decomposition of $\boldsymbol{u}$ fitting the observation $\boldsymbol{v}$:

$$\min_{\boldsymbol{D},\boldsymbol{\alpha}} \|\boldsymbol{D\alpha} - \boldsymbol{v}\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{\alpha}\|_0 < \varepsilon \tag{2.27}$$

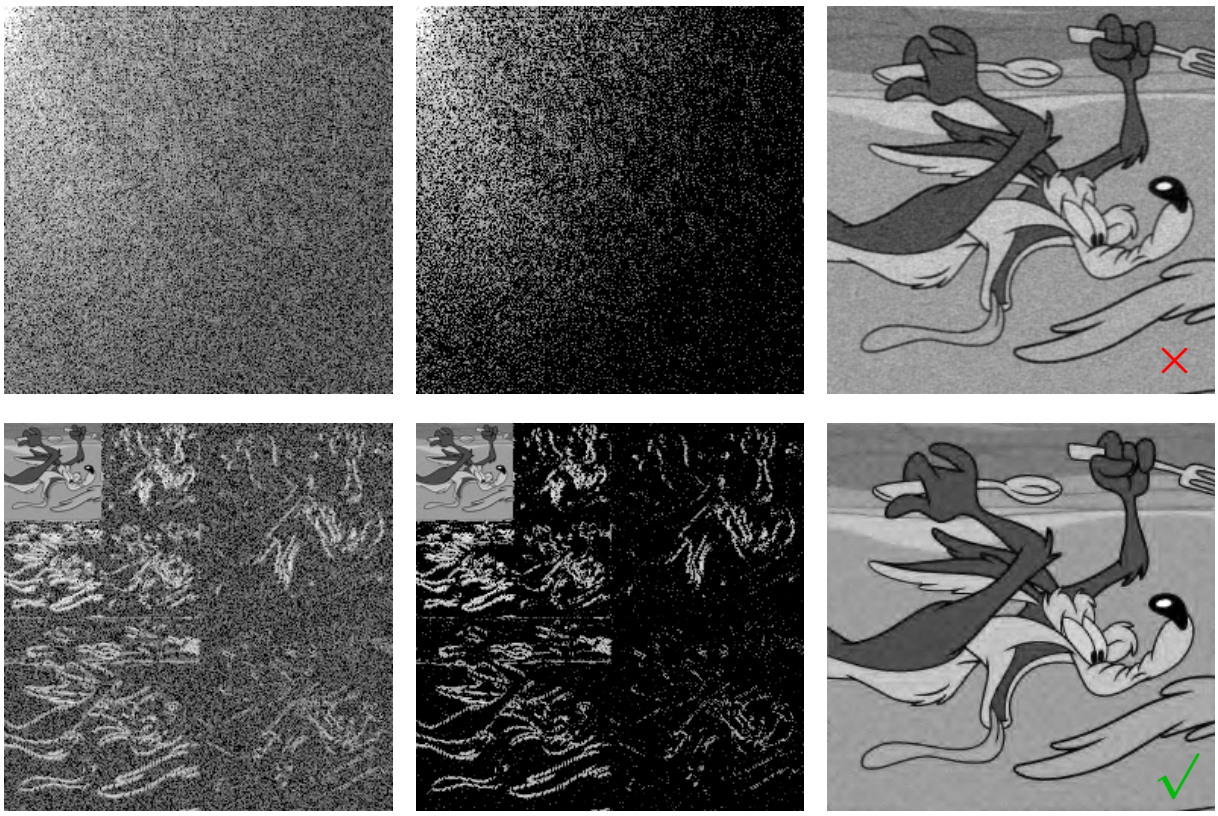
where $\varepsilon > 0$ controls the amount of sparsity in the solution and $\|\boldsymbol{\alpha}\|_0$ is the ℓ_0 pseudo-norm of $\boldsymbol{\alpha}$, i.e., the number of non-zero entries. Eq. (2.27) searches for a sparse vector $\boldsymbol{\alpha}$ and a dictionary $\boldsymbol{D}$ that can synthesize the image $\boldsymbol{u}$. Such a minimization is said to rely on a *sparsity synthesis prior*. By opposition, a *sparsity analysis prior* leads to the minimization of:

$$\min_{\boldsymbol{D},\boldsymbol{u}} \|\boldsymbol{u} - \boldsymbol{v}\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{D}^t\boldsymbol{u}\|_0 < \varepsilon \,. \tag{2.28}$$

This approach rather searches for an image $\boldsymbol{u}$ and a dictionary $\boldsymbol{D}$ such that $\boldsymbol{u}$ can be analyzed as sparse through the dictionary $\boldsymbol{D}$. When $\boldsymbol{D}$ is orthogonal the two approaches are equivalent.

Without restriction on $\boldsymbol{D}$, Davis et al. (1997) show that the minimization of Eq. (2.27) can lead to non-polynomial (NP) hard problems (since it is combinatorial). To relax the problem, the ℓ_0 pseudo-norm is often substituted by other norms. The typical choice is the ℓ_1 norm, leading to the following relaxation:

$$\min_{\boldsymbol{D},\boldsymbol{\alpha}} \|\boldsymbol{D\alpha} - \boldsymbol{v}\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{\alpha}\|_1 < \varepsilon \,. \tag{2.29}$$



(a) Noisy-image coefficients (b) Soft thresholding (c) Denoised result

Figure 2.6: Examples of soft-thresholding on (top) the DCT and (bottom) the Daubechies-DWT. The absolute values of the coefficients are displayed in (a) and (b) such as low values are dark and high values are bright. It appears clearly the the DWT provides a sparser representation of the image resulting in a better smoothing effect and preservation of sharp feature. However, the soft-thresholding of DWT leads to oscillations around edges known as Gibbs phenomena.

When the dictionary is fixed in advance, Eq. (2.29) is known as the LASSO (least absolute shrinkage and selection operator) regularization which also favors solutions with fewer nonzero entries (Tibshirani, 1996). The substitution of the ℓ_0 norm by the ℓ_2 norm leads to the Tikhonov regularization known to provide smooth and non-sparse estimates even when the input signal is sharp and sparse: it does not preserve sparsity.

To solve the optimization problems (2.27) and (2.29), some restrictions have to be made with respect to the nature of the dictionary. It can be fixed in advance (e.g., considering a basis formed of sinusoids, wavelets, ...), pre-determined from the noisy data (e.g., using a principal component analysis), chosen arbitrarily (e.g., in compressed sensing) or it can be obtained by iteratively solving the optimization problem (e.g., Aharon et al., 2006). The dictionary can constitute either an orthogonal basis of the image space or an over-complete family (e.g., an undecimated wavelet family).

2.4.1 Orthonormal decomposition

We consider here that the dictionary is a given orthonormal basis $\boldsymbol{D}$, hence, $K = N$. In this case, each coefficient α_k of the decomposition is obtained by projecting $\boldsymbol{v}$ on each atom $\boldsymbol{d}_k$:

$$\alpha_k = \langle \boldsymbol{v} | \boldsymbol{d}_k \rangle \tag{2.30}$$

where $\langle \boldsymbol{v} | \boldsymbol{d}_k \rangle$ denotes the scalar product of $\boldsymbol{v}$ on $\boldsymbol{d}_k$. For instance cosine and wavelet bases are known to be able to capture most of a signal or image in few coefficients. This property is exploited by compression techniques such as JPEG and JPEG2000.

The discrete cosine transform (DCT) uses a dictionary formed of N orthogonal sinusoidal discrete functions:

$$d_n(f) = \cos\left(\frac{\pi}{2N}(2n+1)f\right) \tag{2.31}$$

where $n = [0, N-1]$ and $f = [0, N-1]$. The DCT represents then any signal as a sum of sinusoids. A similar decomposition is obtained by the discrete Fourier transform (DFT) which encodes a signal on the basis of complex exponential functions. The first main limit of these approaches is that it does not provide a sparse decomposition of sharp features. For instance, an edge has a wideband spectrum. The separability of the signal and noise from the DCT is then a difficult task. Another difficulty by using the DCT is that atoms are not localized in space (see Fig. 2.6). A solution would be to compute the DCT or the DFT on a sliding window, but such representations are redundant and do not constitute an orthonormal decomposition of the image.

While the DCT captures only the frequencies of the signal, the discrete wavelet transform (DWT) captures instead spatial and scale properties of the image. The DWT is said to be localized in both time and frequency. The DWT uses a dictionary derived from a *mother* wavelet $\psi(t)$ and its *daughter* wavelets arising from dyadic translations and dilatation of ψ:

$$d_{j,k}(t) = \sqrt{2}^j \psi\left(2^j t - k\right) \tag{2.32}$$

where j and k are two integers. Unlike sinusoidal functions, the mother wavelet $\psi(t)$ is chosen such that it can easily represent sharp discontinuities, orientations and/or smooth information. Note that specific conditions are required on the chosen mother wavelet to achieve perfect reconstruction. Haar, Daubechies and Gabor wavelets are typical choices verifying these conditions (Daubechies, 1992; Lee, 1996). The 2D DCT and the 2D DWT are obtained by the direct extension of the 1D DCT and 1D DWT localized in time and frequency in both directions. The 2D DWT represents the feature of images in a pyramid of sub-bands corresponding to the different scales and orientations (see Fig. 2.6). Ridgelets (Candès, 1998; Candès and Donoho, 1999), curvelets (Candes and Donoho, 2000; Starck et al., 2002), bandelets (Le Pennec and Mallat, 2005; Mallat and Peyré, 2007), contourlets (Do and Vetterli, 2005), grouplets (Mallat, 2009a) or random basis (i.e., compressed sensing, see Donoho, 2006) are other fixed dictionaries of images (potentially redundant) based on the same ideas. See (Mallat, 2009b) for more details about such representations.

For denoising applications, orthogonal transforms like the wavelet or discrete cosine transforms lead to a separation of signal and noise. Let $\boldsymbol{\beta}$ be the noisy version of the noise-free transformed image $\boldsymbol{\alpha}$. Noise can then be strongly suppressed by zeroing the least significant coefficients. Indeed, in this case, the minimization of (2.27) leads to hard-thresholding (HT) (Mallat, 2009b):

$$\hat{\boldsymbol{u}} = \boldsymbol{D}\hat{\boldsymbol{\alpha}}^{HT} \quad \text{where} \quad \hat{\alpha}_k^{HT} = \beta_k \text{ if } |\beta_k| > \lambda \text{ and } 0 \text{ otherwise} \tag{2.33}$$

where the threshold λ acts on the sparsity of the solution. In the same vein, the minimization of (2.29) leads to soft-thresholding (ST) (Donoho and Johnstone, 1994, 1995):

$$\hat{\boldsymbol{u}} = \boldsymbol{D}\hat{\boldsymbol{\alpha}}^{ST} \quad \text{where} \quad \hat{\alpha}_k^{ST} = \text{sign}(\beta_k) \cdot \max(0, |\beta_k| - \lambda) \ . \tag{2.34}$$

The threshold λ is usually chosen as $\lambda = \sigma\sqrt{2 \log N}$ known as the universal threshold (Donoho and Johnstone, 1994). However, Donoho and Johnstone (1995) show that spatial adaptation of

this threshold can improve the overall quality. Their method, called *SureShrink*, is based on the SURE methodology presented in Sec. 2.6.4. Chang et al. (2000) show next, with *BayesShrink*, that adaptive thresholds can be obtained in a Bayesian framework by modeling the distribution of noise-free coefficients with a generalized Gaussian (or generalized Laplace) distribution (following the idea of Mallat, 1989). All these works are based on the Gaussian noise assumption.

Another solution consists of using the Wiener filter (also referred to as Bayes' filter in the statistical literature) as proposed by Muresan and Parks (2003) and Zhang et al. (2010a). Wiener filter searches at each pixel the linear transform of the coefficient minimizing the mean square error. Under the Gaussian noise assumption, the linearity of orthogonal transforms maps $\boldsymbol{v} = \boldsymbol{u}+\boldsymbol{\varepsilon}$ to the coefficients $\boldsymbol{\beta} = \boldsymbol{\alpha}+\boldsymbol{\xi}$ where $\boldsymbol{\alpha}$ is the transform of $\boldsymbol{u}$, $\boldsymbol{\beta}$ is the transform of $\boldsymbol{v}$ and $\boldsymbol{\xi}$ is the transform of $\boldsymbol{\varepsilon}$. Under independence assumptions, the minimization $\arg\min_w \mathbb{E}\left[(w\alpha_k - \beta_k)^2\right]$ leads to the following solution:

$$\hat{\boldsymbol{u}} = \boldsymbol{D}\hat{\boldsymbol{\alpha}}^{Wiener} \quad \text{where} \quad \hat{\alpha}_k^{Wiener} = \frac{\sigma_{\alpha_k}^2}{\sigma_{\alpha_k}^2 + \sigma_{\xi_k}^2}\beta_k \,. \tag{2.35}$$

where σ_{α_k} and σ_{ξ_k} are respectively the standard deviations of α_k and ξ_k (Scharf and Demeure, 1991). This solution is then a linear rescaling of the input coefficients. Eq. (2.35) is actually a solution of Tikhonov's regularization resulting from the ℓ_2 constraint.

Note that a multitude of thresholding strategies have been proposed depending of the problem at hand. For instance, Abramovich et al. (1998) suggest using the posterior median in a Bayesian context for non-parametric regression.

The above approaches assume that coefficients are statistically independent. Wegmann and Zetzsche (1990); Simoncelli (1997); Buccigrossi and Simoncelli (1999) show that nearby coefficients present high dependencies (see for instance Fig. 2.6). Portilla et al. (2003) suggest modeling the dependencies between neighbor wavelet coefficients in the pyramid of sub-bands at different scales and orientations based on a Gaussian scale mixture (GSM). This approach combined with Bayesian least square (BLS) estimation has led to the BLS-GSM approach considered as one of the most powerful wavelet approaches for image denoising.

Orthonormal basis may not form a suitable dictionary to represent images. In such basis, the thresholding of coefficients leads to the introduction of visual artifacts such as Gibbs phenomena (i.e., large oscillations around edges, see Fig. 2.6). Some artifacts can be attributed to the non-shift invariance of such basis: the coefficients of a shifted image are not a shifted version of those of the original image. Shift-invariance can be reached by using redundant decompositions such as the cycle-spinning decomposition or the undecimated wavelet transform (see Shensa, 1992; Mallat and Zhang, 1993; Coifman and Donoho, 1995; Pesquet et al., 1996). By combining such shift-invariant decompositions with the idea of statistical aggregation (see Sec. 2.6), the solution of Guleryuz (2007) outperforms the BLS-GSM filter.

2.4.2 Redundant or over-complete decomposition

Sparse decompositions with redundant or over-complete or learned dictionaries has been the topic of several works including (Olshausen and Field, 1996; Starck et al., 2002; Aharon et al., 2006; Mairal et al., 2010).

When the dictionary $\boldsymbol{D}$ is redundant, the optimization of the sparsity problem with ℓ_0 pseudo-norm (2.27) leads to a non-polynomial hard problem (Davis et al., 1997). Local-optima can be obtained by greedy algorithms such as the *matching pursuit* algorithm introduced by Mallat and Zhang (1993). Matching pursuit iteratively projects the residual noisy image on the atom minimizing the mean square error. At iteration i, the matching pursuit provides the following

linear expansion of the redundant atoms:

$$\hat{\boldsymbol{u}}^i = \sum_{j=1}^{i} \langle R^j | \boldsymbol{d}_{k^*} \rangle \, \boldsymbol{d}_{k^*} \tag{2.36}$$

$$\text{where} \quad k^* = \arg\max_k \left| \langle R^j | \boldsymbol{d}_k \rangle \right| \ , \tag{2.37}$$

where R^j is the residual image at iteration j, i.e., $R^j = \boldsymbol{v} - \hat{\boldsymbol{u}}^{j-1}$ if $j > 1$ and $R^j = \boldsymbol{v}$ otherwise. The number of performed iterations controls the sparsity of the solution. Extensions of the *matching pursuit* have led, for instance, to the *orthogonal matching pursuit* introduced by Pati et al. (1993). In the same spirit, Chen et al. (1999) introduces the *basis pursuit* under the ℓ_1 constraint.

When redundant and/or over-complete dictionaries are considered, one can be interested in the learning of a dictionary from a collection of noisy data. As mentioned by many authors, the research of a dictionary for sparse representation is intrinsically related to clustering or vector quantization. Vector quantization is an extreme sparse representation where each input vector is represented by only one predefined vector: the *centroïd* of its cluster. Any clustering algorithm can then be used for the learning of redundant and over-complete dictionaries, for instance, the K-means algorithm. Aharon et al. (2006) introduce a generalization of the K-Means to minimize the sparsity problem with ℓ_0 pseudo-norm (see Eq. (2.27)). It is an iterative two stage algorithm:

1. the *sparse coding stage* uses a pursuit algorithm to obtain the sparse coefficients of each noisy sample on the fixed dictionary $\boldsymbol{D}$ and,

2. the *codebook update stage* adapts the dictionary $\boldsymbol{D}$ by sequentially updating each column $\boldsymbol{d}_k$ by performing K single value decompositions (K-SVD) to minimize the error on the group of noisy samples whose sparse decomposition have a non-null coefficient α_k.

Learning a dictionary of images would be a very difficult task due to the high dimension and the high variability of natural images. Instead, the K-SVD uses such decompositions on the collection of the 8×8 small sub-images extracted from the noisy image itself. Due to their small dimensions, such sub-images, called *patches*, present a smaller variability and can be organized in several clusters representing the redundant patterns that occur in the image. An interesting approach to consider a dictionary which encodes spatial relations, called *epitome*, has been recently proposed by (Benoît et al., 2011). The patch representation of images is at the heart of the most recent denoising techniques.

2.5 Patch-based filtering: self-similar images

Patch-based filters model images as a collection of patches (i.e., small windows extracted at different positions) and assume that this collection presents redundancy or clusters. This assumption relies on the self-similarity property of images: the same content can be observed at different positions, thus, most of the patterns occur several times. For instance, consider patches extracted from the same homogeneous area, along edges or on a repetitive texture (see Fig. 2.7). Such models seem to fit well for several natural images and have inspired many image processing approaches including texture synthesis (Efros and Freeman, 2001; Liang et al., 2001; Kwatra et al., 2003), texture classification (Varma and Zisserman, 2003), super-resolution (Freeman et al., 2002), inpainting (Criminisi et al., 2004) and image editing (Cho et al., 2009).

Such approaches represent a noise-free image $\boldsymbol{u}$ as the following ensemble:

$$\{(x, u(\mathrm{P}_x)) \mid x \in \Omega\} \quad \text{where} \quad \mathrm{P}_x \subset \Omega \quad \text{and} \quad u(\mathrm{P}_x) = (u_k)_{x_k \in P_x} \ . \tag{2.38}$$

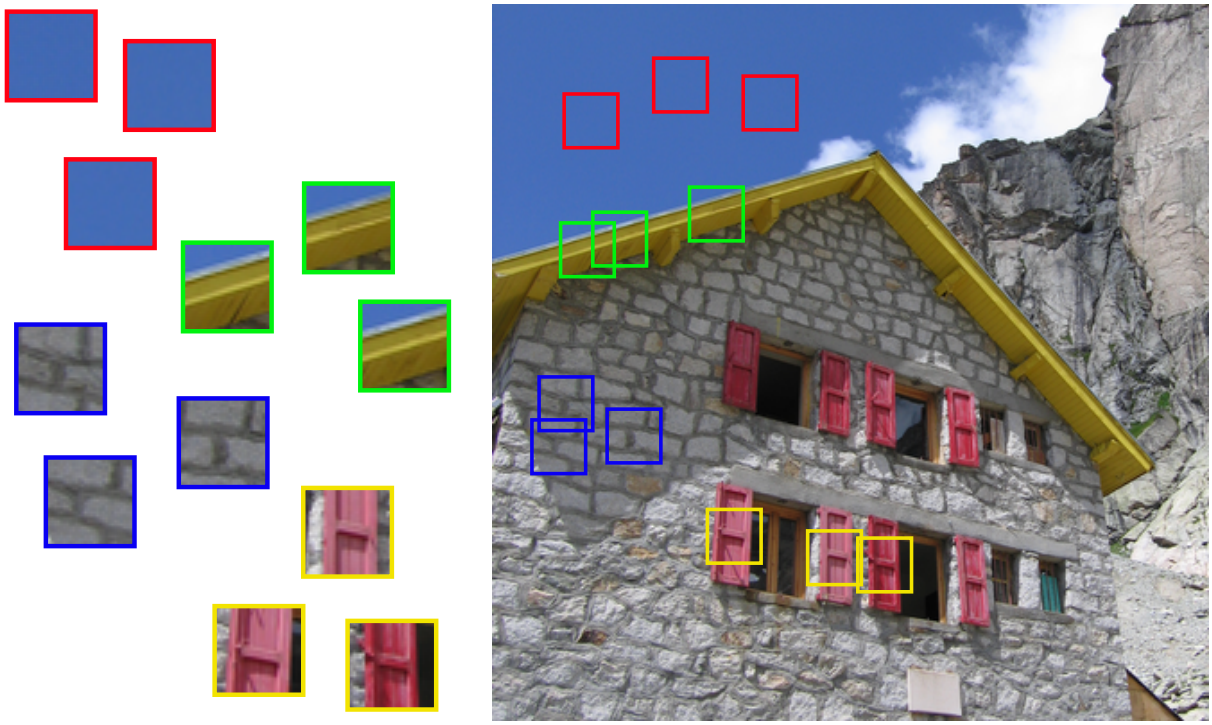


Figure 2.7: Illustration of the self-similarity property of natural images.

For a fixed (odd) width p, the subset P_x usually contains the locations of pixels x' located in the $p \times p$ square window centered on x:

$$\mathrm{P}_x = \left\{ x' = x + \tau \;\middle|\; \tau \in \left[\!\left[-\frac{p-1}{2}, \frac{p-1}{2} \right]\!\right]^2 \right\} \tag{2.39}$$

where $[\![n_1, n_2]\!] = \{n_1, n_1 + 1, \ldots, n_2\}$. The patch P_x can also have a non-square shape, be non spatially-connected, have space-varying shapes and/or space-varying sizes (see Dabov et al., 2009, and Chap. 7). Based on this representation, redundancy of patches can be expressed as:

$$\forall x,\ \#\left\{ x' \mid u(\mathrm{P}_x) \simeq u(\mathrm{P}_{x'}) \right\} \gg 1 \tag{2.40}$$

where $\#\mathcal{E}$ is the cardinal of the set $\mathcal{E}$ and $u(\mathrm{P}_x) \simeq u(\mathrm{P}_{x'})$ denotes that $u(\mathrm{P}_x)$ and $u(\mathrm{P}_{x'})$ are similar in the sense of a given similarity criterion.

From the image $\boldsymbol{u}$ we can build an image of patches such that its content at position x is the vector $u(\mathrm{P}_x)$. The image is then a function from Ω to the patch space of dimension $|P|$ and defined as:

$$\forall x, \quad \boldsymbol{u}(x) = u(\mathrm{P}_x)\ . \tag{2.41}$$

Note that this representation is redundant due to the patch overlap. The problem of reconstructing a 2D image from this image is called the reprojection problem presented in Sec. 2.5.4.

Simultaneously, Awate and Whitaker (2005, 2006) and Buades et al. (2005), respectively with the UINTA filter and the NL means filter, introduce the patch-based model for denoising purposes. This approach based on the self-similarity property of images takes inspiration from the patch-based approach proposed for texture synthesis by Efros and Leung (1999).

2.5.1 Non-local means (NL means) filter

The common idea of the UINTA filter and the NL means filter is to combine the information shared by redundant patches to decrease the noise level (i.e., to select a large set of similar pixels to combine). Compared to local filters, at the expense of spatial relationships, only patch similarities are taken into account. Such approaches are then considered as non-local as pixel values far apart can be combined together. The idea is to replace the noisy value at each pixel position x by the weighted average of the noisy values $v(x')$ with patches $\boldsymbol{v}(x')$ similar to $\boldsymbol{v}(x)$. In kernel regression, such a solution is called the Nadaraya-Watson estimator known to estimate the conditional expectation $\mathbb{E}[V(x)|u(x)]$. In their seminal paper, Buades et al. (2005) define weights by the distance between the noisy values observed in the two $p \times p$ square patches $u(\mathrm{P}_x)$ and $u(\mathrm{P}_{x'})$ (i.e., the patches surrounding the pixel of interest and the candidate pixel):

$$\hat{u}(x) = \frac{\sum_{x'} w(x,x')v(x')}{\sum_{x'} w(x,x')} \tag{2.42}$$

$$\text{where} \quad w(x,x') = \varphi\left(\frac{d^2(v(\mathrm{P}_{x'}), v(\mathrm{P}_x))}{2|P|h^2}\right) \tag{2.43}$$

where φ is a kernel decay function $\mathbb{R}^+ \to [0,1]$, d a distance or a dissimilarity criterion taking its values in $\mathbb{R}^+$, $|\mathrm{P}| = p \times p$ is the size of the patches and $h > 0$ controls the amount of filtering. Note in practice that for computational issues, the research of the pixel candidates x' are limited to a large search window W_x centered on x.

Let us now briefly recall the influence of each parameter (see Duval et al., 2011, for a more extensive discussion on this subject):

The search window size $|\mathrm{W}|$: the summation defined in Eq. (2.42) is usually restricted to a search window around the pixel of interest: it is by convention an $\ell \times \ell$ square window. Defining $\mathrm{W} = [\![-\frac{\ell-1}{2}, \frac{\ell-1}{2}]\!]^2$, the search window centered on each pixel x is then $\mathrm{W}_x = x + \mathrm{W}$. Such a restriction was proposed in the seminal work of Buades et al. (2005) for computational acceleration. Though, some authors have also noticed that locally choosing the best search windows (Kervrann and Boulanger, 2006) or restricting the average over small ones (see the work of Salmon (2010) and Duval et al. (2011) for more details) could benefit the NL means procedure (see also Fig. 2.9).

The patch size $|\mathrm{P}|$: this parameter is generally chosen to be equal to 5, 7 or 9. Using a width $p = 1$ leads to a method close to the Yaroslavsky filter (Yaroslavsky, 1985). This parameter is intrinsically linked to the resolution or the scale of the objects in the image. It controls the redundancy assumption defined in Eq. (2.40). A size too small leads to a wrong selection of candidate pixels leading either to a noise reduction that is too small or a blurring effect with respect to the bandwidth h. A size too large leads instead to the rare patch effect (see Chap. 7): the noise is not reduced enough in regions where such large patches have no similar replica. See for instance Fig. 2.9. The scale of images is inherently space-varying. In most papers the patch width is a global fixed parameter. Few works have tried to handle the difficult task of using several sizes of patches for a single image. To our knowledge, the first attempt was proposed in the context of learning patches by Mairal et al. (2008), using Support Vector Machines (SVM). Another approach using variance control was also considered in (Salmon and Strozecki, 2010). We consider in Chap. 7, a local adaptivity of the shape and the size of the patches.

The bandwidth h: this parameter has a smoothing effect and plays the same role as the bandwidth for kernel methods in statistics (see Wasserman, 2007, for more details). With our parameterization, the larger the bandwidth, the smoother the image becomes. When $h \to 0$, the solution tends toward the noisy image and when $h \to \infty$, the filter tends to the moving average filter (see Fig. 2.9). Choosing this parameter is a difficult task and many solutions have been

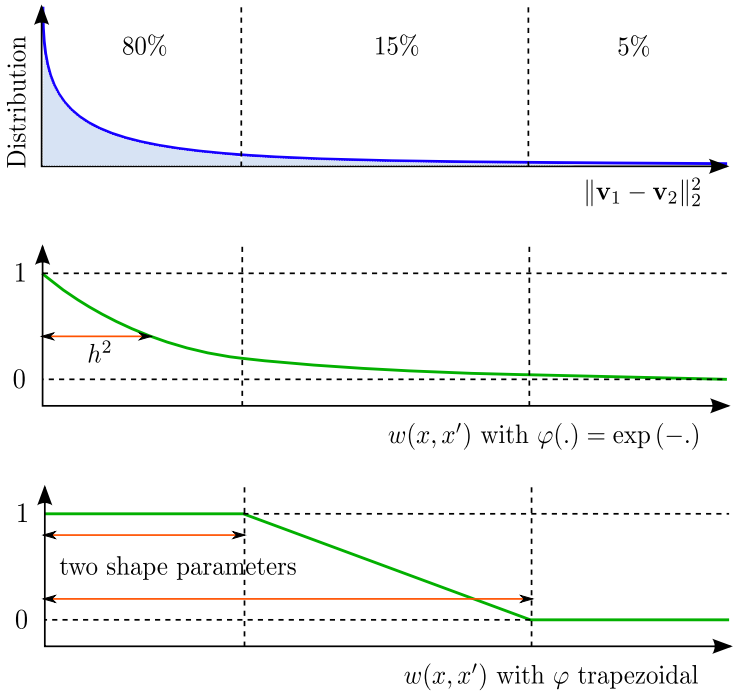


Figure 2.8: (top) χ^2 distribution of the Euclidean distance when the two compared patches share a common noise-free patch. The quantiles of the χ^2 distribution associate to each threshold a probability of false alarm: here 0.20 and 0.05. (middle) According to this probability of false alarm, we can set the bandwidth parameter h of the exponential kernel or (bottom) the two parameters of the trapezoidal kernel.

proposed in the literature. The simplest and most common one is to set a single h for the whole image, whose value is determined by cross validation on a small dataset of images. In (Polzehl and Spokoiny, 2006b; Kervrann and Boulanger, 2006), the authors set this parameter according to the quantiles of a χ^2 distribution, due to the particular metric they consider to compare pixels (or patches) (see Fig. 2.8). Van De Ville and Kocher (2009) calculate an unbiased risk estimate of the NL means to globally select the bandwidth h. Pursuing this idea, Doré and Cheriet (2009); Duval et al. (2011) consider a method based on the same approach but to locally select the bandwidth parameter (see Sec. 2.6).

The kernel φ: the function φ was chosen by Buades et al. (2005) as $t \mapsto \exp(-t)$, but other choices may be considered, such as compactly supported smooth functions. It was noticed by some authors (Goossens et al., 2008) that weights with compact support yield better results. Recent progress in non-local denoising have shown that flat kernels or trapezoidal kernels provide satisfying and competitive results with a lower computing cost than exponential kernels (Buades et al., 2009; Salmon and Strozecki, 2010). In this thesis we have conducted our experiments with the classical exponential kernel and with the trapezoidal kernel as defined in (Buades et al., 2009) (see Fig. 2.8).

The dissimilarity criterion d: this criterion was initially chosen by Buades et al. (2005) as a weighted Euclidean distance between the noisy patches $d(v(\mathrm{P}_{x'}), v(\mathrm{P}_x)) = \|v(\mathrm{P}_{x'}) - v(\mathrm{P}_x)\|_{2,a}$. This corresponds to the Euclidean norm convolved by a Gaussian kernel of bandwidth $a > 0$. The parameter a controls the concentration of the kernel around the central pixel. When $a \to 0$, only the central pixel is taken into account while when $a \to \infty$, all pixel values have the same influence on the Euclidean norm. In this thesis as in many other papers, we consider that the

(a) Noise-free image (b) Noisy image (c) Good parameters

(d) Search window size |W| (e) Patch size |P| (f) Bandwidth h

Figure 2.9: Influence of the three main parameters of the NL means on the solution.

NL means filter uses a standard Euclidean norm leading to several simplifications and to one less parameter.

When the noise level is high, many authors show that results can be significantly improved by refining weights thanks to the similarity between pre-filtered patches or patches extracted from a pre-filtered image. This idea is at the heart of iterative weighted estimation (Polzehl and Spokoiny, 2006a) or the iterative NL means proposed in (Brox et al., 2008; Goossens et al., 2008). In a first step, Dabov et al. (2007) select patches according to the Euclidean distance between their thresholded wavelet coefficients (i.e., between smoothed patches) and, in a second step, between the patches extracted from the estimation obtained in the first step. In (Azzabou et al., 2007a; Tasdizen, 2008; Orchard et al., 2008; Van De Ville and Kocher, 2011), the authors suggest, for acceleration purpose, computing the Euclidean distance between the projection of

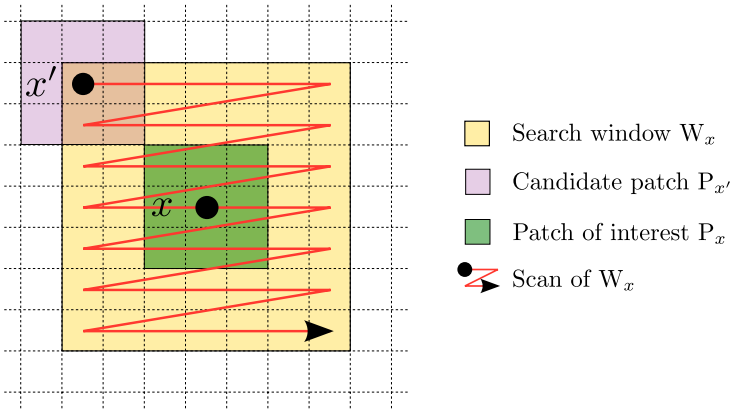


Figure 2.10: The NL means filter combines for each pixel x the noisy value of pixels x' according to the similarity between two patches P_x and $P_{x'}$ centered respectively around the sites x and x'. For complexity reason, the pixels x' are limited to a search window W_x centered around the pixel x.

the patches on their first principal axes (in the sense of the principal component analysis or PCA). If noise is assumed to leave in the span of the axes of small variations, this is equivalent to compare filtered version of the patches. Finally, Louchet and Moisan (2011) compute the Euclidean distance between patches regularized by an adaptive total-variation minimization.

Few authors suggest using the non-Euclidean distance between noisy or pre-filtered patches. After studying a few adaptations of the NL means to non-Gaussian noise in Chap. 3, dissimilarity criteria to compare noisy or pre-filtered patches in the case of non-Gaussian noise will be the main topic of Chap. 4 and 5.

Central weight correction: By using Eq. (2.43), the weight attributed to the pixel candidate $x' = x$ is pre-dominant compared to the other weights in the search window. This would lead to a strong residual noise in the resulting image. To cope with this difficulty, Buades et al. (2009) suggest replacing the weight attributed to the pixel of interest by the maximal weight in the search window:

$$w(x,x) = \max_{x' \in W_x} w(x,x') \; . \tag{2.44}$$

While the noise realizations between pairs of pixel values in $v(P_{x'})$ and $v(P_x)$ are assumed to be decorrelated, it is no more the case when $x = x'$. Doré and Cheriet (2009) and Salmon (2010) suggest then considering that the two patches are two independent noisy realizations $v(P_x)$ and $v'(P_x)$ obtained from the same noise-free patch. Under Gaussian noise assumption, the expectation of this Euclidean distance is $\mathbb{E}\left[\|V(P_x) - V'(P_x)\|_2^2\right] = 2|P|\sigma^2$, leading to the following weight redefinition:

$$w(x,x) = \varphi\left(\frac{\sigma^2}{h^2}\right) \; . \tag{2.45}$$

Doré and Cheriet (2009) propose a solution lying between both solutions and given by:

$$w(x,x) = \max\left[\max_{x' \in W_x} w(x,x'), \varphi\left(\frac{\sigma^2}{h^2}\right)\right] \; . \tag{2.46}$$

In the following, we use the solution (2.45) since it leads to a simpler implementation. Moreover, with this solution, if φ is differentiable almost everywhere, the resulting estimator is also differentiable almost everywhere which will be a property of interest in the following. However, when

noise departs form Gaussian noise, the solution (2.45) does not hold anymore. For some noise distributions and similarity criteria d, the same methodology can be applied. In other cases, we will use the solution (2.44).

Figure 2.10 illustrates the procedure. At each pixel x, the pixels x' are inspected sequentially to produce a weight by comparing the two noisy patches P_x and $\mathrm{P}_{x'}$. Once all weights $w(x,x')$ are computed, an estimate is obtained by the weighted average. Note that for complexity reasons, the pixels x' are restricted to a large window W_x centered around the pixel x.

Related approaches: Although the non-locality principle is intimately related to patch-models, it has been used previously for denoising purposes before the introduction of the *patch* concept (Lee, 1983; Yaroslavsky, 1985; Smith and Brady, 1997; Tomasi and Manduchi, 1998). The sigma filter and the Yaroslavsky filter (Lee, 1983; Yaroslavsky, 1985) can be interpreted as a degenerated version of the NL means where weights are expressed from the comparison of noisy values instead of noisy patches, i.e., the patches are restricted to a single pixel: $p = 1$ and $\mathrm{P}_x = \{x\}$. These approaches already preserve sharp information well. However, under high levels of noise, the comparison of noisy values becomes non robust leading to weights with high variations and to a filtered image with high residual noise (see Fig. 2.9 for a low patch size). If such a non-local model leads to an estimator with high variance, it is probably because its underlying assumption is too flexible: noise-free images are only assumed to be composed of many redundant values. The SUSAN filter (Smith and Brady, 1997) and the bilateral filter (Tomasi and Manduchi, 1998) suggest then restricting the average to similar values with close spatial positions:

$$w(x,x') = \exp\left(-\frac{\|x'-x\|^2}{2h_1^2}\right)\exp\left(-\frac{(v(x')-v(x))^2}{2h_2^2}\right) \tag{2.47}$$

where h_1 and h_2 are two bandwidth parameters controlling respectively the influence of both terms. By exploiting the patch redundancy, the NL means filter outperforms all these previous approaches.

2.5.2 Non-linear filtering in the patch space

The NL means filter is considered as non-local since pixels far apart in the spatial domain can be possibly combined together. However, as noted by Tschumperlé and Brun (2011), if these selected pixels are far apart in the spatial domain, they are neighbors in the patch domain. One can rather proceed to linear denoising in the patch domain instead of in the spatial domain. Such an extension was already proposed in the seminal paper of the NL means by Buades et al. (2005) and referred to as the blockwise NL means. It simply consists of estimating noise-free patches instead of noise-free values:

$$\hat{\boldsymbol{u}}(x) = \frac{\sum_{x'} w(x,x')\boldsymbol{v}(x')}{\sum_{x'} w(x,x')} \tag{2.48}$$

where $\hat{\boldsymbol{u}}(x)$ is an estimate of $\boldsymbol{u}(x)$, $\boldsymbol{v}(x)$ is the noisy patch extracted at location x and $w(x,x')$ are the weights defined in Eq. (2.43). Compared to Eq. (2.20), it clearly appears that the blockwise NL means filter is a linear Gaussian filter acting in the patch space. Since linear filtering is known to damage sharp discontinuities, many authors suggest performing non-linear filtering in the patch space to preserve the structure of the underlying manifold. Let us mention three of these approaches:

Anisotropic diffusion PDE' in the patch space: Based on the initial work of (Tschumperlé and Deriche, 2005), Tschumperlé and Brun (2011) propose to extend the blockwise NL means with

anisotropic diffusion PDE expressed in the patch space:

$$\frac{\partial \boldsymbol{u}(x)}{\partial t} = trace\left(\mathbf{D}(x)\mathbf{H}(x)\right) \tag{2.49}$$

where $\mathbf{H}(x)$ is the Hessian matrix at site x and $\mathbf{D}$ is a spatially variant field of diffusion tensors defined in the patch space.

Non-local graphs based on patch similarities: Non-linear regularization can be performed inside the patch space. Unlike total-variation formulation (see Sec. 2.3.3), such approaches penalize the variations or the transitions in the patch space instead of in the image domain. Variations are measured in the neighborhood of patches (i.e., in the non-local neighborhood) rather than in spatial neighborhoods. Kindermann et al. (2005) first introduce a non-local regularization term based on the similarity of noisy patches and defined as:

$$\sum_{x}\sum_{x'\in W_x}\left(1-\exp\left(-\frac{\|\boldsymbol{u}(x)-\boldsymbol{u}(x')\|^2}{h^2}\right)\right) . \tag{2.50}$$

Intuitively, the minimization of this variational problem leads to solutions verifying (2.40). Next, Gilboa and Osher (2007) propose a similar but convex formulation with quadratic penalty based on a weighted graph able to model, among others, the non-local interactions. Instead of minimizing a quadratic penalty, the solution proposed by Peyré et al. (2008) minimizes the total-variation on a non-local graph leading to the following formulation based on discrete derivatives using the graph gradient operator:

$$\sum_{x}|\nabla^w u(x)| \quad \text{where} \quad |\nabla^w u(x)| = \sqrt{\sum_{x'} w(x,x')(u(x)-u(x'))^2} , \tag{2.51}$$

where $w(x,x')$ follows the patch-based weights definition of the NL means (see Eq. (2.43)). Other energies based on the concept of non-local regularization have been proposed, for instance, in (Mignotte, 2008; Elmoataz et al., 2008; Bougleux et al., 2009; Zhang et al., 2010b). These approaches are especially interesting for image restoration purposes.

Patch dictionary learning with *grouped*-sparsity: Aharon et al. (2006) suggest using a sparse decomposition with a learned redundant dictionary of patches, leading to the following minimization problem:

$$\min_{\boldsymbol{D},\mathbf{A}}\sum_{x}\|\boldsymbol{D}\mathbf{A}(x)-\boldsymbol{v}(x)\|_2^2 \quad \text{subject to} \quad \sum_{x}\|\mathbf{A}(x)\|_0 < \varepsilon . \tag{2.52}$$

where $\boldsymbol{D}$ is a $|\mathrm{P}|\times K$ dictionary of K patches and $\mathbf{A}$ is a $K\times N$ matrix in which each column $\mathbf{A}(x)$ is the sparse representation of the patch located at pixel position x. Such decomposition can be obtained by an iterative two stage algorithm similar to the K-Means (see Sec. 2.4.2). Mairal et al. (2009) noticed that "this procedure implicitly assumes that the patches are *independent* from each other, which is questionable since they overlap". As a consequence, this leads to the following paradox: it is possible that similar patches can have very different sparse representations. As a consequence, the resulting estimator presents high variabilities. The non-local sparse model (NLSM) of Mairal et al. (2009) is an extension of this model using an $\ell_{p,q}$ structural pseudo-norm instead of the ℓ_0 pseudo-norm. The $\ell_{p,q}$ structural pseudo-norm is defined by:

$$\|\mathbf{A}\|_{p,q} = \sum_{k}\|\mathbf{A}^k\|_q^p \tag{2.53}$$

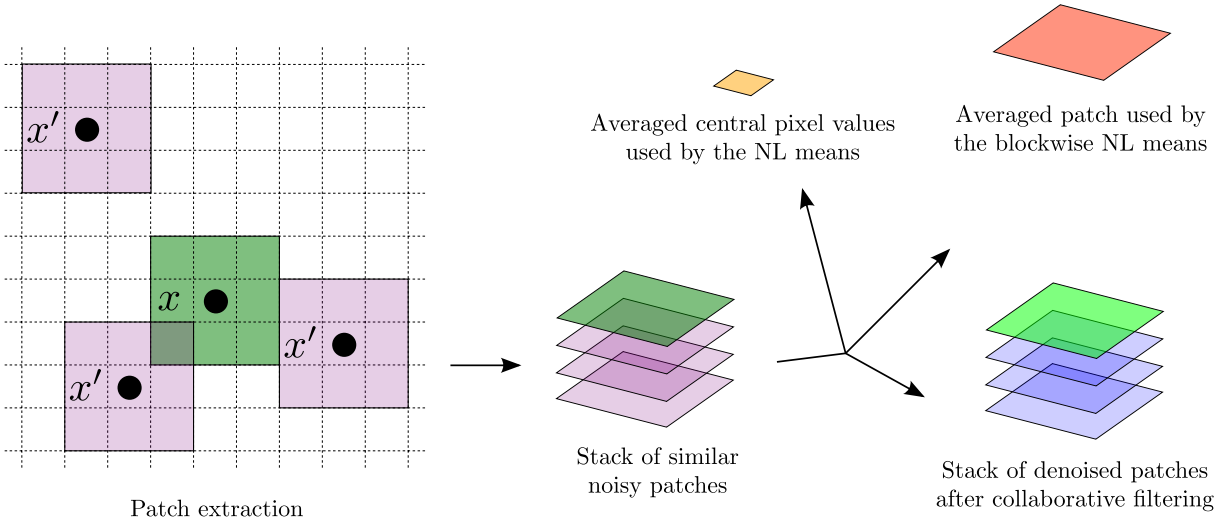


Figure 2.11: Illustration of collaborative filtering compared to the concept of the NL means and the blockwise NL means. The NL means filter averages the central pixel values of similar patches, while the blockwise NL means filter averages all the patches. The collaborative filtering does not average patches, but realizes instead a joint filtering of the selected patches leading to a denoised 3D block.

where $\mathbf{A}^k$ are the rows of the matrix $\mathbf{A}$. This structural norm imposes a *grouped*-sparsity of the patches. For instance, the $\ell_{0,\infty}$ pseudo-norm counts the number of rows with at least one non-zero entry. The NLSM minimizes Eq. (2.52) with the structural norm (2.53) on the set of similar patches following the idea of Buades et al. (2005). Hence, by exploiting the self-similarity property of natural images, such penalties force similar patches to be decomposed on the same atoms solving then the above paradox. By mixing different structural-norms for learning and reconstruction and by using the concept of *collaborative filtering* introduced by Dabov et al. (2007), the NLMS achieves *state-of-the-art* performance.

2.5.3 Collaborative filtering

Collaborative filtering is the main idea of the block matching and 3D collaborative (BM3D) filter introduced by Dabov et al. (2007). At each pixel x, the authors suggest constructing a stack $\mathcal{S}_x$ of the most similar patches:

$$\mathcal{S}_x = \{x' \in \Omega \mid v(\mathrm{P}_{x'}) \simeq v(\mathrm{P}_x)\} \tag{2.54}$$

where $v(\mathrm{P}_x) \simeq v(\mathrm{P}_{x'})$ denotes that $v(\mathrm{P}_x)$ and $v(\mathrm{P}_{x'})$ are similar in the sense of a given similarity criterion. By ordering this stack by degree of similarity, it constitutes a 3D block denoted as $\boldsymbol{v}_{\mathcal{S}_x}$ and presenting strong redundancy. In this 3D block, such content may be described by an extremely sparse representation. Unlike the blockwise NL means, the patches are not averaged together to produce a single denoised patch but they are decomposed on a 3D cosine basis to provide a denoised 3D block $\hat{\boldsymbol{u}}_{\mathcal{S}_x}$ after Wiener thresholding of the 3D coefficients (see Sec. 2.4.1). Thanks to a two pass strategy and a few heuristics, this method has led to the BM3D filter known as one of the most powerful denoising methods. Recent techniques all rely on collaborative filtering either using a stack of local similar patches or by denoising clusters of patches obtained by a pre-classification or segmentation step (e.g., Mairal et al., 2009; Chatterjee and Milanfar, 2011).

The good performance reached by BM3D is essentially due to the collaborative filtering step which provides several estimates for each pixel of the image. Figure 2.11 illustrates the procedure:

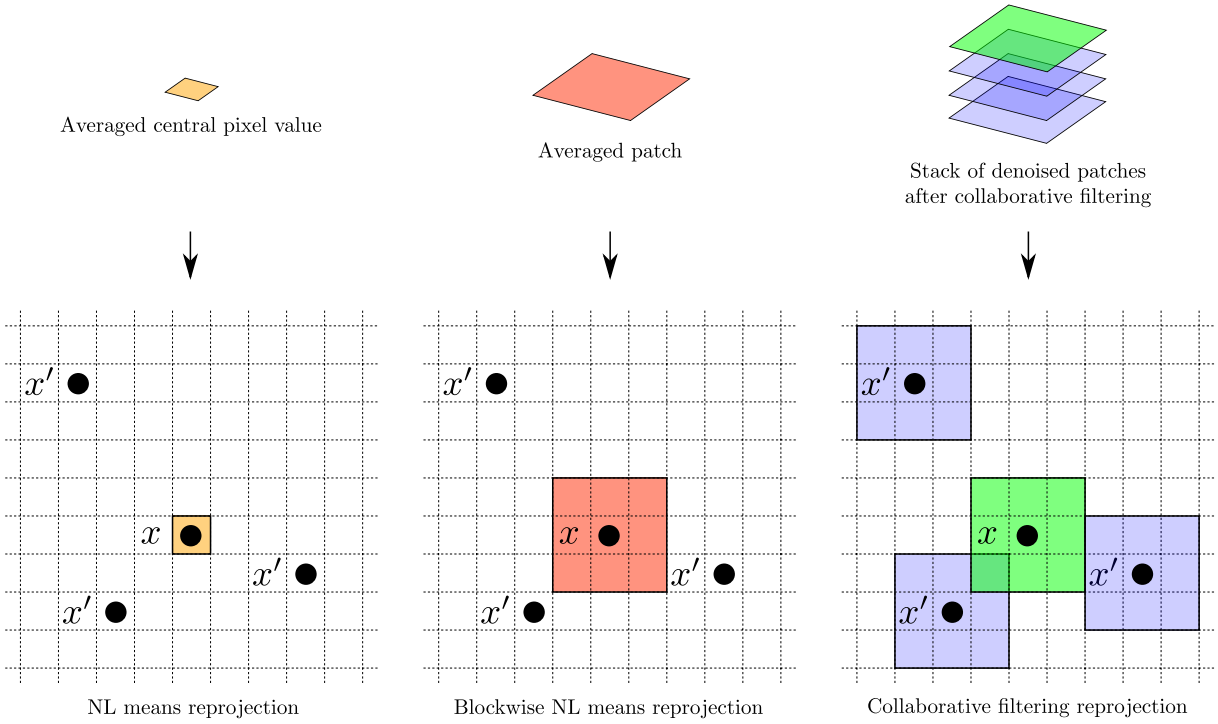


Figure 2.12: Illustration of the reprojection in the case of the NL means, the blockwise NL means and collaborative filtering. For each pixel x, the NL means filter provides one estimate for the pixel x only, the blockwise NL means filter provides an estimate for all pixels included in the patch of interest and collaborative filtering provides an estimate for all pixels included in patches similar to the patch of interest.

while the NL means filter averages the central pixel values of similar patches and the blockwise NL means filter averages all the patches, collaborative filtering does not average patches, but realizes instead a joint filtering of the selected patches leading to a denoised 3D block. As a consequence, since the 3D block processed at location x contains the selected patches located at location x': each patch is denoised several times with respect to its degree of redundancy. The combination of these estimates to produce a final image leads to an impressive bias-variance trade-off: this is the concept of *reprojection*.

2.5.4 From the patch domain to the image domain: the reprojection

After denoising in the patch space, a collection of denoised patches is available. By construction, the patches inherently overlap. As a consequence, there are $|\mathrm{P}|$ estimates available for each pixel value $u(x)$: in each of the denoised patches located at positions x' such that $x' \in \mathrm{P}_x$. Moreover, when collaborative filtering is used, an estimate is also available each time x appears in a patch $\mathrm{P}_{x'}$ selected in a 3D block located at a position x'', i.e., when $x \in \mathrm{P}_{x'}$ and $x' \in \mathcal{S}_{x''}$. The construction of an image $\hat{\boldsymbol{u}}$ from this set of denoised version is called the *reprojection* of patches in the image domain. Figure 2.12 gives an illustration of the different strategies of reprojection performed by the NL means, the blockwise NL means and after collaborative filtering.

The naïve and simple reprojection performed by the blockwise NL means consists of uniformly averaging all available estimates for each pixel x. Dabov et al. (2007) and Salmon and Strozecki (2010) suggest instead combining these estimates using a weighted average driven by the variance associated with each estimator. Reprojection can be seen as a more general problem of combining several estimators. The next section describes several aggregation-based filters used in the general

context of image denoising, essentially, for adaptive window or patch-based denoising.

2.6 Aggregation-based filtering: a combination of *priors*

The aggregation of the responses of several pre-estimates in order to provide an improved response is a typical topic of statistics. The underlying assumption is that among the available pre-estimates, one was built on the suitable *prior*. Usually, the veracity of a *prior* is space varying and then a pre-estimate can be preferable in one region of the image while another pre-estimate should be selected in another region. Aggregation procedures are mostly designed to take local decisions. Among K pre-estimates $\overline{\boldsymbol{u}}_k$ of an unknown information $\boldsymbol{u}$, such methods try to locally select the most relevant one or a combination of them to produce an optimal solution $\hat{\boldsymbol{u}}$.

Aggregations aim at minimizing the bias-variance trade-off, i.e., the mean square error (also called the quadratic risk or, in short, the risk in this context). Selective aggregation locally selects the pre-estimate minimizing the risk:

$$\hat{u}(x) = \underset{\overline{u}_k(x)}{\arg\min}\, R[\overline{U}_k(x)] \tag{2.55}$$

$$\text{where} \quad R[\overline{U}_k(x)] = \mathbb{E}\|\overline{U}_k(x) - u(x)\|_2^2 \,. \tag{2.56}$$

This strategy leads to brutal transitions in the resulting image due to the change of decisions between neighboring pixels. To limit this effect, a more elaborate solution is to locally search for a linear aggregation:

$$\hat{u}(x) = \sum_k \beta_k(x)\overline{u}_k(x) \quad \text{where} \quad (\beta_1(x), \ldots, \beta_K(x)) = \underset{(\beta_1(x),\ldots,\beta_K(x))}{\arg\min} R\left[\sum_k \beta_k(x)\overline{U}_k(x)\right] . \tag{2.57}$$

Different ways of combining the estimators may depend on the theoretical aggregation problem we aim to solve as described by Nemirovski (2000) and Tsybakov (2003). In all cases, the MSE requires the knowledge of the noise-free image $\boldsymbol{u}$ but can still be estimated, for large classes of noise distributions, from the noisy image $\boldsymbol{v}$ alone (or up to an additive constant). In the following, we will present different strategies minimizing an estimation of the mean square error or searching for a bias-variance trade-off from the noisy image $\boldsymbol{v}$ and the pre-estimates $\overline{\boldsymbol{u}}_k$ only.

2.6.1 Maximum selection or variance minimization

Let us consider that we have several pre-estimates $\overline{\boldsymbol{u}}_k$ with a known local residual variance. The crudest approach for the aggregation of these estimators is to select the pre-estimate minimizing the local variance:

$$\hat{u}(x) = \underset{\overline{u}_k(x)}{\arg\min}\, \mathrm{Var}[\overline{U}_k(x)] \,. \tag{2.58}$$

This is equivalent to selecting the pre-estimate that minimizes the MSE by assuming that all the pre-estimates are unbiased, i.e., $R[\overline{U}_k(x)] = \mathrm{Var}[\overline{U}_k(x)]$. When estimates come from a weighted average, the residual local variance is given by:

$$\mathrm{Var}[\overline{U}_k(x)] = \frac{\sum w_k(x,x')^2\, \mathrm{Var}[V(x')]}{\left[\sum w_k(x,x')\right]^2} \,, \tag{2.59}$$

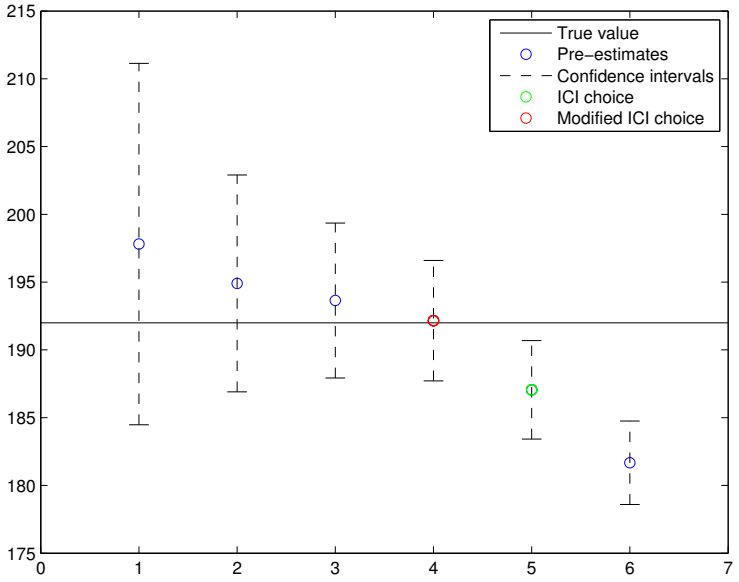


Figure 2.13: Example of the ICI rule (or Lepski method) and the rule selection proposed by Kervrann and Boulanger (2006). The estimates are ordered by their variance. The four first estimates are included in the confidence interval of their predecessor. The first five confidence intervals have non-null intersections. The rule of Kervrann and Boulanger (2006) selects the fifth estimate since it is outside the confidence interval of its predecessor. The ICI rule selects the sixth confidence interval since it does not overlap with all previous intervals.

assuming that weights are independent of the data. Moreover, if the candidate pixel values are independent and identically distributed (i.i.d.) and the weights are binary, it results in:

$$\mathrm{Var}[\overline{U}(x)] = \frac{\mathrm{Var}[V(x)]}{L} \tag{2.60}$$

where L is the number of non-zero weights in the search window. Such approaches choose the estimator which selects the maximum of candidate pixels.

The extension of this approach to linear selection is to find the convex combination minimizing the variance, i.e., the MSE under the assumption that the pre-estimates $\overline{u}_k(x)$ are unbiased and decorrelated:

$$\hat{u}(x) = \min_{\beta_1(x),\ldots,\beta_K(x)} \mathrm{Var}\left[\sum_k \beta_k(x)\overline{U}_k(x)\right] . \tag{2.61}$$

Salmon and Strozecki (2010) showed that for any uncorrelated noise, using the constraint that $\sum_k \beta_k(x) = 1$, the first order condition of the dual Lagrangian problem gives us the following solution:

$$\beta_k(x) = \frac{\mathrm{Var}[\overline{U}_k(x)]^{-1}}{\sum_l \mathrm{Var}[\overline{U}_l(x)]^{-1}} . \tag{2.62}$$

This last strategy leads to smoother aggregation. This solution has been used in (Dabov et al., 2007; Salmon and Strozecki, 2010) for the reprojection of denoised patches. When pre-estimates are biased, the violation of the underlying assumption leads to a bias in the solution. Better strategies can be designed by considering the bias of each estimator.

2.6.2 Nested estimators and the ICI rule

The applications of Lepski's method (Lepski et al., 1997) to kernel smoothing are also known to perform well for denoising and is known in signal processing as the ICI (Intersection of Confidence Intervals) rule (see for instance Katkovnik, 1999; Katkovnik et al., 2002). Assume that we have at each pixel position x a family of K *nested* pre-estimates $\overline{u}_k(x)$, i.e., so that we can order them by their expected variance:

$$\mathrm{Var}(\overline{U}_1(x)) > \mathrm{Var}(\overline{U}_2(x)) > \ldots > \mathrm{Var}(\overline{U}_K(x)) \; . \tag{2.63}$$

The ICI rule starts by building confidence intervals $\mathcal{I}_k(x)$ of the form:

$$\mathcal{I}_k(x) = \left[\overline{u}_k(x) - \gamma \sqrt{\mathrm{Var}[\overline{U}_k(x)]}, \overline{u}_k(x) + \gamma \sqrt{\mathrm{Var}[\overline{U}_k(x)]} \right] \tag{2.64}$$

where γ is the threshold of the confidence intervals. Under particular assumptions on $\overline{\boldsymbol{u}}$, for instance, if the residual noise in $\overline{\boldsymbol{u}}$ is Gaussian with known variance, the value of γ controls the probability that $u(x) \in \mathcal{I}_k(x)$.

Given a value of γ, the ICI rule selects at each pixel position x the estimate with the smaller variance belonging to all previous confidence intervals:

$$\hat{u}(x) = \overline{u}_{k^*}(x) \quad \text{such that} \quad k^* = \sup\{k = 1, \ldots, K \mid \forall l < k \; , \; \bigcup_{l \leq k} \mathcal{I}_l(x) \neq \emptyset\} \; . \tag{2.65}$$

Kervrann and Boulanger (2006) suggest using a slightly different selection rule given by:

$$\hat{u}(x) = \overline{u}_{k^*}(x) \quad \text{such that} \quad k^* = \sup\{k = 1, \ldots, K \mid \overline{u}_k(x) \in \mathcal{I}_l(x)\} \; . \tag{2.66}$$

Since the bias of the pre-estimates usually increases when the variance decreases, such selection procedures lead to minimize the mean square error. Figure 2.13 gives an illustration of the ICI rule. The estimates are ordered by their variance. The first four estimates are included in the confidence interval of their predecessor. The first five confidence intervals have non-null intersections. The fifth estimate is outside the confidence interval of its predecessor and the sixth confidence interval does not overlap with all previous intervals.

In practice, such rules have successfully been applied to image denoising with adaptive weights smoothing (Polzehl and Spokoiny, 2000), shape adaptive windows (Katkovnik et al., 2004), adaptive selection of window sizes (Bioucas-Dias et al., 2008) or non-local filtering with spatially varying search windows (Kervrann and Boulanger, 2006). All these methods assume that the choice of weights/windows are independent of the data, and thus rely on the definition of the residual local variance given by Eq. (2.59). The drawback of such approaches is that the resulting image presents a residual noise looking like a small impulse noise.

2.6.3 Mallows' C_p statistic

The application of Mallows' C_p statistic to image denoising (Mallows, 1973) leads to estimating the mean square error at the pixel position x using an estimate of the local bias and the local variance in x. This estimate of the risk is defined as follows:

$$\hat{R}[\overline{u}(x)] = \sigma^2 C_p(\overline{u}(x)) \triangleq \|v(x) - \overline{u}(x)\|_2^2 - D\sigma^2 + 2p \quad \text{where} \quad p = \mathrm{Var}[\overline{U}(x)] \; . \tag{2.67}$$

where D is the dimension of the data. Under the Gaussian noise assumption, the value p corresponds to the number of regressors averaged together to estimate $u(x)$, i.e., the degree of freedom in the regression. Moreover if the estimates $\overline{u}(x_p)$ are independent of the noisy

data $v(x_p)$, Mallows' C_p statistic provides an unbiased estimate of the risk, i.e., $\mathbb{E}[\sigma^2 C_p(x)] = \mathbb{E}\|v(x) - \overline{u}(x)\|_2^2$. The variance of such an estimator is better as the data dimension D is large.

In the context of the NL means, Doré and Cheriet (2009) use Mallows' C_p statistic in order to select the best bandwidth parameter h. Since grey scale images have a dimension $D = 1$, they estimate the bias between the noisy values and the estimated values located inside the patch P_x surrounding the pixel of interest x. This allows us to decrease the variance of the MSE estimation. As noted by the author, the use of Mallows' C_p statistic requires great care due to the underlying independence assumption. In the case of the NL means, that requires eliminating the pixel of interest from the set of pixel candidates, i.e., the central pixel is removed from the search window. This modification necessarily leads to a loss of punctual targets. The authors estimate the residual local noise variance by using Eq. (2.59) under the assumption that weights are independent of the data. In the case of the Gaussian noise assumption, the resulting Mallows' C_p statistic is given by:

$$\hat{R}[\overline{u}(x)] = \|v(\mathrm{P}_x) - \overline{u}(\mathrm{P}_x)\|_2^2 - |\mathrm{P}|D\sigma^2 + 2\sigma^2 \frac{\sum w(x,x')^2}{[\sum w(x,x')]^2} \,. \tag{2.68}$$

In the next section, we will introduce the Stein unbiased risk estimator (SURE) which, as well as Mallows' C_p statistic, is an estimator of the risk. However, unlike Mallows' C_p statistic, in the case of image denoising, SURE does not rely on a potentially biased estimator of the variance nor does it require that the estimates are uncorrelated to the noisy data.

2.6.4 Stein unbiased risk estimator (SURE)

Stein unbiased risk estimator (SURE) is an estimator of the mean square error which does not require the knowledge of $\boldsymbol{u}$. Let $h(.)$ be an estimator of the noise-free image from a given noisy image, such that $\overline{\boldsymbol{u}} = h(\boldsymbol{v})$. Under Gaussian noise assumption, SURE provides, for all pixels x, an estimator of the MSE (Stein, 1973, 1981) defined by:

$$\hat{R}[\overline{u}(x)] = \|v(x) - \overline{u}(x)\|_2^2 - D\sigma^2 + 2D\sigma^2 \left.\frac{\partial \overline{u}(x)}{\partial v(x)}\right|_{v(x)} \tag{2.69}$$

Compared to Mallows' C_p statistic, such an estimate only requires satisfying the following reasonable relations:

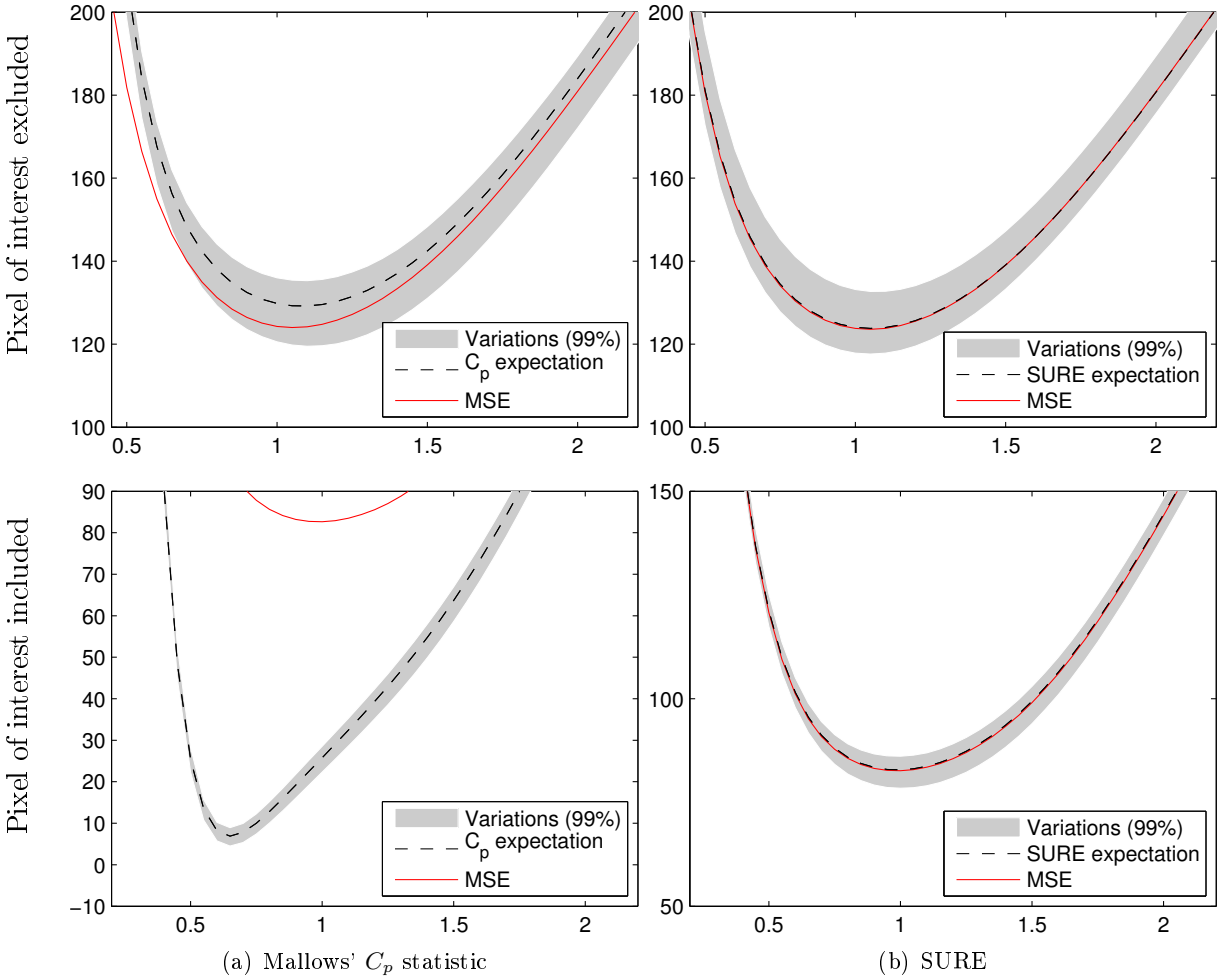
$$\text{i) } lim_{|z|\to\infty} h(u(x) + z)e^{-\frac{z^2}{2\sigma^2}} = 0 \,, \tag{2.70}$$

$$\text{ii) } \mathbb{E}(h(V(x)))^2 < +\infty \,, \text{and} \tag{2.71}$$

$$\text{iii) } \mathbb{E}|h'(V(x))| < +\infty \,. \tag{2.72}$$

where $h'(V(x)) = \left.\frac{\partial h(v(x))}{\partial v(x)}\right|_{V(x)}$. Like Mallows' C_p statistic, the variance of such an estimator is also as good as the data dimension D is large.

Applications of SURE emerged for choosing the smoothing parameter in families of linear estimates (Li, 1985) such as for model selection, ridge regression, smoothing splines, etc. It was then widely used in the wavelet community after the introduction of the SURE-Shrink algorithm (Donoho and Johnstone, 1995). Solo (1996) gave a general form of SURE for an estimator defined as a minimizer of regular energy, especially for least square regression regularized by a Sobolev norm or the total-variation. More recently, Benazza-Benyahia and Pesquet (2005) use Stein's principle for the denoising of multichannel images. Linear combinations of estimates based on SURE were considered (Blu and Luisier, 2007) instead of the selection of a single one. Moreover,



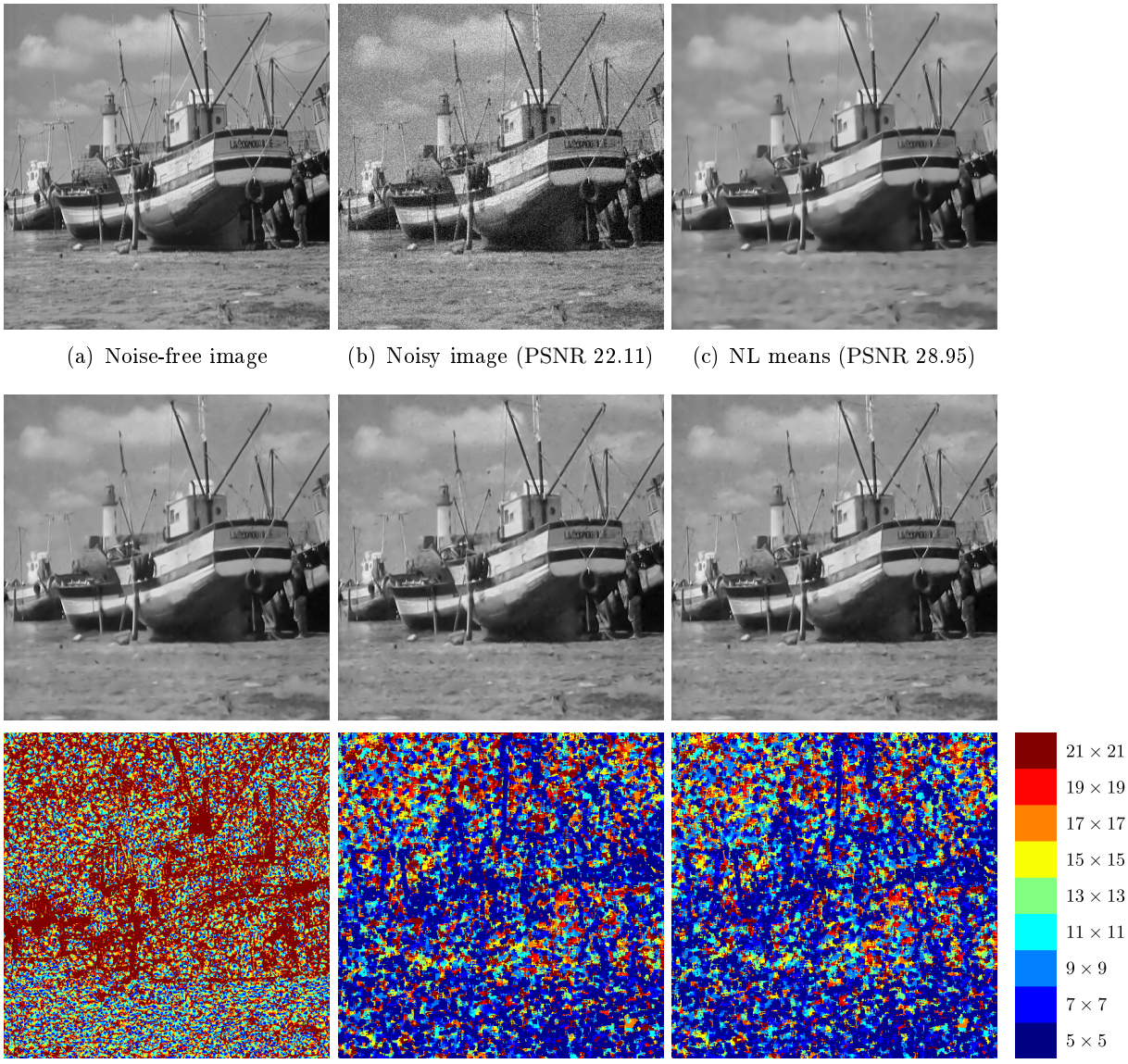
(a) Mallows' C_p statistic (b) SURE

Figure 2.14: Estimation of the MSE of the NL means with respect to the bandwidth parameter h using (a) Mallows' C_p statistic and (b) SURE. First row: the pixel of interest is excluded from the search window. Second row: the pixel of interest is included to the search window. When the pixel of interest is included the denoising performance reaches a lower MSE well estimated by SURE while underestimated by Mallows' C_p statistic.

Ramani et al. (2008) have described a Monte Carlo approach to evaluate SURE when a closed-form expression is not available or too computer-intensive. In the context of the NL means, Van De Ville and Kocher (2009); Van De Ville and Kocher (2011) search for the global bandwidth parameter minimizing the MSE of the resulting image. In this case, due to the law of large numbers, the global SURE approximates well the global MSE. Duval et al. (2011) extended this approach to estimate the local risk in view of setting a local bandwidth parameter.

The limitation of Mallows' C_p statistic and SURE is that they have an extremely large local variance inversely proportional to the dimension D of the image. This variance arises from the term of the square residue (i.e., the square of the method noise): $\|v(x)-\overline{u}(x)\|_2^2$. Under Gaussian noise assumption, this term has a variance of the same order as σ^4. To decrease its variance, we have seen that Doré and Cheriet (2009) evaluate this term on patches. Duval et al. (2011) suggest instead convolving the risk map assuming ergodicity of the risk.

In Figure 2.15, we have compared the local selective aggregations based on ICI rule, Mallows' C_p statistic and SURE. For the two last strategies, the aggregation is based on the convolution



(a) Noise-free image (b) Noisy image (PSNR 22.11) (c) NL means (PSNR 28.95)

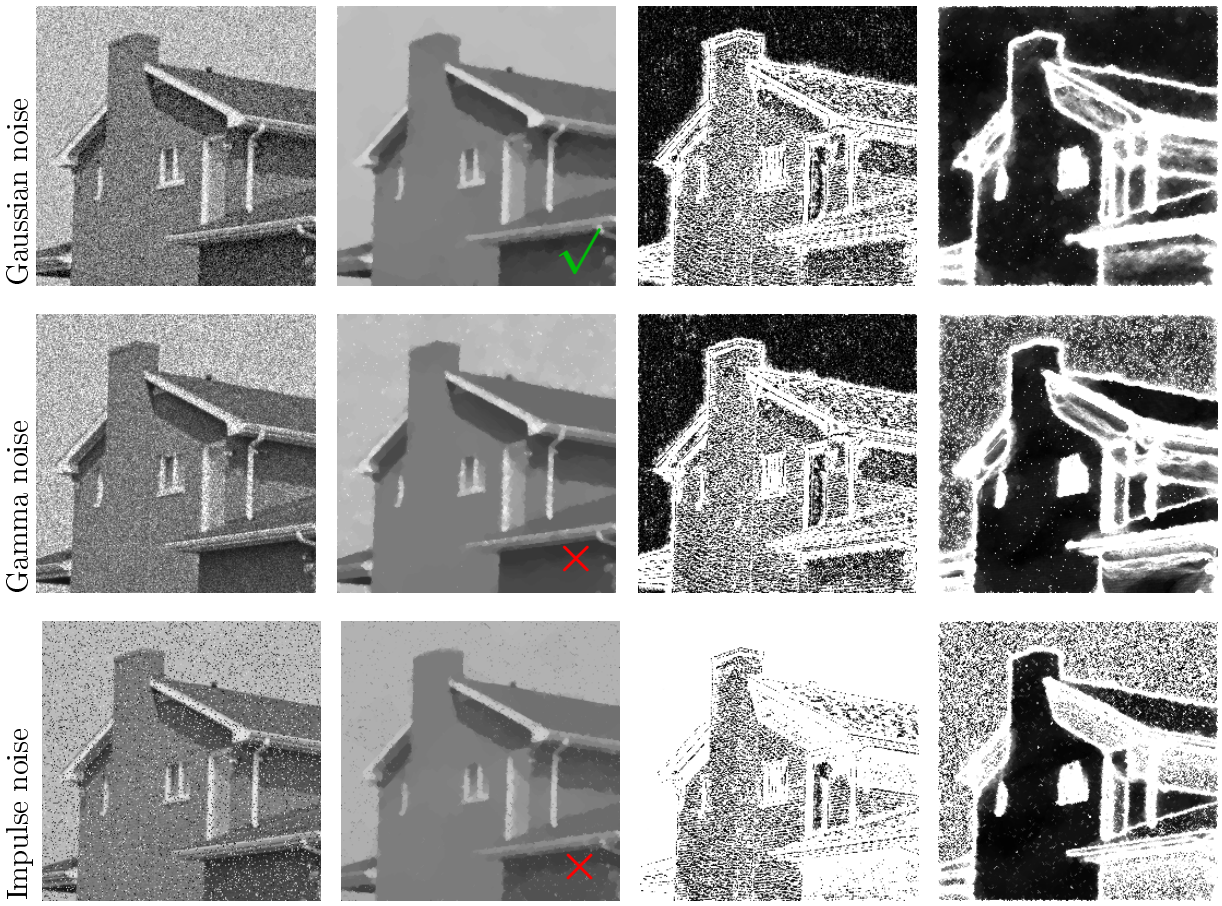
(d) ICI rule (PSNR 29.23) (e) Conv. C_p (PSNR 29.84) (f) Conv. SURE (PSNR 29.85)

Figure 2.15: Selective aggregation between different results of the NL means using nine different sizes for the search window. We consider a local selection based on the ICI rule, the convolution of the map of Mallows' C_p statistic and the convolution SURE map. The aggregation based on SURE provides the best result in this case.

of the local risk map as done in (Duval et al., 2011). In this case, SURE provides the best aggregation result.

2.6.5 Exponential weighted aggregation (EWA)

Mallows' C_p statistic and SURE aim to solve the selective aggregation problem given in Eq. (2.55). However, we have mentioned that it might be better to combine several estimators rather than just selecting one. In particular, it happens to be effective if the best estimators (in term of evaluated risk) are diversified enough or if the risk was wrongly estimated. When an estimate of the risk map is available, a convex aggregation can be obtained using the statistical method



(a) Noisy image (b) TV minimization (c) Bias^2 (d) Relative variance

Figure 2.16: Application of the ROF model on images damaged from (top to bottom) Gaussian, gamma and impulse noise. (a) The noisy image, (b) the filtered image, (c) the local resulting bias and (d) the local *relative variance*. The ROF model only applies well for Gaussian noise: the bias tends towards zero and the variance is equally reduced in all homogeneous areas.

of exponentially weighted aggregation (EWA) as introduced by Leung and Barron (2006). This method has been theoretically studied in (Dalalyan and Tsybakov, 2008) and adapted for patch-based denoising in (Salmon and Le Pennec, 2009). It consists in aggregating the estimators by performing a weighted average with weights based on the confidence attributed to each estimator, measured in term of the risk. More precisely:

$$\hat{u}(x) = \sum_{k=1}^{K} \beta_k \overline{u}_k(x)\,, \tag{2.73}$$

$$\text{with} \qquad \beta_k = \frac{\exp(-\hat{R}[\overline{u}_k(x)]/T)}{\sum_{l=1}^{K} \exp(-\hat{R}[\overline{u}_l(x)]/T)}\,.$$

The temperature parameter $T > 0$ is a smoothing parameter that controls the confidence attributed to the risk estimates. If $T \to \infty$, then the EWA is simply the average of the pre-estimate. Conversely, when $T \to 0$, then the EWA selects the pre-estimate minimizing the risk as discussed before. Most theoretical works about EWA (see Leung and Barron, 2006; Dalalyan and Tsybakov, 2008) recommend a large temperature parameter $T = 4\sigma^2$ under a few assumptions (like independence) on the estimators $\overline{u}_1(x), \cdots, \overline{u}_K(x)$. In practice, since assumptions on the estimators

family may not be satisfied, a smaller value is used, such as $T = 0.4\sigma^2$. Results combining EWA and SURE will be given in Chap. 7.

The current *state-of-the-art* techniques so far consist of a combination of the ideas of similar patches selection with that of sparsification by transforms or learned dictionaries or regularization followed by an aggregation step. We refer the reader to the recent survey by Katkovnik et al. (2010); Milanfar (2011) for a deeper analysis of the connections and evolutions of most of the denoising approaches we mentioned here. All these approaches rely on the Gaussian noise assumption. When noise departs from Gaussian noise, the direct application of these methods can lead to aberrations and results with poor quality.

2.7 Influence of the noise model on the denoising performance

One could think that when noise departs from the Gaussian distribution, the performance of the previously presented denoising techniques are not affected. In Figure 2.16, we use the ROF model[1] (see Sec. 2.3.3) on an image corrupted by Gaussian, gamma and impulse noise. The denoised images, the local bias and the local *relative variance* are given to assess the quality of the denoising (see Sec. 2.1.3). Under Gaussian noise, the filtered image presents the same smoothing effect everywhere in the image: the bias tends towards zero and the variance is equally well reduced in all homogeneous areas. However, under gamma noise, if the solution is still unbiased, the level of noise reduction is higher in bright areas than in dark areas. The filtering does not adapt to the *signal-dependent* nature of gamma noise. As a consequence, the filtering image present many bright residual pixels most of all in bright areas: the model does not take into account that the heavy tail of the gamma distributions is proportional to the signal itself. In the case of impulse noise, bright areas are biased towards lower values while dark areas are biased toward higher values. Visually, the resulting image suffers from a residual impulse noise with high variance.

2.8 Conclusion

We have reviewed several methods to solve the problem of noise reduction. These solutions are based on different models of the underlying signal and the assumption of AWGN. Among them, patch-based filters assume that the image is constituted of several redundant patterns. This *prior* appears as particularly efficient to model natural images and has then led to the *state-of-the-art* techniques. Such a *prior* is parameterized by the patch size the size of the search window and the way to compare patches. The objects in images have by nature different scales and orientations. By using different parameters, aggregation-based filters can be used to combine different *priors* and then adapt to the local content of the scene. Next, we have seen that the direct application of the classical algorithms on images damaged by non-Gaussian noise leads to visual aberrations: the estimation of the noise-free values can be biased and the amount of noise reduction does not adapt to the signal-dependent nature of the noise at hand. Classical approaches have then to be adapted or extended to the statistics of the noise that corrupts a given image. In the case of non-local filtering, non-Gaussian assumptions require us to redefine the combination of noisy samples that have to be selected by a proper similarity criterion. The next chapter focuses on these different aspects.

[1] we use anisotropic TV corresponding to the sum of the ℓ^1 norm of the gradient so that minimization problem can be solved by graph-cuts

Chapter 3

Image denoising beyond Gaussian noise

Image denoising under the classical additive white Gaussian noise (AWGN) assumption has led to multiple algorithms that have been reviewed in Chap. 2. We have seen that when noise departs from the Gaussian distribution, the quality of such algorithms can lead to a poor performance. *Speckle* and *shot* noise are common non-Gaussian phenomena that affect images. Unlike the Gaussian noise, the difficulty emerging is that such noises are *signal-dependent*: the level of the fluctuations spatially varies inside the images according to the intensity of the underlying noise-free image. When the variance depends on the expectation, we say that the noise is *heteroscedastic* (by opposition to the term *homoscedastic* when the variance is constant).

Most of approaches designed for non-Gaussian noise are based on the variance stabilization approach. This approach transforms the input noisy data in a way that the resulting output appears with a noise component approximatively Gaussian and with constant variance. In other words, it maps a *hetereoscedastic* to a *homoscedastic* noise. Hence, it is no necessary to design new algorithms: a common denoising algorithm designed for Gaussian noise can be used instead. If this approach is relevant to deal with multiplicative noise or Poisson noise, it can hardly be generalized to multi-modal or multi-variate distributions (e.g., is it possible to transform the multi-modal distribution of the interferometric phase to approach a Gaussian distribution?). Moreover, such transformations lead to distortions of the underlying signal which can be undesirable for specific algorithms. Finally, if in practice these methods can prove to be powerful, in a theoretical point of view, they are only based on heuristics. To our knowledge, their is no result stating that denoising after a variance stabilization is necessary and optimal to deal with non-Gaussian noise. Hence, in a methodological framework following the discussion in 1.1.3, it is relevant and legitimate to design denoising techniques that directly deal with images suffering from non-Gaussian noise.

Among the standard denoising techniques that we have enumerated in Chap. 2, we have found extensions to non-Gaussian noise for almost all of them. Sometimes, these extensions focus on a given noise distribution or on a family of noise distributions (usually the distributions of the exponential family) or they are general enough to cope with any distributions. Also, it is common that some authors directly propose an original methodology to deal with a given noise distribution without being the direct extension of a previously existing approach. All these methods are based on one or many concepts grounded on the properties of the noise distribution.

Our contributions— This chapter reviews a large number of concepts drawn for the design of denoising algorithms under non-Gaussian noise. We extract from a rich literature on this topic, the main approaches and concepts used for this purpose. We conduct original numerical experiments to show the performance or the drawbacks of some proposed techniques. We also extend methods designed for a specific kind of noise (in particular for *speckle*) to a more general framework based on the understanding of the underlying concepts. Among others: we provide

an example of minimization of the total-variation where the variance stabilization approach is non-optimal and we compare different common estimators under impulse noise. We also applies the common SURE-based extensions of non-Gaussian noise to moving average filters. We show in particular that such extensions fail in the case of gamma noise.

The main contribution of this chapter, is the comparative study of selection rules for different spatially adaptive local filters and non-local filters. In the sight of the proposed evaluation criteria in Sec. 2.1.3, we measure their performance in terms of their bias and their *relative variance*. We introduce a new selection rule based on the generalized likelihood ratio. This selection rule provides the best performance, and as a consequence, it will be at the heart of next chapters.

Organization of the chapter— We first provide in Sec. 3.1 a non-exhaustive enumeration of non-Gaussian noise models that can appear in imagery. Next, we describe in Sec. 3.2, the details of the variance stabilization principle and discuss its limitations and drawbacks. We next mention and organize several concepts extracted from different approaches (sometimes proposed by different communities) and show their links and their differences. Section 3.3 focuses on the extension of moving average filters based on the ergodicity assumption. We present in Sec. 3.4, anisotropic filters for non-Gaussian noise based on edge detectors robust to the noise distribution. We next study extensions of variational models including the total-variation minimization and the sparse decompositions in Sec. 3.5. They rely on a Bayesian framework where the noise distribution is taken into account to define a suitable data fidelity term. In Sec. 3.6, we review some extensions of aggregation-based filters. They are directly based on statistical properties of the noise distribution model at hand. In Sec. 3.7, we present the Bayesian NL means and its variants as extensions of the NL means filter and the UINTA filter for arbitrary noise distributions. They estimate the posterior mean by averaging the values of a pre-filtered image where weights are linked to the likelihood model. In Sec. 3.8, we study and compare selection-based filters that average noisy values and we propose a new one in Sec. 3.9.

3.1 The common noise models in imagery

In this section, we describe the most common models used in the image processing literature. We will see that some of them are just for *academic* purposes while others are used to model realistic data, such as astronomical images and ultrasound images. The common point of these models is that the noise is considered as spatially uncorrelated so that the pdf of $\boldsymbol{V}$ is the product over the image of the pdf of each random variable V_k modeling the observation v_k:

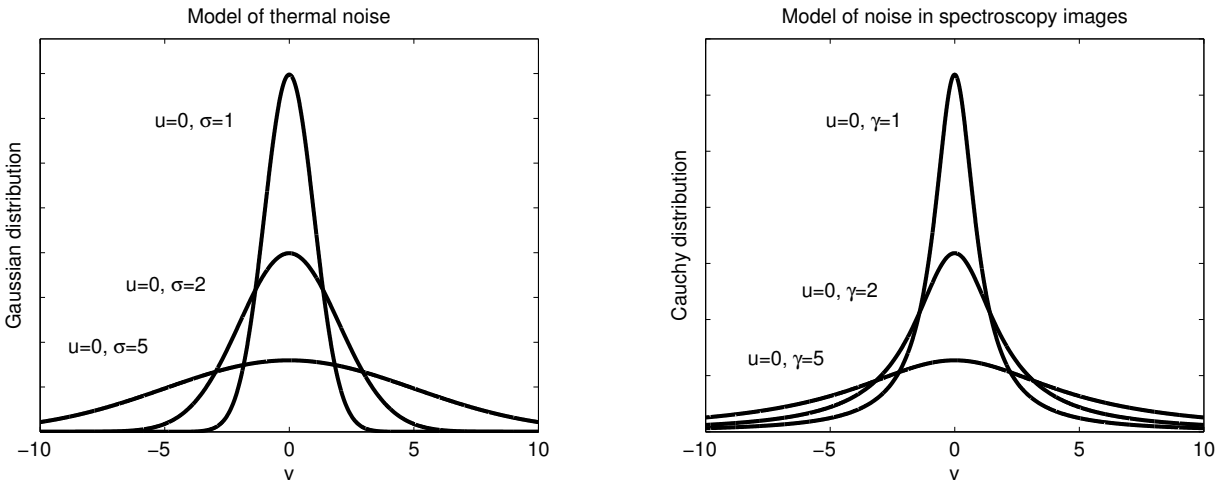
$$p(\boldsymbol{v}|\boldsymbol{u}) = \prod_{k=1}^{N} p(v_k|u_k) \; . \tag{3.1}$$

3.1.1 Gaussian noise

In optical imagery, one of the most common models is the additive white Gaussian noise (in short AWGN or Gaussian noise) model. Given a noise level modeling $\sigma > 0$, and the underlying real value u, a Gaussian random variable V is a real random variable following the pdf:

$$p(v|u) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(v-u)^2}{2\sigma^2}\right] \; , \tag{3.2}$$

with expectation $\mathbb{E}[V] = u$ and variance $\mathrm{Var}[V] = \sigma^2$. Figure 3.1 gives an illustration of Gaussian distributions. Gaussian fluctuations are additive, therefore it is straightforward to show that v can be decomposed as $u + \sigma\varepsilon$ with ε being a realization of a zero mean Gaussian



(a) (b)

Figure 3.1: Two distributions modeling symmetrical and additive noises. (a) Gaussian distributions centered on $u = 0$ with three scale parameters σ. It can represent errors of types $\pm\varepsilon$ and it usually describes the *thermal* noise. (b) Cauchy distributions centered on $u = 0$ with three different scale parameters γ. Due to its heavy tails, this distribution can represent large abberations as it occurs in spectroscopy.

random variable with unit standard-deviation decorrelated from u. Gaussian noise model is suitable for describing symmetrical uncertainties such as when u is known through v at $\pm\varepsilon$ with probability p. Such uncertainties are relevant to model the *thermal* noise of digital systems.

Gaussian models are easy to extend to mutli-dimensional real data, for instance, for color images where each observation u is a 3-dimensional vector. By channel, we denote an entry of this vector. If noise is correlated between channels, a D-dimensional Gaussian random vector V follows the following pdf:

$$p(v|u) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(v-u)^t\Sigma^{-1}(v-u)\right] , \tag{3.3}$$

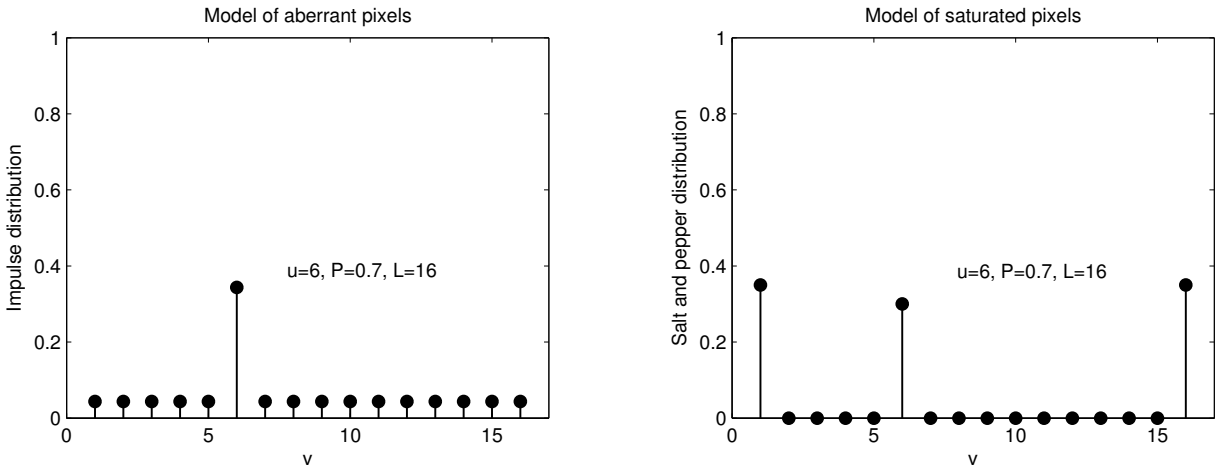
where u is the D-dimensional real vector representing the noise-free information and Σ is a $D \times D$ covariance matrix. Multi-dimensional Gaussian verify the same property mentioned above for mono-dimensional Gaussian distributions.

3.1.2 Cauchy noise

Given a noise level modeling by $\gamma > 0$, a Cauchy random variable V is a real random variable described by the following pdf:

$$p(v|u) = \frac{1}{\pi\gamma\left[1 + \left(\frac{v-u}{\gamma}\right)^2\right]} \tag{3.4}$$

where $u \in \mathbb{R}$ is the location of the mode and γ defines the scale of the distribution. Figure 3.1 gives an illustration of Cauchy distributions. Cauchy fluctuations are also symmetric and additive, therefore it is straightforward to show that V can be decomposed as $u+\gamma\varepsilon$ with ε being a Cauchy random variable with a mode at 0 and an unit scale parameter. Unlike the Gaussian



(a) (b)

Figure 3.2: Two distributions defined on the interval $[1..L]$ modeling errors corrupting a noise-free value u and occuring with a given probability P. Here $u = 6$, $L = 16$ and $P = 0.7$. (a) Impulse errors affect to a pixel an arbitrary value with probability P. (b) Salt-and-pepper errors saturate a pixel value to 1 or L with probability P.

distribution, the particularity of the Cauchy distribution is its very heavy tails. A consequence is that its expectation and variance do not exist: the sample mean and the sample variance do not converge with respect to the number of observations. This kind of noise then defeats all methods based on the average of pixel values.

As far as we know, there is no imaging system delivering images with Cauchy noise. Cauchy distribution are used in spectroscopy. It can be useful to describe an impulse-like noise with unbounded and continuous values.

3.1.3 Impulse noise

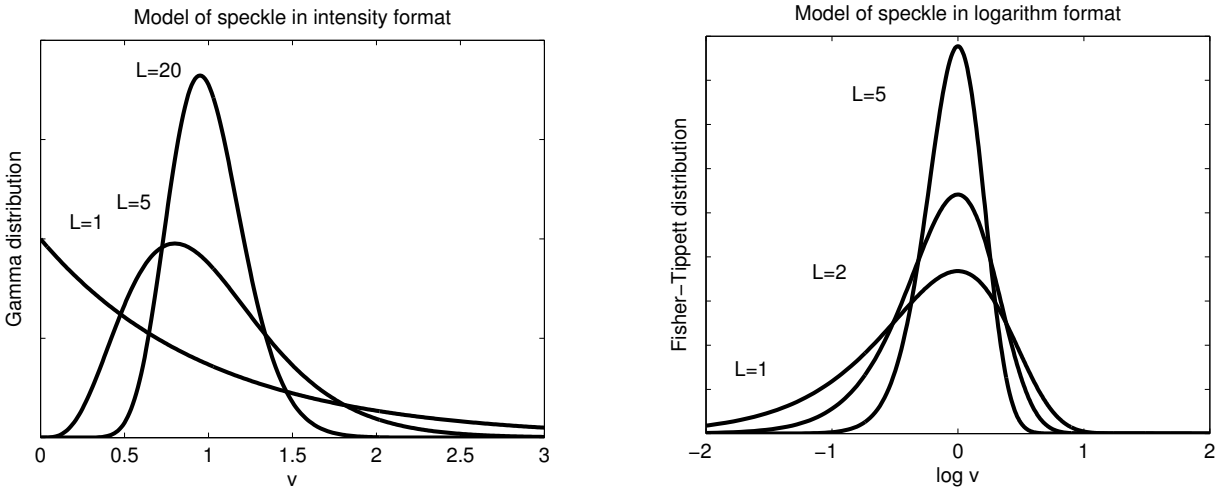
Impulse noise can model random uniform aberrations of imaging systems measuring a discrete information quantified on the interval $[1..L]$. Imaging systems delivering images with such noise fluctuations are relatively uncommon. It can appears in remote applications where bits are corrupted during the transmition of the image. It also constitutes a denoising problem that is particularly difficult and challenging for *academic* purposes. Such images present two types of pixels: either the pixel has the same value as the noise-free value, or it has a value uniformly distributed on the interval $[1..L]$. The degradation rate is given by the parameter $P \in [0, 1]$. Impulse noise is modeled by the probability mass function:

$$p(v|u) = \begin{cases} P/L + 1 - P & \text{if } v = u \\ P/L & \text{otherwise} \end{cases} . \tag{3.5}$$

Figure 3.2 gives an illustration of such a discrete distribution.

3.1.4 Salt-and-pepper noise

Salt-and-pepper noise can be used in any imaging system measuring a discrete information quantified on the interval $[1..L]$. Salt-and-pepper noise models random saturations that occur in physical measurements. Given a parameter $P \in [0, 1]$, modeling the percentage of corrupted



(a) (b)

Figure 3.3: Two distributions modeling *speckle* noise. (a) The gamma distribution describes the *speckle* observed in intensity images. It is defined on positive real values and it describes an *hetereoscedastic* noise since its shape is linked to its mean. Its heavy right-tail explains the highly bright pixels that are present in such images. (b) The Fisher-Tippett distribution describes the *speckle* observed in the logarithm image. Unlike the Gaussian or the Cauchy distribution, this distribution describes an asymmetrical additive noise. Due to its heavy left-tail, this distribution models an image presenting several dark pixels.

pixels, salt-and-pepper noise is modeled by the probability mass function:

$$p(v|u) = \begin{cases} P/2 + (1-P) & \text{if } v = 1 \text{ and } u = 1 \\ P/2 + (1-P) & \text{if } v = L \text{ and } u = L \\ P/2 & \text{if } v = 1 \text{ and } u \neq 1 \\ P/2 & \text{if } v = L \text{ and } u \neq L \\ (1-P) & \text{if } v = u \text{ and } u \notin \{1, L\} \\ 0 & \text{otherwise} \end{cases} . \tag{3.6}$$

Figure 3.2 gives an illustration of such a discrete distribution. With probability P, any pixel value is either affected to the value zero (usually representing black color) or to the value L (usually representing white color). The resulting image is composed of several saturated pixels with white or black colors as if salt-and-pepper have been sprinkled on the image. Like impulse-noise, imaging systems delivering images with such noise fluctuations are relatively uncommon. Such noise can be reduced efficiently if treated as an impainting problem where saturated values are rather considered as missing values.

3.1.5 *Speckle* noise

The case of intensity: In coherent imagery (e.g., radar images, sonar or ultrasound), when measuring the intensity of backscattered wave echo, the interferences of many punctual backscatterers, located in the same resolution cell, lead to *speckle*. *Speckle* produces fluctuations with a distribution depending on the organization inside the resolution cell. As a consequence, *speckle* is signal-dependent since the intensity of fluctuations vary with the underlying information. Given the positive integer $L \in \mathbb{N}^*$, called shape parameter and acting on the noise level,

speckle is usually modeled by a gamma real random variable V described by the following pdf:

$$p(v|u) = \frac{L^L v^{L-1} e^{-L\frac{v}{u}}}{\Gamma(L)u^L}\ , \tag{3.7}$$

where u is a real value modeling the average variations of the punctual back-scatterers. Figure 3.3 gives an illustration of gamma distributions. Its expectation is $\mathbb{E}[V] = u$ and variance $\mathrm{Var}[V] = \frac{u^2}{L}$. The relation $\mathrm{Var}[V] \propto \mathbb{E}[V]^2$ indicates an *heteroscedastic* noise which has moreover a multiplicative behavior. Indeed, it is straightforward to show that V can be decomposed as $u \times S$ with S being a gamma random variable of parameter $u_S = 1$. We then call it multiplicative noise. When $L = 1$, the gamma distribution boils down to the exponential distribution known to have a heavy right-tail and a heavy left-head. Hence the resulting image appears as a dark image with highly bright pixels.

The intensity image v usually has a high dynamic range (mostly when the shape parameter L is small). To decrease the dynamic of such images, for visualization or post-processing purposes, it is common to perform a change of variable.

The case of amplitude: It is common to transform the intensity images to the amplitude format thanks to the square root transform: $\tilde{v} = \sqrt{v}$. In this case, the resulting distribution is called a Nakagami-Rayleigh distribution defined by:

$$p(\tilde{v}|\tilde{u}) = \frac{2L^L \tilde{v}^{2L-1} e^{-L\frac{\tilde{v}^2}{\tilde{u}^2}}}{\Gamma(L)\tilde{u}^{2L}}\ , \tag{3.8}$$

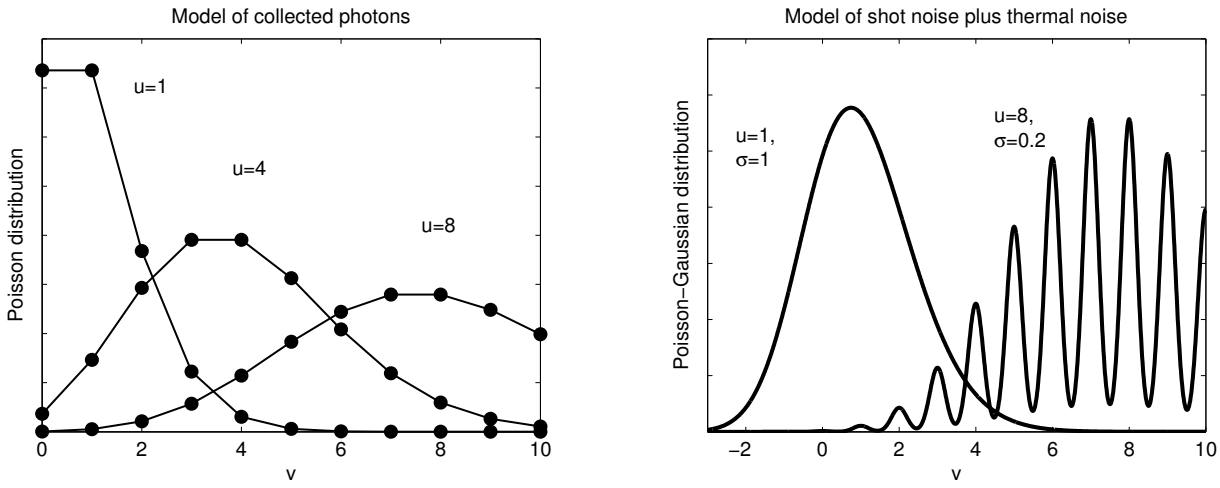
where $\tilde{u} = \sqrt{u}$. Its expectation is $\mathbb{E}[\tilde{V}] = \frac{\Gamma(L+0.5)}{\sqrt{L}\Gamma(L)}\tilde{u}$ and variance $\mathrm{Var}[\tilde{V}] = \left[1 - \left(\frac{\Gamma(L+0.5)}{\sqrt{L}\gamma(L)}\right)^2\right]\tilde{u}^2$. Again, the relation $\mathrm{Var}[\tilde{V}] \propto \mathbb{E}[\tilde{V}]^2$ indicates a multiplicative behavior. Indeed, it is straightforward to show that $\tilde{V}$ can be decomposed as $\tilde{u} \times S$ with S being a Nakagami-Rayleigh random variable of parameter $\tilde{u}_S = 1$. Like the gamma distribution, the Nakagami-Rayleigh has a heavy right-tail.

The case of logarithm transform: Another solution is to use a logarithm transform $\tilde{v} = \log v$. In this case, the resulting distribution is called a Fisher-Tippett distribution (or sometimes, a double exponential distribution) defined by:

$$p(\tilde{v}|\tilde{u}) = \frac{L^L \exp(L\tilde{v}) e^{-L\frac{\exp\tilde{v}}{\exp\tilde{u}}}}{\Gamma(L)\exp(L\tilde{u})}\ , \tag{3.9}$$

where $\tilde{u} = \log u$. Figure 3.3 gives an illustration of Fisher-Tippett distributions. Its expectation is $\mathbb{E}[\tilde{V}] = \tilde{u} + \Psi(L) - \log L$ and variance $\mathrm{Var}[\tilde{V}] = \Psi(1, L)$ where $\Psi(1, L)$ is the first-order polygamma function of degree L (e.g. Xie et al., 2002b). The relation $\mathbb{E}[\tilde{V}] = \tilde{u} + C$ with a variance independent of $\tilde{u}$ indicates an additive behavior: the log-transform of a signal-dependent multiplicative noise results in a signal-independent additive noise. Unlike the previous distribution, the Fisher-Tippet has a heavy left-tail, the resulting image presents then several dark pixels.

If square root or logarithm transforms seem attractive to display or analyze such images, we suggest instead treating noise fluctuations in the intensity format. The statistics of the gamma distribution have simpler expressions than the ones of Nakagami-Rayleigh or Fisher-Tippett distributions. The sum of gamma random variables is a gamma random variable while the sum of Nakagami-Rayleigh random variables is not known in closed-form. Finally unlike Nakagami-Rayleigh or Fisher-Tippett random variables, the sample mean corresponds to the maximum likelihood estimator for gamma distribution.



(a) (b)

Figure 3.4: Two distributions modeling *shot* noises. Both distributions model an *hetereoscedastic* noise since their scale is directly linked to their mean. (a) The Poisson distribution describes the fluctuations arising when counting the number of emited photons received during the exposure time. Because it models the number of photons, the values are discrete and positive. (b) The Poisson-Gaussian distribution models a mixture of fluctuations due to counting and *thermal* noise. It is obtained by convolving a discrete Poisson distribution by a continuous Gaussian distribution with a scale parameter σ. The values are then real and can take negative values. For small level of *thermal* noise, Poisson-Gaussian noise is highly multi-modal.

3.1.6 *Shot* noise or Poisson noise

Shot noise appears in low-light conditions when the number of collected photons is small, such as in fluorescence microscopy or astronomy. *Shot* noise is usually modeled by a Poisson random variable V described by the following probability mass function:

$$p(v|u) = \frac{u^v e^{-u}}{v!} \,. \tag{3.10}$$

where u is a real value modeling the light intensity. Figure 3.4(a) gives an illustration of Poisson distributions. Its expectation is $\mathbb{E}[V] = u$ and variance $\mathrm{Var}[V] = u$. The Poisson noise is then *heteroscedastic*. Note that the relation $\mathrm{Var}[V] = \mathbb{E}[V]$ is non-homogeneous, which is challenging, since, V cannot be related to u through additive or multiplicative decomposition.

3.1.7 Poisson-Gaussian noise

Optical systems can suffer from two sources of noise: *shot* noise and *thermal* noise. They are then described by the sum of two random variables following, respectively, a Poisson and a Gaussian noise distribution. The resulting real random value is modeled by the following pdf obtained by convolving a Poisson distribution by a Gaussian distribution:

$$p(v|u) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=0}^{\infty} \exp\left(-\frac{(v-k)^2}{2\sigma^2}\right) \frac{u^k e^{-u}}{k!} \tag{3.11}$$

where $\sigma > 0$ models the uncertainty due to *thermal* fluctuations. Figure 3.4(b) gives an illustration of Poisson-Gaussian distributions. Its expectation is $\mathbb{E}[V] = u$ and variance $\mathrm{Var}[V] = u+\sigma^2$. It models then an *heteroscedastic* noise.

Poisson-Gaussian noise results from the sum of a discrete random variable taking its values in $\mathbb{N}$ and a continuous random variable taking its values in $\mathbb{R}$. As a consequence, for small values of σ Poisson-Gaussian noise is not mono-modal (see Fig. 3.4(b)). The estimation of u using the maximum likelihood estimator cannot be obtained in closed-form and requires the use of consuming iterative methods prone to fall in local minima such as the split gradient method of Lanteri and Theys (2005).

Since optical sensors provide non-negative discrete values, we suggest that the contribution of *shot* noise and *thermal* noise should be modeled as the sum of two non-negative discrete random variables. While more realistic, such a model should lead to mono-modal distributions which is more suitable to be used with optimization methods. The performance of such approach highly depend on our knownledge about the sources of degradations. A Gaussian distribution with a variance linearly dependent of u can be preferable to model the combination of a *shot* noise with a *thermal* noise when we do not control all sources of degradations.

The enumeration of uncorrelated noise models described above is not exhaustive and we can find many other noise distribution models. In Chap. 6, in the context of coherent images, we will introduce the multi-variate complex circular Gaussian distribution and the complex Wishart distribution modeling the interferences (or correlations) between several backscattered waves.

3.2 Back to the Gaussian world: the variance stabilization

3.2.1 Description and motivations

As described in Chap. 2, Gaussian noise leads to simple formulations and simplifications easing the elaboration of noise reduction techniques. Many algorithms for image denoising are then designed for Gaussian noise only. A classical (cost-less) approach to extend the applicability of filters designed for Gaussian noise to some non-Gaussian noise is to apply a transformation to the noisy data. The transformation is chosen so that the transformed data follows a (close to) Gaussian distribution with constant variance (hence their name: variance-stabilization transforms). These approaches are popular and frequently used, e.g., for density estimation (Brown et al., 2010), wavelet denoising (e.g. Xie et al., 2002a; Achim et al., 2003; Bhuiyan et al., 2007). and patch-based denoising (e.g. Mäkitalo et al., 2010; Boulanger et al., 2010; Mäkitalo and Foi, 2011).

Given an invertible application $\boldsymbol{s}$ which stabilizes the variance for a specific noise pdf and a filter h designed for Gaussian noise, the stabilization-based filtering is given by:

$$\hat{\boldsymbol{u}} = (\boldsymbol{s}^{-1} \circ h \circ \boldsymbol{s})(\boldsymbol{v}) \; . \tag{3.12}$$

The case of gamma noise: This leads for instance to the homomorphic approach which maps, thanks to a logarithm transform, multiplicative noise to additive noise with stationary variance (see Jain, 1989). For instance, in the case of gamma noise:

$$\boldsymbol{s}(V) = \log V \Rightarrow \mathrm{Var}[\boldsymbol{s}(V)] = \mathrm{Var}[\log V] = \Psi(1, L) \tag{3.13}$$

where $\Psi(1, L)$ is the first-order polygamma function of degree L (e.g., Xie et al., 2002b).

The case of Poisson noise: This is also the principle of Anscombe's transform (and its variants) used for Poisson noise:

$$\boldsymbol{s}(V) = 2\sqrt{V + \frac{3}{8}} \Rightarrow (u \gg 0 \Rightarrow \mathrm{Var}[\boldsymbol{s}(V)] = 1) \; . \tag{3.14}$$

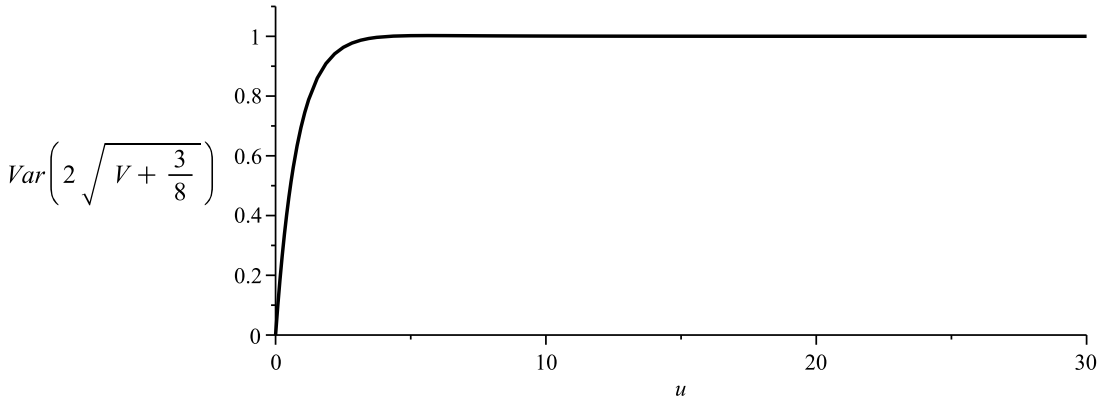


Figure 3.5: Variance of Anscombe's transform of Poisson random variables with respect to parameter u. For u high enough, the variance is independent of u and equal to 1.

Such variance-stabilization transform is only asymptotic and applies well as soon as $u > 4$. Figure 3.5 describes the relationship between u and the variance of Anscombe's transform.

Usually $\boldsymbol{s}$ is non-linear and Eq. (3.12) introduces a bias in the estimation since $\mathbb{E}[V|u] \neq \boldsymbol{s}^{-1}[\mathbb{E}[\boldsymbol{s}(V)|u]]$. A post-precessing step to unbias the estimation is then required.

3.2.2 Post-processing step to unbias the estimation

Equation (3.12) tends to introduce a bias in the estimation. This bias can be estimated and then used to apply a bias correction to $\hat{\boldsymbol{u}}$. This correction is not universal: it depends on the given noise distribution and its stabilization function. Debiasing can be achieved by studying the functional link f between the biased estimate and the noise-free data $\mathbb{E}[\boldsymbol{s}(V)|u] = f(u)$. If f is invertible, this leads to the unbiased stabilization approach given by:

$$\hat{\boldsymbol{u}} = (f^{-1} \circ h \circ \boldsymbol{s})(\boldsymbol{v}) \; . \tag{3.15}$$

The case of gamma noise: The mean of the log transform of a gamma random variable is given by (e.g. Xie et al., 2002b):

$$\mathbb{E}[\boldsymbol{s}(V)|u] = \ln u + \psi(L) - \log L \; . \tag{3.16}$$

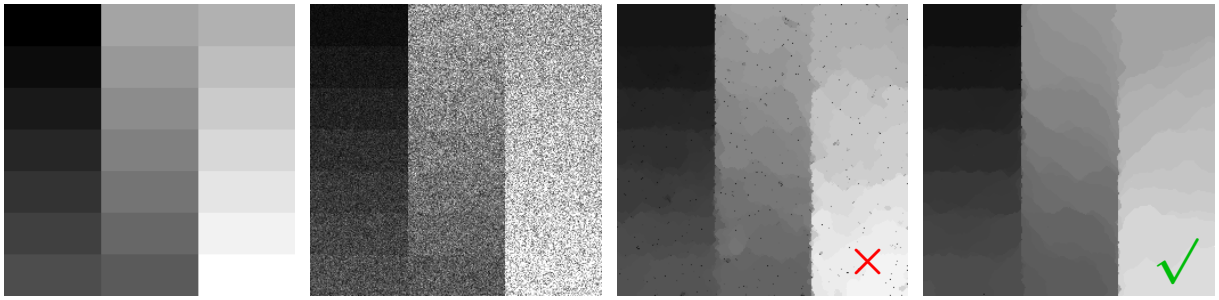
Bias correction can be achieved directly by what Xie et al. (2002a) call the "adjust mean" step:

$$\hat{u}^{(debiased)} = \frac{L}{\exp \psi(L)} \hat{u}^{(biased)} \; . \tag{3.17}$$

The case of Poisson noise: The mean of Anscombe's transform of a Poisson random variable is given by:

$$\mathbb{E}[\boldsymbol{s}(V)|u] = 2 \sum_{v=0}^{\infty} \sqrt{v + 3/8} \frac{u^v e^u}{v!} \; . \tag{3.18}$$

Mäkitalo and Foi (2011) suggest inverting Eq. (3.18) by evaluating numerically the summation for different values of u and using linear interpolation for arbitrary values of u.



(a) Noise-free image (b) Gamma noise (c) Logarithm + ROF (d) (Shi and Osher, 2008)

Figure 3.6: (a) Noise-free image, (b) noisy image corrupted by gamma noise and denoised images using (c) the ROF model applied after a logarithm transform, and (d) the adapted solution of (Shi and Osher, 2008) for gamma noise.

3.2.3 Drawbacks and limitations of variance stabilization approaches

Variance stabilization approaches have two significant advantages: they are simple to design; and they provide a competitive performance due to the important joint effort made by the community to elaborate efficient techniques to deal with Gaussian noise. For instance, *state-of-the-art* approaches for Poisson noise are obtained with three tools: Anscombe's transform, a powerful algorithm designed for Gaussian noise, and an optimal inversion of Anscombe's transform (see for instance Mäkitalo et al., 2010; Boulanger et al., 2010).

A first limitation is the assumption that stabilizing the variance leads to Gaussian noise. In the case of multiplicative gamma noise, we have seen in Sec. 3.1.5 that the logarithm-transform leads to an image corrupted instead by a Fisher-Tippett distribution. This distribution is asymptotically Gaussian when L tends to infinity (due to the law of large numbers). However, for small values of L, this distribution is asymmetric and has a heavy left-tail.

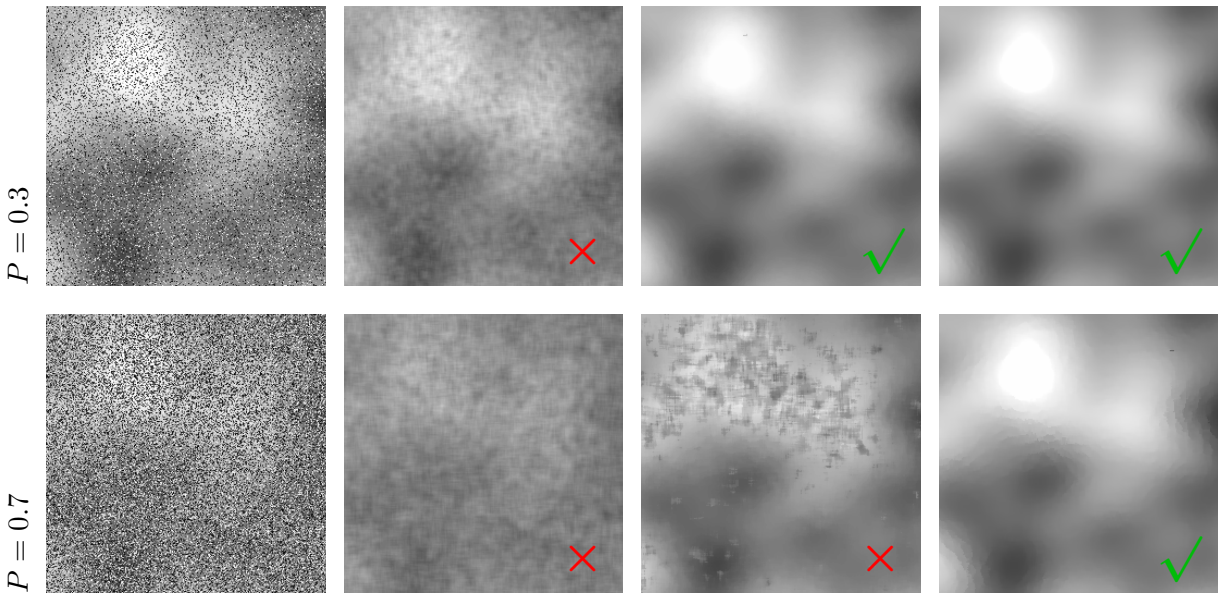
Figure 3.6 compares two approaches based on the minimization of the total-variation[1] (see Sec. 2.3.3) using either the logarithm-transform or a suitable adaptation to the gamma noise that we will present in Sec. 3.5. Since the method based on the logarithm-transform does not take into account the left-tail of the Fisher-Tippett distribution, it leads to the presence of many dark residual pixels while the suitable adaptation does not suffer from this undesirable effect.

Another important limitation lies in the non-linear distortion of noise-free data introduced by $\boldsymbol{s}$. For instance, in the homomorphic approach, the logarithm transforms the contrast of noise-free patches. Prior on the noise-free image are then affected accordingly. For instance, a sparse decomposition on the logarithm transform of an image corrupted by a multiplicative noise is equivalent to considering that the noise-free image can be decomposed as the *product* of a few atoms of the dictionary. A more fundamental limitation is the nonexistence of a variance stabilizing transform $\boldsymbol{s}$ for some distributions.

3.3 Moving average filters and maximum likelihood estimation

When noise departs from Gaussian noise, moving average filters presented in Sec. 2.2.1 can lead to bias or wrong estimation of the underlying smoothed regular noise-free image. For instance, Fig. 3.7 illustrates the result of the moving average filter on an image corrupted by impulse noise. The result is clearly unsatisfactory. It is well-known that for impulse noise, the median filter

[1] we use anisotropic TV corresponding to the sum of the ℓ^1 norm of the gradient so that minimization problem can be solved by graph-cuts



(a) Noisy image (b) Moving average (c) Moving median (d) Moving MLE

Figure 3.7: (a) A smooth image (a) corrupted by impulse noise with two different degradation rates (top) $P = 0.3$ and (bottom) $P = 0.7$. Denoising results obtained by (b) the moving average filter, (c) the moving median filter and (d) the moving MLE. When $P = 0.3$ the moving median filter and (d) the moving MLE provide unbiased values, while when $P = 0.7$ only the MLE filter is unbiased.

(which replaces the value of interest by the median of the values in a moving window) is more robust in dealing with the presence of outliers (see again Fig. 3.7). Why is there such difference of behavior?

Intuition: Assuming local ergodicity, moving average filters correspond to the least square estimate (LSE):

$$\hat{u}(x) = \frac{1}{|W|} \sum_{x' \in W_x} v(x') = \underbrace{\arg\min_u \sum_{x' \in W_x} (v(x') - u)^2}_{\text{LSE}} \ . \tag{3.19}$$

This solution is known to be optimal when the errors (understand noise) have zero mean, are uncorrelated and have equal variance (cf. Gauss-Markov theorem). Therefore, moving average filters are optimal for Gaussian noise. In the case of impulse noise, the LSE is strongly influenced by the aberrant values or outliers (due to the quadratic penalty). A strategy to decrease the influence of outliers is to consider the ℓ_1 penalty instead of the quadratic ℓ_2 norm. It is known that the minimization of the ℓ_1 norm, also referred to as the absolute norm of deviation, leads to the moving median filter:

$$\hat{u}(x) = \text{Median}\left(\{v(x') | x' \in W_x\}\right) = \underbrace{\arg\min_u \sum_{x' \in W_x} |v(x') - u|}_{\ell_1\text{-norm}} \ . \tag{3.20}$$

Explanation: The moving average filter is also the maximum likelihood estimate (MLE) of inde-

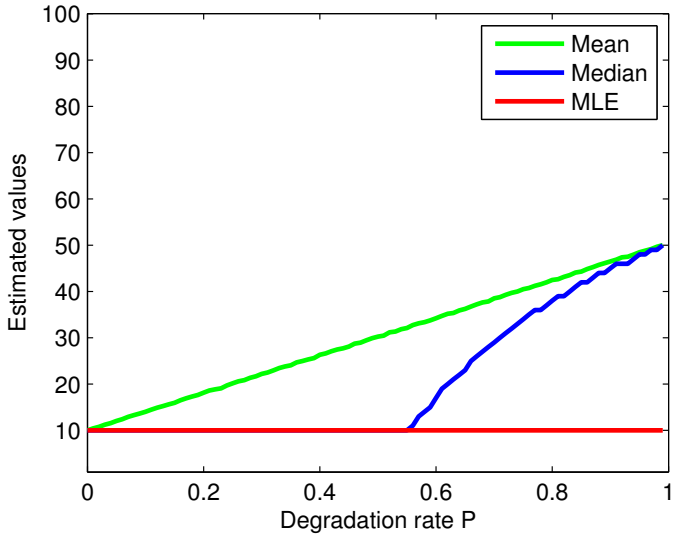


Figure 3.8: Estimation of the underlying value $u = 10$ under impulse noise defined on the range $[1, 100]$ for different degradation rates P. When $P \leq 0.5$, the median filter is the MLE, but, when $P > 0.5$, the median filter departs from the MLE. The average estimator gives poor results in all situations.

pendent and identically distributed (i.i.d.) Gaussian random variables:

$$\hat{u}(x) = \frac{1}{|W|} \sum_{x' \in W_x} v(x') = \underbrace{\arg\max_u \sum_{x' \in W_x} \log p(v(x')|u)}_{\text{MLE}} \tag{3.21}$$

Under Gaussian noise assumption, the optimality of the moving average filter is then ascribed to the consistency and the efficiency of the MLE.

Under impulse noise, the MLE cannot be obtained in closed-form. The MLE can, however, be obtained numerically by exhaustive search. Figure 3.8 compares the asymptotic estimates obtained by the average, the median and the MLE for different degradation rates P. It appears that the bias of the average filter increases with the degradation rate P, the median filter becomes biased for a degradation rate $P > 0.5$ and the MLE is unsurprisingly unbiased. The good behavior of the median filter is then due to the fact that it acts as the MLE for low values of P. Indeed, Fig. 3.7 shows that, for a small value of P, the moving median filter and the moving MLE provide good results compared with the moving average filter. For a high value of P, only the moving MLE provides a good result.

3.4 Anisotropic diffusion and edge detection

The anisotropic diffusion of Perona and Malik (1990) described in Sec. 2.3.1 can lead to aberrant results when noise departs from Gaussian noise. In the case of *speckle*, Yu and Acton (2002) mentioned that "anisotropic diffusion will actually enhance the *speckle*, instead of eliminating the corruption". Few attempts try to extend this approach to non-Gaussian noise. The *speckle* reduction anisotropic diffusion (SRAD) filter of Yu and Acton (2002) uses the following partial differential equation:

$$\begin{cases} \frac{\partial u(x,t)}{\partial t} = \operatorname{div}\left(c(q(x,t))\nabla u(x,t)\right) \\ u(x,0) = v(x) \end{cases} \tag{3.22}$$

where c is a decreasing function of the instantaneous coefficient of variation $q(x,t) = \frac{\sqrt{\mathrm{Var}[u(x,t)]}}{\mathbb{E}[u(x,t)]}$. The instantaneous coefficient of variation is estimated with local statistics which, as a consequence, acts as an edge detector with a constant answer in the homogeneous areas (since *speckle* is multiplicative). The function c is then chosen such that the diffusion is maximal when the local estimate of the coefficient of variation fluctuates around this constant. At the opposite end, the function c gives low answers for high values of the local estimate since it probably indicates the presence of an edge. This idea has been extended in (Aja-Fernández and Alberola-López, 2006; Krissian et al., 2007).

A similar approach could be used for Poisson noise, using the local estimation of $\frac{\mathrm{Var}[u(x,t)]}{\mathbb{E}[u(x,t)]}$ as an edge detector with a constant answer in the homogeneous areas. Nevertheless, note that diffusions realize, in essence, an average. Hence, such an extension should only work for noise distributions whose MLE is the empirical average. To our knowledge, no solution has been proposed thus far to extend the anisotropic diffusion to arbitrary types of noise.

3.5 Variational-based filtering and the Bayesian approach

The minimization of the total-variation (TV) presented in Sec. 2.3.3 and the research of sparse decompositions presented in Sec. 2.4, are two examples of variational-based filters. They relie on the minimization of an energetic (objective) function expressed by two terms:

$$U(\boldsymbol{u}) = U_{data}(\boldsymbol{v},\boldsymbol{u}) + \lambda U_{regularity}(\boldsymbol{u}) \tag{3.23}$$

where $U_{data}(\boldsymbol{v},\boldsymbol{u})$ is usually a quadratic penalty $\|\boldsymbol{v}-\boldsymbol{u}\|_2^2$ measuring the fidelity of the explanation $\boldsymbol{u}$ to the noisy data $\boldsymbol{v}$ and $U_{regularity}(\boldsymbol{u})$ an energetic term modeling the regularity or sparsity of the solution. The Lagrangian multiplier $\lambda > 0$ acts as a trade-off between data fidelity and regularity.

Such a formulation finds its justification in a Bayesian framework. In Sec. 2.1.2 we have seen that denoising can be achieved by maximum *a posteriori* (MAP) estimation:

$$\hat{\boldsymbol{u}} = \arg\max_{\boldsymbol{u}} p(\boldsymbol{u}|\boldsymbol{v}) = \arg\max_{\boldsymbol{u}} p(\boldsymbol{v}|\boldsymbol{u})p(\boldsymbol{u}) \tag{3.24}$$

since the evidence $p(\boldsymbol{v})$ is constant with respect to $\boldsymbol{u}$. In this Bayesian formulation, the noise-free image $\boldsymbol{u}$ is modeled as the realization of a random vector, following a distribution $p(\boldsymbol{u})$, and modeling in a statistical way the *a priori* configuration that can occur. The term $p(\boldsymbol{v}|\boldsymbol{u})$ is the likelihood of $\boldsymbol{u}$ given $\boldsymbol{V} = \boldsymbol{v}$, which is fully described by the given noise distribution model. The minimization of Eq. (3.23) is equivalent to the MAP estimation, thanks to the following identifications:

$$\begin{aligned} p(\boldsymbol{u}|\boldsymbol{v}) &= \exp\left(-U(\boldsymbol{u})\right) \ , \\ p(\boldsymbol{v}|\boldsymbol{u}) &= \exp\left(-U_{data}(\boldsymbol{v},\boldsymbol{u})\right) \ , \\ p(\boldsymbol{u}) &= \exp\left(-\lambda U_{regularity}(\boldsymbol{u})\right) \end{aligned} \tag{3.25}$$

provided that the probability density function does not cancel. Under the Gaussian noise assumption, $U_{data}(\boldsymbol{u},\boldsymbol{v}) = -\log p(\boldsymbol{v}|\boldsymbol{u})$ corresponds to the quadratic penalty. For non-Gaussian noise, we will see, through several examples, that the application of Bayes formula has been used successfully in many approaches in order to derive relevant data fidelity terms.

3.5.1 Maximum *a posteriori* on Markov random fields

Geman and Geman (1984) suggest modeling the local contextual information of noise-free images using a Markov random field (MRF). The MRF approach introduces a local prior model of the noise-free image and searches for a compromise between that prior and the noisy data. Markovian priors describe punctual conditional probabilities at position x as depending only on the local context in the spatial neighborhood $\mathcal{V}(x)$:

$$p(u(x)|u(\Omega/\{x\})) = p(u(x)|u(\mathcal{V}(x)) \; . \tag{3.26}$$

where $u(\Omega/\{x\})$ denotes the set of all noise-free values except the one at position x. Hammersley and Clifford (1971) show that if all configurations are possible, i.e., $\forall \boldsymbol{u}, p(\boldsymbol{u}) > 0$, the prior distribution is necessarily a Gibbs field defined as follows:

$$p(\boldsymbol{u}) = \frac{\exp\left(-\sum_{c\in\mathcal{C}} U_c(\boldsymbol{u})\right)}{Z} \tag{3.27}$$

where $\mathcal{C}$ is the clique system associated to the neighborhood system $\mathcal{V}$, i.e., the set of tuples such that each pair of elements of the same tuple are neighboors with respect to $\mathcal{V}$. The number of elements k of a given clique is called the order of the clique. The energy $U_c > 0$ is the local energy associated to a given clique c and Z is a normalization constant. In this context, the MAP estimation on an MRF writes as:

$$\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}} -\log p(\boldsymbol{v}|\boldsymbol{u}) + \sum_{c\in\mathcal{C}} U_c(\boldsymbol{u}) \; . \tag{3.28}$$

It is crucial for these techniques to define a suitable prior that guarantees both the smoothness of the denoised image and the preservation of its structures. A common prior is to consider only clique of order 2, with the following associated energy:

$$U_{c=(x,x')}(\boldsymbol{u}) = \lambda|u(x) - u(x')| \tag{3.29}$$

which penalizes large transitions (proportionally to $\lambda > 0$) and favors constant regions. Such priors however tend to bias the denoised image, especially when high noise levels are considered. Markovian priors are *local* in essence, and lead to stronger attenuation of several small disconnected regions rather than that of a single region (Strong and Chan, 2003). In practice, an edge-preserving MRF model generally leads to minimization problems with non-smoothness and/or non-convexity issues.

With a clique energy of second-order, Eq. (3.29) can be interpreted as the discrete approximation of the local gradient norm $|\nabla u(x)|$. Hence, the MAP estimation on an MRF leads to the derivation of total-variation models:

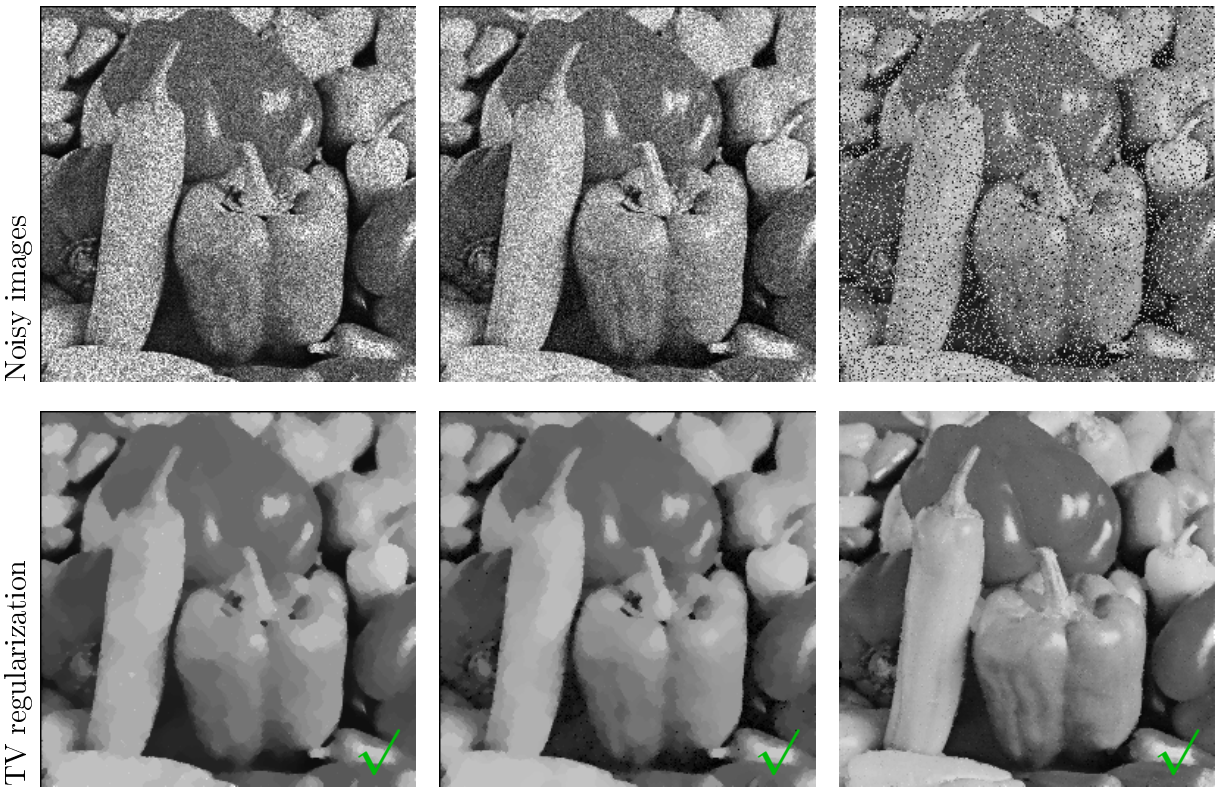
The case of Gaussian noise: The regularization model of Rudin et al. (1992) matches with the MAP estimation on a MRF using Eq. (3.29):

$$\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}} \|\boldsymbol{u} - \boldsymbol{v}\|_2^2 + \lambda \sum_x |\nabla u(x)| \; . \tag{3.30}$$

The parallel between the ROF model and the MAP estimation on a MRF allows us to extend the ROF model to other distributions.

The case of gamma noise: Aubert and Aujol (2008) suggest minimizing the following functional:

$$\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}} \sum_x \left(\log u(x) + \frac{v(x)}{u(x)}\right) + \lambda \sum_x |\nabla u(x)| \tag{3.31}$$



(a) Gamma noise (b) Poisson noise (c) Impulse noise

Figure 3.9: Denoising results obtained on an image corrupted by (a) gamma noise, (b) Poisson noise and (c) impulse noise. The results are obtained by our implementation of the variational approach based on suitable data fidelity term as suggested respectively by (Shi and Osher, 2008), Le et al. (2007) and (Darbon and Sigelle, 2006).

where the data fidelity term corresponds, up to a constant, to the negative log likelihood of the gamma distribution. Denis et al. (2009) suggest using an equivalent data fidelity term but with the total-variation prior defined on the square root of the image (i.e., the amplitude format), whereas, in (Shi and Osher, 2008; Bioucas-Dias and Figueiredo, 2010), the authors use the total-variation prior defined on the logarithm of the image $\tilde{\boldsymbol{u}} = \log \boldsymbol{u}$. This model has then been extended in (Huang et al., 2009) by adding a quadratic penalty to simplify the optimization procedure. In (Durand et al., 2010), the authors use the total-variation prior on the logarithm transform and an ℓ_1 data-fitting term on the curvelet coefficients. Further details about the different data fidelity terms and prior terms that can be used for multiplicative noise reduction can be found in (Steidl and Teuber, 2010). More recently, Xiao et al. (2010); Huang et al. (2010) have introduced another variational model mixing the ideas of Aubert and Aujol (2008) and Shi and Osher (2008) where the term $|\nabla \log u(x)| = \left|\frac{\nabla u(x)}{u(x)}\right|$ is interpreted as the psychological Weber's law known to reflect humans' perceptual sensitivity (see also Shen, 2003).

The case of Poisson noise: Le et al. (2007); Bardsley and Luttman (2009); Figueiredo and Bioucas-Dias (2009); Willett et al. (2010) suggest minimizing the following functional:

$$\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}} \sum_x \left(u(x) - v(x) \log u(x)\right) + \lambda \sum_x |\nabla u(x)| \tag{3.32}$$

where the data fidelity term corresponds, up to a constant, to the negative log likelihood of Poisson distribution. The fidelity term used in Eq. (3.32) has also been justified in the context of

entropy maximization under Poisson noise. Steidl and Teuber (2010) mention that it corresponds, up to a constant, to the generalized Kullback-Leibler divergence.

The case of impulse noise: As mentioned in Sec. 3.3, the median filter is an unbiased estimator for low degradation rates P. Median filters result from the minimization of local ℓ_1 penalty. The TV-L1 model can then be interpreted as an extension of the ROF model relevant for impulse noise. When the degradation rate P increases, the direct application of the MAP estimation on a MRF leads to a non-convex minimization problem solved by graph-cut in (Darbon and Sigelle, 2006).

Figure 3.9 shows that the Bayesian extension of the ROF model[2] applies well for the three kinds of noise we have considered. All results appear as piece-constant images without introducing statistical artifacts. The results present a similar smoothing strenght everywhere in the image in dark and bright areas. This method is then efficient to adapt to signal-dependent noises.

3.5.2 Thresholding as MAP filtering in the sparse domain

We have seen that hard-thresholding, soft-thresholding and Wiener linear rescaling can be interpreted as the minimization of the data fidelity under respectively the ℓ_0, ℓ_1 and ℓ_2 regularity constraints. Thanks to Eq. (3.25), these strategies can be expressed as the MAP estimates where the fidelity term models the distribution of noisy wavelet coefficients and the constraint encodes the *a priori* statistical model of the coefficients of noise-free images:

$$\hat{\boldsymbol{u}} = \boldsymbol{D}\hat{\boldsymbol{\alpha}} \quad \text{where} \quad \hat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} p(\boldsymbol{\beta}|\boldsymbol{\alpha})p(\boldsymbol{\alpha}) \tag{3.33}$$

where $\boldsymbol{\alpha}$ is the transformed image of $\boldsymbol{u}$ and $\boldsymbol{\beta}$ is the transformed image of $\boldsymbol{v}$.

The choice of the prior term $p(\boldsymbol{\alpha})$: In the case of Gaussian noise, the noisy wavelet coefficients also follow a Gaussian distribution. In this case, Simoncelli (1999) shows that the standard thresholding approaches (see Sec. 2.4.1) can be interpreted in a Bayesian manner under the assumption that noise-free coefficients are independent and follow the generalized Gaussian distribution defined by:

$$p(\boldsymbol{\alpha}) = \prod_k p(\alpha_k) = \frac{\exp\left(-\sum_k |\alpha_k/s|^p\right)}{Z(s,p)} \tag{3.34}$$

where the normalization constant $Z(s,p) = 2\frac{s}{p}\Gamma\left(\frac{s}{p}\right)$. The hard, soft and Wiener approaches resemble the MAP estimation using a generalized Gaussian distribution for *prior* density with respective parameters $p = 0$, 1 and 2. Hence, the ℓ_1 norm encodes coefficients following a Laplace distribution while the ℓ_2 norm encodes a Gaussian distribution. The underlying idea is to choose the sparsity constraint according to the distribution of the coefficients of noise-free images $\boldsymbol{\alpha}$. This was the motivation of Mallat (1989); Simoncelli and Adelson (1996); Chang et al. (2000) who first introduced the generalized Gaussian distribution to model the coefficients of optical images. Spatial adaptation of the parameters of Gaussian and generalized Gaussian models have been used next to describe the local distribution of the coefficient of *speckle*-free synthetic aperture radar (SAR) images (Argenti and Alparone, 2002; Argenti et al., 2006; Bianchi et al., 2008). Xie et al. (2002a); Achim et al. (2003); Bhuiyan et al. (2007) suggest instead to model the logarithm transform of *speckle*-free SAR images respectively with a bi-Gaussian distribution, an α-stable distribution and a Cauchy distribution.

[2] we use anisotropic TV corresponding to the sum of the ℓ^1 norm of the gradient so that minimization problem can be solved by graph-cuts

The choice of the likelihood term $p(\boldsymbol{\beta}|\boldsymbol{\alpha})$: The nature of AWGN is preserved by any projection of $\boldsymbol{v}$ on an orthonormal basis $\boldsymbol{D}$. The quadratic term is then a consistent fidelity penalty independent of the choice of the orthonormal image dictionary (cf. Parseval's identity):

$$\|\boldsymbol{v} - \boldsymbol{u}\|_2^2 = \|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_2^2 \,. \tag{3.35}$$

When noise departs from Gaussian noise, this property does not hold anymore. The extension of thresholding wavelet coefficients for non-Gaussian noise has to consider the suitable distributions for both the noisy wavelet coefficients and the noise-free wavelet coefficients.

For simplicity reasons, most approaches designed for non-Gaussian noise still model noisy coefficients with a Gaussian distribution. Wavelet models for *speckle* reduction use either a spatially-varying Gaussian distribution (Argenti and Alparone, 2002) or a single Gaussian distribution after applying a variance-stabilization technique (Xie et al., 2002a; Achim et al., 2003; Bhuiyan et al., 2007) that Durand et al. (2010) justified by the central limit theorem. To our knowledge, only Argenti et al. (2006) attempt to model the noisy coefficients with an adapted spatially-varying generalized Gaussian distribution. However, as indicated by the authors, this model is only heuristic.

Another strategy is to use a data fidelity term in the space domain, for which the likelihood model is known, and a sparsity constraint in the transformed domain. This results in the following optimization problem:

$$\hat{\boldsymbol{u}} = \boldsymbol{D}\hat{\boldsymbol{\alpha}} \quad \text{where} \quad \hat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} p(\boldsymbol{v}|\boldsymbol{D}\boldsymbol{\alpha})p(\boldsymbol{\alpha}) \,. \tag{3.36}$$

This idea is used by Harmany et al. (2009); Raginsky et al. (2010) for the adaptation of compressed sensing methods to images corrupted by Poisson noise. If such a model seems relevant, it leads to non-trivial optimization problems which are not as simple as coefficient thresholding, involving methods such as Douglas-Rachford splitting algorithms.

3.6 Aggregation-based filtering and non-Gaussian noises

In this section, we are interested in the extensions of aggregation-based filters as described in Sec. 2.6. Among K pre-estimates $\overline{\boldsymbol{u}}_k$ of the noise-free image $\boldsymbol{u}$ obtained from a noisy image $\boldsymbol{v}$, such methods try to locally select the most relevant one or to combine the K estimates together to produce an optimal solution $\hat{\boldsymbol{u}}$.

3.6.1 Extension of the variance minimization approach

When the pre-estimate results from the weighted average of noisy values, the aggregation rule proposed by Salmon and Strozecki (2010) and based on the minimization of the residual variance can be used directly for non-Gaussian noise:

$$\hat{u}(x) = \sum_k \beta_k(x)\overline{u}_k(x) \quad \text{where} \quad \beta_k(x) = \frac{\mathrm{Var}[\overline{U}_k(x)]^{-1}}{\sum_l \mathrm{Var}[\overline{U}_l(x)]^{-1}} \,. \tag{3.37}$$

However, since the variance $\mathrm{Var}[\hat{\boldsymbol{U}}_k]$ can be signal dependent, i.e., linked to the unknown input signal $\boldsymbol{u}_k$, Eq. (3.37) cannot be evaluated directly. One can assume that all estimators result from the average of i.i.d. values, hence:

$$\mathrm{Var}[\overline{U}_k(x)] = \mathrm{Var}[V(x)]\frac{\sum w_k(x,x')^2}{[\sum w_k(x,x')]^2} \,. \tag{3.38}$$

In this case, Eq. (3.37) simplifies as:

$$\beta_k(x) = \frac{1}{Z} \frac{\left[\sum w_k(x,x')\right]^2}{\sum w_k(x,x')^2} \tag{3.39}$$

where Z is a normalization constant such that $\sum_k \beta_k(x) = 1$.

3.6.2 Extension of the ICI rule for non-Gaussian noise

Towards a conceptual extension: The ICI rule is based on the fact that if the pre-estimates $\overline{U}_k(x)$ are random values with Gaussian distributions centered on $u_k(x)$ with variance $\mathrm{Var}[\overline{U}_k(x)]$, hence, for all estimator k, the probabilities:

$$\mathbb{P}\left\{ u(x) \in \left[\overline{U}_k(x) - \gamma\sqrt{\mathrm{Var}[\overline{U}_k(x)]}, \overline{U}_k(x) + \gamma\sqrt{\mathrm{Var}[\overline{U}_k(x)]} \right] \right\} \tag{3.40}$$

are equal. Hence, by selecting the estimator with the smallest variance, $\overline{U}_k(x)$ uniformly tends towards $u(x)$ with known probability (cf. the sandwich theorem). The extension of the ICI rule for non-Gaussian noise should be based on the construction of intervals $\mathcal{I}_k(x)$ such that

$$\forall k, l \quad \mathbb{P}\left[u(x) \in \mathcal{I}_k(x)\right] = \mathbb{P}\left[u(x) \in \mathcal{I}_l(x)\right] \tag{3.41}$$

and $\mathcal{I}_k(x)$ tends uniformly towards the singleton $\{u(x)\}$ when $\mathrm{Var}[\overline{U}_k(x)]$ tends towards zero. In practice, the difficulty of such an approach lies on the estimation of $\mathrm{Var}[\overline{U}_k(x)]$. In (Katkovnik et al., 2008; Kervrann and Boulanger, 2008), the authors use Eq. (3.38) as an estimator of the residual variance. In their case, the noise is *homoscedastic*, hence the input noise variance $\mathrm{Var}[V(x)] = \sigma^2$ is independent of x and can be estimated globally from the noisy image. For *heteroscedastic* noise, the noise variance $\mathrm{Var}[V(x)]$ depends on the unknown parameter $u(x)$ which hence interferes with the design of such aggregation rules. As far as we know, direct extensions of the ICI rule for *heteroscedastic* noise have not been proposed so far and could be the topic of future works.

The fitted local likelihood (FLL) solution: Katkovnik and Spokoiny (2008) noticed that the ICI rule in Eq. (2.65) can be rewritten as:

$$\hat{u}(x) = \overline{u}_{k^*}(x) \quad \text{such that} \quad k^* = \sup\left\{ k = 1, \ldots, K \mid \forall l < k\ ,\ \frac{|\overline{u}_k(x) - \overline{u}_l(x)|}{\sqrt{\mathrm{Var}[\overline{U}_k(x)]} + \sqrt{\mathrm{Var}[\overline{U}_l(x)]}} < \gamma \right\}. \tag{3.42}$$

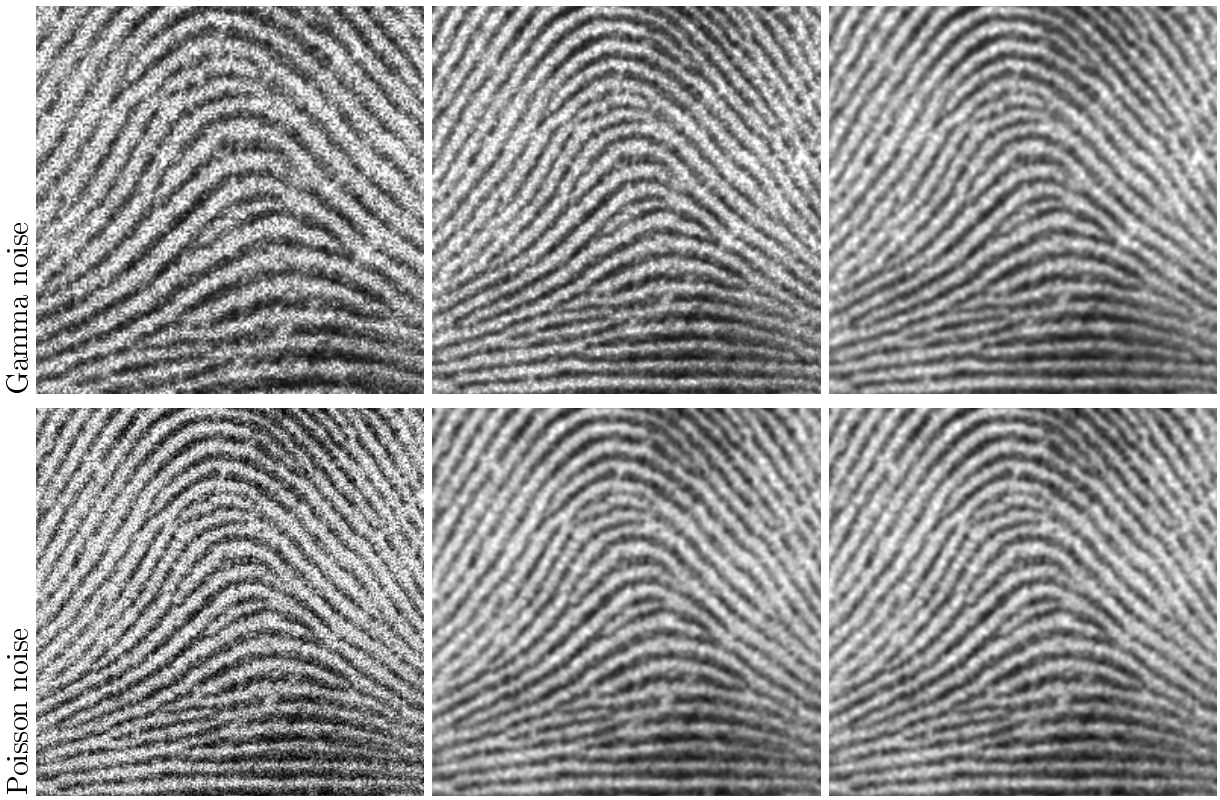
The authors suggest extending this rule for estimates based on the weighted maximum likelihood estimation (WMLE, see Sec. 5.1) under noise of the exponential family. Such estimates boil down to a weighted average, i.e.:

$$\overline{u}_k(x) = \arg\max_t \underbrace{\sum_{x'} w_k(x,x') \log p(v(x')|t)}_{\mathcal{L}_k(x,t)} = \frac{\sum_{x'} w_k(x,x') v(x')}{\sum_{x'} w_k(x,x')} \tag{3.43}$$

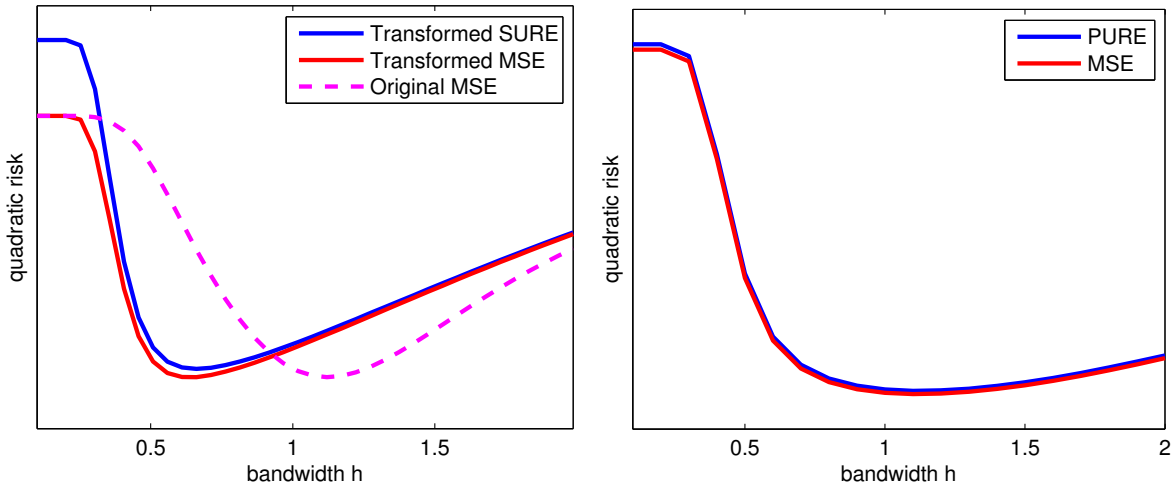
where $w_k(x,x')$ are the weights that the k-th estimator attributes to the noisy value $v(x')$ and $\mathcal{L}_k(x,t)$ is called the local likelihood of t in x. The fitted local likelihood (FLL) selection rule is given by:

$$\hat{u}(x) = \overline{u}_{k^*}(x) \quad \text{such that} \quad k^* = \sup\left\{ k = 1, \ldots, K \mid \forall l < k\ ,\ \mathcal{L}_l(x, \overline{u}_l(x)) - \mathcal{L}_l(x, \overline{u}_k(x)) < \varepsilon \right\}. \tag{3.44}$$

The advantage of such a selection rule, is that it does not require the estimation of the residual variance of each pre-estimate.



(a) Noisy image (b) Minimal estimated risk (c) Minimal true MSE



(d) Risk under gamma noise (e) Risk under Poisson noise

Figure 3.10: (a) Noisy data corrupted by (top) gamma noise and (bottom) Poisson noise. (b) Solutions of the Gaussian filter using the bandwidth parameter h minimizing an estimation of the risk. (c) Solutions of the Gaussian filter using the bandwidth parameter h minimizing the true risk. (d-e) Evolutions of the risk and its estimate for different values of h respectively under gamma and Poisson noise. In the case of gamma noise, the risk is either expressed on transformed data or on the original observation space.

3.6.3 Extension of Stein's unbiased risk estimation

Stein's unbiased risk estimation (SURE) can also be extended to other noise distributions. Let $h(.)$ be an estimator of the noise-free image from a given noisy image $\boldsymbol{v}$, such that $\hat{\boldsymbol{u}} = h(\boldsymbol{v})$, and denote by $h(\boldsymbol{v})|_x$ the estimated value $\hat{u}(x)$. Other unbiased risk estimates, i.e., such that $\mathbb{E}\left[R(\hat{\boldsymbol{U}})\right] = \mathbb{E}\left[\frac{1}{N}\|\boldsymbol{u} - \hat{\boldsymbol{U}}\|_2^2\right]$, are given by the following expressions:

The case of gamma noise: In (Eldar, 2009), the author claims that extensions of SURE for gamma noise can be obtained by a more general extension of SURE for distributions of the exponential family, i.e., of the form $p(v|u) = q(v)\exp(uv - g(u))$ where q is a function of the observation only and g is a function of the parameter only. In this case, SURE extends as follows (Hudson, 1978; Raphan and Simoncelli, 2007; Eldar, 2009):

$$R(\hat{\boldsymbol{u}}) = \sum_x u(x)^2 + \hat{u}(x)^2 + 2\left.\frac{\partial \hat{u}(x)}{\partial v(x)}\right|_{v(x)} + 2\hat{u}(x)\left.\frac{\partial \log q(v(x))}{\partial v(x)}\right|_{v(x)} . \tag{3.45}$$

Note that the true image $\boldsymbol{u}$ is unknown but since it is constant when minimizing $R(\hat{\boldsymbol{u}})$ it can be omitted in practice. Note also that SURE is independent on $g(\tilde{u})$. With the change of variable $\tilde{u}(x) = -\frac{L}{u}(x)$, the gamma distribution belongs to the exponential family with $q(v) = \frac{v^{L-1}}{\Gamma(L)}$ and $g(\tilde{u}) = L\log\tilde{u}$, leading to:

$$R(\hat{\tilde{\boldsymbol{u}}}) = \sum_x \tilde{u}(x)^2 + \hat{\tilde{u}}(x)^2 + 2\left.\frac{\partial \hat{\tilde{u}}(x)}{\partial v(x)}\right|_{v(x)} + 2(L-1)\frac{\hat{\tilde{u}}(x)}{v(x)} . \tag{3.46}$$

However, in this case, the risk is defined on $\tilde{\boldsymbol{u}}$ instead of $\boldsymbol{u}$, therefore minimizing SURE is not equivalent to minimizing the MSE but another objective function defined on the energetic landscape of $\tilde{\boldsymbol{u}}$. We will see that the minimization of this energy leads to poor results.

Eldar (2009) also suggests that this result can be used for Poisson noise. However, Poisson distributions belong to the exponential family but are not continuous distributions. Since Eq. (3.45) requires us to evaluate the variations of the estimator according to the infinitesimal variation of the noise component, another unbiased risk estimator has to be used for the case of Poisson noise.

The case of Poisson noise: The Poisson unbiased risk estimator (PURE) is an estimator of the MSE under Poisson noise (Hudson, 1974; Chen, 1975) defined by:

$$R(\hat{\boldsymbol{u}}) = \sum_x u(x)^2 + \hat{u}(x)^2 - 2v(x)\hat{\bar{u}}(x) \tag{3.47}$$

where $\boldsymbol{u}$ is the unknown true image (which is constant when minimizing $R(\hat{\boldsymbol{u}})$ it can be omitted in practice), $\boldsymbol{v}$ is the input noisy image, $\hat{\boldsymbol{u}}$ is the estimate of $\boldsymbol{u}$ and $\hat{\bar{u}}(x) = h(\bar{v}^x)|_x$ refers to the denoised value obtained by the application of the NL means on the noisy image defined by $\bar{v}^x(x') = \begin{cases} v(x') & \text{if } x' \neq x \\ v(x) - 1 & \text{otherwise} \end{cases}$.

The case of Poisson-Gaussian noise: By using (Hudson, 1974; Chen, 1975) and following (Hudson, 1978; Tsui and Press, 1982), Luisier et al. (2010) extended PURE to the case of Poisson-Gaussian noise, which is given by:

$$R(\hat{\boldsymbol{u}}) = \sum_w \hat{u}(x) - 2v(x)\hat{\bar{u}}(x) + v(x)\bar{v}(x) + 2\sigma^2\left.\frac{\partial \bar{u}(x)}{\partial v(x)}\right|_{v(x)} - \sigma^2 . \tag{3.48}$$

Examples with moving average filters

Here, we suggest finding the best global bandwidth h of the Gaussian kernel of the moving Gaussian filter (see Sec.2.2.1) in the case of images corrupted by gamma and Poisson noise using the previous proposed SURE extensions. Recall that the Gaussian filter is defined by:

$$\hat{u}(x) = \frac{\sum_{x'} w(x,x')v(x')}{\sum_{x'} w(x,x')} \quad \text{using} \quad w(x,x') = \exp\left(-\frac{\|x'-x\|^2}{2h^2}\right) . \tag{3.49}$$

Estimates of the MSE are then obtained as follows:

The case of gamma noise: Following the idea of Eldar (2009), the optimal parameter h is obtained by minimizing:

$$R(\hat{\tilde{\boldsymbol{u}}}) = \sum_w \tilde{u}(x)^2 + \hat{\tilde{u}}(x)^2 + 2L\frac{w(x,x)}{\hat{\tilde{u}}(x)} + 2(L-1)\frac{\hat{\tilde{u}}(x)}{v(x)} . \tag{3.50}$$

Figure 3.10 shows that this objective function is an unbiased estimate of the MSE expressed on the transformed data $\tilde{\boldsymbol{u}}$. Unfortunately, its minimum is not reached for the same value h as the minimum of the MSE. Minimizing Eq. (3.50) leads to non-optimal results.

The case of Poisson noise: The optimal value of h is obtained by minimizing the following objective function:

$$R(\hat{\boldsymbol{u}}) = \sum_x u(x)^2 + \hat{u}(x)^2 - 2(v(x)\hat{u}(x) - w(x,x)) \tag{3.51}$$

where $\boldsymbol{u}$ is the unknown true image (which is constant when minimizing $R(\hat{\boldsymbol{u}})$ it can be omited in practice), $\boldsymbol{v}$ is the input noisy image, $\hat{\boldsymbol{u}}$ is the estimate of $\boldsymbol{u}$. Figure 3.10 shows that this objective function is an unbiased estimate of the MSE, its minimum is reached for the same value h as the minimum of the MSE. Minimizing Eq. (3.51) leads to optimal results.

Extensions of SURE for non-Gaussian noise is a difficult task, and the solutions proposed so far do not apply easily to all situations, in particular in the case of gamma noise. Under Poisson noise, PURE provides appealing performance. The authors of (Luisier et al., 2010) recently suggested using PURE in the context of wavelet denoising. Based on PURE, we will propose in Chap. 5 an automatic setting of the parameters of the NL means (following the idea of Van De Ville and Kocher (2009)) that applies successfully for Poisson noise.

3.7 The Bayesian non-local means filter

Kervrann et al. (2007) propose an extension of the NL means designed to cope with arbitrary uncorrelated noise. The Bayesian NL means filter aims at minimizing the Bayesian risk for each patch P_x:

$$\mathbb{E}\left[\|\boldsymbol{u}(x) - \hat{\boldsymbol{u}}(x)\|_2^2\right] = \int \|\boldsymbol{u}(x) - \hat{\boldsymbol{u}}(x)\|_2^2 \, p(\boldsymbol{u}(x)|\boldsymbol{v}(x)) \, \mathrm{d}\boldsymbol{u}(x) . \tag{3.52}$$

The minimization of Eq. (3.52) leads to the *posterior* mean estimator given by:

$$\hat{\boldsymbol{u}}(x) = \int p(\boldsymbol{u}(x)|\boldsymbol{v}(x))\boldsymbol{u}(x)\mathrm{d}\boldsymbol{u}(x) . \tag{3.53}$$

The *posterior* distribution $p(\boldsymbol{u}(x)|\boldsymbol{v}(x))$ being unknown and having only one image at hand, Kervrann et al. (2007) suggest approaching Eq. (3.53) by the following selection-based estimator:

$$\hat{\boldsymbol{u}}(x) = \frac{\sum_{x'} p(\boldsymbol{v}(x)|\boldsymbol{u}(x'))\boldsymbol{u}(x')}{\sum_{x'} p(\boldsymbol{v}(x)|\boldsymbol{u}(x'))} \; . \tag{3.54}$$

where the prior $p(\boldsymbol{u}(x'))$ is considered uniform. The underlying value $\boldsymbol{u}$ being unknown, the authors propose in a first step to substitute $\boldsymbol{u}(x')$ by $\boldsymbol{v}(x')$ in Eq. (3.54). This first step then relies on the strong assumption that $\boldsymbol{v}(x')$ is a good estimate of $\boldsymbol{u}(x')$. This first-obtained pre-estimate $\overline{\boldsymbol{u}}$ is next used in a second iteration as a refinement of $\boldsymbol{u}(x')$. This yields the following estimator:

$$\hat{\boldsymbol{u}}(x) = \frac{\sum_{x'} p(\boldsymbol{v}(x)|\overline{\boldsymbol{u}}(x'))^{1/h}\overline{\boldsymbol{u}}(x')}{\sum_{x'} p(\boldsymbol{v}(x)|\overline{\boldsymbol{u}}(x'))^{1/h}} \tag{3.55}$$

where h acts as the filtering parameter of the NL means and its introduction is justified to counterbalance the invalidity of the patch independence assumption.

In the case of Gaussian noise, the first step boils down to the original blockwise NL means (Buades et al., 2005) and then it can be viewed as an extension of the NL means for non-Gaussian noise. The second step follows in the same spirit as the UINTA filter (Awate and Whitaker, 2005), where pre-estimated values are averaged and known to minimize an entropy-based criterion. Coupé et al. (2008) use the Bayesian NL means in the case of *speckle* modeled by an image-dependent Gaussian distribution with mean $u(x)$ and variance $u(x)\sigma^2$. More recently, Zhong et al. (2011) propose to use this approach for *speckle* described by a gamma distribution. However, due to the strong assumption of the first step, the authors suggest using another pre-estimation step based on the improved sigma filter that we will present in the next section. The authors show that their method leads to the *state-of-the-art* performance for *speckle* reduction.

In general, when noise departs from the Gaussian distribution, the performance of the Bayesian NL means decreases due to the poor estimate provided in the first step. Indeed, when the average is performed on the noisy values $v(x)$, the posterior mean should be replaced by the MLE (following the discussion in Sec. 3.3). Moreover, we will see that even if the sample mean is the solution of the MLE (e.g., for Poisson and gamma distributions), the Bayesian weights given in the first step by $p(\boldsymbol{v}(x)|\boldsymbol{v}(x'))^{1/h}$ lead to strong artifacts. We call a filter that combines selected or weighted noisy values a selection-based filter. We will show in the next section that in this case the Bayesian weights behave poorly and that other selection rules should be used instead.

3.8 Selection-based filtering and non-Gaussian noises

Filtering with adaptive windows presented in Sec. 2.3.2 and non-local approaches presented in Sec. 2.5.1 are two examples of selection-based filters. Selection-based filters try to collect as many samples as possible in the image such that the combination of these samples reduce the noise level while preserving the information of interest (i.e., minimize the bias-variance trade-off). There are then two ingredients in selection-based filters: the selection rule and the combination strategy. Under Gaussian noise, adaptive filters usually select samples according to their Euclidean distance and combine samples using a weighted average. If selected samples are assumed to be i.i.d., then a combination rule relevant for non-Gaussian noise can be based on the maximum likelihood estimation (MLE) as presented in Sec. 3.3. A weighted version of the MLE will be also discussed in Chap. 5. In the following, we will first see why the Euclidean distance leads to a good selection rule under Gaussian noise. Next, we will give examples of different proposed attempts including our own solution to select noisy samples for a given noise distribution.

3.8.1 Selection with Gaussian noise: why does it work?

In this section, we will try to interpret why the Euclidean distance is a good selection rule under the Gaussian noise assumption. Then, we will explain the relative performance of the sigma filter, the Yaroslavsky filter and the NL means filters.

The sigma filter

To estimate the pixel value $u = u(x)$, the sigma filter of Lee (1983) selects the pixels at location x' having a noisy value $v' = v(x')$ in the neighborhood of the value $v = v(x)$, i.e., such that:

$$v' \in [v - \gamma\sigma, v + \gamma\sigma] \tag{3.56}$$

where σ refers to the standard-deviation of the noise and $\gamma > 0$ controls the amount of filtering. It is straightforward to show that the selection rule in Eq. (3.56) is equivalent to selecting noisy candidate values with respect to the Euclidean distance with the noisy value of interest, i.e.:

$$\frac{(v' - v)^2}{\sigma^2} \leq \gamma^2 \quad \Leftrightarrow \quad \frac{|v' - v|}{\sigma} \leq \gamma \, . \tag{3.57}$$

This relation explains why selection rules are usually driven by the Euclidean distance between noisy values. Note that under the Gaussian noise assumption, the interval $\mathcal{I}(v) = [v - \gamma\sigma, v + \gamma\sigma]$ is a confidence interval, i.e., we can control the probability that $u \in \mathcal{I}(v)$. When $\gamma = 2$, this probability is given by $\mathbb{P}(u \in \mathcal{I}(v)) = 0.95\%$. This selection rule is appealing under the Gaussian noise assumption since it verifies two properties related to the evaluation criteria introduced in Sec. 2.1.3:

1. It leads to an *unbiased* estimator. Assume that the sigma filter finds candidate pixels x' verifying (3.56) such that $V' = V(x')$ are i.i.d. with $V = V(x)$ (i.e., $u(x) = u(x') = u$). The final estimate $\hat{u}$ is then unbiased:

$$\mathbb{E}[\hat{U}] = \frac{\displaystyle\int_{-\infty}^{\infty} \int_{v-\gamma\sigma}^{v+\gamma\sigma} v' p(v'|u) \, \mathrm{d}v' \, p(v|u) \, \mathrm{d}v}{\displaystyle\int_{-\infty}^{\infty} \int_{v-\gamma\sigma}^{v+\gamma\sigma} p(v'|u) \, \mathrm{d}v' \, p(v|u) \, \mathrm{d}v} = u \tag{3.58}$$

   whatever the value of γ. In practice, the number of samples is limited and the samples are not necessarily i.i.d. Hence, the smaller the value of γ, the fewer number of samples selected and the higher the resulting variance of the sigma filter. One would choose a high value for γ. However, this could lead to mix samples arising from different populations leading to bias. The parameter γ acts on the bias-variance trade-off.

2. It leads to an estimator with a *stationary relative variance*. Assume that we have N i.i.d. samples of same variance, the resulting variance is reduced by a factor independent of u:

$$\frac{\mathrm{Var}[V]}{\mathrm{Var}[\hat{U}]} = N \underbrace{\int_{-\infty}^{\infty} \int_{v-\gamma\sigma}^{v+\gamma\sigma} p(v'|u) \, \mathrm{d}v' \, p(v|u) \, \mathrm{d}v}_{\text{Percentage of noise reduction } \varepsilon} = N \int_0^{\frac{\gamma^2}{2}} \chi_1^2(t) \, \mathrm{d}t \tag{3.59}$$

   where χ_1^2 is the chi-square distribution with 1 degree of freedom. When $\gamma \to \infty$, all candidate values are averaged and the variance is reduced by a factor N. When $\gamma = 0$, none of the candidate values are selected and the variance reduction is null. Hence, ε lies

in the interval $[0, 1]$ and represents the percentage of variance reduction compared to the maximal possible variance reduction. Note that Eq. (3.59) shows that the level of variance reduction decreases when γ increases. Since the level of the noise reduction is independent of u, the estimator have a *stationary relative variance*: for a given value of γ the same level of noise reduction can be maintained everywhere in the image. This result is well known in detection theory and usually ascribed to the distribution of the Euclidean distance in (3.57). The level of variance reduction is then linked to the false-alarm rate of the similarity criterion (see Chap. 4).

Yaroslavsky's filter and the NL means

Instead of using a binary selection, the Yaroslavsky filter combines noisy values thanks to data-driven weights based on the Euclidean distance between noisy pixel values (Yaroslavsky, 1985):

$$w(x', x) = \phi\left(\frac{(v(x') - v(x))^2}{2h^2}\right) \tag{3.60}$$

where $h > 0$ controls the amount of filtering in the same vein as γ for the sigma filter. See Sec. 2.5.1 for more details about this filter. Under the Gaussian noise assumption, as soon as the weights are symmetric, the resulting estimator is unbiased and maintains the same level of noise reduction (this last property is straightforward to prove using the changes of variables $\boldsymbol{v} \to \boldsymbol{v} - \boldsymbol{u}$ and $\boldsymbol{v}' \to \boldsymbol{v}' - \boldsymbol{u}$). Hence, the efficiency of the Yaroslavsky filter can be also ascribed to the fact that it is unbiased with a *stationary relative variance*. The same properties can be shown for the NL means algorithm. The main difference is that the NL means will be able to maintain a higher level of noise reduction for the same value of h. As a consequence, the NL means filter is more robust when samples are no longer i.i.d. but rather come from a mixture of populations (see Sec. 5.3 for more details on this aspect).

When noise departs from Gaussian noise, Eq. (3.58) and (3.59) no longer hold true. The choice of a suitable selection rule for non-Gaussian noise is a difficult task that has been explored in several works. We will enumerate few of them in the next sections and highlight their advantages and drawbacks.

3.8.2 Selection rules under non-Gaussian noise

In the following we focus on the different selection rules that have been proposed for non-Gaussian noise. They have been introduced for sigma filtering, the NL means filtering or filtering with adaptive windows. However, all of them are grounded on the comparison of a candidate value v' with a value of interest v, or equivalently, on the fact that v' lies in a neighborhood $\mathcal{I}(v)$ of v:

$$v' \in \mathcal{I}(v)\ . \tag{3.61}$$

In order to have a fair comparison of these selection rules, we will study their behaviors in the case of gamma noise and Poisson noise. Note that in these two cases the average is the MLE. In the sight of the evaluation criteria proposed in Sec. 2.1.3, we suggest evaluating selection rules by studying the expectation of the resulting estimator:

$$\mathbb{E}[\hat{U}] = \frac{\displaystyle\iint_{\mathcal{I}(v)} v' p(v'|u)\ \mathrm{d}v'\ p(v|u)\ \mathrm{d}v}{\displaystyle\iint_{\mathcal{I}(v)} p(v'|u)\ \mathrm{d}v'\ p(v|u)\ \mathrm{d}v} \quad \text{should be equal to } u \tag{3.62}$$

and its percentage of noise reduction:

$$\varepsilon = \iint_{\mathcal{I}(v)} p(v'|u)\,\mathrm{d}v'\; p(v|u)\,\mathrm{d}v \quad \text{should be independent of } u\ . \tag{3.63}$$

Let us first present three selection procedures based on an oracle providing the unknown true value u. We will see that even in this case the design of a selection rule leading to an unbiased estimate which maintains the same level of noise reduction is not trivial.

Direct extension of the sigma filter

Given the variance $\mathrm{Var}[V]$ provided by the oracle, the direct extension of Eq. (3.57) for an arbitrary noise distribution would lead to the following selection rule:

$$\frac{|v'-v|}{\sqrt{\mathrm{Var}[V]}} \leq \gamma\ . \tag{3.64}$$

This rule selects pixel candidates whose noisy values lie in the following neighborhood of v:

$$\mathcal{I}(v) = \left[v - \gamma\sqrt{\mathrm{Var}[V]}, v + \gamma\sqrt{\mathrm{Var}[V]}\right]\ . \tag{3.65}$$

The case of gamma noise: Assume a gamma distribution with a shape parameter $L = 1$ (cf. Sec. 3.1.5). In this case, the variance is given by $\mathrm{Var}[V] = u^2$ resulting in the following selection rule:

$$\mathcal{I}(v) = [v - \gamma u, v + \gamma u]\ . \tag{3.66}$$

The case of Poisson noise: Assume a Poisson distribution. In this case, the variance is given by $\mathrm{Var}[V] = u$ resulting in the following selection rule:

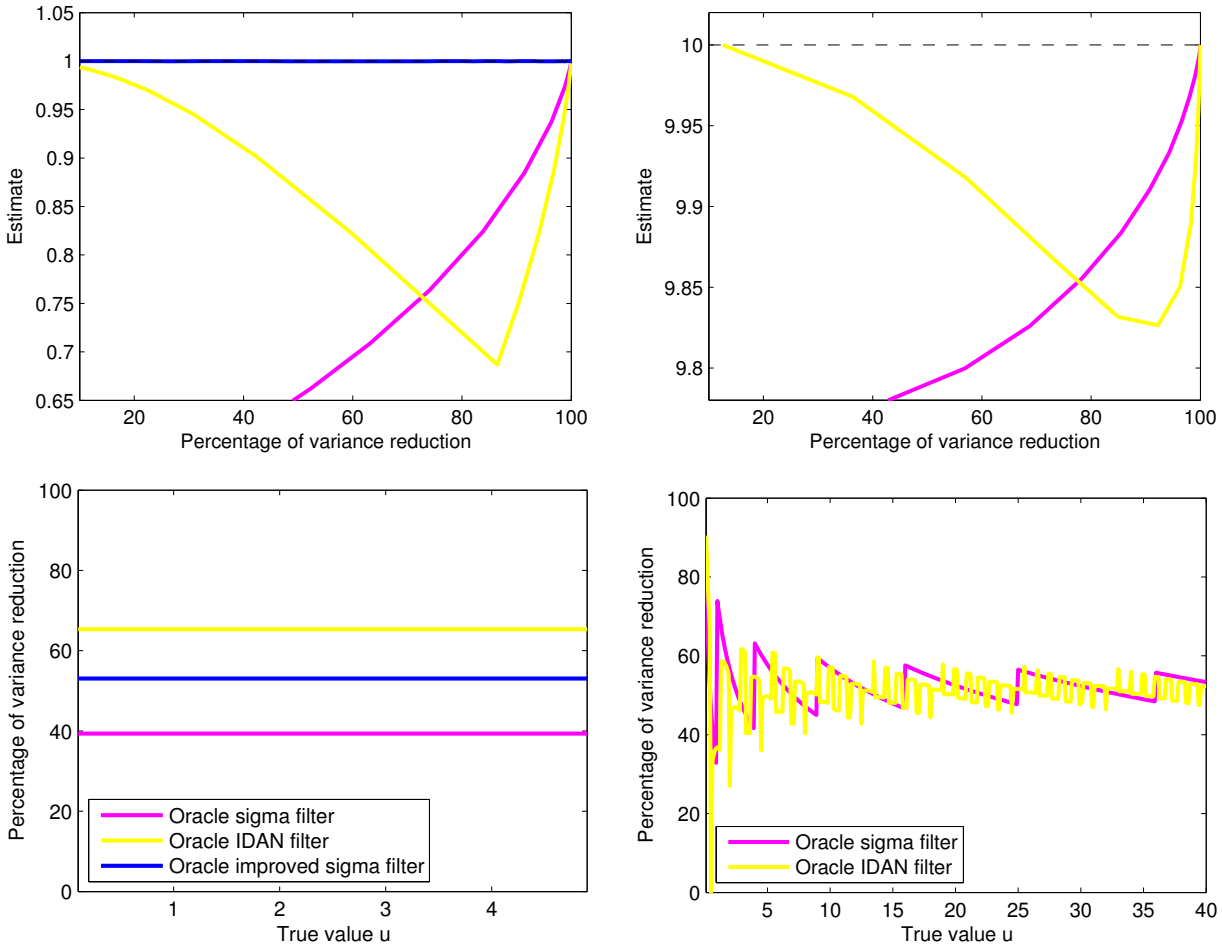
$$\mathcal{I}(v) = \left[v - \gamma\sqrt{u}, v + \gamma\sqrt{u}\right]\ . \tag{3.67}$$

This selection rule seems relevant for symmetric distributions. However, when the distribution is not symmetric, Figure 3.11 shows (thanks to numerical integrations of Eq. (3.62)) that the estimation of u is significantly underestimated under gamma and Poisson noise. This rule leads to an estimator with a *stationary relative variance* for gamma noise, whereas for Poisson noise, the percentage of noise reduction varies with u. Note that, the transitions in this percentage are due to the discrete nature of Poisson noise: the number of integer values inside the interval $\mathcal{I}(v)$ abruptly changes with respect to u.

Intensity driven adaptive neighborhood (IDAN)

The intensity driven adaptive neighborhood (IDAN) filter is a filtering method originally proposed for images damaged by gamma noise (Vasile et al., 2006). Unlike the previous solution, IDAN does not select pixel values $v' = v(x')$ with similar values as $v = v(x)$, but, with similar values as $u = u(x)$:

$$\frac{|v'-u|}{\sqrt{\mathrm{Var}[V]}} < \gamma \quad \Leftrightarrow \quad \mathcal{I}(v) = \left[u - \gamma\sqrt{\mathrm{Var}[V]}, u + \gamma\sqrt{\mathrm{Var}[V]}\right]\ . \tag{3.68}$$



(a) Gamma noise (b) Poisson noise

Figure 3.11: (top) The expectation $\mathbb{E}[\hat{U}]$ compared to the underlying value u as a function of the percentage of variance reduction ε. In (a) $u = 1$ and in (b) $u = 10$. (bottom) The percentage of variance reduction ε as a function of the underlying value u. The curves are given in (a) under a gamma noise and (b) under a Poisson noise. The selection rules are those of the direct extension of the sigma filter, the IDAN filter and the improved Lee's filter, all using an oracle providing the unknown value u.

The case of gamma noise: Assume a gamma distribution with a shape parameter $L = 1$ (cf. Sec. 3.1.5). In this case, the variance is given by $\mathrm{Var}[V] = u^2$ resulting in the following selection rule:

$$\mathcal{I}(v) = [u(1-\gamma), u(1+\gamma)] \ . \tag{3.69}$$

When $\gamma = 2/\sqrt{L}$, this selection rule has a probability of false alarm (or, equivalently, a percentage of noise reduction) of 95%.

The case of Poisson noise: Assume a Poisson distribution. In this case, the variance is given by $\mathrm{Var}[V] = u$ resulting in the following selection rule:

$$\mathcal{I}(v) = \left[u - \gamma\sqrt{u}, u + \gamma\sqrt{u}\right] \ . \tag{3.70}$$

Surprisingly, as illustrated in Fig. 3.11, the IDAN filter is also biased towards lower values. Under gamma noise, this rule leads to an estimator with a *stationary relative variance*, while

under Poisson noise, the level of noise reduction varies with u. Again, the transitions in the percentage of noise reduction in the case of Poisson noise are due to its discrete nature: the number of integer values inside the interval $\mathcal{I}(v)$ abruptly changes with respect to u.

Improved sigma filter

Recently, Lee et al. (2009) suggested improving their sigma filter for *speckle* by calculating the bound of an interval $[u\zeta_1, u\zeta_2]$ such that the estimate $\hat{u}$ converges towards the true value u, i.e.:

$$\mathbb{E}[\hat{U}] = \frac{\displaystyle\int_{u\zeta_1}^{u\zeta_2} v' p(v'|u)\, \mathrm{d}v'}{\displaystyle\int_{u\zeta_1}^{u\zeta_2} p(v'|u)\, \mathrm{d}v'} = u\ . \tag{3.71}$$

Unfortunately, it is not possible to solve Eq. (3.71) directly, i.e., to get a closed-form expression for ζ_1 and ζ_2. However, a numerical search technique can be used with an iterative algorithm to estimate ζ_1 and ζ_2 for different levels of false-alarms. Note that in the case of the gamma distribution, the use of the improved sigma filter to replace the first step of the Bayesian filter (see Sec. 3.7) has led to appealing results (Zhong et al., 2011).

Under the gamma noise assumption, this method relies on the fact that thanks to the multiplicative behavior of gamma distribution, the bounds of the optimal interval are proportional to the underlying true value u. In the case of Poisson noise, this relation does not hold anymore, and, one should compute these bounds for all possible values of u and levels of false alarms. For this reason, we have only implemented the case of the gamma distribution. Fig. 3.11 shows that, in this case, this improved sigma filter gives an unbiased estimate with a constant level of noise reduction.

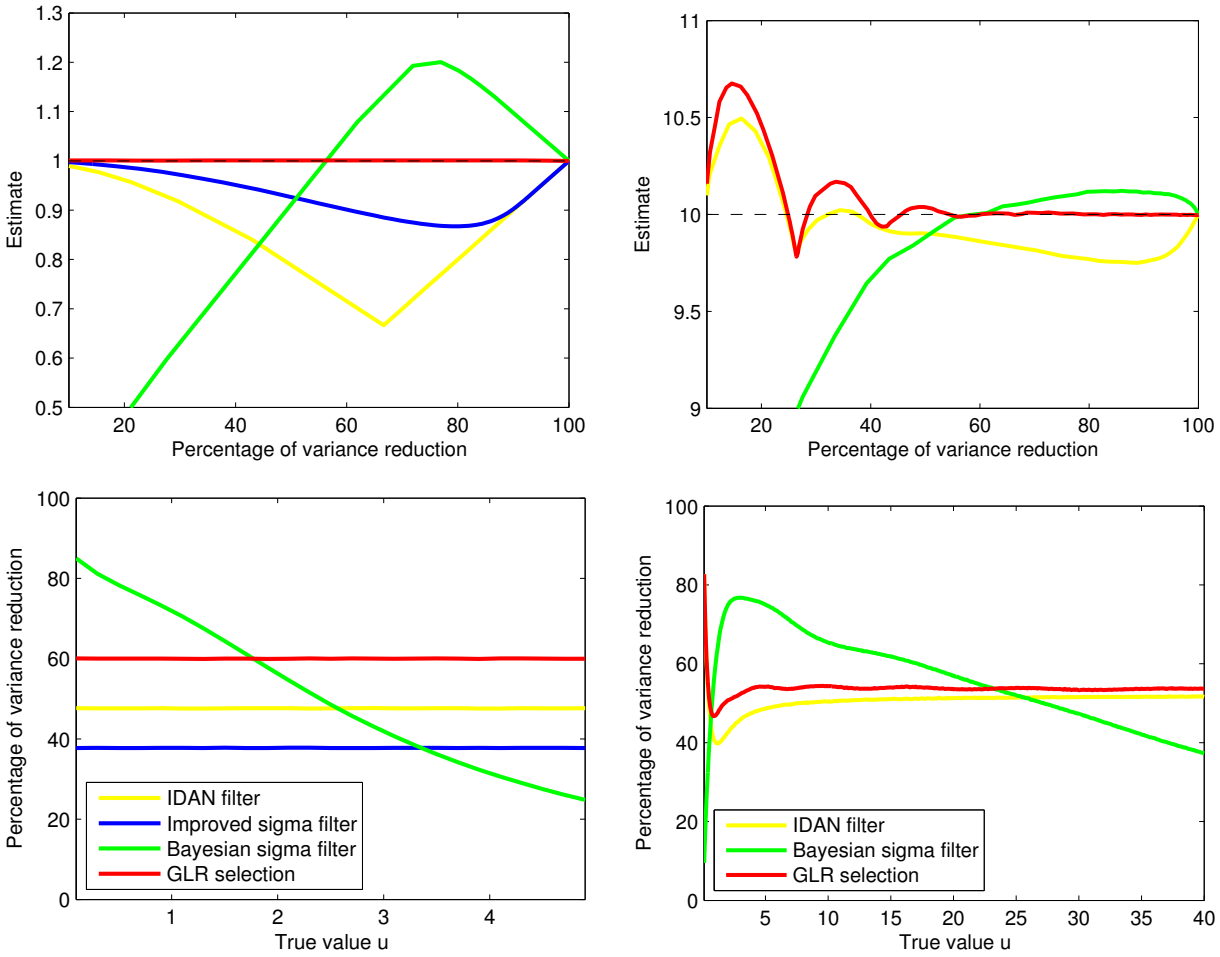
Of course, in the three last scenarios, we are in a favorable condition where the unknown noise-free value u is provided by an oracle in order to perform the selection rule. Lee et al. (2009) suggest replacing the unknown u by the estimate obtained thanks to a 3×3 boxcar filter. Vasile et al. (2006) use rather a region growing algorithm driven by the selection rule where u is substituted by its estimation inside the current region of connected pixels (see also Sec. 2.3.2). However, the substitution of u by a pre-estimate $\overline{u}$ can introduce bias or artifacts due to its possible poor quality. From a statistical point of view, this substitution also prevents us from properly studying the behavior of such approaches. In the following, we focus only on selection rules, or equivalently on similarity criteria, that do not depend on the unknown value u or an estimate $\overline{u}$ of this parameter.

Misled selection driven by the noisy data

We have studied the previous methods by substituting the unknown value u by its noisy realization v. The numerical simulations given in Fig. 3.12 show that this substitution introduces an important bias for all these approaches including the improved sigma filter. Under gamma noise, all these approaches still maintain the same level of noise reduction. Under Poisson noise, the IDAN extension succeeds in maintaining the same level of noise reduction as soon as $u > 4$. Note that the value 4 corresponds also to the threshold after which the Anscombe transform succeeds to stabilize the variance (see Sec. 3.2).

First step selection of the Bayesian NL means

The first step of the Bayesian NL means presented in Sec. 3.7 can be viewed as a selection-based filter averaging noisy values. In order to compare this approach to other selection rules, let us



(a) Gamma noise (b) Poisson noise

Figure 3.12: (top) The expectation $\mathbb{E}[\hat{U}]$ compared to the underlying value u as a function of the percentage of variance reduction ε. In (a) $u = 1$ and in (b) $u = 10$. (bottom) The percentage of variance reduction ε as a function of the underlying value u. The curves are given in (a) under a gamma noise and (b) under a Poisson noise. The selection rules are those of the IDAN filter, the improved sigma filter the bayesian sigma filter and the GLR based filter when only noisy observations are available.

simplify it with a binary selection and a selection based on the comparison of noisy pixel values instead of noisy patches. In other words, we simplify the Bayesian NL means to a "Bayesian sigma filter". This yields a selection rule that selects pixel values according to the negative log-likelihood $-\log p(v|v')$ rather than the Euclidean distance $|v - v'|$. The resulting selection rules for gamma noise and Poisson noise are given as follows:

The case of gamma noise: The application of the Bayesian methodology leads to the following selection rule (see also Zhong et al., 2011):

$$-\frac{L-1}{L}\log v + \log v' + \frac{v}{v'} < \gamma \tag{3.72}$$

Note that in this case the corresponding interval $\mathcal{I}(v)$ cannot be obtained in closed-form. Contrary to all other selection rules that we have mentioned so far, this selection rule is not invariant up to a multiplicative constant. If v' is selected when observing v, it does not mean that $\alpha v'$

will be selected when observing αv. Gamma noise being multiplicative, this phenomenon sounds counter-intuitive and can explain the bad performance of such a selection rule.

The case of Poisson noise: The application of the Bayesian methodology leads to the following selection rule:

$$-v \log v' + v' + \log v! < \gamma \; . \tag{3.73}$$

Note that in this case the corresponding interval $\mathcal{I}(v)$ cannot be obtained in closed-form.

The numerical simulations presented in Fig. 3.12 show that this approach suffers from an important bias, which is, moreover non-linearly dependent of the underlying value u. Worse, this rule does not lead to an estimator with a *stationary relative variance* neither for gamma noise nor Poisson noise.

3.9 A new selection rule based on the generalized likelihood ratio

With the idea of comparing noisy values directly instead of a noisy value against a (pre-estimated) noise-free value, we derive in Chap. 4 a similarity criterion grounded on the generalized likelihood ratio (GLR) and resulting in the following statistical selection rule:

$$-\log \frac{\sup_u p(v|u)p(v'|u)}{(\sup_u p(v'|u))\,(\sup_u p(v|u))} < \gamma \; . \tag{3.74}$$

This selection rule aims at selecting candidate noisy values v' coming from the same population as the value of interest v. In Chap. 4, we will see that the underlying similarity criterion is asymptotically a constant false-alarm rate detector, meaning that asymptotically to the dimension D of the input data, it ensures to maintain the same level of noise reduction whatever the distribution at hand.

The case of gamma noise: The application of GLR for the case of gamma distributions leads to the following selection rule:

$$\log\left(\sqrt{\frac{v}{v'}} + \sqrt{\frac{v'}{v}}\right) < \gamma \quad \Leftrightarrow \quad \mathcal{I}(v) = \left[v\left(\frac{\eta - \sqrt{\eta^2-4}}{2}\right)^2, v\left(\frac{\eta + \sqrt{\eta^2-4}}{2}\right)^2\right] \tag{3.75}$$

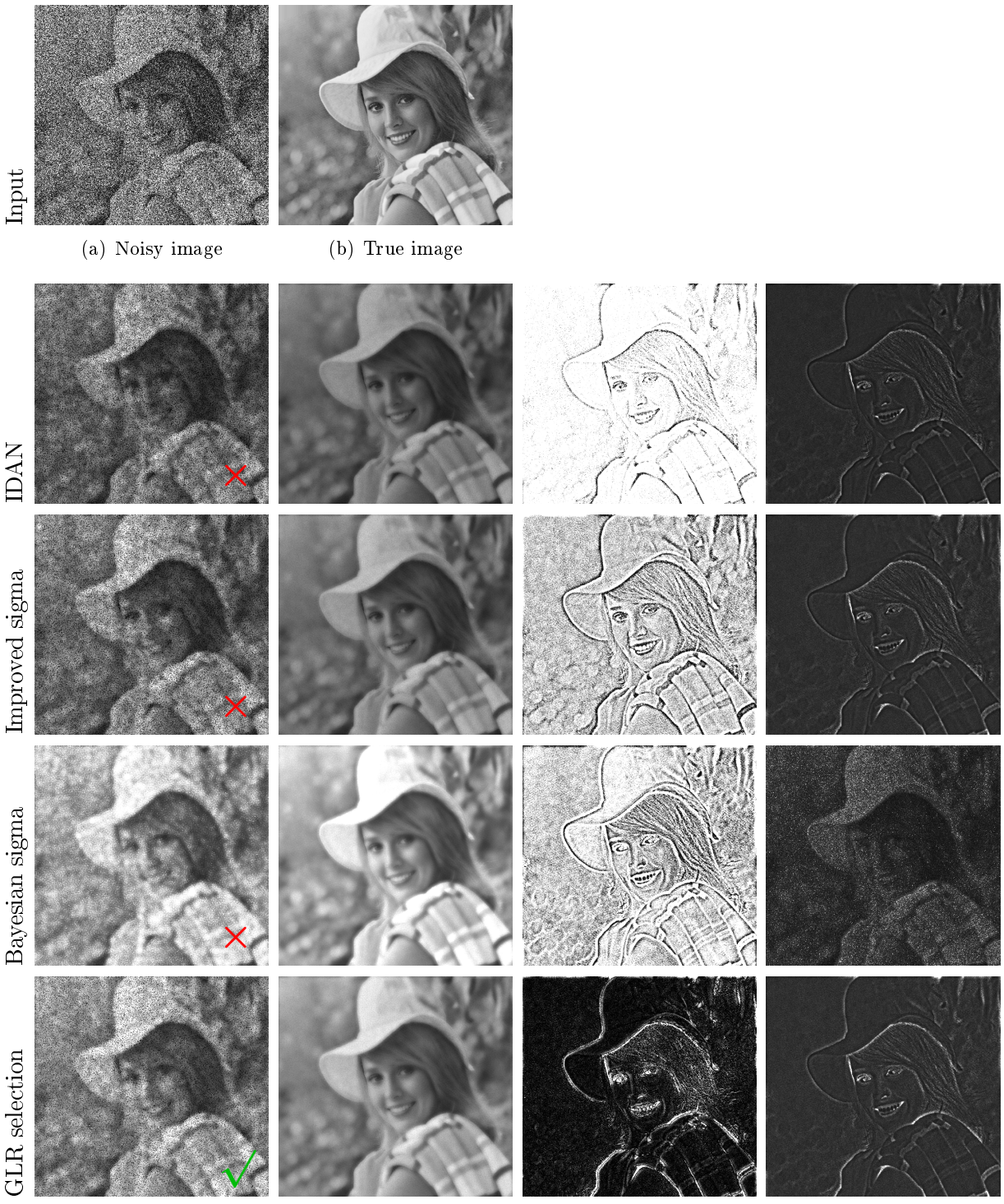
where $\eta = \exp(\gamma)$. Unlike Eq. (3.72), this rule is invariant up to a multiplicative constant affecting both v and v'. In the context of non-local filtering, we have first derived Eq. (3.75) in (Deledalle et al., 2009b) as the Bayesian joint likelihood rather than the GLR (see Chap. 4 for details). After this publication, this selection rule has been successfully applied for the extension of the BM3D procedure in (Parrilli et al., 2010) and for the estimation of polarimetric data in (Chen et al., 2011). Teuber and Lang (2011) have also derived other extensions of our work in the Bayesian joint likelihood framework.

The case of Poisson noise: The application of GLR for the case of Poisson distributions leads to the following selection rule:

$$v \log v + v' \log v' - (v + v') \log\left(\frac{v + v'}{2}\right) < \gamma \; . \tag{3.76}$$

Note that in this case the corresponding interval $\mathcal{I}(v)$ cannot be obtained in closed-form.

Figure 3.12 shows that GLR outperforms all other selection rules driven by the noisy data. The bias is null for gamma noise while it tends to zero for Poisson noise. The level of variance



(a) Noisy image (b) True image

(c) Estimation (d) Expectation (e) Relative bias^2 (f) Variance reduction

Figure 3.13: (a) Noisy realizations of a gamma noise corrupting (b) a noise-free image. (c) The estimates obtained by the four studied selection rules. (d) The expectation $\mathbb{E}[\hat{\boldsymbol{U}}]$, (e) the relative square bias, i.e., $(\mathbb{E}[\hat{\boldsymbol{U}}] - \boldsymbol{u})^2/\boldsymbol{u}^2$ and (f) the *relative variance*, i.e., $\mathrm{Var}[\hat{\boldsymbol{U}}]/\boldsymbol{u}^2$ estimated over 100 noisy realizations.

reduction is maintained in the case of gamma noise. Under the Poisson noise assumption, this level is constant as soon as $u > 4$. It seems that the value 4 is an inherent lower bound of Poisson noise below which denoising is much more difficult.

Note that this result is in contradiction with those of (Zhong et al., 2011) where they show that our filter is biased towards higher values. We assume that the authors used on an intensity image our online implementation designed for data in amplitude format. We can indeed prove that such a misuse of our methodology leads to a result biased towards higher values.

Figure 3.13 compares the performance of the selection-based filters using the four aforementioned selection rules on an image corrupted by gamma noise. All of them consist of averaging the selected noisy pixels in a sliding window according to the given selection rule. The filtering parameter γ has been chosen such that all methods reduce the noise variance by 80% in average. With such a level of noise reduction, as predicted by the study in Fig. 3.12, the IDAN filter and the improved Lee filter are biased towards lower values, the Bayesian sigma filter is biased towards higher values and the GLR selection leads to an unbiased result in homogeneous areas. The bias factor of the IDAN filter and the improved filter is the same in the different homogeneous areas of the image (even if the underlying noise-free values are different). Hence, in the same vein as the post-processing step performed in the context of the variance stabilization (see Sec. 3.2), a debiasing post-processing step could be performed by a simple rescaling of the estimated values by a multiplicative constant. This simple strategy could not be used for the Bayesian sigma filter whose bias factor depends non-linearly on the underlying noise-free value. Also predicted by the study in Fig. 3.12, the IDAN filter, the improved Lee filter and the GLR-based filter have a *stationary relative variance* (same level, here 80% of noise reduction in homogeneous areas) while the Bayesian sigma filter has clearly a non-stationary *relative variance*.

3.10 Conclusion

In this chapter we have reviewed many extensions of standard denoising approaches to deal with non-Gaussian noises. After having illustrated that the simple stabilization approach can lead to non-optimal solutions, we have showed through several examples that good solutions can be obtained if the statistics of the noise are taken into account. According to the given filter, its extension considering the noise statistics can be more or less simple. For instance, the moving average and the ROF models extend easily to non-Gaussian noise while suitable unbiased risk estimates are more complicated to derive when noise departs from Gaussian noise.

Next, we have studied the behavior of selection rules based on the comparisons of punctual values under non-Gaussian noise. As performance criteria, we studied the bias of the estimates and the *stationarity* of their *relative variance* (the ability to maintain the same level of noise reduction in homogeneous areas) of different selection rules under the assumption that samples are i.i.d. We have proven that most proposed solutions are biased especially with Poisson noises. We have quantified their bias as a function of the level of noise reduction. We have then proposed a new selection rule based on the GLR test. This solution has an appealing behavior in terms of both *bias* and *relative variance* in homogeneous areas. We have not studied the behavior of such criteria in non-homogneneous areas. Maximizing the probability of detection of similarity criteria is essential to avoid bias when samples are no longer i.i.d. We will see in Chap. 4 that GLR also provides good probabilities of detection. We will explain in Chap. 4 and 5 how selections based on the comparisons of patches and suitable similarity criteria improve the robustness of selection rules under non-Gaussian noise.

Chapter 4

Robust patch-similarity under non-Gaussian noise

Estimated patch-similarity (or patch-dissimilarity) is at the heart of numerous image processing methods, e.g., region-based methods for image registration (Zitova and Flusser, 2003), matching in stereo-vision (Scharstein and Szeliski, 2002) or block selection for denoising (Buades et al., 2005). Similarity between pixel values has been defined in many different ways in the literature, depending on the vision problem at hand, the noise model and the prior knowledge. While the shape and size of patches should adapt to the multi-scale and anisotropic behaviour of natural images (Dabov et al., 2008; Deledalle et al., 2011c), the choice of the similarity criterion is rather a problem related to the nature of noise. When comparing noisy patches, adaptation to noise distribution is essential for robust similarity evaluation.

A fundamental difficulty when comparing two patches from "real" data is to decide whether the differences should be ascribed to noise or intrinsic dissimilarity. Gaussian noise assumption leads to the classical definition of patch similarity based on the square differences of intensities. For the case where noise departs from the Gaussian distribution, several similarity criteria have been proposed in the literature of image processing, detection theory and machine learning.

We focus in the following on how to compare noisy values, and how similarity criteria can be derived from a given noise distribution. The comparison of noise-free patches (design of a suitable metric in noise-free patches space) and the similarity between a noise-free and a noisy version of a patch (template matching) are out of the scope of this thesis.

There have been few attempts to define a methodology for the derivation of patch-similarity criteria adapted to given noise distributions. In the context of image block matching, Alter et al. (2006) were among the first to address this problem. They have shown that their criterion, based on maximum likelihood estimation, improves over the classical Euclidean-distance. This criterion has later been refined by Matsushita and Lin (2007) to avoid the maximum likelihood estimation step and to better take into account the shape of the likelihood distributions. This corresponds also to the approach considered in our previous work on patch-based denoising with non-Gaussian noise, for multiplicative noise (Deledalle et al., 2009b), impulsive noise (Deledalle et al., 2009a) and multi-dimensional complex data with circular complex Gaussian likelihood (Deledalle et al., 2011a).

Matsushita's approach has, however, several limitations: the criterion is hard to derive in closed-form, it requires defining a *prior* model and its performance depends heavily on the choice of the representation domain of the observations. The latter limitation has recently been pointed out by Teuber and Lang (Teuber and Lang, 2011) who showed that the criterion we proposed

for multiplicative noise in (Deledalle et al., 2009b) leads to different expressions whether it is derived for squared data or log-transformed data. Depending on the transformation choice, such criteria can lead to the following paradox: two different values can be more similar than two identical values. It appears that this result has been known since 1995 in the community of pattern recognition and information theory. Indeed, Matsushita's criterion can be traced back to the *stochastic equivalence predicate* introduced by Yianilos (1995) on metric learning where the above paradox is referred to as *the self-recognition paradox*.

At the end of the 90s, Minka (2000) exhibited an equivalence between *canonical distance measures*, developed in (Baxter, 1995; Baxter and Bartlett, 1998), and the work of Yianilos, thanks to a Bayesian formulation based on *prior* distributions. He referred to his criterion as the evidence ratio and linked it to mutual information (Minka, 1998). Concurrently, in the context of machine learning, Seeger (2002) introduced the mutual information kernel as an inner product in a high dimensional space. As he stated himself, his kernel can be also interpreted as a Bayesian extension of Yianilos' criterion. Compared to (Yianilos, 1995; Alter et al., 2006; Matsushita and Lin, 2007), their methodology provides criteria with unchanged expression whatever the representation of the observations, and, as we show in Sec. 4.2, Seeger's criterion solves *the self-recognition paradox*. A common limitation to all these approaches is the introduction of a *prior* model on the distribution of the underlying noise-free values.

Recently, we have introduced another criterion used in the case of Poisson noise in (Deledalle et al., 2010c) which can be viewed as a combination or unification of (Minka, 2000; Seeger, 2002; Alter et al., 2006). Independently, Chen et al. (2011) proposed a similar definition for complex Wishart distributions. This methodology is *prior*-less, independent of the representation of the observations and solves *the self-recognition paradox* under reasonable assumptions. In this chapter, we show that it corresponds to the generalized likelihood ratio test.

Our contributions— We address the problem of defining patch similarity under non-Gaussian noise. We first propose to express formally patch dissimilarity as a statistical test. In the light of this test, we describe several similarity criteria proposed in the literature and discuss their theoretical grounding. The definition of patch dissimilarity as a statistical test provides a new point of view on these criteria driven by many years of research on detection theory.

We consider the properties that a satisfying similarity criterion should fulfill and discuss which properties each criterion fulfills. This provides arguments in favour of well-behaved criteria.

We then turn to a task-based evaluation of the criteria. We compare the ability of each criterion to discriminate patches from a dictionary learnt on a natural image. The performance of each criterion is assessed for non-local denoising under Poisson and gamma noises. We illustrate the use of non-quadratic matching costs in stereo matching when the stereo pair is corrupted by non-Gaussian noise. In a motion-tracking problem for glacier monitoring, we show the superiority of a similarity criterion designed for the multiplicative speckle noise that occurs in synthetic aperture radar (SAR) images.

We advocate that generalized likelihood ratio offers a flexible yet powerful way to generalize patch similarity to non-Gaussian noises. Beyond dissimilarity detection, task-specific weighting of the similarity criterion is required to reach optimal performance. For low to moderate noise levels, quadratic difference computed on stabilized-variance data proves preferable to unweighted use of other criteria.

Organization of the chapter— In Sec. 4.1, we introduce seven criteria by expressing patch (dis)similarity as a detection test under a given noise model. Some desirable properties of similarity criteria necessary for comparing patches are then discussed in Sec. 4.2. Task-based evaluation of the criteria is performed in Sec. 4.3, using four different tasks: patch discrimination, image denoising, stereo-matching and motion-tracking under gamma and Poisson noises. The generalized

likelihood ratio is shown to be both easy to derive and powerful in these diverse applications. We discuss the importance of adapting patch similarity to noise models in Sec. 4.4 and draw some conclusions from our comparisons of similarity criteria.

4.1 Patch similarity criteria

In this section, we propose to express the similarity between noisy patches based on the detection of dissimilarity. Noisy patches are modeled in a probabilistic way in order to take into account the noise statistics. The notations are given as well as the fundamental concepts of detection theory. Seven criteria, extracted from the fields of image processing, detection theory and machine learning, are studied. Their concepts, origins and motivations are given. Their theoretical performance and limitations to solve our detection problem are then discussed.

By $\boldsymbol{v}$ we denote a patch, i.e., a collection of N observations (pixel values) extracted in the same local neighborhood. We do not specify here a shape for the patch but consider that the values in vector $\boldsymbol{v}$ are ordered so that when two patches $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are compared, values with identical index are in (spatial) correspondence.

We assume that noise is modeled by a given distribution so that a noisy patch $\boldsymbol{v}$ is a realization of an N-dimensional random vector $\boldsymbol{V}$ characterized by the probability density function (pdf) $p(\boldsymbol{v}|\boldsymbol{u})$. The vector of parameters $\boldsymbol{u}$ of that pdf is referred in the following to as the noise-free patch[1].

Detecting dissimilarity: a pair of (noisy) patches $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ is considered similar (i.e., in-match) when $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are realizations of independent random variables $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ following the same parametric distribution of common parameter $\boldsymbol{u}_{12}$ (i.e., the underlying noise-free patch). The evaluation of the similarity between noisy patches can then be rephrased as the following hypothesis test (i.e., a parameter test):

$$\mathscr{H}_0 : \boldsymbol{u}_1 = \boldsymbol{u}_2 \equiv \boldsymbol{u}_{12} \qquad \text{(null hypothesis)}, \tag{4.1}$$

$$\mathscr{H}_1 : \boldsymbol{u}_1 \neq \boldsymbol{u}_2 \qquad \text{(alternative hypothesis)}. \tag{4.2}$$

A similarity criterion $\mathscr{C}_{\boldsymbol{V}_1,\boldsymbol{V}_2}$ (written $\mathscr{C}$ in short) defines a mapping from a pair of noisy patches $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ to a real value. The larger the value of $\mathscr{C}(\boldsymbol{x}_1, \boldsymbol{x}_2)$, the more similar $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are considered to be. For a given similarity criterion $\mathscr{C}$, the *probability of false alarm* (to decide $\mathscr{H}_1$ under $\mathscr{H}_0$) and the *probability of detection* (to decide $\mathscr{H}_1$ under $\mathscr{H}_1$) are defined as:

$$P_{FA}^{\mathscr{C}}(\tau) = \mathbb{P}(\mathscr{C}(\boldsymbol{V}_1, \boldsymbol{V}_2) < \tau | \boldsymbol{u}_{12}, \mathscr{H}_0), \tag{4.3}$$

$$P_{D}^{\mathscr{C}}(\tau) = \mathbb{P}(\mathscr{C}(\boldsymbol{V}_1, \boldsymbol{V}_2) < \tau | \boldsymbol{u}_1, \boldsymbol{u}_2, \mathscr{H}_1). \tag{4.4}$$

where τ is a real threshold value. Note that the inequality symbols are reversed compared to usual definitions since we consider detecting dissimilarity based on similarity measure $\mathscr{C}$.

According to the Neyman-Pearson theorem, the optimal criterion, i.e., the criterion which maximizes P_D for any given P_{FA}, is the likelihood ratio test (see Kay, 1998):

$$\mathcal{L}(\boldsymbol{v}_1, \boldsymbol{v}_2) = \frac{p(\boldsymbol{v}_1, \boldsymbol{v}_2 | \boldsymbol{u}_{12}, \mathscr{H}_0)}{p(\boldsymbol{v}_1, \boldsymbol{v}_2 | \boldsymbol{u}_1, \boldsymbol{u}_2, \mathscr{H}_1)}. \tag{4.5}$$

The application of the likelihood ratio test requires the knowledge of the parameters $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ and $\boldsymbol{u}_{12}$ (the noise-free patches) which are, in practice, unavailable. The problem is thus a *composite hypothesis problem*.

[1] the vector of parameters $\boldsymbol{u}$ may have a different number of dimensions than noisy patches $\boldsymbol{v}$

Kendall and Stuart (1979) showed that there is no uniformly most powerful (UMP) detector for such *composite hypothesis problem*, i.e, any criterion $\mathscr{C}$ can be defeated by another criterion $\mathscr{C}'$ at a specific false alarm rate:

$$\forall \mathscr{C}, \exists \mathscr{C}', \tau, \tau' \qquad \text{such that} \qquad P_{FA}^{\mathscr{C}}(\tau) = P_{FA}^{\mathscr{C}'}(\tau') \text{ and } P_{D}^{\mathscr{C}}(\tau) < P_{D}^{\mathscr{C}'}(\tau') \,. \tag{4.6}$$

The research of a universal similarity criterion is then futile. We review in the following seven similarity criteria in the light of dissimilarity detection. We then turn to task-based evaluation of the criteria on natural images.

4.1.1 Euclidean distance and Gaussian kernel

The usual way to measure the similarity between two noisy patches is to consider their Euclidean distance:

$$\mathcal{D}(\boldsymbol{v}_1, \boldsymbol{v}_2) = \|\boldsymbol{v}_1 - \boldsymbol{v}_2\|_2^2. \tag{4.7}$$

$\mathcal{D}$ is minimal when the two patches $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are identical. It is common to use an exponential kernel of bandwidth $h > 0$, leading to the following similarity criterion:

$$\mathcal{G}(\boldsymbol{v}_1, \boldsymbol{v}_2) = \exp\left(-\tfrac{1}{h}\|\boldsymbol{v}_1 - \boldsymbol{v}_2\|_2^2\right), \tag{4.8}$$

or if noise is correlated with covariance matrix $\boldsymbol{\Gamma}$, by substituting $\mathcal{D}$ with the Mahalanobis distance:

$$\mathcal{G}(\boldsymbol{v}_1, \boldsymbol{v}_2) = \exp\left[-\tfrac{1}{h}(\boldsymbol{v}_1 - \boldsymbol{v}_2)^t \boldsymbol{\Gamma}^{-1}(\boldsymbol{v}_1 - \boldsymbol{v}_2)\right]. \tag{4.9}$$

Under the assumption of Gaussian noise, all the similarity criteria we consider in the following boil down to this same expression. There is then more than one way to justify or interpret the expression of the similarity criterion $\mathcal{G}$ in that case. For this reason and its link with Gaussian kernels, $\mathcal{G}$ will be referred as the Gaussian kernel.

Under Gaussian noise assumption, the distribution[2] of $\mathcal{G}$ can be used to choose a threshold τ with a given P_{FA} value. It is a *constant false alarm rate detector* (CFAR), which means that a constant P_{FA} can be maintained with a given τ independently of the underlying noise-free patches.

The performance of this criterion however drops when noise departs from a Gaussian distribution. While parameter h in Eq. (4.8) could be set globally from the noise variance, difficulties arise when the variance is signal-dependent, and therefore varies between and inside patches. A classical approach to extend the applicability of Euclidean distance to some non-Gaussian noise distributions is to stabilize the variance with a suitable mapping of the input noisy data (see Sec. 3.2). This approach has been used for patch selection (i.e., block-matching) in many denoising algorithms (e.g. Mäkitalo et al., 2010; Boulanger et al., 2010; Mäkitalo and Foi, 2011).

Given an invertible application $\boldsymbol{s}$ which stabilizes the variance for a specific noise pdf, the similarity is computed from the transformed patches:

$$\mathcal{S}(\boldsymbol{v}_1, \boldsymbol{v}_2) = \mathcal{G}(\boldsymbol{s}(\boldsymbol{v}_1), \boldsymbol{s}(\boldsymbol{v}_2)). \tag{4.10}$$

An important limitation lies in the non-linear distortion of noise-free patches introduced by $\boldsymbol{s}$. For instance, in the homomorphic approach, the logarithm transforms the contrast of noise-free patches; the performance is affected accordingly. A more fundamental limit is the nonexistence of a variance stabilizing transform $\boldsymbol{s}$ for some distributions. See Sec. 3.2 for a discussion on the limits of the stabilization-transform approaches.

[2] $\log(\mathcal{G})$ follows a Chi square distribution

4.1.2 Likelihood ratio extensions

Motivated by optimality guarantees of the likelihood ratio test $\mathcal{L}$ given in Eq. (4.5), similarity criteria can be defined from statistical tests designed for *composite hypothesis problems*.

The Bayesian likelihood ratio $\mathcal{L}_B$ considers noise-free patches as realizations of random vectors with known *prior* pdf:

$$\mathcal{L}_B(\boldsymbol{v}_1, \boldsymbol{v}_2) = \frac{p(\boldsymbol{v}_1, \boldsymbol{v}_2|\mathscr{H}_0)}{p(\boldsymbol{v}_1, \boldsymbol{v}_2|\mathscr{H}_1)} = \frac{\int p(\boldsymbol{v}_1|\boldsymbol{u}_{12}=\boldsymbol{t})p(\boldsymbol{v}_2|\boldsymbol{u}_{12}=\boldsymbol{t})p(\boldsymbol{u}_{12}=\boldsymbol{t})\,\mathrm{d}\boldsymbol{t}}{\int p(\boldsymbol{v}_1|\boldsymbol{u}_1=\boldsymbol{t}_1)p(\boldsymbol{u}_1=\boldsymbol{t}_1)\mathrm{d}\boldsymbol{t}_1 \int p(\boldsymbol{v}_2|\boldsymbol{u}_2=\boldsymbol{t}_2)p(\boldsymbol{u}_2=\boldsymbol{t}_2)\mathrm{d}\boldsymbol{t}_2}. \quad (4.11)$$

With perfect knowledge of *prior* pdf $p(\boldsymbol{u}_1)$, $p(\boldsymbol{u}_2)$ and $p(\boldsymbol{u}_{12})$, Eq. (4.11) leads to an optimal Neyman-Pearson detector (see Kay, 1998).

This criterion has been used in the context of classification: Minka (2000) exhibits a relationship between $\mathcal{L}_B$ and the *canonical distance measure* minimizing errors in nearest neighborhood classifiers. He also relates $\mathcal{L}_B$ to mutual information: the more additional knowledge is brought by $\boldsymbol{v}_2$ compared to the observation of $\boldsymbol{v}_1$ alone, the more dissimilar the underlying parameters are (Minka, 1998).

Despite its theoretical performance, this approach suffers from two drawbacks in practice. First, it requires computation of integrals which, depending on the distributions, may not be known in closed-form and therefore are time-consuming to evaluate numerically. Second, it requires knowledge of the *prior* pdf. In the absence of a statistical model of noise-free patches, a *non-informative prior* can be used. Jeffreys' *prior* is independent upon the choice of the noise-free patch representation (e.g., testing that two gamma random values share identical standard deviations $u_{12} = \sigma$ or identical variances $u_{12} = \sigma^2$ leads to the same expression of $\mathcal{L}_B$ when Jeffreys' *prior* are used).

Rather than modeling noise-free patches as random variables, the generalized likelihood ratio $\mathcal{L}_G$ replaces $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ and $\boldsymbol{u}_{12}$ in Eq. (4.5) by their maximum likelihood estimates (MLE) under each hypothesis:

$$\mathcal{L}_G(\boldsymbol{v}_1, \boldsymbol{v}_2) = \frac{\sup_{\boldsymbol{t}} p(\boldsymbol{v}_1, \boldsymbol{v}_2|\boldsymbol{u}_{12} = \boldsymbol{t}, \mathscr{H}_0)}{\sup_{\boldsymbol{t}_1, \boldsymbol{t}_2} p(\boldsymbol{v}_1, \boldsymbol{v}_2|\boldsymbol{u}_1 = \boldsymbol{t}_1, \boldsymbol{u}_2 = \boldsymbol{t}_2, \mathscr{H}_1)} = \frac{p(\boldsymbol{v}_1|\boldsymbol{u}_1 = \hat{\boldsymbol{t}}_{12})p(\boldsymbol{v}_2|\boldsymbol{u}_2 = \hat{\boldsymbol{t}}_{12})}{p(\boldsymbol{v}_1|\boldsymbol{u}_1 = \hat{\boldsymbol{t}}_1)p(\boldsymbol{v}_2|\boldsymbol{u}_2 = \hat{\boldsymbol{t}}_2)}. \quad (4.12)$$

For low levels of noise, the MLE is very close to the true value and $\mathcal{L}_G$ approaches $\mathcal{L}$. As a consequence, the distribution of $\mathcal{L}_G$ is asymptotically known for low noise levels. It results that P_{FA} values associated to any given threshold τ are known: $\mathcal{L}_G$ is asymptotically CFAR (asymptotically to vanishing levels of noise). $\mathcal{L}_G$ is also asymptotically UMP among all invariant tests (see Sec. 4.2 and Lehmann, 1959).

This criterion has been introduced in Chap. 3 in the context of selection-based filtering. It appeared that it leads to an *unbiased* estimator with *stationary relative variance* in homogeneous areas. Stability is a behaviour directly related to the CFAR property of GLR. Now, with the asymptotical UMP property of GLR, we can also claim that this estimator performs the best bias-variance trade-off for any given level of noise reduction.

Compared to the Bayesian likelihood ratio $\mathcal{L}_B$, the generalized likelihood ratio $\mathcal{L}_G$ is easier to implement, since it requires only the computation of the MLE (generally known in closed-form, or estimated in few iterations), and does not require (nor rely on) the definition of a *prior* model.

The main drawback of $\mathcal{L}_G$ lies in the lack of theoretical guarantees on how it behaves in low signal-to-noise ratio (SNR) conditions (i.e., for too small patches according to the noise level). It is known that, for low SNR and specific applications, $\mathcal{L}_G$ can be defeated by other invariant detectors (Kim and Hero III, 2001). This limitation is due to its dependency on MLE which behaves poorly for low SNR (e.g., the $\mathcal{L}_G$ that two random Gaussian vectors share an identical covariance matrix $\boldsymbol{u}_{12}$ is undefined since MLE of $\boldsymbol{u}_1$ from $\boldsymbol{v_1}$ only would not be positive definite).

	Max. self sim.	Eq. self sim.	Id. of indiscernible	Invariance	Asym. CFAR	Asym. UMPI
$\mathcal{Q}_B$	×	×	×	×	×	×
$\mathcal{Q}_G$	×	×	×	×	×	×
$\mathcal{L}_B$	×	×	×	√	×	×
$\mathcal{L}_G$	√	√	√(†)	√	√	√
$\mathcal{K}_B$	√	√	√(‡)	√	×	×
$\bar{\mathcal{G}}$	√	√	√	×	×	×
$\mathcal{S}$	√(⋆)	√(⋆)	√(⋆)	√(⋆)	√(⋆)	×

Table 4.1: Properties of the different studied criteria. Legend: (√) the criterion holds, (×) the criterion does not hold. Holds only if the observations are statistically identifiable (†) through their MLE or (‡) through their likelihood (such assumptions are frequently true). (⋆) Holds only for an exact variance stabilizing transform $s(\cdot)$ (such an assumption is usually wrong). A sketch of the proofs of all these properties is given in Appendix A.

4.1.3 Joint likelihood criteria

Other criteria use the joint likelihood of observations under $\mathscr{H}_0$ to evaluate similarities between noisy data. This leads to the Bayesian joint likelihood criteria (Yianilos, 1995; Seeger, 2002; Matsushita and Lin, 2007; Deledalle et al., 2009b):

$$\mathcal{Q}_B(\boldsymbol{v}_1, \boldsymbol{v}_2) = p(\boldsymbol{v}_1, \boldsymbol{v}_2|\mathscr{H}_0) = \int p(\boldsymbol{v}_1|\boldsymbol{u}_1 = \boldsymbol{t}) p(\boldsymbol{v}_2|\boldsymbol{u}_2 = \boldsymbol{t}) p(\boldsymbol{u}_{12} = \boldsymbol{t}) \, \mathrm{d}\boldsymbol{t} \tag{4.13}$$

or, following the simplification of the generalized likelihood ratio, the maximum joint likelihood (Alter et al., 2006):

$$\mathcal{Q}_G(\boldsymbol{v}_1, \boldsymbol{v}_2) = \sup_{\boldsymbol{t}} p(\boldsymbol{v}_1, \boldsymbol{v}_2|\boldsymbol{u}_{12} = \boldsymbol{t}, \mathscr{H}_0) = p(\boldsymbol{v}_1|\boldsymbol{u}_1 = \hat{\boldsymbol{t}}_{12}) p(\boldsymbol{v}_2|\boldsymbol{u}_2 = \hat{\boldsymbol{t}}_{12}) \, . \tag{4.14}$$

Such criteria have been designed to measure the likelihood of sharing a common parameter. However, the likelihood provides relative information compared to the likelihoods of other hypotheses. The evaluation of the joint likelihood under $\mathscr{H}_0$ cannot provide information if it is not confronted against the alternative hypothesis $\mathscr{H}_1$. This leads to non-invariance issues and to the violation of the maximal self-similarity property (Property 2, Sec. 4.2) as pointed out recently in (Teuber and Lang, 2011). Yianilos (1995) already referred to this problem as *the self-recognition paradox*: "there are queries which do not recognize themselves, i.e., even if the query is in the database, some other element may be preferred.".

However, $\mathcal{Q}_B$ offers a useful property: it corresponds to an inner product in the space of functions $\boldsymbol{u} \mapsto \mathbb{R}$, the feature of $\boldsymbol{v}$ being $(p(\boldsymbol{v}|\boldsymbol{u} = \boldsymbol{t}))_{\boldsymbol{t}}$ (Seeger, 2002). The "mutual information" kernel follows this interpretation.

4.1.4 Mutual information kernel

Given the Bayesian joint distribution $\mathcal{Q}_B(\boldsymbol{v}_1, \boldsymbol{v}_2)$, Seeger (2002) defines a covariance kernel related to the sample mutual information between $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$:

$$\mathcal{K}_B(\boldsymbol{v}_1, \boldsymbol{v}_2) = \frac{\mathcal{Q}_B(\boldsymbol{v}_1, \boldsymbol{v}_2)}{\sqrt{\mathcal{Q}_B(\boldsymbol{v}_1, \boldsymbol{v}_1) \mathcal{Q}_B(\boldsymbol{v}_2, \boldsymbol{v}_2)}} . \tag{4.15}$$

Since $\mathcal{Q}_B$ can be seen as an inner product in the feature space, $\mathcal{K}_B$ corresponds to a cosine in the feature space $\mathcal{K}_B(\boldsymbol{v}_1, \boldsymbol{v}_2) = \dfrac{\langle \boldsymbol{v}_1 | \boldsymbol{v}_2 \rangle}{\|\boldsymbol{v}_1\| \|\boldsymbol{v}_2\|}$. Seeger shows that it is a kernel covariance matrix and coins it the mutual information kernel. Algorithms can be adapted to the noise pdf using the so-called *kernel tricks*, i.e., by considering higher dimensional space while never mapping

the data in practice. This leads for instance to non-linear support vector machines or non-linear principal component analysis. Note that a *prior*-less extension using MLE would lead to the generalized likelihood ratio $\mathcal{L}_G$. Compared to $\mathcal{L}_G$, the main limitation of the mutual information kernel is its dependency on the *prior* pdf and the lack of asymptotic results.

Among criteria involving probability densities, $\mathcal{L}_B$, $\mathcal{L}_G$ and $\mathcal{K}_B$ have no dimension thanks to their definitions as ratios of likelihoods (in terms of dimensional analysis), which is not the case for $\mathcal{Q}_B$ and $\mathcal{Q}_G$. We show in Sec. 4.2 that similarity criteria that are not adimensional lack some important properties. For this reason, we will refer to $\mathcal{L}_B$, $\mathcal{L}_G$ and $\mathcal{K}_B$ as normalized criteria and $\mathcal{Q}_B$ and $\mathcal{Q}_G$ as unnormalized criteria.

4.2 Desirable properties for similarity criteria

Beyond the theoretical grounding of each of the criteria described in the previous section, there are some desirable properties that are necessary to compare together the given similarity criteria.

It is natural to require that the similarity between two patches does not depend on the order in which the patches are compared:

Property 1 (Symmetry)**.** *The similarity between patch $\boldsymbol{v}_1$ and patch $\boldsymbol{v}_2$ is equal to the similarity between patch $\boldsymbol{v}_2$ and patch $\boldsymbol{v}_1$:*

$$\mathscr{C}(\boldsymbol{v}_1, \boldsymbol{v}_2) = \mathscr{C}(\boldsymbol{v}_2, \boldsymbol{v}_1).$$

All previously considered criteria are symmetrical.

For some criteria, it may occur that a distinct pair $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ is more similar than the pair formed by repeating observation $\boldsymbol{v}_1$: $(\boldsymbol{v}_1, \boldsymbol{v}_1)$. This phenomenon is called *the self-recognition paradox* (Yianilos, 1995). It is desirable to ask for maximal self-similarity:

Property 2 (Maximal self-similarity)**.** *No distinct pair $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ can be more similar than the observed patch $\boldsymbol{v}_1$ is similar to itself:*

$$\forall \boldsymbol{v}_1, \boldsymbol{v}_2,\ \mathscr{C}(\boldsymbol{v}_1, \boldsymbol{v}_2) \leq \mathscr{C}(\boldsymbol{v}_1, \boldsymbol{v}_1).$$

Joint likelihood criteria do not verify property 2:
Consider a noise distribution with a variance depending on the signal level, like gamma distribution that models speckle noise. For the pixel-based comparison, we have (Tab. 4.2 with $L = 1$): $\mathcal{Q}_B(v_1, v_2) = (v_1 + v_2)^{-2}$. Choose observation v_1 to be $v_1 = 2v_2$. Since $\mathcal{Q}_B(v_1, v_2) = (3\,v_2)^{-2}$ is larger than $\mathcal{Q}_B(v_1, v_1) = (4\,v_2)^{-2}$, property 2 is violated. □

Most criteria with a normalization like generalized likelihood ratio $\mathcal{L}_G$ and mutual information kernel $\mathcal{K}_B$ fulfill property 2 (see Tab. 4.1).

Property 2 does not guarantee that a pair $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ of distinct patches is always less similar than a pair formed by the repetition of a third observation $(\boldsymbol{v}_3, \boldsymbol{v}_3)$. A supplementary property is needed:

Property 3 (Equal self-similarities)**.** *Two pairs of identical patches always have equal similarity:*

$$\forall \boldsymbol{v}_1, \boldsymbol{v}_2,\ \mathscr{C}(\boldsymbol{v}_1, \boldsymbol{v}_1) = \mathscr{C}(\boldsymbol{v}_2, \boldsymbol{v}_2).$$

Criteria $\mathcal{L}_G$ and $\mathcal{K}_B$ verify both property 2 and 3 and their self-similarities are always equal to one (see Tab. 4.1).

Additionally, one may ask that the criterion is maximal only in case of strict patch equality, and for every comparison between identical patches:

Property 4 (Identity of the indiscernibles). *The similarity reaches its maximum if and only if the compared patches are identical:*

$$\forall \boldsymbol{v}_1, \boldsymbol{v}_2,\ \mathscr{C}(\boldsymbol{v}_1, \boldsymbol{v}_2) = \max_{\boldsymbol{v}} \mathscr{C}(\boldsymbol{v}, \boldsymbol{v}) \quad \textit{iif} \quad \boldsymbol{v}_1 = \boldsymbol{v}_2.$$

Note that this condition involves the noisy patches themselves (not the noiseless patches $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$).

For likelihood based criteria, it is clear that property 4 cannot be verified if two different observations lead to the same likelihoods. We need then to require that the observations be statistically identifiable through their likelihood:

$$\forall \boldsymbol{u},\ p(\boldsymbol{v}_1|\boldsymbol{u}) = p(\boldsymbol{v}_2|\boldsymbol{u}) \qquad \text{iif} \qquad \boldsymbol{v}_1 = \boldsymbol{v}_2\ . \tag{4.16}$$

Provided that observations are statistically identifiable through their likelihood, property 4 is fulfilled by the mutual information kernel $\mathcal{K}_B$. For $\mathcal{L}_G$ we require that the observations are statistically identifiable through their MLE, i.e., that the likelihood has a unique maximum and:

$$\arg\max_u p(\boldsymbol{v}_1|\boldsymbol{u}) = \arg\max_u p(\boldsymbol{v}_2|\boldsymbol{u}) \quad \text{iif} \quad \boldsymbol{v}_1 = \boldsymbol{v}_2\ . \tag{4.17}$$

The statistical answer of a similarity criterion should not depend on the choice of a specific noisy patch representation:

Property 5 (Invariance). *Let $\boldsymbol{g}$ be an invertible and differentiable function mapping random vectors $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ to random vectors $\boldsymbol{V}_1' = \boldsymbol{g}(\boldsymbol{V}_1)$ and $\boldsymbol{V}_2' = \boldsymbol{g}(\boldsymbol{V}_2)$. Let $\mathscr{C}_{\boldsymbol{V}_1,\boldsymbol{V}_2}$ and $\mathscr{C}_{\boldsymbol{V}_1',\boldsymbol{V}_2'}$ be, respectively, the similarity criteria derived from the family of parametric distributions followed by $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ (resp. $\boldsymbol{V}_1'$ and $\boldsymbol{V}_2'$). An invariant similarity criterion leads to the same similarity whether it is evaluated with $\mathscr{C}_{\boldsymbol{V}_1,\boldsymbol{V}_2}$ on $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ or with $\mathscr{C}_{\boldsymbol{V}_1',\boldsymbol{V}_2'}$ on $(\boldsymbol{g}(\boldsymbol{v}_1), \boldsymbol{g}(\boldsymbol{v}_2))$:*

$$\forall \boldsymbol{v}_1, \boldsymbol{v}_2,\ \mathscr{C}_{\boldsymbol{V}_1,\boldsymbol{V}_2}(\boldsymbol{v}_1, \boldsymbol{v}_2) = \mathscr{C}_{\boldsymbol{V}_1',\boldsymbol{V}_2'}(\boldsymbol{g}(\boldsymbol{v}_1), \boldsymbol{g}(\boldsymbol{v}_2)).$$

Due to their unnormalization, joint likelihood criteria $\mathcal{Q}_B$ and $\mathcal{Q}_G$ typically do not have the invariance property. Transforming the patches by, for example, taking their squared value leads to modified probability densities with different dimensions. The change of variables leads to a similarity criterion with a different scaling from the original one. Normalized criteria, defined as a ratio of probability densities, are the only ones to fulfil property 5.

Deciding for dissimilarity is done by thresholding the similarity: patches $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are considered dissimilar if $\mathscr{C}(\boldsymbol{v}_1, \boldsymbol{v}_2) < \tau$. The associated probability of false alarm $P_{FA}^{\mathscr{C}}$ is the probability that $\mathscr{C}(\boldsymbol{v}_1, \boldsymbol{v}_2) < \tau$ although $\boldsymbol{u}_1 = \boldsymbol{u}_2\ (= \boldsymbol{u}_{12})$, i.e., that the detected dissimilarity is only due to noise.

Property 6 (Constant false alarm rate). *For all threshold τ, the probability of false alarm $P_{FA}^{\mathscr{C}}$ of similarity criterion $\mathscr{C}$ is independent on the noise-free patch $\boldsymbol{u}_{12}$:*

$$\forall \tau,\ P_{FA}^{\mathscr{C}}(\tau) \textit{ does not depend on } \boldsymbol{u}_{12}.$$

name	pdf	$\mathcal{Q}_B$	$\mathcal{Q}_G$	$\mathcal{L}_B$	$\mathcal{L}_G$	$\mathcal{K}_B$	$\mathcal{S}$	$\mathcal{G}$
Gaussian	$\frac{e^{-\frac{(x-u)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}$	$e^{-(v_1-v_2)^2}$						
Gamma	$\frac{L^L x^{L-1} e^{-\frac{Lx}{u}}}{\Gamma(L) u^L}$	$\frac{1}{v_1 v_2}\left(\frac{v_1 v_2}{(v_1+v_2)^2}\right)^L$		$\frac{v_1 v_2}{(v_1+v_2)^2}$			$e^{-\left(\log \frac{v_1}{v_2}\right)^2}$	
Poisson	$\frac{u^x e^{-u}}{x!}$	$\frac{\Gamma'(v_1+v_2)}{2^{v_1+v_2} v_1! v_2!}$	$\frac{(v_1+v_2)^{v_1+v_2}}{(2e)^{v_1+v_2} v_1! v_2!}$	$\frac{\Gamma'(v_1+v_2)}{2^{v_1+v_2}\Gamma'(v_1)\Gamma'(v_2)}$	$\frac{(v_1+v_2)^{v_1+v_2}}{2^{v_1+v_2} v_1^{v_1} v_2^{v_2}}$	$\frac{\Gamma'(v_1+v_2)}{\sqrt{\Gamma'(2v_1)\Gamma'(2v_2)}}$	$e^{-\left(\sqrt{v_1+a}-\sqrt{v_2+a}\right)^2}$	

Table 4.2: Instances of the seven criteria for Gaussian, gamma and Poisson noise (parameters σ and L are fixed and known). All Bayesian criteria are obtained with Jeffreys' *priors* (resp. $1/\sigma$, $\sqrt{L}/u$, $\sqrt{1/u}$). All constant terms which do not affect the detection performance are omitted. For clarity reason, we define $\Gamma'(x) = \Gamma(x + 0.5)$ and the Anscombe constant $a = 3/8$.

The Gaussian kernel $\mathcal{G}$ is an example of a criterion which does not guarantee property 6. For instance in case of two Poisson noisy values v_1 and v_2, $\mathbb{E}\left[\|v_1 - v_2\|_2^2 \mid \mathscr{H}_0\right] = 2u_{12}$, hence, the distribution of $P_{FA}^{\mathcal{G}}$ is clearly dependent on u_{12}. Due to the efficiency of MLE with respect to the noise level, $\mathcal{L}_G$ is asymptotically CFAR (see Kay, 1998).

Based on the properties presented so-far, a proper similarity criterion can be selected. However, it is also important to compare the relative performance of similarity criteria. While we mentioned in Sec. 4.1 that there is no UMP detector for the considered *composite hypothesis problem*, the optimality can be studied on a subset of similarity criteria.

Property 7 (Uniformly Most Powerful Invariant)**.** *A similarity criterion $\mathscr{C}$ is said to be the uniformly most powerful invariant (UMPI) if it is an invariant criterion (property 5) and its probability of detection is larger than that of all other criteria for any given false-alarm rate:*

$$\forall \mathscr{C}',\ \tau,\ \tau' \quad P_{FA}^{\mathscr{C}}(\tau) = P_{FA}^{\mathscr{C}'}(\tau') \Rightarrow P_D^{\mathscr{C}}(\tau) \geq P_D^{\mathscr{C}'}(\tau')\ . \tag{4.18}$$

Asymptotically to the noise level, $\mathcal{L}_G$ is UMPI (see Lehmann, 1959). All other invariant criteria are then asymptotically defeated by $\mathcal{L}_G$.

Table 4.1 summarizes the properties of each of the seven considered criteria. The unnormalized criteria $\mathcal{Q}_B$ and $\mathcal{Q}_G$ fulfil none of the properties while the generalized likelihood ratio $\mathcal{L}_G$ fulfil all of them. Note that some properties require that observations are statistically identifiable. Such assumptions are generally true, except, e.g., for multi-modal distributions or when two different observations lead to the same likelihood function (e.g., a Gaussian distribution with zero mean and unknown variance leads to the same likelihood function for the observation of v or $-v$). Finally, note that $\mathcal{S}$ verifies most of these properties when the function $\boldsymbol{s}$ exists, which is generally not the case, e.g., there is no exact variance stabilization for the Poisson distribution or the Cauchy distribution.

4.3 Evaluation of similarity criteria

All criteria have been derived[3] in the case of gamma or Poisson noise (Tab. 4.2). In practice, Bayesian criteria are more difficult to obtain due to integrations over the noise-free patch space. While all criteria are equivalent for Gaussian noise, there are four different expressions for gamma noise and they are all different for Poisson noise. The distinction seems to emerge with the "complexity" induced by the noise distribution (by considering that gamma noise is more challenging than Gaussian noise, and that Poisson noise is even more challenging).

[3] the complete derivations are available in Appendix A

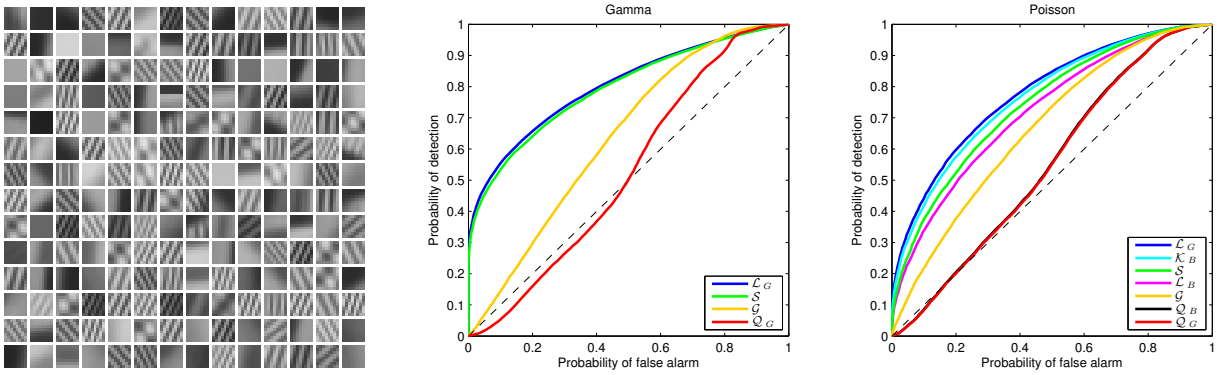


Figure 4.1: (left) Patch dictionary. (center) ROC curve obtained under gamma noise and (right) ROC curve obtained under Poisson noise. In both experiments, the SNR over the whole dictionary is about 1 dB.

4.3.1 Performance for patch discrimination

We evaluate the relative performance of the seven aforementioned criteria on a dictionary composed of 196 noise-free patches of size $N = 8 \times 8$. The noise-free patches have been obtained using the k-means on patches extracted from the classical 512×512 *Barbara* image. The noisy patches are noisy realizations of the noise-free patches under gamma or Poisson noise with an overall SNR of about 1 dB. All criteria are evaluated for all pairs of noisy patches. The process is repeated 200 times with independent noise realizations.

Numerically, the performance of the similarity criteria is given in term of their *receiver operating characteristic* (ROC) curve, i.e., the curve of P_D with respect to P_{FA}. Results are given in Fig. 4.1. For small P_{FA}, the generalized likelihood ratio (GLR) is the most powerful followed by the mutual information kernel, the Bayesian likelihood ratio and the variance stabilization criteria. Other criteria behave poorly for such a low SNR. Such behaviors agree with the theoretical predictions. The poor performance of the joint likelihood based criteria (worse than a detector that would not make use of the data) can arise from their non-invariance and the induced self-similarity paradox. The low performance of $\mathcal{G}$ is certainly due to its non-adaptivity to either the target noise or the target noise variance. The variance stabilization criteria are always defeated by GLR, due to the distortions of the noise-free patches as well as the consideration of the noise variance only, instead of the full noise pdf. The worse performance of Bayesian criteria compared to criteria that use MLE may be due to the low quality of the *prior* pdf (non-informative Jeffreys' prior have been used).

4.3.2 Application to image denoising

As we have seen in Chap. 2, patch correspondence is at the heart of most recent image denoising approaches since the introduction of the NL means (Buades et al., 2005; Dabov et al., 2007; Mairal et al., 2009). Most attempts to adapt such approaches for non-Gaussian noise relies on variance stabilization (e.g. Mäkitalo et al., 2010; Boulanger et al., 2010; Mäkitalo and Foi, 2011). Few authors try to extend the NL means by directly considering non-Gaussian noise distributions (Kervrann et al., 2007; Deledalle et al., 2009b, see also Sec. 3).

While local filters lead to biases and resolution loss, non-local techniques are known to efficiently reduce noise and preserve structures. Instead of combining neighboring pixels, the non-local means average similar pixels. Let $\boldsymbol{v}(x)$ be the observed noisy patch at pixel $x \in \Omega$ and $\boldsymbol{u}(x)$ its underlying noise-free version. The NL means filter defines the estimate $\hat{\boldsymbol{u}}(x)$ as a

		Noisy	$\mathcal{Q}_G$	$\mathcal{L}_G$	$\mathcal{S}$	$\mathcal{G}$	Noisy	$\mathcal{Q}_B$	$\mathcal{Q}_G$	$\mathcal{L}_B$	$\mathcal{L}_G$	$\mathcal{K}_B$	$\mathcal{S}$	$\mathcal{G}$
		Gamma					Poisson							
Strong noise levels	barbara	5.86	20.25	**20.97**	20.90	20.33	5.68	20.25	20.25	20.52	**20.68**	20.65	20.59	20.42
	boat	5.32	20.90	**21.47**	21.42	20.97	5.23	20.90	20.90	21.11	**21.21**	21.19	21.15	21.04
	bridge	6.09	18.44	**19.21**	19.16	18.49	5.83	18.36	18.36	18.65	**18.81**	18.78	18.72	18.53
	cameraman	5.54	18.56	**20.88**	20.87	7.48	5.59	18.61	18.61	19.17	**19.56**	19.49	19.37	19.01
	couple	5.98	20.93	**21.54**	21.51	20.99	5.55	20.91	20.91	21.11	**21.20**	21.18	21.15	21.04
	fingerprint	4.60	15.34	**16.30**	16.22	15.57	4.87	15.48	15.48	16.18	**16.41**	16.38	16.30	15.96
	hill	6.35	20.18	**20.68**	20.61	20.20	5.88	20.13	20.13	20.41	**20.54**	20.52	20.47	20.31
	house	4.84	20.54	**21.20**	21.13	20.64	4.94	20.48	20.49	20.81	**20.97**	20.94	20.88	20.67
	lena	5.64	22.14	**22.89**	22.83	22.23	5.44	22.14	22.15	22.44	**22.59**	22.56	22.49	22.30
	man	6.47	21.56	**22.16**	22.10	21.64	5.89	21.55	21.55	21.77	**21.89**	21.87	21.82	21.69
	mandril	5.52	20.22	**20.44**	20.41	20.27	5.31	20.23	20.23	20.34	**20.38**	20.37	20.36	20.30
	peppers	5.56	18.59	**20.44**	20.43	18.65	5.46	18.55	18.56	19.09	**19.46**	19.38	19.25	18.88
Medium noise levels	barbara	14.34	22.61	25.66	**25.67**	23.83	14.43	23.59	23.57	25.43	25.40	25.41	**25.44**	24.79
	boat	13.78	23.40	**25.50**	**25.50**	24.06	13.99	24.00	23.98	25.28	25.26	25.27	**25.29**	24.74
	bridge	14.58	20.17	**22.36**	**22.36**	21.01	14.58	21.06	21.04	22.30	22.29	22.30	**22.31**	21.84
	cameraman	13.96	23.88	**25.04**	25.01	14.93	14.33	23.63	23.57	25.01	25.02	25.02	**25.03**	24.22
	couple	14.37	23.19	**25.08**	25.06	23.68	14.31	23.54	23.52	**24.88**	24.85	24.86	**24.88**	24.29
	fingerprint	13.00	18.37	21.88	**21.89**	20.27	13.62	20.59	20.58	22.03	21.99	22.00	**22.04**	21.60
	hill	14.80	21.46	**24.24**	**24.24**	22.47	14.62	22.49	22.48	**23.98**	23.96	23.97	**23.98**	23.36
	house	13.35	22.52	26.33	**26.34**	24.36	13.73	24.36	24.34	**26.58**	26.57	26.57	**26.58**	25.76
	lena	14.09	24.61	27.71	**27.72**	25.61	14.20	25.57	25.55	**27.40**	27.37	27.38	**27.40**	26.58
	man	14.88	23.49	26.00	**26.01**	24.50	14.64	24.08	24.06	25.66	25.65	25.66	**25.67**	25.09
	mandril	14.02	21.61	**23.20**	**23.20**	22.22	14.03	22.18	22.17	23.03	23.01	23.02	**23.04**	22.68
	peppers	14.02	22.95	**25.54**	25.51	23.41	14.20	23.38	23.35	**25.45**	25.41	25.43	**25.45**	24.41

Table 4.3: PSNR values obtained by the NL means denoising using different similarity criteria on 13 standard images corrupted by gamma noise and Poisson noise with (top) strong noise levels and (bottom) medium noise levels.

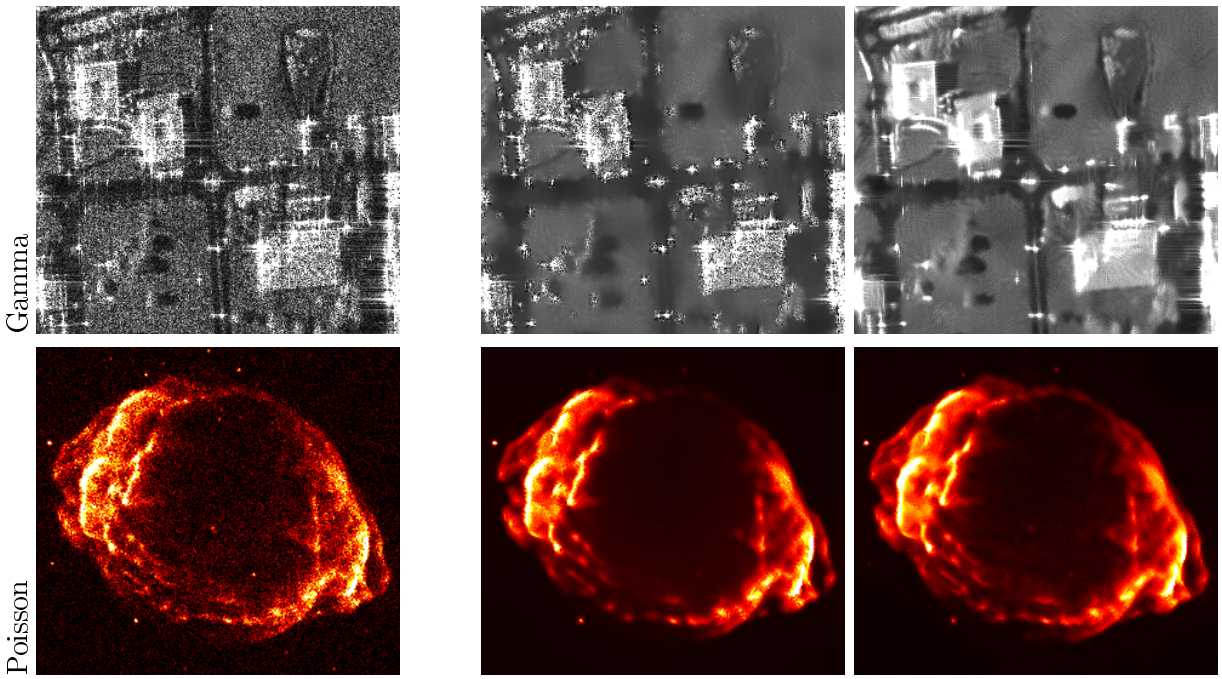
weighted average:

$$\hat{\boldsymbol{u}}(x) = \frac{\sum_x \mathscr{C}(\boldsymbol{v}(x), \boldsymbol{v}(x'))^{1/h} \boldsymbol{v}(x')}{\sum_x \mathscr{C}(\boldsymbol{v}(x), \boldsymbol{v}(x'))^{1/h}} \tag{4.19}$$

where x' is a pixel index located in a search window centered on x, and $h > 0$ is a filtering parameter. The similarity criterion $\mathscr{C}(\boldsymbol{v}(x), \boldsymbol{v}(x'))$, through the power function $(.)^{1/h}$, plays the role of a data-driven weight depending on the similarity between two patches centered around pixels of indices x and x' respectively. While patch-similarity is originally defined by the Gaussian kernel $\mathcal{G}$, we suggest comparing the denoising performance of the NL means when substituting the similarity criterion by one of the seven aforementioned criteria.

We evaluate first the denoising performance of the NL means obtained using each of the 7 similarity criteria on 13 standard images synthetically damaged by gamma or Poisson noise. The NL means is used with a 21×21 search window and 7×7 patches. The filtering parameter h as well as the central weight $\mathscr{C}(\boldsymbol{v}(x), \boldsymbol{v}(x))$ should be selected from the statistics of the similarity criterion $\mathscr{C}$ under $\mathscr{H}_0$ (Kervrann and Boulanger, 2008; Salmon, 2010). Unfortunately, such solutions cannot be investigated here since some of the studied criteria are not CFAR: the statistics vary locally with respect to $\boldsymbol{u}(x)$. The central weight should rather be replaced with the maximum of the weights in the search window, following the solution proposed in (Buades et al., 2009). Here, since our motivation is to compare patch similarity criteria, we have decided to use the true noise-free image $\boldsymbol{u}$ to select the best value of h for each criteria. In practice, we apply a gradient descent on h to optimize the mean square error $\|\boldsymbol{u} - \hat{\boldsymbol{u}}\|_2^2$. This allows us to compare similarity criteria in the most favorable case when each denoiser reaches its optimal performance.

Denoising performance is given in terms of the peak signal to noise ratio (PSNR) defined in Sec. 2.1.3. Table 4.3 displays the obtained PSNR values. Two levels of noise are considered, the

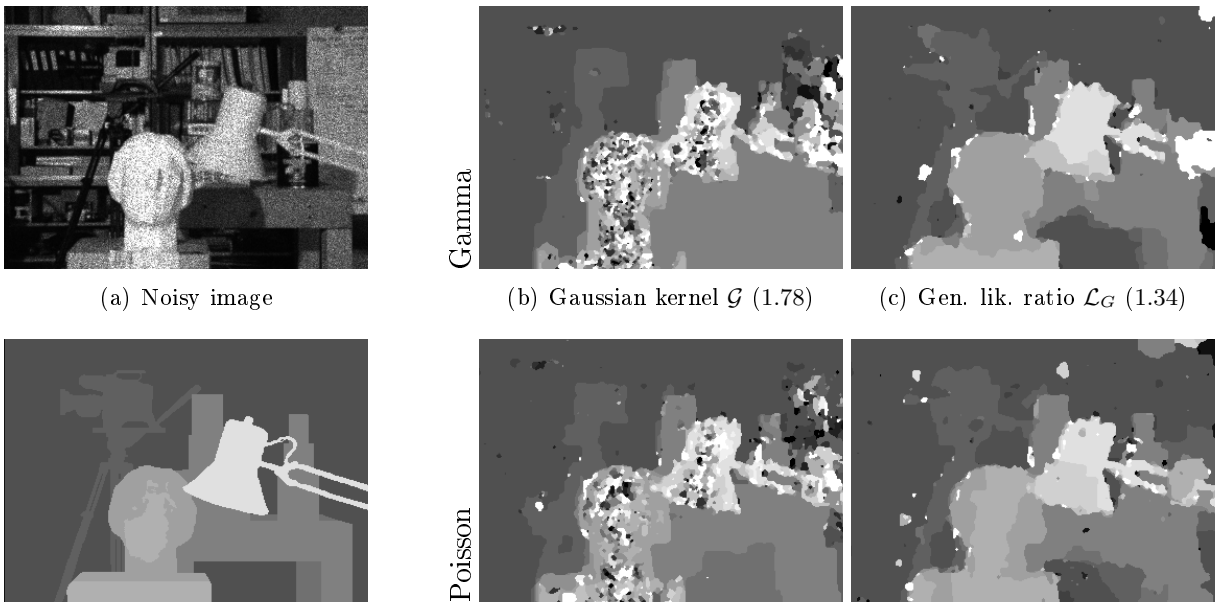


(a) Noisy image (b) Gaussian kernel (c) Generalized likelihood ratio

Figure 4.2: Results of the NL means on (a) noisy images using (b) the Gaussian kernel ($\mathcal{G}$) and (c) the generalized likelihood ratio ($\mathcal{L}_G$). The images are (top) a SAR image of two buildings suffering from gamma noise (©ONERA, CNES) and (bottom) an X-ray image of a supernova explosion in the Milky Way of the supernova remnant G1.9+0.3 suffering from Poisson noise (with a colormap varying smoothly from black through shades of red, orange, and yellow, to white) (image courtesy to Chandra X-ray Observatory – data identifier: `ADS/ Sa.CXO#Contrib/ ChandraDeepField`).

first one, very strong, leads to a noisy image with a PSNR around 5dB, and the second one, medium, provides a PSNR around 14dB. For strong noise levels, the generalized likelihood ratio $\mathcal{L}_G$ outperforms all other similarity criteria while for medium noise levels, the criterion based on variance stabilization works generally better. In medium/low level of noise, the variance stabilization based criterion $\mathcal{S}$ can outperform $\mathcal{L}_G$. When the noise level is low, the problem of weight definition is less a problem of detecting identical patches under noise than a matter of selecting patches with "close" noise-free patches (the noise component becomes negligible). Compared to $\mathcal{L}_G$, the properties provided by Euclidean distances can then be preferable in this context, since it defines a reasonable metric on the space of noiseless patches. A generalized likelihood ratio testing that $\boldsymbol{u}_1$ is close to $\boldsymbol{u}_2$ could be more adapted to the denoising problem, i.e.: $\mathscr{H}_0 : \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_2^2 < \epsilon$, where ϵ is a real positive value. This different definition of similarity could be the topic for future work.

Figure 4.2 provides a visual comparison of the use of the Gaussian kernel $\mathcal{G}$ and the generalized likelihood ratio $\mathcal{L}_G$ on real data. The first one is a synthetic aperture radar (SAR) image of two buildings. SAR data suffers from speckle noise modeled by a gamma distribution. The second one is an X-ray image of a supernova explosion in the Milky Way of the supernova remnant G1.9+0.3. Due to low-light conditions, such images suffer from Poisson noise. Without knowledge of $\boldsymbol{u}$, the methodology of Van De Ville and Kocher (2009) has been used to automatically select the value of h that maximizes an estimate of the mean square error. We will propose in Chap. 5 an extension of this approach for Poisson noise (initially proposed in (Deledalle et al., 2010c)) in the



(a) Noisy image (b) Gaussian kernel $\mathcal{G}$ (1.78) (c) Gen. lik. ratio $\mathcal{L}_G$ (1.34)

(d) Ground truth (e) Gaussian kernel $\mathcal{G}$ (1.50) (f) Gen. lik. ratio $\mathcal{L}_G$ (1.37)

Figure 4.3: Results of a stereo vision approach on a standard pair of noisy stereo views. (a) One of the noisy input images, (d) the ground truth (i.e. the target disparity map) and the estimated disparity maps obtained on the pair damaged by (b-c) gamma and (e-f) Poisson noise. The method is based on energy minimization using either (b,e) the Gaussian kernel $\mathcal{G}$ or (c,f) the generalized likelihood ratio $\mathcal{L}_G$. The minimum mean square error (RMSE) according to the regularization parameter is given in brackets.

same vein following (Hudson, 1978). In both cases $\mathcal{G}$ blurs dark areas and leaves noise in bright areas, GLR allows to reduce the noise level everywhere in the image with a similar amount of smoothing.

Note that the results provided here could be improved by refining weights using the similarity between pre-estimated patches as proposed in Chap. 5. Chapter 6 gives a deep analysis for the spetial case of the non-local denoising for (multi-variate complex) SAR data. Our motivation here is only to provide a fair comparison between similarity criteria, and therefore we have chosen not to refine weights to avoid interferences with pre-estimation procedures. Note that the performance of GLR for denoising SAR images has also been demonstrated in collaborative filtering (Parrilli et al., 2010).

4.3.3 Application to stereo-vision

Stereo-vision is one of the tasks in computer vision which extensively uses patches. Given two images of the same scene, the purpose is to estimate the depth of the image parts. Using epipolar geometry, each pixel $x \in \Omega$ of one image has a corresponding pixel x' at the same line in the other image (omitting the occlusion issues). The horizontal shift between these two pixels is called the disparity. The initial problem is then reduced to the estimation of a disparity map $\boldsymbol{d}$ (see Hartley and Zisserman, 2000). Given the disparity map, each patch $\boldsymbol{v}_1(x)$ should be similar to the patch $\boldsymbol{v}_2(x + d(x)\overrightarrow{h})$ and $\overrightarrow{h}$ is a unit vector directed on the horizontal orientation.

The definition of patch similarity is then central to stereo-vision. Note however that two patches $\boldsymbol{v}_1(x)$ and $\boldsymbol{v}_2(x')$ can be similar while p and q are not corresponding pixels (e.g. in homogeneous regions or on repetitive patterns). As a consequence, many works introduce a

prior knowledge on the solution to regularize the disparity map. Boykov et al. (1998) suggest that disparity maps are piece-wise constant. An estimate of the disparity map can then be obtained by solving the following optimization problem:

$$\hat{\boldsymbol{d}} = \arg\max_{\boldsymbol{d}} \sum_{x} -\log \mathscr{C}(\boldsymbol{v}_1(x), \boldsymbol{v}_2(x + d(x)\overrightarrow{h})) - \lambda \sum_{x \sim x'} \delta(d(x) - d(x')). \tag{4.20}$$

where $x \sim x'$ denotes two neighboring pixels, $\delta(.)$ is the Dirac delta function and the Lagrangian multiplier $\lambda > 0$ acts as a regularization parameter. Thanks to the patch similarity criteria $\mathscr{C}$, the first term measures the data fidelity of the solution. The second term assesses the regularity of the solution: it corresponds to the Potts model which penalizes transitions in $\boldsymbol{d}$. Satisfying solutions of such discrete optimization problems can be iteratively obtained by graph cuts with the α-β swap strategy described in (Boykov et al., 2001).

While the patch-similarity is usually defined by the Gaussian kernel $\mathcal{G}$, or equivalently by the Euclidean distance usually referred to as the sum of square differences (SSD), we suggest comparing stereo-vision performance of the model of Eq. (4.20) when the similarity criterion is replaced by the generalized likelihood ratio $\mathcal{L}_G$.

Figure 4.3 shows the visual comparison on a standard pair of stereo views damaged by gamma or Poisson noise. In both cases, the use of SSD leads to a disparity map over-regularized in dark areas and under-regularized in bright areas: there is no global regularization parameter λ offering the same amount of smoothing everywhere in the image. Since GLR is CFAR, we get the same level of regularization both in dark and bright areas for a global regularization parameter λ. As a numerical performance criterion, we have also computed the root mean square error (RMSE), defined by

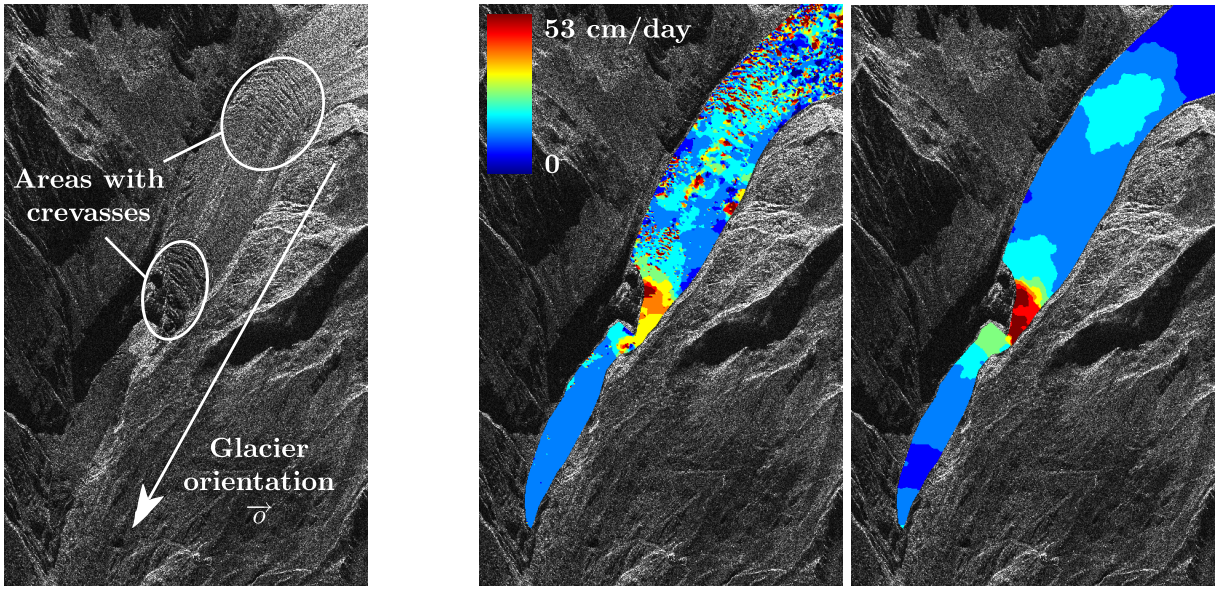
$$RMSE(\hat{\boldsymbol{d}}, \boldsymbol{d}) = \sqrt{\frac{1}{|\Omega|}\|\boldsymbol{d} - \hat{\boldsymbol{d}}\|_2^2} \tag{4.21}$$

for the results obtained by the use of both similarity criteria. For the same reason as with the image denoising task, we have decided to use the true disparity map $\boldsymbol{d}$ to select the best possible value of λ for each criterion. In practice, an exhaustive research has been done. This allows comparing similarity criteria in the most favorable case when each estimator reaches its optimal performance. The minimum root mean square error is in favour of GLR.

4.3.4 Application to motion tracking

Motion tracking, object tracking or optical flow estimation are classical problems involving the matching of image parts (e.g. Horn and Schunck, 1981; Lowe, 1992; Comaniciu et al., 2003). Here, we focus on the velocity estimation problem of a flowing Alpine glacier using a pair of synthetic aperture radar (SAR) images. SAR images provide scattering information which can be used under any weather conditions for glacier monitoring. Such images present a multiplicative speckle noise commonly modeled by gamma distributions (Goodman, 1976). The use of a similarity criterion robust to the statistics of the SAR intensity is then essential for the estimation of the displacement field.

Given two registered images of the same glacier sensed at different dates, the purpose is to estimate a displacement field characterizing at each position the local velocity of the glacier. Assuming that the movement is collinear to the glacier orientation, we only have to estimate the magnitude of the velocity. This quantity can be estimated by researching the patches of one acquisition which are similar to those present in the other acquisition along the glacier movement direction.



(a) Noisy image (b) Gaussian kernel $\mathcal{G}$ (c) Generalized lik. ratio $\mathcal{L}_G$

Figure 4.4: Results of motion tracking on a pair of SAR images of the glacier of Argentière (acquired by TerraSAR-X ©DLR). (a) One of the noisy input images and (b-c) the estimated magnitudes of the vector field. The method is based on energy minimization using either (b) the Gaussian kernel $\mathcal{G}$ or (c) the generalized likelihood ratio $\mathcal{L}_G$. The estimated speeds have an average over the surface of 12.27 cm/day and a maximum of 41.12 cm/day in the breaking slope (called "serac falls") for the estimation with the generalized likelihood ratio compared to 20.7 cm/day with a maximum of 67.2 cm/day for the Gaussian kernel $\mathcal{G}$.

For the same reasons as in the stereo-vision problem, the solution has to be regularized. Since glacier movement is assumed to be smooth, we propose here to use the total-variation (TV) model[4] whose penalization depends on the height of the transitions. This leads to the following optimization problem:

$$\hat{\boldsymbol{d}} = \arg\max_{\boldsymbol{d}} \sum_{x} -\log \mathscr{C}\big(\, \boldsymbol{v}_1(x),\ \boldsymbol{v}_2(x + d(x)\overrightarrow{o})\,\big) \\ +\lambda \sum_{x\sim x'} |d(x) - d(x')| \tag{4.22}$$

where $p \sim q$ denotes two neighboring pixels, the Lagrangian multiplier $\lambda > 0$ acts as a regularization parameter and $\overrightarrow{o}$ is a unit vector directed along the glacier orientation. Optimal solutions of such discrete optimization problems can be obtained by graph cuts using the graph construction described in (Ishikawa, 2003).

We suggest now comparing the quality of the estimated displacement fields obtained by solving (4.22), when using either the Gaussian kernel $\mathcal{G}$ or the generalized likelihood ratio $\mathcal{L}_G$.

Figure 4.4 shows the estimated magnitude of the displacement field obtained on two SAR images of the lower part of the glacier of Argentière (French Alps) sensed by TerraSAR-X on September 29th, 2008 and October 21th, 2008 respectively. The two SAR images have been previously co-registered on static areas. They have a resolution cell of 1.36×2.04 meters in line of sight and azimuth directions respectively. The displacement along the orientation $\overrightarrow{o}$ is searched

[4] we use anisotropic TV corresponding to the sum of the ℓ^1 norm of the gradient of $\boldsymbol{d}$ so that minimization problem (4.22) can be solved by graph-cuts

in a range of magnitude from 0 to 10 pixels. This corresponds to a maximum displacement of about 111 cm/day. Patches of size 3×3 were chosen, i.e. about 4 m and 6 m in ground geometry. A binary mask was provided to localize the glacier surface. Only corresponding pixels which are both on the glacier surface are used in patch comparisons. At each position is represented the magnitude of the local displacement estimated by both similarity criteria. According to experts and GPS measurements, the estimated velocities obtained with the generalized likelihood ratio $\mathcal{L}_G$ better reflects the ground truth with an average over the surface of 15.4 cm/day and a maximum of 53.8 cm/day in the breaking slope (called "serac falls") compared to 20.7 cm/day with a maximum of 67.2 cm/day for the Gaussian kernel $\mathcal{G}$. The use of $\mathcal{G}$ leads to a vector field over-regularized in dark areas and under-regularized in bright areas: there is no regularization parameter λ offering the same amount of smoothing everywhere in the vector field. Once again, since the generalized likelihood ratio $\mathcal{L}_G$ is CFAR, we get the same amount of regularization of the field map both in dark and bright areas for a global regularization parameter λ.

Finally let us mention that no criterion is optimal for this task due to illumination variations between the two observations. Correlation-based criteria could then be more adapted for such a task or a generalized likelihood ratio testing that $\boldsymbol{u}_1$ is within an affine transform of $\boldsymbol{u}_2$, i.e.: $\mathscr{H}_0 : \boldsymbol{u}_1 = \alpha \boldsymbol{u}_2 + \beta$, where α and β are unknown real values considered as nuisance parameters. Such an extension of the definition of similarity could be the topic of future work.

4.4 Conclusion

We have presented and compared seven similarity criteria taken from different research fields. Their theoretical grounding has been discussed as well as the different properties that they fulfil. In particular, it has been shown that some criteria are not invariant to the choice of the data space, and should thus be discarded. Others require signal-adaptive thresholds which restricts their usability in image processing applications. It has then been shown on patches extracted from a natural image that, under high levels of gamma or Poisson noise, the similarity criterion based on generalized likelihood ratio (GLR) is the most powerful. It also leads to the best denoising performance when used as the criterion for patch similarity in the NL means filtering, as assessed on a denoising benchmark made of twelve standard images synthetically damaged with strong gamma or Poisson noise. While GLR clearly outperforms techniques based on variance stabilization (such as the homomorphic approach or Anscombe transform) for low SNR images, our experiments show that variance stabilization is preferable for better SNR. With high SNR, patch comparison probably requires further modeling of noiseless patch distances. In the absence of such a model, the Euclidean distance used after variance stabilization is probably the best choice.

We have illustrated the improvements brought by a suitable similarity criterion to denoise real-world images: a synthetic aperture radar image corrupted by multiplicative speckle noise, and an X-ray image of a supernova explosion with Poisson noise. With a similarity criterion adapted to the noise distribution, noise is smoothed out equally well in dark and bright regions. We then illustrated the wide applicability of the proposed similarity criterion in vision by considering a stereo-vision reconstruction problem and the estimation of displacement of a glacier with remote sensing.

Based on this study, we recommend a broader use of GLR for measuring patch similarity in computer vision. This criterion is both easy to implement and theoretically well grounded. With its very general definition based on hypothesis testing, this criterion is flexible and can easily be adapted to other problems of matching image parts. Two extensions could be derived in future work. Similarity criteria invariant to some transforms of the noise-free patch (e.g., change of illumination) could be derived, which would increase robustness in application such as motion

tracking, stereo vision or flickering reduction. The modeling of a metric in the space of noise-free patches could also improve denoising performance, as our experiments with high SNR suggest.

Although GLR appears as the best similarity criterion to compare noisy patches, denoising performance can be improved by considering also the similarity of pre-filtered patches. The idea is to refine the evaluation of the hypothesis test when such extra information is available. The definition of similarity between noisy and noise-free patches linked to the problem of weighted combination under non Gaussian noise will be at the heart of the next chapter. In this context, we will define a general methodology to extend efficiently non-local approaches to any given statistical model of an uncorrelated noise.

Chapter 5

Selection-based filtering under non-Gaussian noise

An extension of the non-local (NL) means is proposed for images damaged by an uncorrelated noise described by a given pdf. The proposed method is guided by the noisy image and a pre-filtered image and is adapted to the statistics of the noise model. The main ingredient of the proposed approach is the use of the weighted maximum likelihood estimator where weights are defined from the generalized likelihood ratio based criterion between patches extracted from the noisy image (as given in Chap. 4) and the Kullback-Leibler divergence between patches extracted from a pre-filtered image. The influence of both images can be tuned using two filtering parameters. We propose an automatic setting to select these parameters based on the minimization of the estimated risk (mean square error). This selection uses an estimator of the MSE for NL means and Newton's method to find the optimal parameters in few iterations.

The adaptation of selection-based filters for non-Gaussian noise has been the topic of several works. Most of them are designed for a particular noise model: Lee's filter and the intensity-driven adaptive-neighborhood filter try to adapt to the multiplicative nature of gamma noise (Lee, 1981; Vasile et al., 2006); He and Greenshields (2009) propose an *ad hoc* solution for magnetic resonance images (MRI); and the SAFIR and BM3D filters stabilize the variance of Poisson noise before filtering (Boulanger et al., 2008; Mäkitalo and Foi, 2011). The Bayesian NL means filter of Kervrann et al. (2007) is a more general approach which has been used for Gaussian noise, ultra-sound *speckle* reduction in (Coupé et al., 2008) and SAR *speckle* reduction in (Zhong et al., 2011). However, we have seen that such a generalization can fail in some particular situations (see Sec. 3.7). On the contrary, other approaches are too general since they assume an unknown noise model (Awate and Whitaker, 2006; Brox et al., 2008; Azzabou et al., 2007b) and, as a consequence, they do not take the greatest advantage of the noise model when it is available.

Our contributions— This chapter presents a new approach for image denoising in the case of an uncorrelated noise described by a given pdf. The proposed filter is an extension of the NL means algorithm introduced by Buades et al. (2005), which performs a weighted average of the values of similar pixels. Pixel similarity is defined in the NL means as the Euclidean distance between patches (rectangular windows centered on each two pixels). Following the study in Chap. 4, we suggest replacing the Euclidean distance by the generalized likelihood ratio (GLR) which is statistically well grounded and general to cope with any noise distribution model. The denoising process is expressed as a weighted maximum likelihood estimation problem where the weights are derived in a data-driven way. These weights can be refined based on both the similarity between noisy patches (with GLR) and the similarity of patches extracted from the previous estimate (using the Kullback-Leibler divergence). The influence of both images can be

tuned using the different filtering parameters. We propose an unsupervised strategy to select these parameters. We show that this refinement strategy noticeably improves the denoising performance, especially in the case of low signal-to-noise ratio images. Numerical experiments illustrate that this technique can be successfully applied not only to the classical case of additive Gaussian noise but also to cases such as multiplicative *speckle* noise or Poisson noise. The proposed denoising technique seems to challenge on the *state-of-the-art* performance in these latter cases.

Organization of the chapter— We give in Sec. 5.1 an interpretation of our selection-based filters in the framework of weighted maximum likelihood. In Sec. 5.2, performance of oracle-based selection is studied when bias is introduced. This allows us to interpret, in Sec. 5.3, the NL means as an estimator of the oracle-based selection under Gaussian noise. Our approach based on detection theory will allow us to define weights able to deal with non-Gaussian noise in Sec. 5.4. Drawing from Chap. 4 and a study of optimal oracle-based weights, an hypothesis test will be used to define weights based on the joint similarity between noisy and noise-free patches, leading to a two step or iterative filter with two important filtering parameters. We then suggest setting these parameters with an unsupervised approach in Sec. 5.5. Section 5.6 provides results compared with the *state-of-the-art* techniques on images damaged with Gaussian, gamma and Poisson noise.

5.1 Weighted maximum likelihood estimation (WMLE)

This section introduces the proposed denoising method in the framework of weighted maximum likelihood estimation (WMLE) as investigated in (Polzehl and Spokoiny, 2006a).

As discussed in Chap. 2, a denoised image is an estimate $\hat{\boldsymbol{u}}$ of an unknown noise-free image $\boldsymbol{u}$ from its noisy observed version $\boldsymbol{v}$. The images are defined over a discrete regular grid Ω of N pixels and we denote by $v_k = v(x_k)$ a pixel value located at $x_k \in \Omega$. We consider an uncorrelated noise model defined by a parametric noise distribution, such that $p(\boldsymbol{v}|\boldsymbol{u}) = \prod_k p(v_k|u_k)$, where the noise-free image $\boldsymbol{u}$ plays the role of a space-varying unknown parameter. Denoising an image is then equivalent to find the best estimate $\hat{\boldsymbol{u}}$ of $\boldsymbol{u}$.

At each location x, the maximum likelihood estimator (MLE) defines an estimate $\hat{u}(x)$ of the underlying parameter $u(x)$ from the set $\mathcal{S}_{u(x)}$ of independent and identically distributed random variables present in the image:

$$\hat{u}(x) = \arg\max_t \sum_{x' \in \mathcal{S}_{u(x)}} \log p(v(x')|t) \tag{5.1}$$

$$= \arg\max_t \sum_{x'} \delta_{\mathcal{S}_{u(x)}}(x') \log p(v(x')|t), \tag{5.2}$$

with $\delta_{\mathcal{S}_{u(x)}}$ being the indicator function of $\mathcal{S}_{u(x)}$ (i.e., $\delta_{\mathcal{S}_{u(x)}}(x') = 1$ if $u(x') = u(x)$, 0 otherwise). The MLE is unbiased and asymptotically efficient with respect to $|\mathcal{S}_{u(x)}|$. In practice, the sets $\mathcal{S}_{u(x)}$ for each $x \in \Omega$ are unknown since it requires the knowledge of the underlying image $\boldsymbol{u}$. Hence, we only approximate $\delta_{\mathcal{S}_{u(x)}}(x')$ by data-driven weights $w(x, x') \geq 0$. This leads to the weighted maximum likelihood estimation (WMLE) given by:

$$\hat{u}(x) = \arg\max_t \sum_{x'} w(x, x') \log p(v(x')|t). \tag{5.3}$$

WMLE is known to reduce the mean square error by reducing the variance of the estimate at the cost of a bias introduced by samples that follow a distribution with a parameter $u(x')$ different to

$u(x)$ (Fan et al., 1998). The WMLE framework was first applied to image denoising by Polzehl and Spokoiny (2006a).

In the particular case of the additive white Gaussian noise model, the corresponding WMLE estimate is defined by the weighted average:

$$\hat{u}(x) = \frac{\sum_{x'} w(x,x')v(x')}{\sum_{x'} w(x,x')}. \tag{5.4}$$

This is consistent with the numerous denoising methods existing in image processing based on weighted averages, such as moving average filters, the Yaroslavsky filter and the NL means filter. Hence WMLE can be seen as a generalization of the weighted average for non-Gaussian noise.

Note that the Bayesian NL means filter of Kervrann et al. (2007), minimizes a Bayesian risk driven by the noisy data instead of maximizing the weighted likelihood (see Sec. 3.7). In the particular case of Gaussian noise, the same solution (5.4) is obtained by WMLE and Bayesian risk minimization driven by the noisy data.

The definition of the weights $w(x,x')$ is the main problem addressed in this chapter. As noted in (Polzehl and Spokoiny, 2006a), a well-chosen definition of the weights constitutes the key to the success of WMLE filters.

5.2 Oracle-based selection: a bias-variance trade-off

The weights used to approximate the indicator function can be seen as membership values over a fuzzy set version of $\mathcal{S}_{u(x)}$ (with proper weight normalization). The optimal unknown fuzzy set introduces a bias in the estimation since similar noisy values coming from different distributions are incorporated. However, this drawback is counterbalanced by a decrease in the variance of the estimation (Fan et al., 1998). In fact, more pixel values are included in the fuzzy set which decreases the variance of the estimation (note that for pixel values defined on a continuum, the probability measure of the event $u(x') = u(x)$ is zero, which means that we almost never find two pixels following the same distribution, thus we do not average pixel values therefore leaving the noisy image unchanged).

According to this bias-variance trade-off, WMLE can outperform MLE for well-chosen weights. Here, we suggest comparing the results obtained by MLE and WMLE both using an oracle based selection, i.e., the pixel values to average are chosen using the knowledge of the true image $\boldsymbol{u}$. Using this information, we have implemented both Oracle-MLE (Eq. (5.1)) and Oracle-WMLE (Eq. (5.3)). For both, the candidate pixels x' have been restricted to a limited circular search window W centered on x. While MLE averages only pixels with the same noise-free values $u(x') = u(x)$, we have chosen that the WMLE averages all pixels with the following weights:

$$w(x',x) = \exp\left(-\frac{(u(x') - u(x))^2}{2h^2}\right) \tag{5.5}$$

where $h > 0$ controls the width of the fuzzy set. This weight definition is designed such that when $h \to 0$, WMLE tends towards MLE and when $h \to \infty$, WMLE tends to a moving average filter.

Both filters have been used on a gray level image where the true values u_k are integers between 0 and 255 while the noisy values $\boldsymbol{v}_k$ are real values resulting from the application of a white Gaussian noise on $\boldsymbol{u}$ with standard deviation $\sigma = 20$. Figure 5.1 shows these images with the resulting images obtained by MLE and WMLE. It appears clearly that MLE suffers from a residual noise (i.e., the estimator as a high variance) while the underlying information seems to have been preserved (i.e., the estimator is unbiased). In contrast to MLE, WMLE has

(a) Noisy image (b) Noise-free image used by the oracle

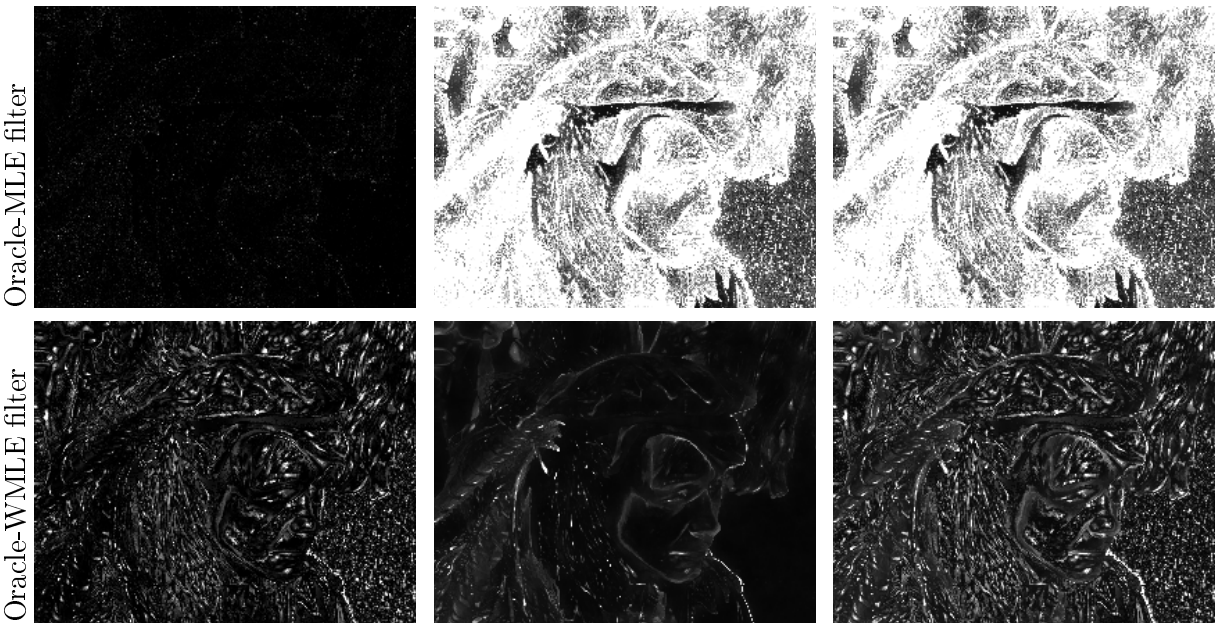
(c) Oracle-MLE filter (d) Oracle-WMLE filter

Figure 5.1: Denoising results obtained by (c) MLE and (d) WMLE on (a) a noisy image damaged by additive white Gaussian noise by using an oracle knowing (b) the underlying noise-free image. The result obtained by MLE suffers from residual noise while the underlying information seems to have been preserved. In contrast to MLE, WMLE has no residual noise but seems to be over-smoothed: the underlying information has been lost.

no residual noise but seems to be over-smoothed (e.g., look at the background texture): the underlying information has been lost (i.e., the estimator is biased).

To measure the bias-variance trade-off of both methods, we have computed the images of the local bias, the local relative variance and the local mean square error using Monte-Carlo simulations with 100 noise realizations (see Sec. 2.1.3 for a description of such an evaluation technique). The results are given on Fig. 5.2 where it appears clearly that MLE is unbiased (non-zero values are only ascribed to the Monte-Carlo simulations) but has a very high variance, while WMLE is biased but has a smaller variance. The resulting mean square error is then in favor of the WMLE. Note that the bias corresponds to a resolution loss (different populations of noisy values have been mixed) and the variance corresponds to a residual noise (too few candidates with similar noise-free values have been found).

In conclusion, even with the knowledge of the noise-free image, it is preferable to mix samples coming from different populations. Selection-based filters necessarily mix different populations in order to reduce enough the noise variance and usually to introduce a bias (except for the ultimate oracle-LMMSE, see Sec. 5.4.2). In the next section, we will study weight definitions driven only by noisy data in the case of Gaussian noise. This study will guide us in adapting this weight definition to the noise model. Next, we will show that the oracle-WMLE using Eq. (5.5) is not optimal and in particular it does not adapt to the noise distribution. This indicates that even for comparing noise-free values it is important to take into account the noise distribution



(a) Bias^2 (b) Relative variance (c) Mean square error

Figure 5.2: Bias-variance trade-off of (top) MLE versus (bottom) WMLE. (a) The square bias, (b) the relative variance and (c) the mean square error obtained on Monte-Carlo simulations with 100 noise realizations. In all images, dark colors correspond to small values (zero is black) and bright colors correspond to high values (white is reached at 50). MLE is unbiased but has a very high variance, while WMLE is biased but has a smaller variance. The resulting mean square error is then in favor of the WMLE.

in selection-based filtering. Finally, it will be helpful to introduce a weight refinement based on the similarity of pre-estimated images.

5.3 Patch-based weights with Gaussian noise

We have seen that weights should select pixels with identical or similar noise-free values. This information being latent or hidden behind a noise component, we have to find a robust method allowing us to select pixels with almost identical noise-free values. In this section, we will study the Yaroslavsky filter and the NL means filter as special cases of WMLE filters under the Gaussian noise assumption. We will statistically study their weight definition and the oracle weights given in Eq. (5.5) under Gaussian noise assumption. We will also show that patches allow us to obtain weights robust to noise fluctuations.

5.3.1 Statistical study of the Yaroslavsky filter

Instead of defining $w(x, x')$ in the spatial domain as done by moving average filters, we can use data-driven weights based on the comparisons of noisy pixel values. The Yaroslavsky filter (described in Sec. 2.5.1) increases the weight $w(x, x')$ when the noisy values $v(x)$ and $v(x')$ become more similar (Yaroslavsky, 1985).

The Yaroslavsky filter is a WMLE-based filter where the weights $w(x', x)$ are defined from the similarity of the noisy values $v(x')$ and $v(x)$. It assumes that if two noisy values are similar, their average is consistent: they probably come from the same population. The formulation of

(a) Oracle-WMLE filter (b) Yaroslavsky filter (c) NL means filter

Figure 5.3: Denoising results obtained by (a) the oralce-WMLE, (b) the Yaroslavsky filter and (c) the NL means filter on a noisy image damaged by additive white Gaussian noise. The result obtained by the Yaroslavsky suffers from a strong residual noise which has been reduced properly by NL means. The NL means filter reaches performance challenging the oracle-WMLE.

(a) Bias^2 (b) Relative variance (c) Mean square error

Figure 5.4: Bias-variance trade-off of (top) the Yaroslavsky filter versus (bottom) the NL means filter. (a) The square bias, (b) the relative variance and (c) the mean square error obtained on Monte-Carlo simulations with 100 noise realizations.

the weights is given as follows:

$$w(x', x) = \exp\left(-\frac{(v(x') - v(x))^2}{2h^2}\right) \tag{5.6}$$

where $h > 0$ controls the amount of filtering. This expression is similar to Eq. (5.5) where the unknown noise-free information $\boldsymbol{u}$ has been substituted by the available noisy information $\boldsymbol{v}$. If, furthermore, we overload the weights attributed to the pixel of interest by $w(x, x) = \exp\left(-\sigma^2/h^2\right)$, following (Doré and Cheriet, 2009; Salmon, 2010), we can show that Eq. (5.6) is actually an estimate of Eq. (5.5) under the Gaussian noise assumption. Indeed, it is straightforward to show

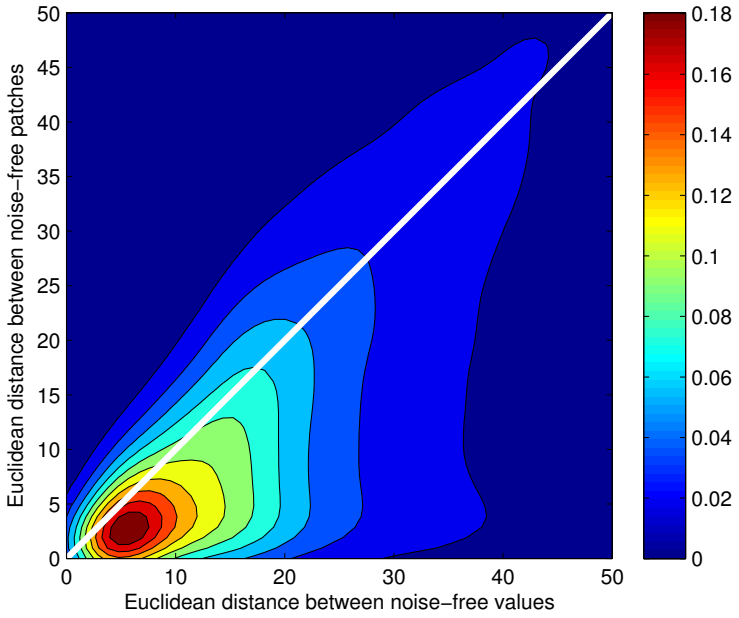


Figure 5.5: Two dimensional histogram of the Euclidean distance between noise-free patches with respect to the Euclidean distance between noisy values. The underlying assumption of patch-based denoising is that all the points are located on the white line.

that in this case:

$$\mathbb{E}[(V(x') - V(x))^2] = (u(x') - u(x))^2 + 2\sigma^2 \tag{5.7}$$

where $\boldsymbol{V}$ is the random vector describing the observed image $\boldsymbol{v}$. This estimator has, however, a large variance given by:

$$\mathrm{Var}[(V(x') - V(x))^2] = 8\sigma^2(u(x') - u(x))^2 + 8\sigma^4 \,. \tag{5.8}$$

This high variance produces weights with high fluctuations. To decrease this large residual noise, a high value of h has to be used, but it results in introducing a large bias. These two effects are illustrated in Fig. 5.3 and Fig. 5.4.

5.3.2 The NL means: towards patch based weights

The main limitation of the Yaroslavsky filter is its large variance resulting from the comparison of noisy values. In order to decrease this variance, i.e. to be robust to noise fluctuations, Buades et al. (2005) suggest comparing instead the two small square windows (i.e. the patches) surrounding the two pixels of interest. This leads to the NL means described in Sec. 2.5.1, and relies on the self similarity property of natural image. The weights are defined from the Euclidean distance between noisy patches according to:

$$w(x', x) = \varphi\left(\frac{\|v(\mathrm{P}_{x'}) - v(\mathrm{P}_x)\|_2^2}{2|P|h^2}\right) \tag{5.9}$$

where φ is a kernel decay function and $\mathrm{P}_x \in \Omega$ defines the subset of pixel positions belonging to the patch extracted at the position x and $|P|$ denotes the number of pixels in a patch (see Sec. 2.5.1). Under the Gaussian noise assumption, this substitution involves rather an estimator of the Euclidean distance between noise-free patches:

$$\mathbb{E}\left[\frac{1}{|P|}\|V(\mathrm{P}_{x'}) - V(\mathrm{P}_x)\|_2^2\right] = \frac{1}{|P|}\|u(\mathrm{P}_{x'}) - u(\mathrm{P}_x)\|_2^2 + 2\sigma^2 \,. \tag{5.10}$$

Contrary to the Yaroslavsky filter, this estimator has a smaller variance depending on the size of the patch:

$$\mathrm{Var}\left[\frac{1}{|P|}\|V(\mathrm{P}_{x'}) - V(\mathrm{P}_x)\|_2^2\right] = \frac{1}{|P|}\left[8\sigma^2\frac{1}{|P|}\|u(\mathrm{P}_{x'}) - u(\mathrm{P}_x)\|_2^2 + 8\sigma^4\right] . \tag{5.11}$$

According to Eq. (5.11), the larger the patch, the smaller the weight fluctuations. This allows us to drastically decrease the large residual noise present in the Yaroslavsky solution such that a smaller value of h can be used to limit the blurring effects. This is also illustrated on Fig. 5.3 and Fig. 5.4.

Note that with this substitution, the NL means approach the WMLE solution as soon as the distance between the patches $u(\mathrm{P}_{x'})$ and $u(\mathrm{P}_x)$ is linked to the distance between the pixel values $u(x')$ and $u(x)$. We experimentally measure the validity of this assumption on a natural image (a similar study can be found in Duval et al., 2011). For each pair of pixels (x, x') we calculate the Euclidean distance between the pixel values $|u(x')-u(x)|$ and the Euclidean distance between patches $\frac{1}{N}\|u(\mathrm{P}_{x'}) - u(\mathrm{P}_x)\|_2$. We then construct a two dimensional histogram using the Parzen–Rosenblatt window method. The histogram is given on Fig. 5.5. It appears that the assumption is well verified with a tendency that $\frac{1}{N}\|u(\mathrm{P}_{x'})-u(\mathrm{P}_x)\|_2$ underestimates $|u(x')-u(x)|$: noise-free patches are more similar than noise-free pixel values. This tendency implies that NL means will generally introduce more bias than the Yaroslavsky filter. For instance, punctual features will be blurred by the NL means while the Yaroslavsky filter will better preserve them for the same value of h. The less frequent and opposite effect is when $\frac{1}{N}\|u(\mathrm{P}_{x'}) - u(\mathrm{P}_x)\|_2$ overestimates $|u(x')-u(x)|$: noise-free patches are less similar than noise-free pixel values. In that case, for the same value of h, the NL means will leave more residual noise than the Yaroslavsky filter (it appears for example around edges with high contrast, see for instance the problem of noise halos discussed in Chap. 7).

5.4 Patch-based weights with non-Gaussian noise

We have seen that the Euclidean distance between noisy patches allows to obtain a good approximation of the oracle-WMLE under the Gaussian noise assumption. When noise departs from Gaussian noise, the expected square difference is given by the following relation:

$$\begin{aligned}\mathbb{E}\left[\frac{1}{|P|}\|V(\mathrm{P}_{x'}) - V(\mathrm{P}_x)\|_2^2\right] =& \frac{1}{|P|}\|\mathbb{E}[V(\mathrm{P}_{x'})] - \mathbb{E}[V(\mathrm{P}_x)]\|_2^2 \\ &+ \frac{1}{|P|}\|\mathrm{Var}[V(\mathrm{P}_{x'})]\|_1 + \frac{1}{|P|}\|\mathrm{Var}[V(\mathrm{P}_x)]\|_1 .\end{aligned} \tag{5.12}$$

According to (5.12), as soon as the expectation of noisy values does not converge to the noise-free value or when the variance is signal-dependent, the distance between noisy patches is no longer an estimate of the distance between noise-free patches. The interpretation of the NL means as an approximation of the oracle-WMLE does not hold anymore and in practice results can be of poor quality.

Rather than defining weights based on the Euclidean distance between noisy patches to estimate the distance between noise-free patches to approach Eq. (5.5), let us come back to the original problem. According to the previous comments, the weights can be seen as a membership value over the fuzzy set version of $\mathcal{S}_{u(x)} = \{x'|u(x') = u(x)\}$. The membership value can then be interpreted as a statistical hypothesis test measuring the validity that x' belongs to $\mathcal{S}_{u(x)}$. Due to noise fluctuations, we follow the same idea as that of the NL means and assume equal values for the central pixel of two statistically close image patches. The hypothesis test is then

performed on patches:

$$\begin{aligned} &\mathscr{H}_0 : u(\mathrm{P}_{x'}) = u(\mathrm{P}_x) && \text{(null hypothesis)}, \\ &\mathscr{H}_1 : u(\mathrm{P}_{x'}) \neq u(\mathrm{P}_x) && \text{(alternative hypothesis)}. \end{aligned} \tag{5.13}$$

The Bayesian NL means filter of Kervrann et al. (2007) substitutes instead the Euclidean distance based by the conditional probability $p\left(v(\mathrm{P}_x)|u(\mathrm{P}_x) = v(\mathrm{P}_{x'})\right)$. This approach assumes that $v(\mathrm{P}_{x'})$ provides a good approximation on the true parameter $u(\mathrm{P}_x)$. Since the Bayesian NL means filter makes this strong assumption, the authors proposed a two step algorithm to refine the weights (see Sec. 3.7). We suggest that the weights based on (5.13) should be more suitable since they do not make such a strong assumption.

Similarity criteria based on the hypothesis test (5.13) have been presented in Chap. 4. This will allow us to directly propose a first extension of non-local filters for non-Gaussian noise based on the similarity of noisy patches using the generalized likelihood ratio presented in Chap. 4. Next, we will come back to an oracle study in order to refine the weights in the case of low signal-to-noise ratio images. This will lead to a second extension based on the similarity of noise-free patches.

5.4.1 Similarity between noisy patches

In Chap. 3, it has been shown that a selection rules based on the generalized likelihood ratio leads to an *unbiased* estimator with *stationary relative vararianc*e. In Chap. 4, it was moreover concluded that the evaluation of GLR between the noisy patches $v(\mathrm{P}_{x'})$ and $v(\mathrm{P}_x)$ provides good detection performance for the hypothesis test in Eq. (5.13) and it can be succesfully applied in several computer vision tasks. Based on these arguments, we suggest extending the NL means for non-Gaussian noise by using a WMLE filter where weights are based on the GLR between noisy-patches. Another argument is that, under the Gaussian noise assumption, WMLE and GLR boil down respectively to the weighted average and to the Gaussian kernel (i.e., Eq. (5.9)): the NL means appears then as a special case of the proposed extension.

We suggest expressing the weights from the generalized likelihood ratio when the two noisy patches share identical noise-free values. This leads to the following weight definition:

$$w(x, x') = \varphi\left[-\log \mathcal{L}_G(v(\mathrm{P}_x), v(\mathrm{P}_{x'})\right] \tag{5.14}$$

$$\text{where } \mathcal{L}_G(\boldsymbol{v}_1, \boldsymbol{v}_2) = \frac{\sup_{\boldsymbol{t}}\left[p(\boldsymbol{v}_1|\boldsymbol{u}_1 = \boldsymbol{t})p(\boldsymbol{v}_2|\boldsymbol{u}_2 = \boldsymbol{t})\right]}{\left[\sup_{\boldsymbol{t}} p(\boldsymbol{v}_1|\boldsymbol{u}_1 = \boldsymbol{t})\right]\left[\sup_{\boldsymbol{t}} p(\boldsymbol{v}_2|\boldsymbol{u}_2 = \boldsymbol{t})\right]} \tag{5.15}$$

where φ is the same kernel decay function as the one of the NL means presented in Sec. 2.5.1. Its shape or scale is controlled respectively by one or two parameters which are usually set according to the distribution of the Euclidean distance (Polzehl and Spokoiny, 2006a; Kervrann and Boulanger, 2006), or, in our context, the distribution of GLR under $\mathscr{H}_0$ (i.e. according to the probabilities of false alarm). This method based on probabilities of false alarm requires that GLR has a constant false alarm rate (CFAR) which is not always the case for patches with a finite size (e.g., for Poisson noise).

We have derived in closed-form expressions the weights between noisy patches for Gaussian, gamma and Poisson noise. These expressions are directly obtained from the expressions given in Table 4.2:

The case of Gaussian noise: Noise is additive and GLR is linked to the square difference of the noisy patches. It corresponds to the Gaussian kernel presented in Sec. 4.1 and it leads to weights

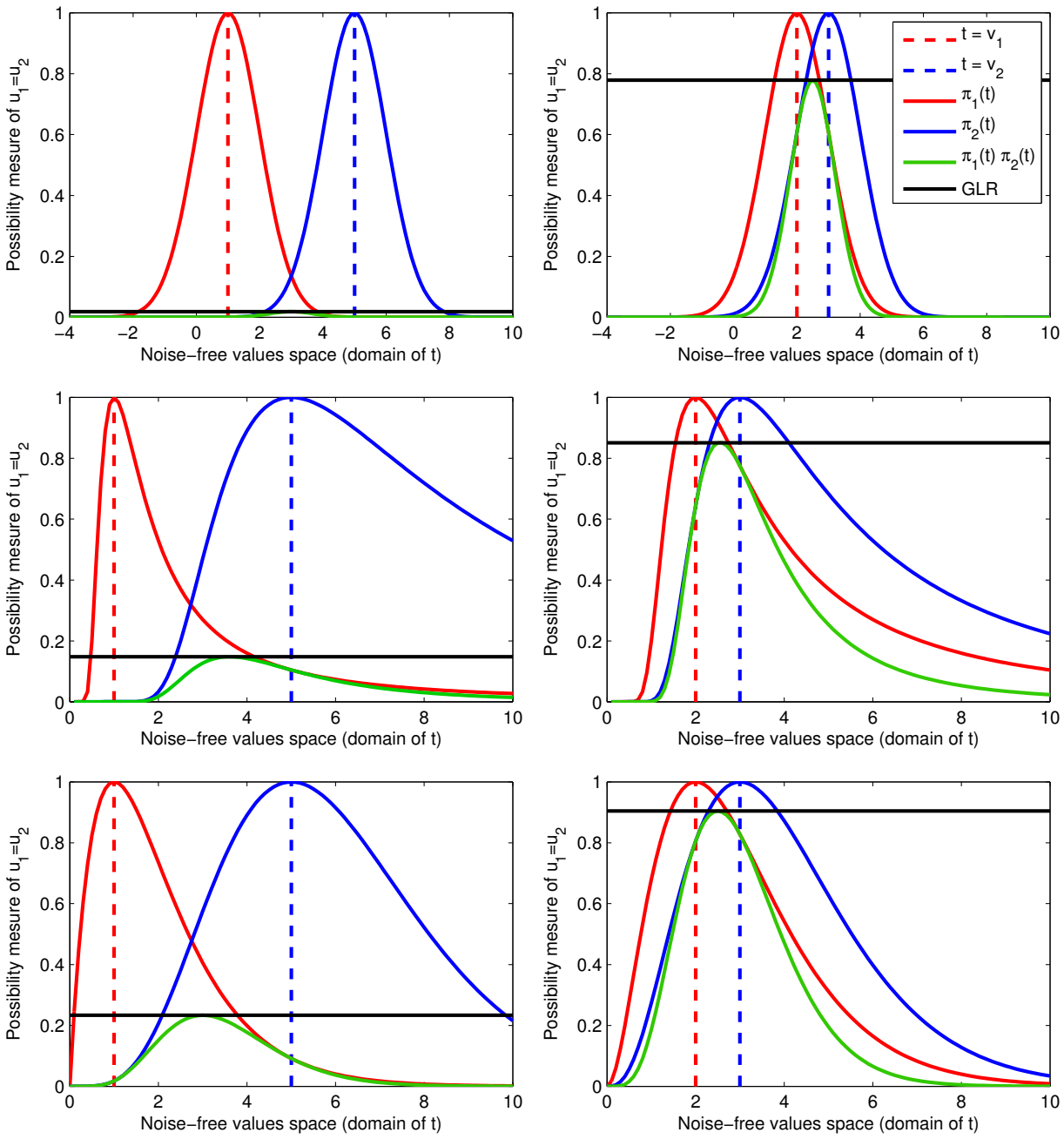


Figure 5.6: Illustration of GLR in the point of view of fuzzy aggregation, under (from top to bottom) Gaussian noise, gamma noise and Poisson noise.

defined from the Euclidean distance:

$$w(x,x') = \varphi\left(\frac{\|v(\mathrm{P}_x) - v(\mathrm{P}_{x'})\|_2^2}{4\sigma^2}\right) . \tag{5.16}$$

The case of gamma noise: Noise is multiplicative and GLR is linked to the ratio of the noisy patches:

$$w(x,x') = \varphi\left[2L\sum_{\tau\in\mathrm{P}}\log\left(\sqrt{\frac{v(x+\tau)}{v(x'+\tau)}} + \sqrt{\frac{v(x'+\tau)}{v(x+\tau)}}\right) - 2L|P|\log 2\right] . \tag{5.17}$$

Note that if we consider the log-transform $\tilde{\boldsymbol{v}} = \log \boldsymbol{v}$, whose noise component is additive, GLR rewrites as the difference of the log-transform of the noise-free patches as follow:

$$w(x,x') = \varphi\left[2L\sum_{\tau\in \mathrm{P}} \log\cosh\left(\frac{\tilde{v}(x+\tau)-\tilde{v}(x'+\tau)}{2}\right)\right] . \tag{5.18}$$

The case of Poisson noise: Noise is neither additive nor multiplicative and GLR provides the following weight definition:

$$w(x,x') = \varphi\left[2\sum_{\tau\in \mathrm{P}}\left(\frac{g(v(x+\tau))+g(v(x'+\tau))}{2} - g\left(\frac{v(x+\tau)+v(x'+\tau)}{2}\right)\right)\right] \tag{5.19}$$

where $g(x) = x\log x$ if $x > 0$ and $g(0) = 0$.

In Chap. 4, we give an interpretation of GLR based on detection theory. Here, we suggest reinterpreting Eq. (5.14) in a point of view based on the theory of fuzzy sets and possibilities (Zadeh, 1965; Dubois and Prade, 1988). GLR can be seen as a degree of possibility on the hypothesis $u(\mathrm{P}_{x'}) = u(\mathrm{P}_x)$. Indeed, consider the quantities:

$$\pi_1(\boldsymbol{t}) = \frac{p(\boldsymbol{v}_1|\boldsymbol{u}_1=\boldsymbol{t})}{\sup_{\boldsymbol{t}} p(\boldsymbol{v}_1|\boldsymbol{u}_1=\boldsymbol{t})} \qquad \pi_2(\boldsymbol{t}) = \frac{p(\boldsymbol{v}_2|\boldsymbol{u}_2=\boldsymbol{t})}{\sup_{\boldsymbol{t}} p(\boldsymbol{v}_2|\boldsymbol{u}_2=\boldsymbol{t})} . \tag{5.20}$$

Since $\pi_1(\boldsymbol{t})$ and $\pi_2(\boldsymbol{t})$ are two functions defined from the noise-free space to $[0,1]$, they can be interpreted as the respective possibility distributions of $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, or, as the fuzzy number versions of $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. They model the uncertainty we have on these unknown values. The product $\pi_1(\boldsymbol{t})\pi_2(\boldsymbol{t})$ is called a triangular norm and is known to model the intersection of the two fuzzy sets associated to $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. In some way, it models the fuzzy set of the assumed shared value $\boldsymbol{u}_{12}$. By taking the maximum value $\mathcal{L}_G(\boldsymbol{v}_1,\boldsymbol{v}_2) = \sup_{\boldsymbol{t}} \pi_1(\boldsymbol{t})\pi_2(\boldsymbol{t})$, GLR evaluates the possibility measure that $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ share the same noise-free patch $\boldsymbol{u}_{12}$. Fig. 5.6 gives an illustration of this fuzzy interpretation of GLR between two noisy values v_1 and v_2 when noise follows either a Gaussian, gamma or Poisson distribution. It appears clearly that not only the closeness of the noisy values impact on the possibility measure that their underlying noise-free values are identical, but also the shape of the possibility distributions $\pi_1(\boldsymbol{t})$ and $\pi_2(\boldsymbol{t})$. In Chap. 4, we mentioned that GLR is a *prior*-less version of the Bayesian likelihood ratio. In this framework, this *prior* ignorance on the noise-free patches could be modeled by a possibility distribution $\pi_{prior}(\boldsymbol{t}) = 1$. From the possibility measure provided by GLR, the function φ defines the membership function to the fuzzy set version of $\mathcal{S}_{u(x)}$ which is our original motivation.

In previous sections, we have seen that to obtain the best performance, selection-based filters should introduce bias in order to significantly decrease the noise level. However, GLR tries to select only pixels with identical values. In practice, weights based on GLR will also introduce bias. In a detection point of view, this is due to the fact that GLR has a non-zero probability of false-alarm. In a fuzzy point of view, this results from the uncertainty modeled by $\pi_1(\boldsymbol{t})$ and $\pi_2(\boldsymbol{t})$. The shape and the scale of the kernel function φ play also an important role in the bias-variance trade-off. For instance, the setting of the bandwidth h of the exponential decay function is a crucial parameter controlling the quality of the result. Based on this property, we will propose an unsupervised setting of this parameter in Sec. 5.5.

In the case of images with low signal to noise ratio (i.e., with a high level of noise), GLR still has a high variance resulting in weights with high fluctuations. In a detection point of view, this is due to the fact that, for the same probability of false alarm, the probability of detection of

GLR decreases when the noise level increases. Based on the non-local approaches proposed in (Kervrann and Boulanger, 2006; Polzehl and Spokoiny, 2006a; Brox et al., 2008; Dabov et al., 2007; Goossens et al., 2008; Louchet, 2008), in the next section, we suggest refining the weights in this case by using the information provided by a pre-estimate of the noise-free image.

5.4.2 Similarity between noise-free patches

This section presents a refinement of the weights based on the evaluation of the hypothesis test (5.13) using a pre-estimate of the noise-free image. This refinement seems to be the more relevant on images with low signal to noise ratio. In (Kervrann and Boulanger, 2006; Brox et al., 2008; Dabov et al., 2007; Goossens et al., 2008; Louchet, 2008), the authors show that weights based on the Euclidean distance between filtered noise-free patches improve the denoising performance for such strong noise levels. Based on an oracle study, we will see in this section that even for comparing noise-free patches, the Euclidean distance has to be substituted by a similarity criterion adapted to the noise distribution. Then based on Polzehl and Spokoiny (2006a), we will propose instead to use the Kullback-Leibler divergence between filtered noise-free values. According to the confidence we have in our pre-estimate, we will show that it can be preferable to use jointly a similarity criterion based on noisy patches (using GLR) and noise-free patches (using the Kullback-Leibler divergence). A Bayesian interpretation will be given to this joint criterion.

Oracle-LMMSE: the ultimate weights

We considered in Sec. 5.2 an oracle that use the true image $\boldsymbol{u}$ to define weights of the form (5.5). We now relax this constraint and consider linear minimum mean square error (LMMSE) denoising for a given true image $\boldsymbol{u}$. Consider x to be fixed and define $u = u(x)$ and $\boldsymbol{w}$ being the vector such that $w_k = w(x, x_k)$. The oracle-LMMSE weights are implicitly defined as:

$$\boldsymbol{w}^{(\text{LMMSE})} = \arg\min_{\boldsymbol{w}} \mathbb{E}\left[(\boldsymbol{w}^t\boldsymbol{v} - u)^2\right], \tag{5.21}$$

i.e., we consider the linear combination of noisy values $\boldsymbol{w}^t\boldsymbol{v}$ that leads to the best estimate of u, in the sense of expected quadratic loss. Note that our problem is slightly different from the LMMSE approach of Chatterjee and Milanfar (2011). They consider minimizing the MSE in the patch domain knowing only the patch $\boldsymbol{v}(x)$ and the first and second order statistics of the patch $\boldsymbol{u}(x)$ under AWGN. We consider rather minimizing the MSE in the image domain knowing all pixel values $v(x_k)$ and all true values $u(x_k)$ whatever the noise model.

Let the noisy image $\boldsymbol{v}$ be modeled by its mean $\mathbb{E}[\boldsymbol{v}] = \boldsymbol{u}$ and its covariance matrix $\mathbb{E}[(\boldsymbol{v} - \boldsymbol{u})(\boldsymbol{v} - \boldsymbol{u})^t] = \boldsymbol{\Gamma}$ (e.g., Gaussian, gamma and Poisson noise). The MSE can be expanded as follows:

$$\begin{aligned}\mathbb{E}\left[(\boldsymbol{w}^t\boldsymbol{v} - u)^2\right] &= \boldsymbol{w}^t\mathbb{E}\left[\boldsymbol{v}\boldsymbol{v}^t\right]\boldsymbol{w} - 2u\boldsymbol{w}^t\mathbb{E}\left[\boldsymbol{v}\right] + u^2 &(5.22)\\ &= \boldsymbol{w}^t(\boldsymbol{\Gamma} + \boldsymbol{u}\boldsymbol{u}^t)\boldsymbol{w} - 2u\boldsymbol{w}^t\boldsymbol{u} + u^2. &(5.23)\end{aligned}$$

The optimum weights are obtained by setting the gradient of (5.23) with respect to $\boldsymbol{w}$ equal to zero:

$$\boldsymbol{w}^{(\text{LMMSE})} = u(\boldsymbol{\Gamma} + \boldsymbol{u}\boldsymbol{u}^t)^{-1}\boldsymbol{u}. \tag{5.24}$$

This expression involves inverting the sum of a covariance matrix plus a rank-one matrix. If noise is not correlated, covariance matrix $\boldsymbol{\Gamma}$ is diagonal and straightforward to invert. By use of

Sherman-Morisson formula, we get the following expression which requires only the inversion of $\boldsymbol{\Gamma}$:

$$\boldsymbol{w}^{(\mathrm{LMMSE})} = u\left(\boldsymbol{\Gamma}^{-1} - \frac{\boldsymbol{\Gamma}^{-1}\boldsymbol{u}\boldsymbol{u}^t\boldsymbol{\Gamma}^{-1}}{1+\boldsymbol{u}^t\boldsymbol{\Gamma}^{-1}\boldsymbol{u}}\right)\boldsymbol{u} \tag{5.25}$$

$$= \frac{u}{1+\boldsymbol{u}^t\boldsymbol{\Gamma}^{-1}\boldsymbol{u}}\boldsymbol{\Gamma}^{-1}\boldsymbol{u}. \tag{5.26}$$

For a diagonal covariance matrix $\boldsymbol{\Gamma} = \mathrm{diag}(\gamma_1^2, \ldots, \gamma_N^2)$, the k^{th} weight is written as:

$$w_k^{(\mathrm{LMMSE})} = \frac{u\, u_k/\gamma_k^2}{1+\sum_l u_l^2/\gamma_l^2}. \tag{5.27}$$

Note that when $N = 1$, Eq. (5.27) boils down to the Wiener shrinkage presented in Sec. 2.4.1:

$$w_k^{(\mathrm{Wiener})} = \frac{u^2}{\gamma_k^2 + u^2}. \tag{5.28}$$

The expression of the obtained weights is interesting because it corresponds to the optimal weights (independent of the noisy data) for a given true image. The oracle-LMMSE we just derived provides ultimate performance, i.e., a lower bound on the achievable MSE for a given search window of size N (see Fig. 5.8). The oracle considered is largely favored compared to a real denoising scenario. It can indeed combine samples v_k in such a way that their biases cancel. The drawback is that if we have only a pre-estimate instead of the true image, the use of weights defined in Eq. (5.26) will provide a result too close to the pre-estimate (there will be no improvement compared to the pre-estimate itself, see Fig. 5.9). Moreover, the oracle-LMMSE does not provide a linear combination where coefficients are normalized and then it is not a WMLE-based filter. The expression (5.27) is not a good candidate to define similarity between pre-estimates of the noise-free patches but it sheds light on why the similarity should adapt to the noise distribution.

Optimal normalized weights with oracle-similarity

We now consider a more realistic situation in which the oracle selects normalized weights, i.e., $\sum_k w_k = 1$ (note that a similar problem is derived in (Lee et al., 2011) where they consider an empirical MSE driven by patch similarities rather than the true MSE). While the oracle-LMMSE, provides the best selection-based filter, this normalized filter will be the best WMLE-based filter. In this case, the bias $\boldsymbol{w}^t\boldsymbol{u} - u$ of estimator $\boldsymbol{w}^t\boldsymbol{v}$ can be expressed with respect to $\boldsymbol{d}$:

$$\boldsymbol{w}^t\boldsymbol{u} - u = \boldsymbol{w}^t(\boldsymbol{u} - u\mathbb{1}) = \boldsymbol{w}^t\boldsymbol{d}, \tag{5.29}$$

where $\boldsymbol{d}$ is the vector of all differences (or bias): $d_k = (u_k - u)$. The MSE can be rewritten as follows:

$$\mathbb{E}\left[(\boldsymbol{w}^t\boldsymbol{v} - u)^2\right] = \underbrace{\boldsymbol{w}^t\boldsymbol{\Gamma}\boldsymbol{w}}_{\text{variance}} + \underbrace{\left(\boldsymbol{w}^t\boldsymbol{u} - u\right)^2}_{\text{square bias}} = \boldsymbol{w}^t\boldsymbol{\Gamma}\boldsymbol{w} + (\boldsymbol{w}^t\boldsymbol{d})^2. \tag{5.30}$$

Based on the vector $\boldsymbol{d}$, the normalized LMMSE (NLMMSE) is implicitly defined as the solution of:

$$\boldsymbol{w}^{(\mathrm{NLMMSE})} = \underset{\boldsymbol{w}}{\arg\min} \quad \boldsymbol{w}^t\boldsymbol{\Gamma}\boldsymbol{w} + (\boldsymbol{w}^t\boldsymbol{d})^2 \quad \text{s.t.} \quad \mathbb{1}^t\boldsymbol{w} = 1 \tag{5.31}$$

The Lagrangian function for problem (5.31) is:

$$\mathcal{L}(\boldsymbol{w}, \lambda) = \boldsymbol{w}^t\boldsymbol{\Gamma}\boldsymbol{w} + (\boldsymbol{w}^t\boldsymbol{d})^2 - \lambda\,(\mathbb{1}^t\boldsymbol{w} - 1) \tag{5.32}$$

where λ is a Lagrange multiplier (i.e., dual variable). The first-order optimality conditions (Lagrange conditions) for $\boldsymbol{w}^{(\mathrm{NLMMSE})}$ to be a solution of (5.31) are:

$$\nabla_{\boldsymbol{w}} \mathcal{L}(\boldsymbol{w}, \lambda) = 0 \qquad \text{(stationarity)} \tag{5.33}$$

$$\mathbb{1}^t \boldsymbol{w} = 1 \qquad \text{(primal feasibility)}. \tag{5.34}$$

Stationarity condition (5.33) applied to the Lagrangian (5.31) gives:

$$\nabla_{\boldsymbol{w}} \mathcal{L}(\boldsymbol{w}, \lambda) = \boldsymbol{\Gamma}\boldsymbol{w} + \boldsymbol{d}\boldsymbol{d}^t\boldsymbol{w} - \lambda\mathbb{1} = 0 \tag{5.35}$$

so

$$\boldsymbol{w}^{(\mathrm{NLMMSE})} = \lambda \left(\boldsymbol{\Gamma} + \boldsymbol{d}\boldsymbol{d}^t\right)^{-1} \mathbb{1}. \tag{5.36}$$

Lagrange multiplier λ is obtained from primal feasibility condition (5.34):

$$\lambda = \frac{1}{\mathbb{1}^t \left(\boldsymbol{\Gamma} + \boldsymbol{d}\boldsymbol{d}^t\right)^{-1} \mathbb{1}}. \tag{5.37}$$

The optimal weights are then, from (5.36) and (5.37):

$$\boldsymbol{w}^{(\mathrm{NLMMSE})} = \frac{\left(\boldsymbol{\Gamma} + \boldsymbol{d}\boldsymbol{d}^t\right)^{-1} \mathbb{1}}{\mathbb{1}^t \left(\boldsymbol{\Gamma} + \boldsymbol{d}\boldsymbol{d}^t\right)^{-1} \mathbb{1}}. \tag{5.38}$$

As done in the previous paragraph, Sherman-Morisson formula can be applied to compute the matrix inversion in (5.38). For a diagonal covariance matrix $\boldsymbol{\Gamma} = \mathrm{diag}(\gamma_1^2, \ldots, \gamma_N^2)$, the k^{th} weight writes:

$$w_k^{(\mathrm{NLMMSE})} = \lambda \left(\frac{1}{\gamma_k^2} - \frac{\frac{(u_k - u)}{\gamma_k^2} \sum_l \frac{(u_l - u)}{\gamma_l^2}}{1 + \sum_l \frac{(u_l - u)^2}{\gamma_l^2}} \right), \tag{5.39}$$

with Lagrange multiplier λ computed by imposing normalization constraint (5.34).

In order to have a better understanding of Eq. (5.39), let us consider the simple case of two pixels x_1 and x_2 (i.e., $N = 2$). Assume that we want to estimate $u = u_1 = u(x_1)$ using the best combination $(1 - \rho)v(x_1) + \rho v(x_2)$. In this case, the optimal solution given by (5.39) simplifies to the following expression:

$$\rho = \frac{\gamma_1^2}{\gamma_1^2 + \gamma_2^2 + (u_1 - u_2)^2} \tag{5.40}$$

In this simple case, the optimal weights depend on the distance between noise-free values and the noise variance at pixel with index $_1$ and $_2$. We have derived in closed-form expressions the optimal weights between two noise-free values when averaging two noisy values damaged by Gaussian, gamma and Poisson noise. These expressions are directly obtained from Eq. (5.40):

The case of Gaussian noise: Noise is additive and the optimal weights depend on the Euclidean distance between noise-free values and the noise level σ:

$$\rho = \frac{\sigma^2}{2\sigma^2 + (u_1 - u_2)^2}\,. \tag{5.41}$$

using the fact that $\gamma_1 = \gamma_2 = \sigma^2$.

The case of gamma noise: Noise is multiplicative and the optimal weights depend on the ratio of noise-free values:

$$\rho = \frac{1/2}{1 + \frac{u_2^2}{u_1^2} - \frac{u_2}{u_1}} \, . \tag{5.42}$$

using the fact that $\gamma_1 = u_1^2$ and $\gamma_2 = u_2^2$.

The case of Poisson noise: Noise is neither additive nor multiplicative and the optimal weights provides the following weight definition:

$$\rho = \frac{1}{1 + \frac{u_2}{u_1} - \frac{(u_1 - u_2)^2}{u_1}} \, . \tag{5.43}$$

using the fact that $\gamma_1 = u_1$ and $\gamma_2 = u_2$.

The expression (5.39) is interesting because it corresponds to the optimal WMLE filter (independent of the noisy data) for a given true image. Its expression only depends on the differences $u_k - u$ and the variance γ_k^2, and is then **dependent on the noise model**. As well as the oracle-LMMSE, the oracle-NLMMSE provides remarkable performance. Since we did not impose the weights to be positive, oracle-NLMMSE is also able to cancel the bias by taking a non-convex combination of samples v_k (see Fig. 5.8). Again, this filter will provide a result too close to the pre-estimate (see Fig. 5.9).

In conclusion, the optimal weights defined only from the noise-free patches adapt to the noise distribution. The underlying idea is that to minimize the bias-variance trade-off we have to select noisy samples whose average is an estimate of the noise-free value. If we have two close noise-free values modeling two very different noisy populations, mixing these populations will lead to a poor estimate (see Fig. 5.7). In the framework of WMLE, we want to select noisy samples $v(x')$ whose distributions are close to the pdf $p(.|u(x))$. Weights should use the knowledge provided by pre-estimates to measure the similarity of the noisy realizations that they describe.

Oracle with Kullback-Leibler based weights

We have seen that, providing a pre-estimate of the noise-free images, weights can be refined if, when measuring the similarity between the patches extracted form this pre-estimate, the noise distribution is considered. Instead of selecting pixels with similar noise-free values, weights should select pixels whose noisy realizations follow a similar distribution. Figure 5.7 gives an illustration with Rayleigh noise that when considering similarity between noise-free values from the Euclidean distance, we select a mixture of populations whose resulting distribution can be far from the target distribution.

Polzehl and Spokoiny (2006a) used the symmetrical Kullback-Leibler divergence between the estimates[1] as a statistical measure of our hypothesis test (5.13). For simplicity, we will in the following speak about the Kullback-Leibler (KL) divergence to denote the symmetrical Kullback-Leibler divergence. The KL divergence is a measure between distributions and, as a consequence, allows us to select almost identically distributed noisy samples. The KL divergence between two noise-free patches $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ is given by:

$$\mathcal{D}_{KL}(\boldsymbol{u}_1, \boldsymbol{u}_2) = \int (p(\boldsymbol{v}|\boldsymbol{u}_1) - p(\boldsymbol{v}|\boldsymbol{u}_2)) \log \frac{p(\boldsymbol{v}|\boldsymbol{u}_1)}{p(\boldsymbol{v}|\boldsymbol{u}_2)} \, d\boldsymbol{v} \, . \tag{5.44}$$

[1] in their work noise, they considered distributions from the exponential family

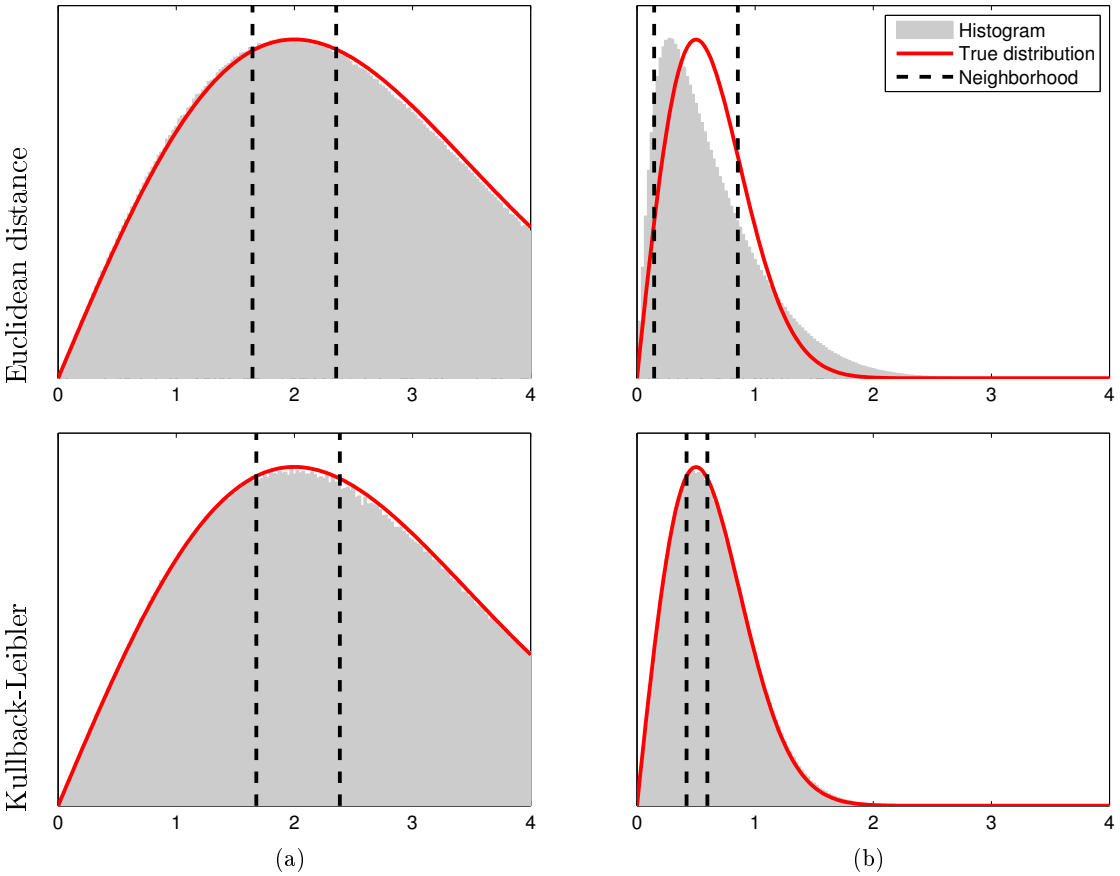


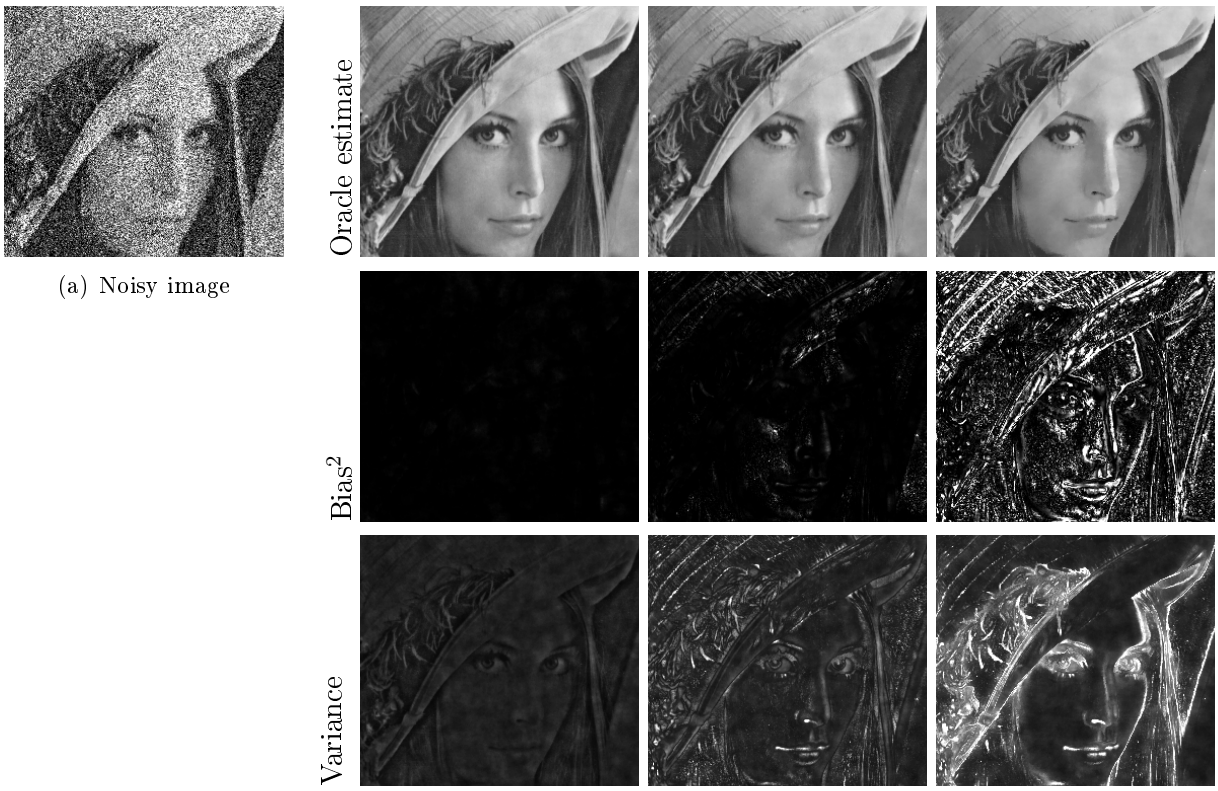
Figure 5.7: Distribution of the mixture of the populations following a Rayleigh distribution. When the mixture is located in a neighborhood such that (top) the populations have noise-free values u_2 close to u_1: $(u_1 - u_2)^2 < \zeta$, and (bottom) the populations are statistically closed: $D_{KL}(u_1 \| u_2) < \eta$. (a) For a given noise-free value $u_1 = 2$, the thresholds ζ and η have been tuned such that the mixture of the populations follows almost the same distribution as the one of v_1, but (b) by using the same thresholds when $u_1 = 0.5$, only the selection based on Kullback-Leibler succeeds to select the good populations. (In red the true distribution for $u_1 = 2$ and $u_1 = 0.5$, in gray the histogram of the distribution of the values selected in the neighborhood defined by the Euclidean distance or the KL divergence – dashed lines –).

When noise is uncorrelated, the KL divergence decomposes as $\mathcal{D}_{KL}(\boldsymbol{u}_1, \boldsymbol{u}_2) = \sum \mathcal{D}_{KL}(u_{1,k}, u_{2,k})$.

There are several examples where the Euclidean distance between noise-free patches cannot be used. For instance, in the case where the noise-free data are non-scalar entities composed of several parameters of different nature. In this case, the Euclidean distance needs to take into account the dynamics of the different channels introducing useless parameters. Wrapped phase, impulse noise and salt-and-pepper noise are other examples where the use of the Euclidean distance is a non-sense. The use of the KL divergence allows us to avoid such difficulties since the KL divergence is defined on the parameter space.

The KL divergence is defined from the parameter space to $\mathbb{R}^+$. It cancels if and only if the two distributions are equal, i.e., $\boldsymbol{u}_1 = \boldsymbol{u}_2$. The larger the KL divergence the more dissimilar the underlying distributions. Since weights should take their values in $[0, 1]$, we suggest using the following weight definition:

$$w(x, x') = \varphi\left[\mathcal{D}_{KL}(u(\mathrm{P}_x), u(\mathrm{P}_{x'}))\right] \tag{5.45}$$



(a) Noisy image

(b) Oracle-LMMSE (c) Oracle-NLMMSE (d) Oracle-KL

Figure 5.8: Comparisons of oracle-based filtering driven by the true image on a noisy version of the *lena* image damaged by synthetic Poisson noise. The oracle-LMMSE filter provides the best result with a quasi zero bias and a quasi optimal variance reduction (the factor of reduction is directly linked to the size of the search window). The oracle-NLMMSE filter provides the optimal WMLE performance while oracle-KL filter has a lower performance in terms of bias-variance trade-off.

where φ is the kernel decay function $\mathbb{R}^+ \to [0,1]$ as defined in Sec. 5.4.1. This definition matches with the KL divergence based kernel proposed in (Moreno et al., 2004) for machine learning. Following (Moreno et al., 2004; Polzehl and Spokoiny, 2006a), we have chosen to use the KL divergence. The Bhattacharyya could have also been considered following (Goudail et al., 2004).

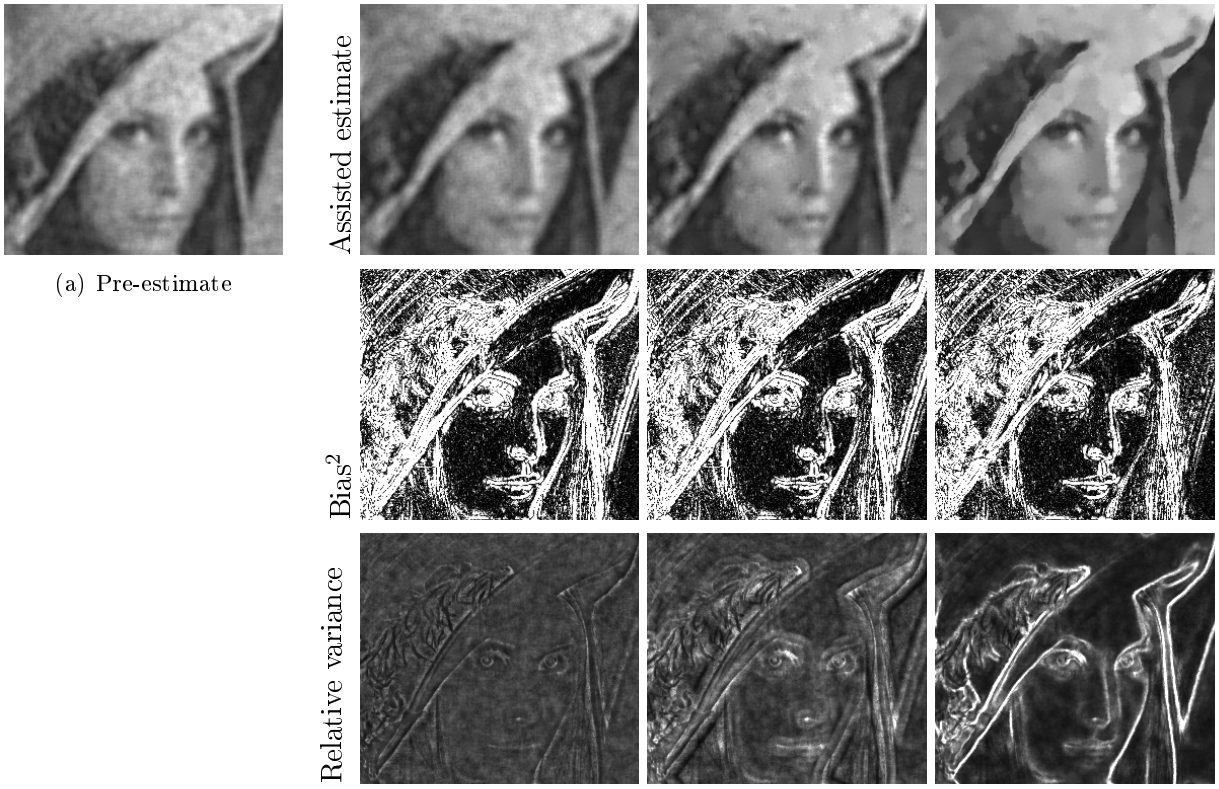
We have derived, in closed-form expressions, the weights between two noise-free patches for Gaussian, gamma and Poisson noise. From Eq. (5.45), the following expressions have been obtained:

The case of Gaussian noise: Noise is additive and the KL divergence is linked to the difference between the pre-filtered patches. It corresponds to the Gaussian kernel leading to weights defined from the Euclidean distance:

$$w(x,x') = \varphi\left(\frac{\|u(\mathrm{P}_x) - u(\mathrm{P}_{x'})\|_2^2}{\sigma^2}\right) . \tag{5.46}$$

The case of gamma noise: Noise is multiplicative and the KL divergence is linked to the ratio of the noisy patches:

$$w(x,x') = \varphi\left[L \sum_{\tau \in \mathrm{P}} \left(\frac{u(x+\tau)}{u(x'+\tau)} + \frac{u(x'+\tau)}{u(x+\tau)} - 2\right)\right] . \tag{5.47}$$



(a) Pre-estimate (b) Assisted-LMMSE (c) Assisted-NLMMSE (d) Assisted-KL

Figure 5.9: Comparisons of assisted-based filtering driven by a pre-estimate on a noisy version of the *lena* image damaged by synthetic Poisson noise. The assisted-KL filter provides the best result in terms of bias-variance trade-off. Assisted-LMMSE or NLMMSE filters give too much confidence on the pre-estimate and there is no improvement compared to the pre-estimate itself.

Note that if we consider the log transform $\tilde{\boldsymbol{v}} = \log \boldsymbol{v}$, whose noise component is additive, and if we perform the same transform on the noise-free image $\tilde{\boldsymbol{u}} = \log \boldsymbol{u}$, KL rewrites as the difference between the log-transform of the noise-free patches:

$$w(x,x') = \varphi\left[2L \sum_{\tau \in \mathrm{P}} \big(\cosh(\tilde{u}(x+\tau) - \tilde{u}(x'+\tau)) - 1\big)\right] . \tag{5.48}$$

The case of Poisson noise: Noise is neither additive nor multiplicative and the KL divergence provides the following weight definition:

$$w(x,x') = \varphi\left[\sum_{\tau \in \mathrm{P}} (u(x+\tau) - u(x'+\tau)) \log \frac{u(x+\tau)}{u(x'+\tau)}\right] . \tag{5.49}$$

Compared to the optimal weights definitions given in Eq. (5.26) and (5.39), these expressions are easier to evaluate. Unlike the two other oracles, the oracle-KL filter is a WMLE with positive weights. Compared to oracle-LMMSE or NLMMSE, the Kullback-Leibler divergence does not give too much confidence in $\boldsymbol{u}$ such that an estimate $\hat{\boldsymbol{u}}$ can be used instead to drive the WMLE-based filtering (see Fig. 5.9).

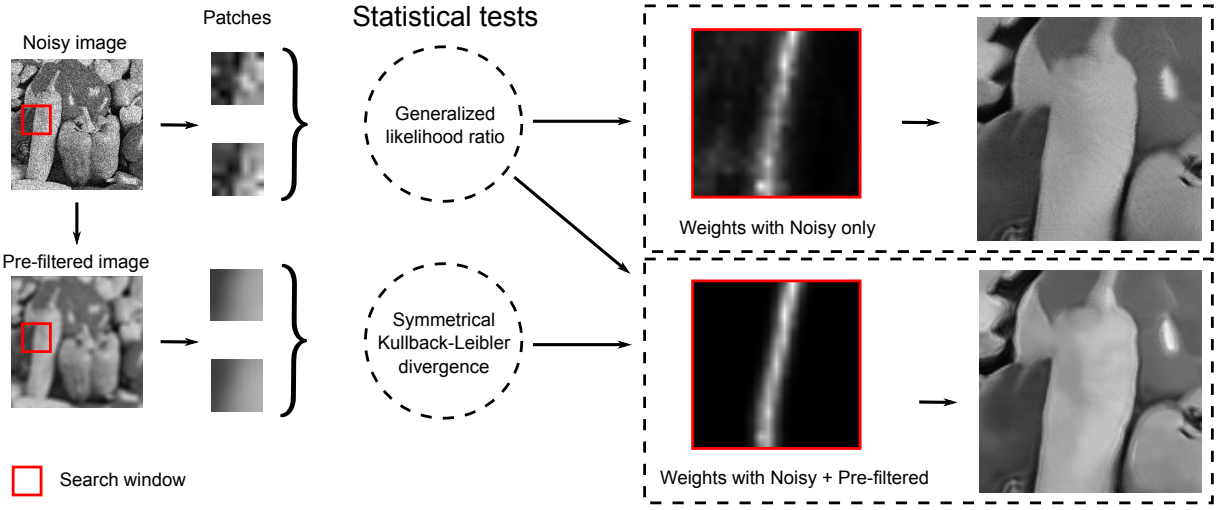


Figure 5.10: Illustration of the improvements obtained by refining the weights. Under a high noise level, the similarity between noisy patches (given by the generalized likelihood ratio) is not robust enough. When the weights are refined using the similarity between pre-filtered patches (given by the symmetrical Kullback-Leibler divergence), we can decrease the noise more while preserving the edges.

5.4.3 Joint similarity between noisy and noise-free patches

According to the noise level or to the quality of the pre-estimate, one can define weights from the similarity between noisy patches or from the similarity between the pre-estimated patches. In an intermediate case, for instance, under a low noise level with a good pre-estimate or under a high noise level with a bad pre-estimate, we could take advantage of both similarity criteria. Taking a convex combination of both would lead to the following definition:

$$w(x,x') = \varphi\left[(1-\lambda) f\left(-\log \mathcal{L}_G(v(\mathrm{P}_x), v(\mathrm{P}_{x'}))\right) + \lambda g\left(\mathcal{D}_{KL}(\hat{u}(\mathrm{P}_x), \hat{u}(\mathrm{P}_{x'}))\right)\right] \tag{5.50}$$

where the parameter $\lambda \in [0,1]$ controls the confidence we have in the pre-estimate and f and g are two increasing affine transforms chosen such that both criteria answer with the same dynamic. The amount of filtering is controlled by φ, for instance, by the bandwidth parameter h used in the exponential decay kernel. Note that Eq. (5.50) is over-parameterized here and, in the case of the exponential decay kernel, the number of parameters can be reduced to two parameters $\alpha > 0$ and $\beta > 0$:

$$w(x,x') = \exp\left(-\frac{-\log \mathcal{L}_G(v(\mathrm{P}_x), v(\mathrm{P}_{x'}))}{\alpha} - \frac{\mathcal{D}_{KL}(\hat{u}(\mathrm{P}_x), \hat{u}(\mathrm{P}_{x'}))}{\beta}\right) . \tag{5.51}$$

The advantage of Eq. (5.50) is the better control and interpretation on the behavior of each parameters while in Eq. (5.51), the two parameters α and β jointly influence on both the confidence we have in the pre-estimate and the amount of filtering (then, they are harder to set in a supervised way). However, Eq. (5.51) involves only two parameters that we will be able to set automatically (see Sec. 5.5)

Weights defined from the sum of these two terms can be interpreted in a Bayesian framework. If we consider our original hypothesis test (5.13), we can consider the following *a posteriori* test:

$$\frac{p(\mathscr{H}_0|\boldsymbol{v}_1,\boldsymbol{v}_2)}{p(\mathscr{H}_1|\boldsymbol{v}_1,\boldsymbol{v}_2)} = \underbrace{\frac{p(\boldsymbol{v}_1,\boldsymbol{v}_2|\mathscr{H}_0)}{p(\boldsymbol{v}_1,\boldsymbol{v}_2|\mathscr{H}_1)}}_{\text{Likelihood ratio}} \times \underbrace{\frac{p(\mathscr{H}_0)}{p(\mathscr{H}_1)}}_{\text{Extra information}} . \tag{5.52}$$

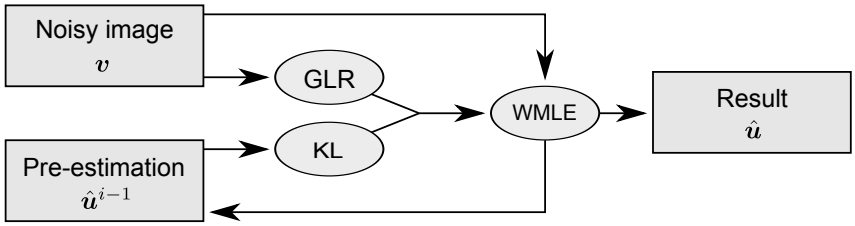


Figure 5.11: Scheme of the iterative filtering. The weights $w(x, x')$ are computed using the noisy image $\boldsymbol{v}$ and the estimate $\hat{\boldsymbol{u}}^{i-1}$. The WMLE computes the new parameters $\hat{\boldsymbol{u}}^{i}$ by using the weights $w(x, x')$ and the noisy image $\boldsymbol{v}$. The procedure is repeated until there is no longer any change between two consecutive estimates.

The likelihood ratio term corresponds to the noisy data fidelity measured in our case by GLR. The second term gives an *a priori* knowledge brought by an extra information. For instance, bilateral filters can be explained by Eq. (5.52) with the extra information that close pixels usually have similar noise-free values. In our case the extra information is the available pre-filtered image: noisy values following similar pre-estimated distribution, in the sense of the KL divergence, can be used together to estimate the noise-free value.

Figure 5.10 illustrates the improvements obtained by refining the weights. The maps of the weights obtained in the same search window are compared with and without weight refinement. The resuling images are also given. When the weights are refined using the similarity between pre-filtered patches, we can decrease the noise more while preserving the edges.

5.4.4 Two step or iterative filtering

This refining procedure of the weights can be done either in two steps or iteratively. In a two step strategy, the image $\hat{\boldsymbol{u}}$ is estimated first from the noisy image $\boldsymbol{v}$ (e.g., using a moving average filter). The result depends on the quality of the pre-filtering with respect to the noise level. In an iterative strategy, at iteration $i-1$, the estimate $\hat{\boldsymbol{u}}$ provides the estimate $\hat{\boldsymbol{u}}^{i-1}$ used at iteration i. Since all the pixels in $\hat{\boldsymbol{u}}^{i-1}$ are updated before moving to the next iteration, this corresponds to a synchronous local iterative method (see Bratsolis and Sigelle, 2003). This kind of algorithm converges to a solution depending on the initialization $\hat{\boldsymbol{u}}^{1}$.

Figure 5.11 illustrates the iterative procedure:

1. first, the weights $w(x, x')$ are estimated by using patches extracted from the noisy image $\boldsymbol{v}$ and by using patches extracted from the pre-estimated image $\hat{\boldsymbol{u}}^{i-1}$ (see Eq. (5.50));
2. next, the WMLE provides the new image $\hat{\boldsymbol{u}}^{i}$ by using the weights $w(x, x')$ and the noisy image $\boldsymbol{v}$ (see Eq. (5.3));
3. steps 1 and 2 are repeated until there is no longer any change between two consecutive estimates.

In this iterative filtering, the weights are defined by two terms. The first one, the data fidelity term, depends on the original noisy image and considers its pixel values as a realization of the noise generative model. The second term is calculated from the previously estimated image and considers its pixel values as the "true" parameters of the noise generative model. This idea is different from the iterative NL means versions defined in (Awate and Whitaker, 2006; Kervrann and Boulanger, 2008; Brox et al., 2008; Goossens et al., 2008), where only previously estimated parameters are used to compute the similarity criterion (a large confidence is attributed to the previous iterations). Moreover, in (Awate and Whitaker, 2006; Goossens et al., 2008) a weighted

average is performed on the previously estimated image instead of the noisy image: they do not fit the WMLE definition. Averaging noisy values and keeping weights driven by the noisy image seems to be a better strategy to stay closer to the noise-free image.

This iterative scheme can be related to the Expectation-Maximization (EM) procedure of Dempster et al. (1977). The EM algorithm is a two step iterative algorithm which converges to a local optimum depending on the initial estimate. The first step (E-Step) evaluates a complete-data likelihood expectation by computing sufficient parameters using a previous estimate, while we evaluate a weighted likelihood by computing similarity probabilities using the previous estimate $\hat{\boldsymbol{u}}^{i-1}$. The second step (M-Step) maximizes the complete-data likelihood expectation, while we maximize the weighted likelihood. As in the EM procedure, the filter also considers the previous estimate as "true" parameters. According to our experiments, this consideration involves the model stability over the different iterations and provides the convergence of our method. Nevertheless, our function is not related to a complete-data likelihood expectation over our latent variable $\delta_{\mathcal{S}_{u(x)}}(x')$. The similarity between two patches is a good indication that their central values are close (as demonstrated by the performance of the NL means). Dissimilar patches however do not provide any clue on the difference or closeness between the central values. The complete-data likelihood expectation that should be computed in a normal E-Step is therefore less relevant in our context. Finally, our latent variable definition makes the algorithm locally defined for all pixels x. Thus, the filter is a synchronous local iterative method whereas an EM algorithm would try to solve iteratively the problem directly on the global image.

5.5 Unsupervised setting of the parameters

The setting of the parameters in the case of non-Gaussian noise is maybe a more critical problem than in the Gaussian case. In (Buades et al., 2005; Polzehl and Spokoiny, 2006a; Kervrann and Boulanger, 2006), the authors propose to define the parameters according to the variance or the quantiles of the similarity criterion when it is subject to identical and independent distributed random variables (see Sec. 2.5.1). Unfortunately, for some distributions such as Poisson noise, these quantities depend on the unknown image $\boldsymbol{u}$ since GLR is not CFAR in this case. Van De Ville and Kocher (2009) propose a risk minimization approach for Gaussian noise. Their method selects the parameters minimizing an unbiased estimate of the quadratic risk (without any specific assumption on the underlying image $\boldsymbol{u}$). This kind of approach seems relevant in the case of non-Gaussian noise as long as one can provide a relevant estimate of the quadratic risk.

The parameters of the denoising technique can be selected as those that minimize the expected mean square error (MSE):

$$\mathbb{E}\left[\frac{1}{N}\|\boldsymbol{u}-\hat{\boldsymbol{U}}\|_2^2\right] = \frac{1}{N}\sum_k \left(u_k^2 + \mathbb{E}\left[\hat{U}_k^2\right] - 2\mathbb{E}\left[u_k\hat{U}_k\right]\right) \tag{5.53}$$

where N is the image size. Searching the estimator which minimizes the MSE enables us to find the trade-off between bias and variance reduction. Since the MSE requires the knowledge of $\boldsymbol{u}$, unbiased estimators $R[\hat{\boldsymbol{u}}]$ of the MSE can be used instead, such that, Stein's unbiased risk estimator (SURE), the generalized SURE or the Poisson unbiased risk estimator (PURE) (see Sec. 2.6.4 and 3.6.3).

5.5.1 Risk minimization for our extension of the NL means

Selecting parameters that minimize an estimator of the risk gives parameters close to that minimizing the MSE. In the case of the classical NL means, the authors of (Van De Ville and Kocher,

2009; Duval et al., 2011; Luisier et al., 2010; Van De Ville and Kocher, 2011) compute the optimal parameters by exhaustive search on a predefined grid. Optimization techniques can be applied to reach the optimal parameters in few iterations. In (Zhang and Luo, 1999), a gradient descent is performed to optimize SURE for wavelet shrinkage, while Doré and Cheriet (2009) use Newton's method to select the bandwidth parameter of the NL means minimizing Mallow's C_p statistics. We follow such a strategy here to optimize $R(\hat{\boldsymbol{u}})$ for our extension of the NL means using the joint similarity between noisy and noise-free patches as defined in Eq. (5.51). We apply Newton's method on the joint filtering parameters α and β. Newton's method iteratively refines α and β with the update:

$$\begin{pmatrix} \alpha^{(n+1)} \\ \beta^{(n+1)} \end{pmatrix} = \begin{pmatrix} \alpha^{(n)} \\ \beta^{(n)} \end{pmatrix} - H(R)^{-1}\nabla R \tag{5.54}$$

$$\text{with } H(R)^{-1}\nabla R = \begin{pmatrix} \frac{\partial^2 R(\hat{\boldsymbol{u}}^{(n)})}{\partial \alpha^2} & \frac{\partial^2 R(\hat{\boldsymbol{u}}^{(n)})}{\partial \alpha \partial \beta} \\ \frac{\partial^2 R(\hat{\boldsymbol{u}}^{(n)})}{\partial \beta \partial \alpha} & \frac{\partial^2 R(\hat{\boldsymbol{u}}^{(n)})}{\partial \beta^2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial R(\hat{\boldsymbol{u}}^{(n)})}{\partial \alpha} \\ \frac{\partial R(\hat{\boldsymbol{u}}^{(n)})}{\partial \beta} \end{pmatrix}$$

where n is the current iteration index, $H(R)$ the hessian of the risk R and ∇R the gradiant of the risk. To perform the optimization procedure in (5.54), the closed-form expressions of the first and second order differentials are required.

Newton's method finds in few iterations the best trade-off between the information brought by the noisy image and the pre-estimated image to define the weights. For instance, β will get a high value when the pre-estimated image has a poor quality, resulting in weights determined only from the noisy image. Reciprocally, α will get a high value when the pre-estimated image has a high quality: the weights will be determined only from the well pre-estimated image. Figure 5.12 illustrates the influence of the parameters α and β on the solution of our extension of the NL means.

The main result in (Van De Ville and Kocher, 2009) is that SURE for the NL means can be obtained in closed-form. From its closed-form expression, the first and second order differentials can be obtained and Newton's method (5.54) can be performed for images damaged by Gaussian noise. Next, we will consider the case of Poisson noise, and provide the corresponding closed-form expressions.

5.5.2 Application of the methodology for Poisson noise

Let us consider the case of our extension of the NL means given by:

$$\begin{aligned} \hat{u}(x_k) &= \frac{\sum_l w_{k,l} v_l}{\sum_l w_{k,l}} \\ \text{with} \quad w_{k,l} &= w(x_k, x_l) = \exp\left(-\frac{F_{k,l}}{\alpha} - \frac{G_{k,l}}{\beta}\right), \\ F_{k,l} &= \sum_\tau f(v(x_k+\tau), v(x_l+\tau)), \\ \text{and} \quad G_{k,l} &= \sum_\tau g(\hat{u}(x_k+\tau), \hat{u}(x_l+\tau)) \end{aligned} \tag{5.55}$$

where $f(v_1, v_2) = -\log \mathcal{L}_G(v_1, v_2)$ and $g(\hat{u}_1, \hat{u}_2) = \mathcal{D}_{KL}(\hat{u}_1, \hat{u}_2)$. Here, we restrict our study to the case of Poisson noise. Note however that the same methodology could be used for Gaussian noise, gamma noise or for any noise model for which we have an estimate of the risk. The closed-form expression of the risk estimator for Poisson noise, i.e., PURE, on our extension of the NL

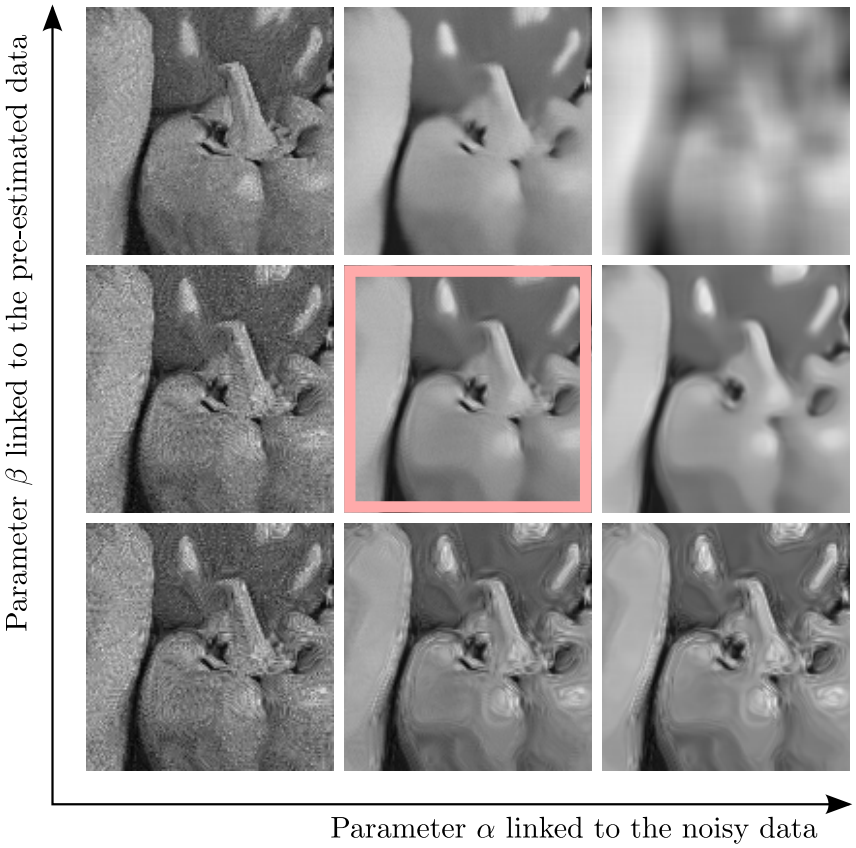


Figure 5.12: Illustration of the influence of the parameters α and β on the solution of our extension of the NL means. When α and β are too small, the redundancy is minimal and the noise is not reduced. At the opposite, when α and β are too big, the redundancy is maximal and the resulting image is blurry. Too much confidence on the pre-estimate will introduce artifacts. The best α-β trade-off (at the center) is the one which minimizes the mean square error (MSE).

means (5.55) is given by Eq. (3.47) that can rewrite as:

$$R(\hat{\boldsymbol{u}}) = \sum_x u(x)^2 + \hat{u}(x)^2 - 2v(x)\hat{\overline{u}}(x) \tag{5.56}$$

$$\text{with} \quad \hat{\overline{u}}(x_k) = \frac{\sum_l \overline{w}_{k,l} v_l}{\sum_l \overline{w}_{k,l}}, \tag{5.57}$$

$$\overline{w}_{k,l} = \overline{w}(x_k, x_l) = \exp\left(-\frac{\overline{F}_{k,l}}{\alpha} - \frac{G_{k,l}}{\beta}\right),$$

$$\overline{F}_{k,l} = \sum_\tau f(\overline{v}(x_k+\tau), \overline{v}(x_l+\tau)) \ .$$

Note that (5.57) holds by assuming that $G_{k,l}$ (i.e the pre-estimate $\hat{\boldsymbol{u}}$) does not depend on the noise component of $\boldsymbol{v}$. To satisfy this assumption, the noise variance in $\hat{\boldsymbol{u}}$ has to be reduced significantly. This assumption simplifies drastically the expression of $\overline{\boldsymbol{u}}$.

In terms of time complexity, we note as in (Van De Ville and Kocher, 2009) that the computation time is unchanged since the computation of PURE can be incorporated within the core of the NL means. Moreover, the scan of the patches of $\overline{\boldsymbol{v}}$ can be avoided thanks to the following

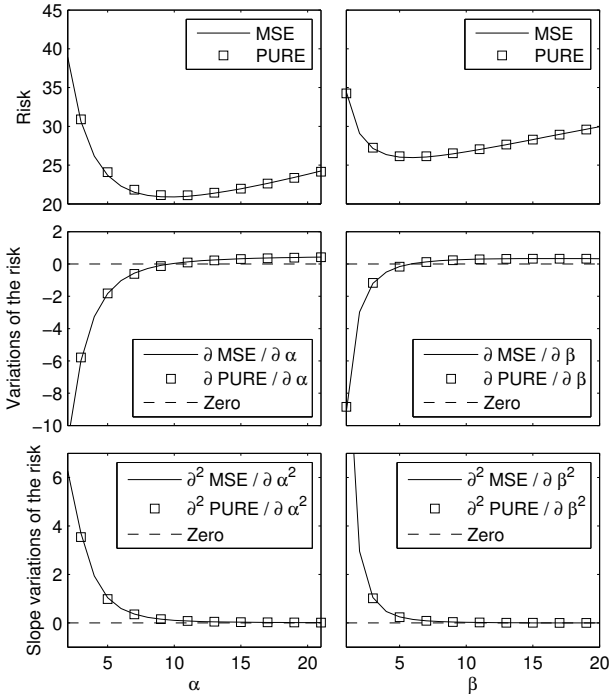


Figure 5.13: The risk (MSE and PURE) and their two first order variations (from top to bottom) with respect to the parameters α (left) and β (right).

relation:

$$\overline{F}_{k,l} = F_{k,l} + \begin{cases} f(\overline{v}_k, \overline{v}_k) - f(v_k, v_k), & \text{if } x_k = x_l, \\ f(\overline{v}_k, v_l) - f(v_k, v_l) & \\ \quad + f(v(2x_k - x_l), \overline{v}_k) - f(v(2x_k - x_l), v_k), & \text{if } x_k \in \mathrm{P}_{x_l}, \\ f(\overline{v}_k, v_l) - f(v_k, v_l), & \text{otherwise.} \end{cases}$$

Newton's method can then be performed using the following expressions given by substituting μ and ν by α or β in the following equations:

$$\begin{aligned} \frac{\partial R(\hat{\boldsymbol{u}})}{\partial \mu} =& \frac{2}{N} \sum_k \hat{u}_k \frac{\partial \hat{u}_k}{\partial \mu} - \frac{2}{N} \sum_k v_k \frac{\partial \hat{\overline{u}}_k}{\partial \mu}, \\ \frac{\partial^2 R(\hat{\boldsymbol{u}})}{\partial \mu \partial \nu} =& \frac{2}{N} \sum_k \hat{u}_k \frac{\partial^2 \hat{u}_k}{\partial \mu \partial \nu} + \frac{2}{N} \sum_k \left(\frac{\partial \hat{u}_k}{\partial \mu} \right) \left(\frac{\partial \hat{u}_k}{\partial \nu} \right) - \frac{2}{N} \sum_k v_k \frac{\partial^2 \hat{\overline{u}}_k}{\partial \mu \partial \nu}, \end{aligned} \tag{5.58}$$

with:

$$\begin{aligned} \frac{\partial \hat{u}_k}{\partial \mu} =& \frac{\sum X_{k,l} w_{k,l} (k_t - \hat{u}_k)}{\mu^2 \sum w_{k,l}}, \\ \frac{\partial^2 \hat{u}_k}{\partial \mu^2} =& \frac{\sum X_{k,l}^2 w_{k,l} (k_t - \hat{u}_k)}{\mu^4 \sum w_{k,l}} - 2 \frac{\partial \hat{u}_k}{\partial \mu} \frac{\sum (X_{k,l} + \mu) w_{k,l}}{\mu^2 \sum w_{k,l}}, \\ \frac{\partial^2 \hat{u}_k}{\partial \mu \partial \nu} =& \frac{\sum X_{k,l} Y_{k,l} w_{k,l} (k_t - \hat{u}_k)}{\mu^2 \nu^2 \sum w_{k,l}} - \frac{\partial \hat{u}_k}{\partial \mu} \frac{\sum Y_{k,l} w_{k,l}}{\nu^2 \sum w_{k,l}} - \frac{\partial \hat{u}_k}{\partial \nu} \frac{\sum X_{k,l} w_{k,l}}{\mu^2 \sum w_{k,l}} \end{aligned}$$

where $X = F$ (resp. $Y = F$) when $\mu = \alpha$ (resp. $\nu = \alpha$) and $X = G$ (resp. $Y = G$) when $\mu = \beta$ (resp. $\nu = \beta$). The differentials for $\overline{\boldsymbol{u}}$ are the same with respect to $\overline{k}$, $\overline{w}$ and $\overline{F}$. The resulting filter has been coined *Poisson NL means*.

Gaussian noise	$\sigma=60$	$\sigma=40$	$\sigma=20$	$\sigma=10$
Barbara				
Noisy image	0.04	3.09	8.80	14.73
K-SVD	09.29	13.01	17.43	21.02
BM3D	**12.14**	**14.59**	**18.38**	**21.48**
NL means	10.24	12.85	16.97	19.85
Our filter (25×)	10.99	13.49	15.96	18.69
Boat				
Noisy image	-1.49	1.63	7.42	13.41
K-SVD	9.04	11.78	15.62	18.87
BM3D	**10.55**	**12.83**	**16.09**	**19.09**
NL means	8.96	11.06	14.63	17.59
Our filter (25×)	9.50	11.63	14.51	17.19
House				
Noisy image	-1.62	1.45	7.26	13.27
K-SVD	10.22	14.36	18.31	21.15
BM3D	**13.28**	**15.78**	**18.94**	**21.77**
NL means	10.40	13.33	17.55	20.25
Our filter (25×)	11.57	14.20	17.03	19.59
Lena				
Noisy image	-1.25	1.81	7.60	13.59
K-SVD	11.09	14.18	17.81	20.93
BM3D	**13.05**	**15.33**	**18.42**	**21.27**
NL means	11.33	13.66	17.10	20.12
Our filter (25×)	11.99	14.20	16.90	19.50

Gamma noise	$L=1$	$L=2$	$L=4$	$L=16$
Barbara				
Noisy image	-1.09	1.69	4.61	10.57
WIN-SAR	8.82	10.48	12.04	15.82
MAP-UWD-S	9.65	11.44	13.28	16.93
Our filter (1×)	9.79	11.88	**14.05**	**17.83**
Our filter (25×)	**10.58**	**12.51**	13.98	16.59
Boat				
Noisy image	-2.99	-0.18	2.70	8.67
WIN-SAR	8.57	10.65	12.14	15.17
MAP-UWD-S	9.26	10.68	**12.31**	**15.71**
Our filter (1×)	8.71	10.49	12.22	15.33
Our filter (25×)	**9.43**	**10.91**	12.25	15.10
House				
Noisy image	-3.55	-0.76	2.11	8.10
WIN-SAR	8.69	11.42	13.15	16.24
MAP-UWD-S	10.34	11.97	13.72	17.24
Our filter (1×)	9.06	11.61	14.29	**18.27**
Our filter (25×)	**10.46**	**12.98**	**14.50**	17.42
Lena				
Noisy image	-2.45	0.34	3.25	9.19
WIN-SAR	10.35	13.00	14.72	17.90
MAP-UWD-S	11.87	13.53	15.14	**18.65**
Our filter (1×)	11.05	13.20	15.18	18.61
Our filter (25×)	**12.16**	**13.95**	**15.25**	18.10

Table 5.1: SNR values of estimated images using different denoising methods for images corrupted by Gaussian noise with different standard deviations and by gamma noise with different equivalent numbers of looks.

Figure 5.13 shows the risk and its two first order differentials with respect to α and β. These curves have been computed by applying the proposed method on a 150×150 noisy image for different values of the parameters. The MSE and its differentials have been computed from the noise-free image and finite differences. All estimates seem to fit the ground truth well.

5.6 Experiments and results

5.6.1 Simulations

This section presents visual and numerical results obtained on three synthetic images corrupted by Gaussian noise, gamma noise and Poisson noise. The corrupted images are obtained from three classical noise-free images: *Barbara*, *Boat* and *House*. On all noisy images, our extension of the NL means has been applied (iteratively for Gaussian and gamma noise and in a two step strategy with unsupervised setting for Poisson noise). A search window of size $|W| = 21 \times 21$ px^2 and patches of size $|\mathrm{P}| = 7 \times 7$ px^2 have been used. The iterative filtering is initialized with the result of our extension of the NL means using the similarity between noisy patches only. The two pass filtering used a pre-estimate provided by a moving average filter with a convolution disk of radius 5 px. Denoising parameters α and β have been determined in a supervised way using the quantile method for Gaussian and gamma noise (in this case GLR is CFAR). Since GLR is not CFAR for Poisson noise, we used instead the unsupervised setting based on risk minimization. Some comparison with the latest *state-of-the-art* filters are provided. For Gaussian noise, the comparisons have been performed with the original NL means (Buades et al., 2005), the K-singular value decomposition (K-SVD) (Aharon et al., 2006) and the block-matching and 3D collaborative filtering (BM3D) (Dabov et al., 2007). Note that the NL means

Poisson noise	$p = 5$	$p = 10$	$p = 20$	$p = 150$
Barbara				
Noisy image	-2.82	0.13	3.16	11.88
P-HaarTI	9.08	9.75	10.56	15.64
P-BM3D	**11.30**	**13.23**	**15.17**	**19.86**
P-SAFIR	9.14	11.05	13.03	17.46
Our filter (2×)	9.97	11.72	13.65	18.63
Boat				
Noisy image	-4.81	-1.80	1.20	9.95
Haar	9.20	10.45	11.72	15.82
P-BM3D	**9.97**	**11.58**	**13.04**	**17.40**
P-SAFIR	8.76	9.91	11.68	16.04
Our filter (2×)	9.20	10.57	12.06	16.21
House				
Noisy image	-4.95	-1.93	1.09	9.84
P-HaarTI	10.62	12.10	13.74	17.80
P-BM3D	**11.92**	**13.94**	**15.87**	**19.97**
P-SAFIR	9.24	10.93	13.14	18.10
Our filter (2×)	10.96	12.57	14.50	19.09
Lena				
Noisy image	-4.23	-1.23	1.75	10.52
P-HaarTI	11.57	13.06	14.51	18.49
P-BM3D	**12.16**	**14.02**	**15.74**	**20.02**
P-SAFIR	10.39	11.61	13.67	18.45
Our filter (2×)	11.73	13.22	14.89	19.28

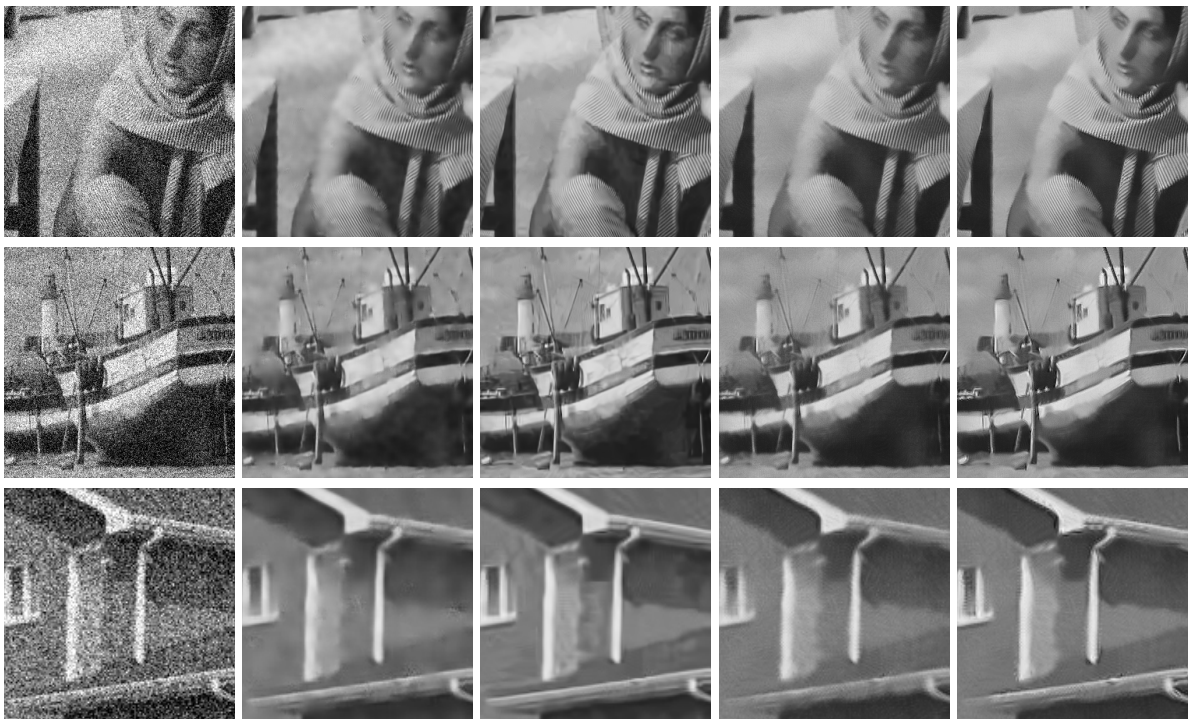
Table 5.2: SNR values of estimated images using different denoising methods for images corrupted by Poisson noise with different maximum peak values.

filter corresponds here to a non-iterative version of our filter. For gamma noise, the comparisons have been performed with the wavelet-based image-denoising non-linear SAR (WIN-SAR) filter (Achim et al., 2003) and the MAP filter based on undecimated wavelet decomposition and image segmentation (MAP-UWD-S) (Bianchi et al., 2008). For Poisson noise, the comparisons have been performed with an approach based on translation invariant Haar-wavelet transform for Poisson noise (P-HaarTi) (Willett and Nowak, 2004), the BM3D filter using the optimal inversion of Anscombe transform (P-BM3D) (Mäkitalo and Foi, 2011) and the SAFIR filter using also the optimal inversion of Anscombe transform (P-SAFIR) (Boulanger et al., 2008).

Figures 5.14, 5.15 and 5.16 present the obtained denoised images for the images corrupted respectively by Gaussian noise with a standard deviation $\sigma = 40$, by gamma with an equivalent number of $L = 3$ looks and by Poisson noise with a maximum peak of 20 photons. Note that these three noise levels have been chosen because they provide comparable levels of signal-to-noise ratio (SNR, definition given in Sec. 2.1.3). Note that for synthetic SAR images the square roots of the images are displayed for a better visual assessment.

In the case of Gaussian noise, the images obtained with our filter seem to be well smoothed with better edge and shape preservation than the NL means; refining weights is necessary for high level of noise. The images denoised by the K-SVD and the BM3D filters present some artifacts while our filter provides smoother regions with comparable edge preservation. However, our filter seems to attenuate the image contrast and thin and dark structures such as the mouth of *Lena*, the eyes of *Barbara* and the ropes of the *Boat*, while the BM3D filter preserves these structures. This phenomenon can in part be explained by the high values of α-quantile chosen to get a (qualitatively satisfying) low variance in homogeneous regions. It could also be reduced by considering smaller search windows, at the cost of larger remaining noise variance.

In the case of gamma noise, the images denoised by the WIN-SAR and the MAP-UWD-S filters are less smoothed than the images obtained by our filter. Moreover, the WIN-SAR filter



(a) Noisy image (b) K-SVD (c) BM3D (d) NL means (e) Our filter $\times 25$

Figure 5.14: (a) From top to bottom, corrupted images of *Barbara*, *Boat* and *House* by Gaussian noise with standard deviation $\sigma = 40$. Denoised images using (b) the K-SVD filter, (c) the BM3D filter, (d) the NL means filter and (e) 25 iterations of our filter.

blurs the edges and the MAP-UWD-S filter introduces some artifacts in the neighborhood of the edges. The gain of using pre-estimate, i.e., a joint similarity criteria, is similar in the gamma case as in the Gaussian case. In the case of Poisson noise, P-BM3D provides the best visual result. Our approach better reduces the residual noise than HaarTI and SAFIR but the result is maybe too smooth.

To quantify the denoising qualities, Table 5.1 and 5.2 presents numerical results for images corrupted by Gaussian noise with standard deviations $\sigma = 10, 20, 40$ and 60, gamma noise with equivalent number of looks $L = 1, 2, 4$ and 16 and Poisson noise with a maximum peak of $p = 5, 10, 20$ and 150 photons. The performance criterion used is the signal to noise ratio (SNR) defined in Sec. 2.1.3. We observe that iterative filters improve non-iterative filters for low SNR images. High SNR images (standard deviation $\sigma \leq 30$ or equivalent number of looks $L \geq 4$) do not require iterative refinement of the weights. In the case of Gaussian noise, our approach is better than the K-SVD filter for low SNR images, but is out-performed by the BM3D filter at all SNR values. In the case of gamma noise, our filter out-performs all the *state-of-the-art* filters considered for low SNR images. Nevertheless, the PPB filter provides comparative results to the MAP-UWD-S filter for high SNR images (i.e., with a large equivalent number of looks L). The iterative filter is then more relevant since images damaged by gamma noise are generally provided for a low equivalent number of looks such as $L = 1, 2, 3$ or 4 (e.g. radar or sonar images). In the case of Poisson noise, our approach is out-performed by P-BM3D for all values of maximum peak, but improves on P-HaarTI and P-SAFIR.

Finally, note that our purpose is not to provide the best denoising algorithm, but to find a

(a) Noisy image (b) WinSAR (c) MAP-UWD-S (d) Our filter ×1 (e) Our filter ×25

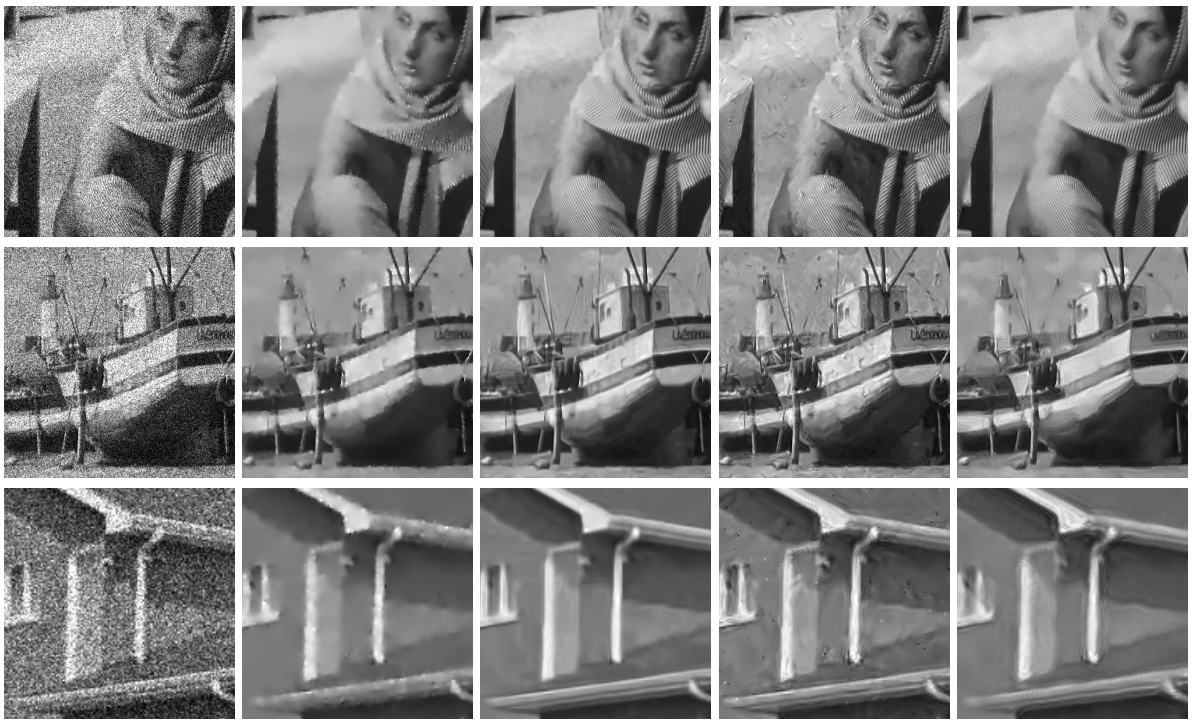
Figure 5.15: (a) From top to bottom, corrupted images of *Barbara*, *Boat* and *House* by gamma noise with equivalent number of looks $L = 3$. Denoised images using (b) the WIN-SAR filter, (c) the MAP-UWD-S filter, (d) our non-iterative filter and (e) 25 iterations of our filter.

rigorous methodology to adapt non-local filtering to an uncorrelated noise described by a given pdf. Our filter seems to be working equally well for Gaussian, gamma and Poisson noise when the noise levels are similar, whereas, most of the other filters are specially designed to cope with a fixed noise model. Thus, our filter seems to be an efficient extension of the NL means filter to take into account different noise degradation models.

5.6.2 Results on real data

This section presents different results obtained on real data damaged by gamma or Poisson noise. We use the same *state-of-the-art* filters as the ones used above and our filter. In all experiments, the algorithms are executed with the same parameters described in the previous section.

Figure 5.17 presents two single-look SAR acquisitions identified as *Bayard* and *Cheminot* from Saint-Pol-sur-Mer (France), sensed in 1996 by RAMSES of ONERA, and one single-look SAR acquisition identified as *Toulouse* of CNES in Toulouse (France) sensed also by RAMSES and provided by CNES. All these images are assumed to follow the gamma noise model. The obtained denoised images for the different real SAR images and the different denoising filters are given. The results obtained with our iterative filter seem to be well smoothed with a better edge and shape preservation than other filters. The *speckle* effect is strongly reduced and the spatial resolution seems to be well preserved: buildings, sidewalks, streets and fields are well restored. Moreover, the bright scatterers (numerous in urban areas) are well restored. Unfortunately, our filter seems to attenuate thin and dark structures existing in the SAR image, such as the thin streets in *Cheminot* and *Toulouse*.



(a) Noisy image (b) P-HaarTI (c) P-BM3D (d) P-SAFIR (e) Our filter ×2

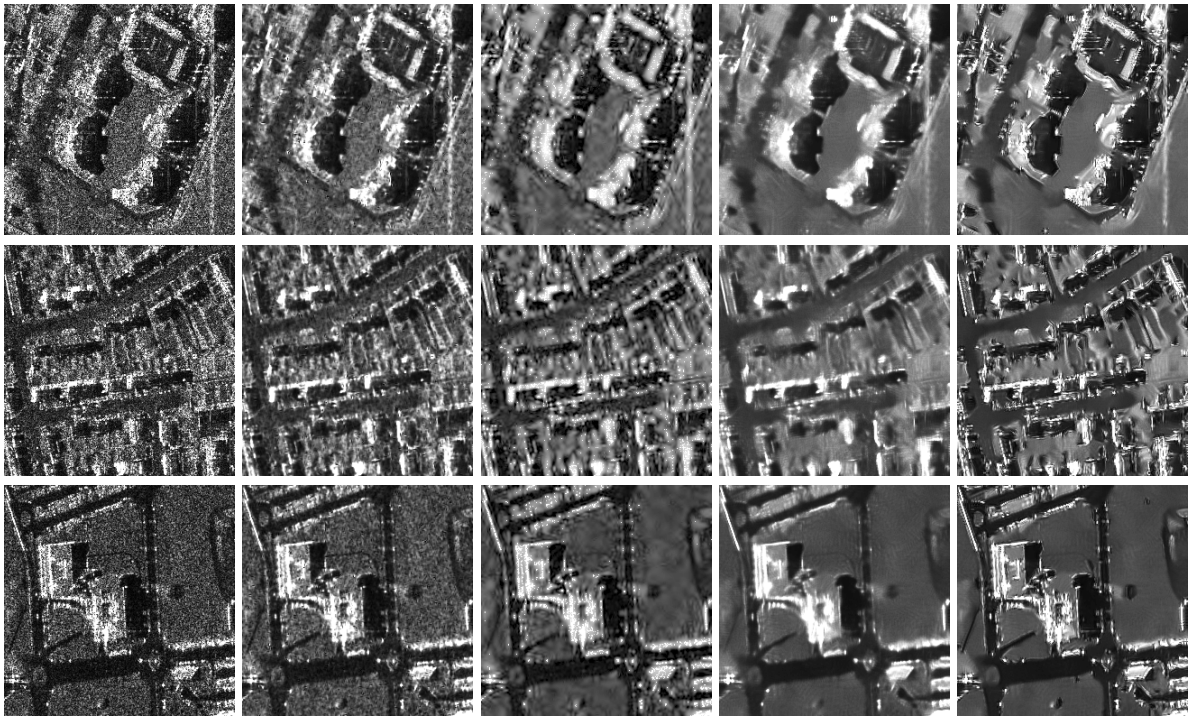
Figure 5.16: (a) From top to bottom, corrupted images of *Barbara*, *Boat* and *House* by Poisson noise with a maximum peak of 20 photons. Denoised images using (b) the P-HaarTI filter, (c) the P-BM3D filter, (d) the P-SAFIR filter and (e) our two step filter.

Figure 5.18 presents an image[2] of a mitochondrion sensed in low-light conditions by confocal fluorescence microscopy (Pelloux et al., 2006), and an X-ray image of a supernova explosion in the Milky Way of the supernova remnant G1.9+0.3. Both images are assumed to follow the Poisson noise model. The obtained denoised images for these different images and the different denoising filters are also given. The visual results of our filter challenge the *state-of-the-art* results.

5.7 Conclusion

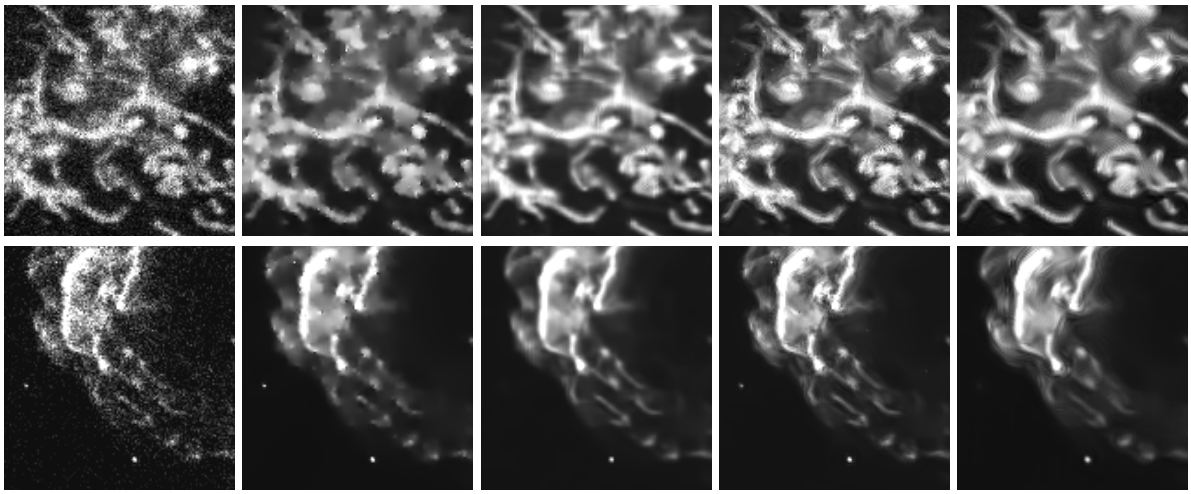
An extension of the NL means was proposed for image denoising when noise is non-Gaussian and the noise distributions is known. Our extension is a weighted maximum likelihood estimator where weights are determined by a joint similarity criterion. It is based on a statistical test used to compare noisy patches and patches of a pre-estimated image. The use of a pre-estimated image to refine weights is proposed to enhance the denoising quality in the case of low signal to noise ratio images and its efficiency has been shown on Gaussian noise, gamma noise and Poisson noise. Thanks to oracle studies, we have shown the relevance of the proposed joint similarity criterion. It provides a new framework for image denoising when the uncorrelated noise model is known. A risk estimator for the NL means has been derived for Poisson noise. This risk estimator is used in an optimization method to automatically select the filtering parameters in few iterations. Numerical results as well as visual results support the efficiency of this extended method.

[2] image courtesy of Y. Tourneur



(a) Noisy image (b) WinSAR (c) MAP-UWD-S (d) Our filter ×1 (e) Our filter ×25

Figure 5.17: (a) From top to bottom, SAR images of *Bayard* (France) ©DGA ©ONERA, *Cheminot* (France) ©DGA ©ONERA and *Toulouse* (France) ©DGA ©ONERA. Denoised images using (b) the WIN-SAR filter, (c) the MAP-UWD-S filter, (d) our non-iterative filter and (e) 25 iterations of our filter.



(a) Noisy image (b) P-HaarTI (c) P-BM3D (d) P-SAFIR (e) Our filter ×2

Figure 5.18: (a) From top to bottom, an image of a mitochondrion sensed in low-light conditions by confocal fluorescence microscopy (Pelloux et al., 2006), an X-ray image of a supernova explosion in the Milky Way of the supernova remnant G1.9+0.3 (image courtesy to Chandra X-ray Observatory – data identifier: `ADS/ Sa.CXO#Contrib/ ChandraDeepField`). Denoised images using (b) the P-HaarTI filter, (c) the P-BM3D filter, (d) the P-SAFIR filter and (e) our two step filter.

Chapter 6

Applications of non-local approaches to SAR images

Synthetic aperture radar (SAR) images are increasingly used in remote sensing, for a broad variety of applications ranging from crisis management to biomass study. Several new high resolution airborne and spaceborne sensors with full polarimetric and/or interferometric capabilities are now operating (F-SAR, TerraSAR-X, ...).

Prior to their analysis, SAR images generally undergo processing steps that degrade their resolution. Due to strong *speckle* in SAR images, local smoothing is performed to mitigate the fluctuations in homogeneous regions. Furthermore, the computation of the interferometric and polarimetric signature of a radar scene requires estimating local covariance matrices. Though a *speckle* reduction step is unavoidable in many applications, special care must be taken to limit blurring of significant structures in SAR images.

The crudest approach to *speckle* reduction and covariance estimation, referred to as *multi-looking* in radar community, computes a simple moving average with a (typically rectangular) window (see Sec. 2.2.1). A satisfying smoothing of homogeneous regions comes at the cost of a resolution loss.

Several improvements to multi-looking have been proposed in the radar literature. The common underlying idea is to adapt the selection of pixels used in each covariance matrix estimation in order to prevent mixing pixels belonging to different structures (e.g., blurring edges and strong scatterers by averaging them with their surrounding background). Several approaches for adaptive selection have been considered:

- Lee et al. (1999, 2003) suggested locally selecting the best window among a few pre-defined windows (a rectangular window and 8 edge-aligned oriented windows, see Sec. 2.3.2 and Fig. 2.5). Window selection is based on the gradient of the intensity image. This leads to good preservation of straight edges. However, abrupt change in the decision (from one window to another) at neighboring pixels creates artifacts. The limited number of pre-defined windows considered restricts the adaptation to complex structures or textures, and thus the ability to correctly restore them.

- Vasile et al. (2006) proposed to build locally by region-growing an adaptive neighborhood restricted to similar pixels. Adjacent pixels are aggregated incrementally based on their intensity (hence the name "IDAN": intensity-driven adaptive-neighborhood). This approach is therefore more flexible than the use of pre-defined windows, and leads to better resolution preservation, for a given amount of smoothing, than the previous methods. By construction, adaptive neighborhoods are necessarily a set of connected pixels, and all values are given the same weight in the estimation. The method is known to suffer from a selection

bias. We show in Sec. 3.8.2 that due to *speckle*, intensities follow a heavy-tailed distribution; the selection of pixels with similar intensity however discards large values which biases the subsequent maximum likelihood estimation (see Sec. 3.8).

- The approach for pixel selection can be further generalized by considering extended non-local (i.e., non connected) neighborhoods, and by weighting the relative importance of pixels. This is the idea considered in previous chapter. Independently, Lee et al. (2009) designed an extension of its (non-local) sigma filter (see Sec. 3.8.2) while we proposed a first adaptation of the NL means in (Deledalle et al., 2009b), both of these methods aim at reducing *speckle* in amplitude SAR images. We then extended our methodology to interferometric SAR (Deledalle et al., 2011a) and polarimetric SAR (Deledalle et al., 2010d). At the same time, a similar approach was independently described in (Chen et al., 2011) for polarimetric SAR images. Our proposed methodology inspired Parrilli et al. (2010) to adapt the BM3D filter of Dabov et al. (2007) to amplitude denoising and Teuber and Lang (2011) to derive extensions of our work. Recently, Zhong et al. (2011) mixed the ideas of Lee et al. (2009) and Kervrann et al. (2007) and showed the efficiency of their methods to reduce *speckle* in intensity images.

Our contributions— In this chapter, we describe a unified framework, NL-SAR, for non-local denoising of amplitude (SAR), interferometric (InSAR), polarimetric (PolSAR) or polarimetric and interferometric (PolInSAR) radar images. We address several issues that are crucial in practice for resolution-preserving denoising:

1. Polarimetric or interferometric images are either processed after multi-looking (Chen et al., 2011), or with filters driven by the (span) intensity only (Vasile et al., 2006; Lee et al., 2003). Following the general methodology of Chap. 5, we describe a non-local filter for SAR data (NL-SAR) driven by similarities jointly estimated on all channels and preserving resolution by processing directly single-look images.

2. We discuss the benefits of *iterative* denoising with refined similarities.

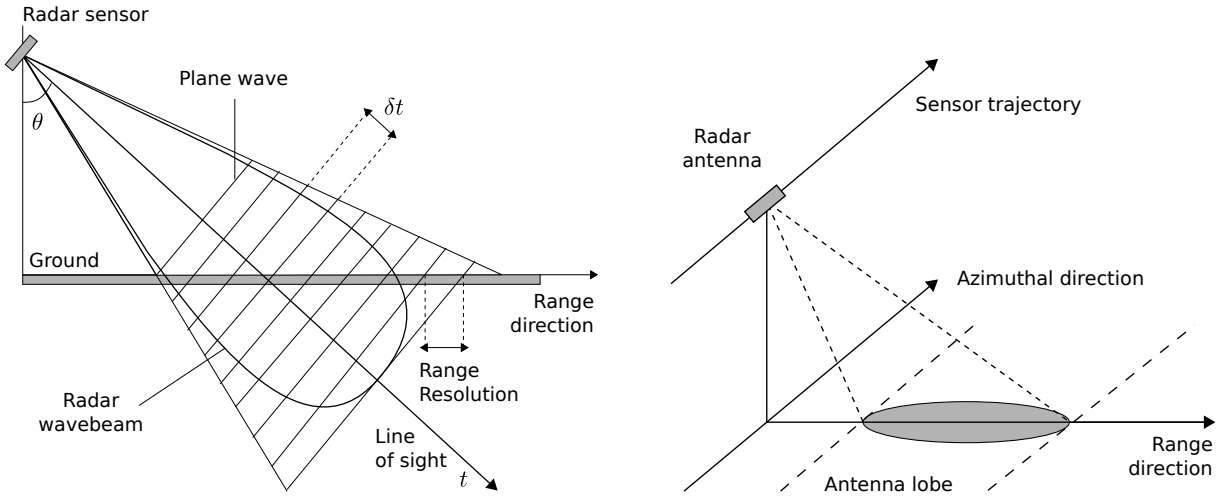
3. The issue of correlated data is addressed by adapting the kernel used to define the weights.

4. A semi-supervised approach for setting the filtering parameter is described.

5. The method is validated on RAMSES (ONERA), AIRSAR (NASA/JPL-Caltech) and recent aerial (F-SAR) data from the German Aerospace Center (DLR).

Together with this method, we release under public license the source code of NL-SAR. The technical documentation of NL-SAR is given in Appendix B.

Organization of the chapter— Section 6.1 gives an overview of SAR technologies from the physical aspects to the mathematical representations of SAR amplitude, InSAR, PolSAR and PolInSAR images. The statistics of SAR images are detailed in Sec. 6.2 under the fully developed *speckle* assumption. Based on these distributions, Sec. 6.3 reviews the basic estimators proposed to retrieve the underlying SAR parameters. In Sec. 6.4, the methodology proposed in Chap. 5 is used to derive a non-local filter (NL-SAR) relevant for *speckle* described by a complex Wishart distribution. We see that the application of this methodology requires special attention to deal with the multi-channel complex SAR data especially when the equivalent number of looks is low. Experiments and results in Sec. 6.5 show the efficiency of the proposed approach in SAR interferometry and SAR polarimetry.



(a) Radar imagery in *range*

(b) Radar imagery in *azimuth*

Figure 6.1: (a) Principle of line construction in a radar image. Thanks to the use of a non-null incidence angle θ, objects located at different positions on the ground in the *range direction* will be projected at different positions on the line of sight. (b) Principle of column construction in a radar image. The radar sensor acquires the different lines by moving in the *azimuthal* direction.

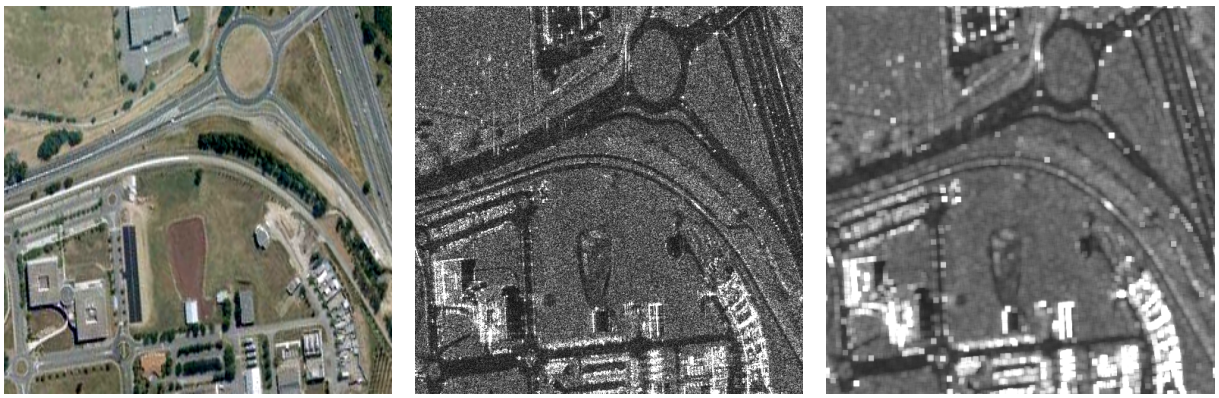
6.1 The different modalities of synthetic aperture radar images

6.1.1 Synthetic aperture radar (SAR) imagery

Radar (radio detection and ranging) technology consists of emitting an electromagnetic wave characterized by its wavelength and of measuring the echo of the backscattered wave. When the wave meets the ground the wave is diffused in multiple directions. The amplitude and the phase of the electromagnetic wave are then affected according to the nature of the ground (up to attenuations). The main factors are the roughness, the soil moisture, the ground permittivity and the local slope. The backscattered wave, i.e., the part of the wave remitted in the direction of the emission, is then acquired by the sensor.

The time delay between the emission and the reception is linked to the distance between the object and the sensor. The wave is emitted with a non-null incidence angle in a direction called the *range* direction (see Fig. 6.1). The echos of two objects located at different positions in the *range* direction will be received at different times. Based on this principle, a radar imaging system sends an impulsional wave considered as plane with respect to the observed scene. It collects a set of measurements separated by a time step δt and then located at different positions in the *range* direction. The relation between time and distance allows us to build one line of the image. Having moved forward in the orthogonal direction, called the *azimuthal* direction, the sensor acquires a new line. The set of lines forms the radar image. Figure 6.1 illustrates the formation of a radar image in *range* and *azimuth*.

According to Fig. 6.1, the resolution in *range* is linked to the time step δt: in order to distinguish two objects, we should be able to dissociate them in the temporal signal. In *azimuth*, the resolution is linked to the wavelength, the distance from the sensor to the object and the length of the antenna. To reach a good trade-off in terms of ground permittivity and surface roughness, the wavelengths are usually chosen in bands L, C or X, i.e., on the order of centimeters. The distance R can range from 2 km for airborne sensors to 800 km for spaceborne sensors. To reach a resolution below the order of meters, the length of the antenna should be on the order of several kilometers which is impossible in practice. Synthetic aperture radar (SAR)



(a) Optical image (b) Single-look amplitude image (c) 7×7-look amplitude image

Figure 6.2: (a) An optical image, (b) the single-look amplitude and (c) the multi-look amplitude of a SAR image of the same urban area of Toulouse (France). The optical image has been sensed by Quickbird (©DigitalGlobe) and the SAR image by RAMSES (©ONERA). The single-look amplitude image presents high fluctuations while the multi-look amplitude image presents smaller fluctuations but with blurry features. In average, the amplitude is low in smooth areas such as streets, it is higher in rough areas such as grass or vegetation and is extremely high around man-made structures, vehicles or tree trunks.

imagery is a technique based on the coherent emission along track to synthesize a large virtual antenna while in practice using an antenna on the order of few meters.

A SAR imaging system provides images such that each pixel corresponds to an area on the ground, called "resolution cell", whose dimensions depend on the resolution in *range* and *azimuth*. A pixel contains the amplitude and the phase of the backscattered wave observed for the corresponding resolution cell. For this reason, a SAR image is a complex image whose pixel values can be decomposed as $z = Ae^{j\varphi}$ where j is the imaginary number such that $j^2 = -1$. We denote such an image as a single-look complex (SLC) image. The modulus $A = |z|$ is the amplitude and the quantity $I = A^2$ is the intensity. The argument $\varphi = \arg z$ is the phase and, without extra information, it is meaningless since its values are uniformly distributed in the range $]-\pi, \pi]$ and uncorrelated.

For a given resolution cell, the observed backscattered echo results from the interferences between many elementary scatterers. Due to these interferences, SLC SAR images suffer from high fluctuations inherent to the random organization of the elementary scatterers inside the resolution cell (see Sec. 6.2.1). These fluctuations are referred to as *speckle*. Due to the high variability caused by *speckle*, SAR images have long been spatially averaged at the price of a loss of resolution. Under the ergodicity assumption, such techniques rely on the fact that each sample in the local neighborhood can be interpreted as another look, i.e., a realization of the pixel value of interest. By averaging all these samples, the noise can be reduced significantly (see also Section 2.2.1). Multi-look images result from the computation of the mean intensity of L scattering complex values z_i over a sliding window:

$$I = \frac{1}{L} \sum_{i=1}^{L} |z_i|^2, \tag{6.1}$$

where L is referred to as the equivalent number of looks. The multi-look amplitude image is obtained by taking the square root of the multi-look intensity image: $A = \sqrt{I}$.

Figure 6.2 gives an example of a SAR image compared to an optical image. The average amplitude is low in smooth areas such as streets, higher in rough areas such as grass or vegetation and extremely high around man-made structures, vehicles or tree trunks. The high fluctuations due to *speckle* strongly corrupt the signal of interest. The multi-look amplitude image presents smaller fluctuations but with a loss of resolution: the edges of objects appear blurry.

6.1.2 Interferometric SAR (InSAR) imagery

The phase φ can be decomposed in two terms as follows:

$$\varphi = \varphi_{proper} + \varphi_{topographic} \,. \tag{6.2}$$

The topographic phase $\varphi_{topographic}$ is linked to the propagation time of the wave from the target to the sensor and is equal to:

$$\varphi_{topographic} = \frac{4\pi}{\lambda} R \tag{6.3}$$

where λ is the wavelength and R the distance between the target and the sensor. The proper phase φ_{proper} depends on the coherent summation of complex signals backscattered by many elementary scatterers located in the same resolution cell. The proper phase φ_{proper} is then ascribed to *speckle* which is stable for different acquisitions under the assumption that the nature of the surface has not changed. The phase is then deterministic but the configuration of the resolution cell is unpredictable. Hence, the phase appears as random with a uniform distribution in $]-\pi, \pi]$.

Because of the random behavior of the *proper* phase, the observed phase φ seems to be non-exploitable. In a stereo-vision framework, Graham (1974) suggests using two SAR images of the same area sensed with quasi-identical conditions and a slightly different incidence angle. As a consequence, the term φ_{proper} is unchanged and only $\varphi_{topographic}$ differs between the two acquisitions. The phase difference – the interferometric phase — between two images z_1 and z_2 acquired in close locations gives the path delay between the two waves and is then independent on the *proper* phase:

$$\phi = \varphi_1 - \varphi_2 = \varphi_{topographic,1} - \varphi_{topographic,2} \,. \tag{6.4}$$

Under reasonable geometrical assumptions, Massonnet and Rabaute (1993) show that the height z can be linked to the interferometric phase ϕ as follows:

$$\mathrm{z} = \left(\frac{\sqrt{X^2 + H^2}}{b_x} \tan\theta \right) (R_1 - R_2) = \left(\frac{\lambda}{4\pi} \frac{\sqrt{X^2 + H^2}}{b_x} \tan\theta \right) \phi \tag{6.5}$$

where the different involved quantities are defined in Fig. 6.3. There is proportionality between the height z and the interferometric phase ϕ. Note that under other geometrical assumptions, other relations can be used to link the height z of the scene with the interferometric phase ϕ (Prati and Rocca, 1990; Lin et al., 1992). Depending on the system, the phase can be measured with a precision from 2 to 5% of the wavelength. Since the wavelengths are chosen on the order of centimeters, interferometry leads to a precision on the order of meters.

A first problem in InSAR is that the interferometric phase ϕ is known modulo 2π. This results in the presence of fringes in the interferometric phase image. In order to use Eq. (6.5), it is then necessary to unwrap the phase to retrieve the underlying height. Phase unwrapping is a difficult task addressed by several authors, see for instance (Goldstein et al., 1988; Ghiglia

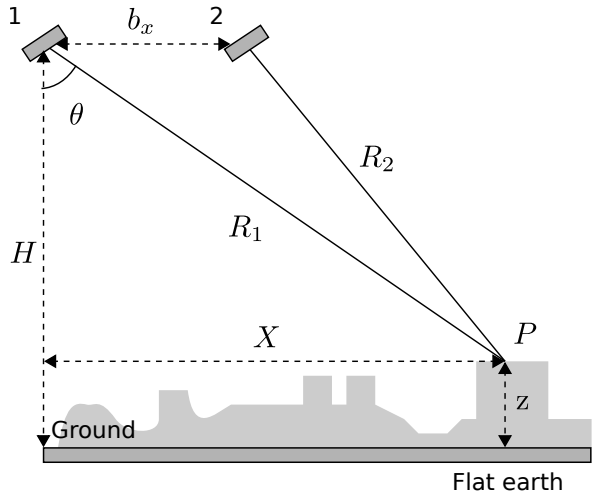


Figure 6.3: Principle of interferometric SAR imagery. Using two acquisitions separated by a baseline b_x. The height z of a point P can be retrieved from geometrical considerations where R_1 and R_2 to are the distance of P to the two respective sensors, X is its location on the ground and H the altitude of the sensor.

and Romero, 1989; Pascazio and Schirinzi, 2001; Bioucas-Dias and Valadão, 2007; Bioucas-Dias et al., 2008). The phase unwrapping problem is out of the scope of this thesis.

A second problem in InSAR is that the interferometric phase contains an orbital component due to the flat earth assumption (see Fig. 6.3). The orbital phase produces a linear phase ramp resulting in orbital fringes. In some cases, when the sensor parameters are well-chosen and with reasonable assumptions, pre-processing methods can efficiently suppress these orbital fringes (Rosen et al., 2000). In the following, this component is considered to be removed by a pre-processing step, so that an horizontal area always appears with a constant interferometric phase, up to fluctuations due to noise.

Due to temporal and spatial variations, the *speckle* components between the two acquisitions (i.e., the underlying scattering processes) can present a decorrelation which affects the interferometric phase (Hanssen, 2001). The interferometric phase ϕ and the empirical coherence d are, respectively, the phase and the magnitude of the normalized complex hermitian product between the two acquisitions z_1 and z_2. To reduce errors, interferograms are also commonly built by averaging L samples over a sliding window:

$$de^{j\phi} = \frac{\sum_{i=1}^{L} z_{1,i} z_{2,i}^*}{\sqrt{\sum_{i=1}^{L} |z_{1,i}|^2 \sum_{i=1}^{L} |z_{2,i}|^2}} \,. \tag{6.6}$$

The empirical coherence d appears as a crucial indicator of the reliability of the observed interferometric phase ϕ. The empirical coherence is comprised between 0 and 1. When it is close to 1, the two acquisitions are coherent and the phase difference ϕ is directly related to the path delay (modulo an uncertainty due to the phase wrapping). When d is close to 0, the two acquisitions are decorrelated and the phase ϕ does not contain any information. The level of noise in the interferometric phase ϕ increases when the empirical coherence d decreases.

Fig. 6.4 gives an example of an interferogram built on an urban area acquired with two simultaneous acquisitions (no temporal decorrelation). In this configuration (very small baseline), the phase is inside one fringe, then no unwrapping step is necessary. The ground has a uniform height while buildings, vegetation and hills appear at higher levels. Phase errors can be observed in the single-look interferogram. They are all the more important when the coherence is low. The

(a) Single-look interferogram (b) 7×7-look interferogram (c) 7×7-look emp. coherence

Figure 6.4: (a) Single-look interferogram, (b) multi-look interferogram and multi-look empirical coherence obtained from a pair of co-registered SAR images. The SAR images have been sensed by RAMSES (©ONERA). The phase is linked to the height of the scene up to errors arising in region with low coherence. The multi-look interferogram present smaller errors at the expense of a resolution loss.

multi-look interferogram presents smaller errors at the expense of a resolution loss. Note that without multi-looking, the empirical coherence would always be maximum: $d = 1$ when using $L = 1$ in Eq. (6.6). In order to properly estimate the level of coherence between two acquisitions, one needs to proceed to multi-looking. The level of coherence is then always measured with a loss of resolution.

6.1.3 Polarimetric SAR (PolSAR) imagery

Polarimetric SAR images are obtained by sensing the horizontal and vertical polarization components of the back-scattered wave, when a wave with vertical or horizontal polarization is emitted. A common representation, called the lexicographic representation, is to build a scattering vector as $\boldsymbol{k} = (z_{hh}, z_{vv}, \sqrt{2}z_{hv})^t$ where z_{hh}, z_{vv} and z_{hv} are the backscattered returns from respectively horizontal emission and horizontal reception, vertical emission and vertical reception, and horizontal emission and vertical reception. Note that the element z_{vh} is omitted since, using mono-static polarimetry, it can be assumed that $z_{vh} = z_{vh}$ (cf. the target reciprocity assumption). The factor $\sqrt{2}$ is then introduced to preserve the norm of the vector. Polarimetric SAR images inform us on both the intensities of the different polarizations and the complex hermitian product between each pair of different polarizations. These quantities are respectively given by the diagonal elements and the off-diagonal elements of the empirical complex covariance matrix $\boldsymbol{C}$ of the scattering vector $\boldsymbol{k}$. Multi-looking can also be performed to decrease *speckle* such that:

$$\boldsymbol{C} = \frac{1}{L} \sum_{i=1}^{L} \boldsymbol{k}_i \boldsymbol{k}_i^{\dagger} \tag{6.7}$$

where † indicates the hermitian transpose. The complex cross correlations between the polarimetric channels (i.e., the phase differences and the empirical coherences) depend on the polarimetric nature of the scene, e.g., the kind of bounce, the heterogeneity of the back scatterers or the wave incidence angle. Several works aim at extracting the semantics of these parameters to provide a physical description of the scattering process (Huynen, 1970; Krogager, 1990; Cameron and Leung, 1990; Freeman and Durden, 1998; Touzi, 2004).

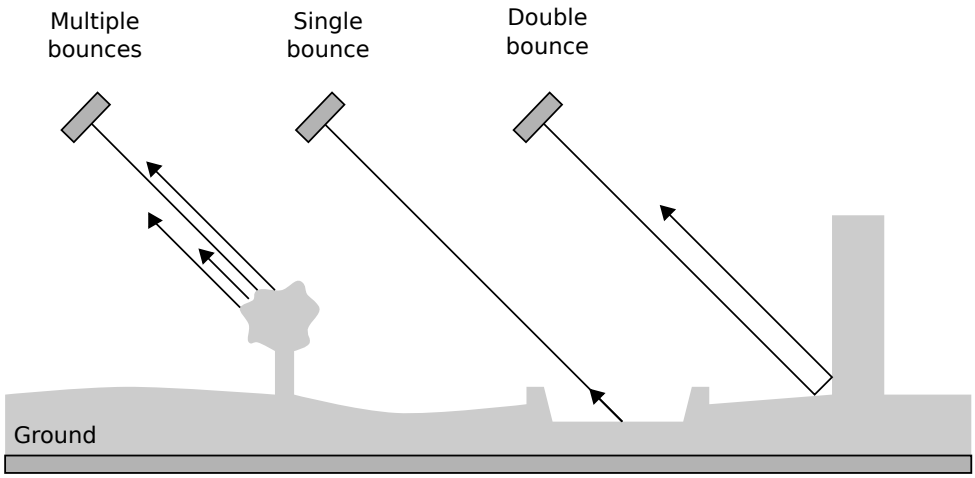


Figure 6.5: Principle of polarimetry SAR imagery. The information brought by the use of different polarimetric waves informs us on the kinds of bounces. The echos measured on a smooth surface, a man-made structure or the forest canopy results respectively from a single bounce, double bounce and multiples bounces.

A common way to emphasize the physical phenomena of wave scattering is to consider the Pauli basis to represent the scattering vectors (Cloude and Pottier, 1995). The scattering vector becomes:

$$\boldsymbol{k}^{(Pauli)} = \frac{1}{\sqrt{2}} \begin{pmatrix} z_{hh} + z_{vv} \\ z_{hh} - z_{vv} \\ 2z_{hv} \end{pmatrix} = \frac{z_{hh}}{\sqrt{2}} \underbrace{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}}_{\overrightarrow{\boldsymbol{v}}_1} + \frac{z_{vv}}{\sqrt{2}} \underbrace{\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}}_{\overrightarrow{\boldsymbol{v}}_2} + \frac{2z_{hv}}{\sqrt{2}} \underbrace{\begin{pmatrix} 0 \\ 0 \\ \sqrt{2} \end{pmatrix}}_{\overrightarrow{\boldsymbol{v}}_3}, \tag{6.8}$$

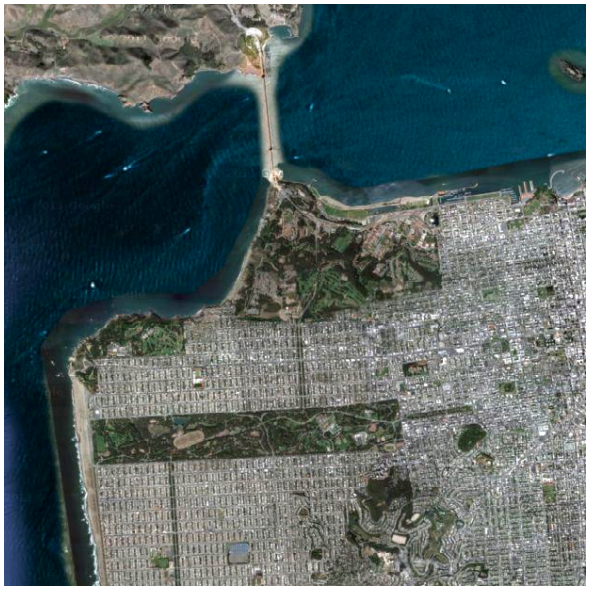
and the empirical covariance matrix $\boldsymbol{C}^{(Pauli)}$ is defined by substituting $\boldsymbol{k}$ by $\boldsymbol{k}^{(Pauli)}$ in Eq. (6.7). The relations between the lexical and the Pauli representation are given by:

$$\boldsymbol{k}^{(Pauli)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix} \boldsymbol{k}\,, \tag{6.9}$$

$$\boldsymbol{C}^{(Pauli)} = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix} \boldsymbol{C} \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix}. \tag{6.10}$$

This representation relies on the fact that we have an interpretation of the scattering mechanisms described by $\overrightarrow{\boldsymbol{v}}_1$, $\overrightarrow{\boldsymbol{v}}_2$ and $\overrightarrow{\boldsymbol{v}}_3$. The vector $\overrightarrow{\boldsymbol{v}}_1$ describes single or odd bounces (e.g., the ocean surface). The vector $\overrightarrow{\boldsymbol{v}}_2$ describes the double or even bounces (e.g., urban areas and man-made structures). The vector $\overrightarrow{\boldsymbol{v}}_3$ describes scatters returning the orthogonal polarization (e.g., the forest canopy). Figure 6.5 illustrates these different types of bounces. Such a representation is used to visualize PolSAR images with an RGB colorization built such that the red is linked to $\overrightarrow{\boldsymbol{v}}_1$, the green to $\overrightarrow{\boldsymbol{v}}_3$ and the blue to $\overrightarrow{\boldsymbol{v}}_2$. In Fig. 6.6, the ocean appears in blue, forest in green and urban areas in red.

Another solution to ease the interpretation of the scattering phenomena is to consider a decomposition of $\boldsymbol{C}$ invariant of the chosen basis. Inspired by the work of Huynen (1970), Cloude and Pottier (1996) suggest diagonalizing the covariance matrix in terms of its eigenvectors and

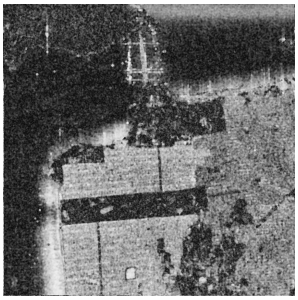


(a) Optical image

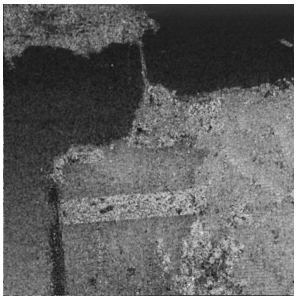
(b) Polarimetric SAR image

(c) Multi-look PolSAR

(d) Entropy



(e) Anisotropy



(f) Alpha angle

Figure 6.6: (a) An optical image and (b) the single-look, (c) multi-look, (d) entropy, (e) anisotropy and (f) alpha angle of a polarimetric SAR image of San Francisco (California, USA). The optical image has been sensed by Quickbird (©DigitalGlobe) and the SAR image by AirSAR (©NASA/JPL). The polarimetric SAR image is displayed using an RGB representation based on the Pauli basis. The red color is representative of urban areas, the green of vegetation areas and the blue of the ocean surface.

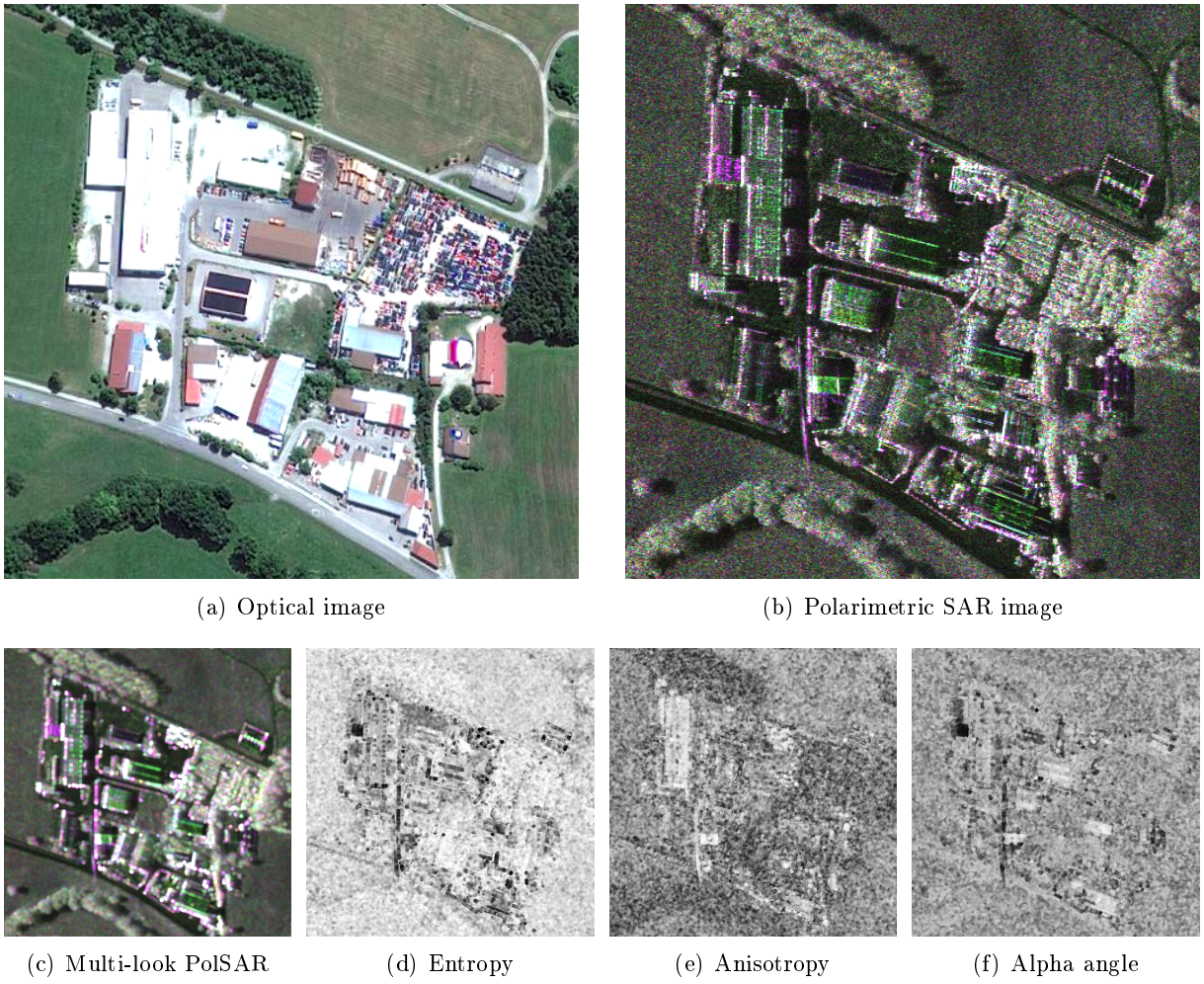
eigenvalues such that:

$$\boldsymbol{C} = \boldsymbol{U} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \boldsymbol{U}^{-1} \qquad \text{where} \qquad \boldsymbol{U} = (\vec{\boldsymbol{e}}_1 \ \vec{\boldsymbol{e}}_2 \ \vec{\boldsymbol{e}}_3) \tag{6.11}$$

and $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ are the eigenvalues and $\vec{\boldsymbol{e}}_1$, $\vec{\boldsymbol{e}}_2$ and $\vec{\boldsymbol{e}}_3$ the respective eigenvectors. Each eigenvector encodes a scattering mechanism and the associated eigenvalue its contribution. Since the eigenvectors are unitary vectors, they can be decomposed as follows:

$$\vec{\boldsymbol{e}}_i = \begin{pmatrix} \cos\alpha_i \ e^{j\zeta_i} \\ \sin\alpha_i \cos\beta_i \ e^{j\zeta_i} \\ \sin\alpha_i \sin\beta_i \ e^{j\gamma_i} \end{pmatrix} . \tag{6.12}$$

Based on these decompositions, Cloude and Pottier (1997); Cloude et al. (2001); Hajnsek et al. (2003) extract three important physical features describing the underlying physical phenomena:



(a) Optical image (b) Polarimetric SAR image

(c) Multi-look PolSAR (d) Entropy (e) Anisotropy (f) Alpha angle

Figure 6.7: (a) An optical image, (b) the single-look, (c) multi-look, (d) entropy, (e) anisotropy and (f) alpha angle of a polarimetric SAR image of an urban area in Kaufbeuren (Germany). The optical image has been sensed by Quickbird (©DigitalGlobe) and the SAR image by F-SAR (©DLR). The polarimetric SAR image is displayed using an RGB representation based on the Pauli basis.

- The entropy $H \in [0,1]$ (defined in the Von Neumann sense):

$$H = -\sum_{i=1}^{3} p_i \log_3 p_i \qquad \text{where} \qquad p_i = \frac{\lambda_i}{\sum_{i=1}^{3} \lambda_i} \tag{6.13}$$

and p_i are referred to as the scattering probabilities. The entropy measures the randomness of the scattering process (the statistical disorder of the scene). When $H = 0$, there is only one single mechanism involved since $\lambda_2 = \lambda_3 = 0$. It corresponds to a pure target, for instance, a smooth surface such as a sea ice surface or an ocean surface. When $H = 1$, three pure random mechanisms are involved since $\lambda_1 = \lambda_2 = \lambda_3$. We say that we have a distributed target, depolarized for which no information can be extracted. For other values of H, we have a combination of three random mechanisms whose contributions are given by the eigenvalues.

- The anisotropy $A \in [0, 1]$:

$$A = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3} \tag{6.14}$$

  provides complementary information to the entropy. When the entropy is non-zero, typically $H > 0.7$, the anisotropy indicates if either two pure or several mechanisms are involved. When $A = 1$, there are two pure targets since $\lambda_3 = 0$. When $A = 0$, there are three pure targets since $\lambda_2 = \lambda_3$.

- The mean scattering angle $\alpha \in]-\pi, \pi]$:

$$\alpha = p_1\alpha_1 + p_2\alpha_2 + p_3\alpha_3 \tag{6.15}$$

  is a direct measure of the underlying scattering process. When $\alpha = 0$, it corresponds to a single bounce scattering produced by a rough surface. When $\alpha = \pi/4$, it corresponds to volume scattering. When $\alpha = \pi/2$, it corresponds to double bounce scattering. This parameter is also influenced by the moisture of the surface.

Figure 6.6 and 6.7 give two examples of polarimetric images. The first one is an image of San Francisco in 1988 sensed by L-band AIRSAR (NASA/JPL-Caltech) with a low resolution of about 10 meters/pixel in range and azimuth. The second one is an image of Kaufbeuren (Germany) in June 2011 sensed by S-band F-SAR (DLR) with a high resolution of 0.5 meter/pixel in azimuth and 0.64 meter/pixel in range. For each image, we display their entropy, anisotropy and mean scattering angle. Note that without multi-looking, these quantities cannot be estimated since they require that the empirical covariance matrix $\boldsymbol{C}$ is non-singular. A minimum of 3 equivalent number of looks is then required.

6.1.4 Polarimetric Interferometric SAR (PolInSAR) imagery

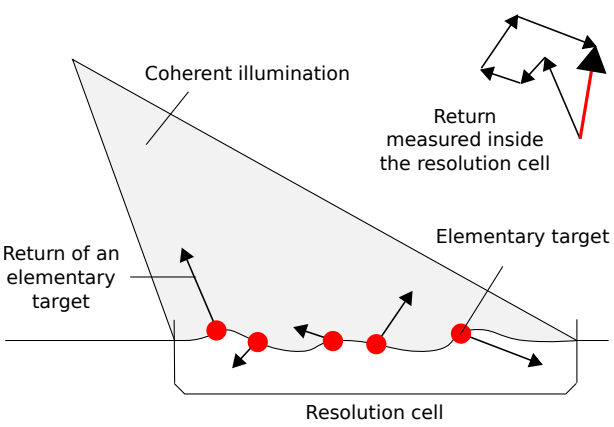
When two polarimetric images are acquired in an interferometric configuration, the resulting 6-dimensional scattering vector $\boldsymbol{k}$ is referred to as the polarimetric interferometric SAR (PolInSAR) vector (two acquisitions of three different polarizations). The empirical covariance matrix informs us on both path delays and polarimetric characteristics. PolInSAR is getting much attention, for two related reasons: the increasing availability of PolInSAR data and the appealing richness of information it captures in particular for biomass applications.

In all these different modalities SAR images suffer from a strong *speckle* effect and *decorrelation* errors. The use of multi-looking allows us to decrease the noise level at the expense of a resolution loss. In the next section, we will see that, under Goodman's assumptions, all these images can be described by a common model that can be considered to perform noise reduction with resolution preservation.

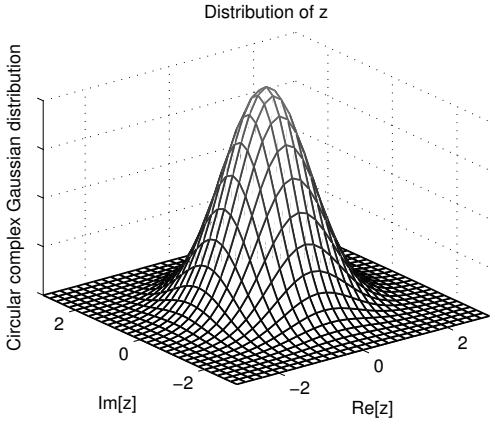
6.2 Statistics of SAR images

6.2.1 Goodman's model of SLC images

Due to interferences between the elementary scatterers inside the same resolution cell, SLC SAR images suffer from fluctuations inherent to the complex geometrical organization of the elementary scatterers. As illustrated in Fig. 6.8, the observed complex value results from the coherent summation of these elementary scatterers. Goodman (1963) assumes that each punctual target can be modeled as a complex value such that:



(a) Coherent summation inside a resolution cell



(b) Circular complex Gaussian distribution

Figure 6.8: (a) An illustration of the coherent summation of the several elementary scatterers located in the same resolution cell. (b) According to Goodman's model, the distribution of the observed complex echo follows a circular complex Gaussian distribution.

1. its real and its imaginary parts are independent and identically distributed random variables,

2. it is independent on other punctual targets,

3. its phase is uniformly distributed and independent of the amplitude.

SAR images having a low resolution compared to the scale of the punctual targets, the number of elementary scatterers inside each resolution cell can be assumed to be large. The application of the central limit theorem under Goodman's assumptions leads to model the resulting complex value z as following a zero-mean complex circular Gaussian distribution defined as:

$$p(z|\sigma^2) \triangleq p(\mathsf{Re}[z], \mathsf{Im}[z]|\sigma^2) = \frac{1}{\pi\sigma^2} \exp\left(-\frac{|z|^2}{\sigma^2}\right) \tag{6.16}$$

where $\sigma^2 = \mathbb{E}[|z|^2]$ is a quantity linked to the backscattering coefficient and thus to the radar cross section per unit volume (Bamler and Hartl, 1998). It characterizes the surface roughness, the surface moisture and the scattering material perceived at location x (Sprague, 1972; McKinney et al., 2000). In the following, we call the unknown parameter of interest σ^2 the surface reflectivity. Note that Eq. (6.16) is deduced from the multi-variate Gaussian distribution presented in Eq. (3.3) by using $\mathsf{Re}[z]$ and $\mathsf{Im}[z]$ as independent random variables with a zero-mean Gaussian distribution of variance $\sigma^2/2$. This model is referred to as the fully developed *speckle* model which has been shown to be valid for different modalities of SAR images (Sarabandi, 1992).

Note that Eq. (6.16) is independent on the phase φ. The phase is thus as predicted uniformly distributed, and only the modulus $|z|$, the amplitude, is informative. By the change of variable $(\mathsf{Re}[z], \mathsf{Im}[z]) \mapsto (I = |z|^2, \varphi = \arg z)$ in Eq. (6.16) followed by integration over φ, it results in the intensity $I = |z|^2$ being distributed according to an exponential law of parameter σ^2, and hence, its amplitude $A = |z|$ follows a Rayleigh distribution of parameter σ (see Sec. 3.1.5). These distributions describe the multiplicative fluctuations of *speckle* observed in SAR images.

6.2.2 Multi-variate model of a collection of SLC SAR images

When K co-registered SAR images are available (e.g., $K = 2$ in InSAR, $K = 3$ in PolSAR or $K = 6$ in PolInSAR imagery), we can form a K-dimensional scattering vector $\mathbf{k} = (z_1, \ldots, z_K)$ at each pixel x, with entries corresponding to the complex values of the different acquisitions at the same location x. Under Goodman's model, the scattering vector $\mathbf{k}$ follows a K-dimensional circular complex Gaussian distribution:

$$p(\boldsymbol{k}|\boldsymbol{\Sigma}) = \frac{1}{\pi^K |\boldsymbol{\Sigma}|} \exp\left(-\boldsymbol{k}^\dagger \boldsymbol{\Sigma}^{-1} \boldsymbol{k}\right) \tag{6.17}$$

where $\boldsymbol{\Sigma} = \mathbb{E}\{\boldsymbol{k}\boldsymbol{k}^\dagger\}$ is a $K \times K$ complex covariance matrix and † indicates the hermitian transpose. The matrix $\boldsymbol{\Sigma}$ can be decomposed as follows:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_1\sigma_k\rho_{1,k} & \cdots & \sigma_1\sigma_K\rho_{1,K} \\ \vdots & \ddots & \vdots & & \vdots \\ \sigma_k\sigma_1\rho_{1,k}^* & \cdots & \sigma_k^2 & \cdots & \sigma_k\sigma_K\rho_{k,K} \\ \vdots & & \vdots & \ddots & \vdots \\ \sigma_K\sigma_1\rho_{1,K}^* & \cdots & \sigma_K\sigma_k\rho_{k,K}^* & \cdots & \sigma_K^2 \end{pmatrix} \tag{6.18}$$

$$\text{where} \qquad \sigma_k^2 = \mathbb{E}[|z_k|^2] \qquad \text{and} \qquad \rho_{k,l} = \frac{\mathbb{E}[z_k z_l^*]}{\sqrt{\mathbb{E}[|z_k|^2]\mathbb{E}[|z_l|^2]}}\,. \tag{6.19}$$

The diagonal elements σ_k^2 are the surface reflectivities. The off-diagonal elements $\rho_{k,l} = D_{k,l}e^{j\beta_{k,l}}$ define the complex correlation between each pair of channels. The quantity $D_{k,l} < 1$ quantifies the degree of *coherence* between the different acquisitions/channels and is an indicator of the reliability of how the noisy phase $\phi_{k,l} = \arg z_k z_l^*$ is close to the true phase $\beta_{k,l} = \arg \rho_{k,l}$. In general, decorrelations occur from temporal variations (the scene changed between the different acquisitions), the use of different incidence angles (baseline variation), the use of different polarizations and also atmospherical perturbations, registration errors or imperfections in the sensor trajectories.

When $K = 1$, Eq. (6.17) boils down to Eq. (6.16). When $K = 2$, the distribution of $\boldsymbol{k}$ depends only on the amplitudes A_1, A_2 and the phase difference $\phi = \varphi_1 - \varphi_2$. With a suite of changes of variables and integrations, Goodman (1984, 2006) derives from Eq. (6.17) the joint distribution of the triplet (A_1, A_2, ϕ) given by:

$$p(A_1, A_2, \phi|\sigma_1^2, \sigma_2^2, \beta, D) = \frac{2A_1A_2}{\pi\sigma_1\sigma_2(1-D^2)} \exp\left(-\frac{A_1^2}{\sigma_1^2} + \frac{A_2^2}{\sigma_2^2} - \frac{2DA_1A_2\cos(\phi-\beta)}{\sigma_1\sigma_2(1-D^2)}\right)\,. \tag{6.20}$$

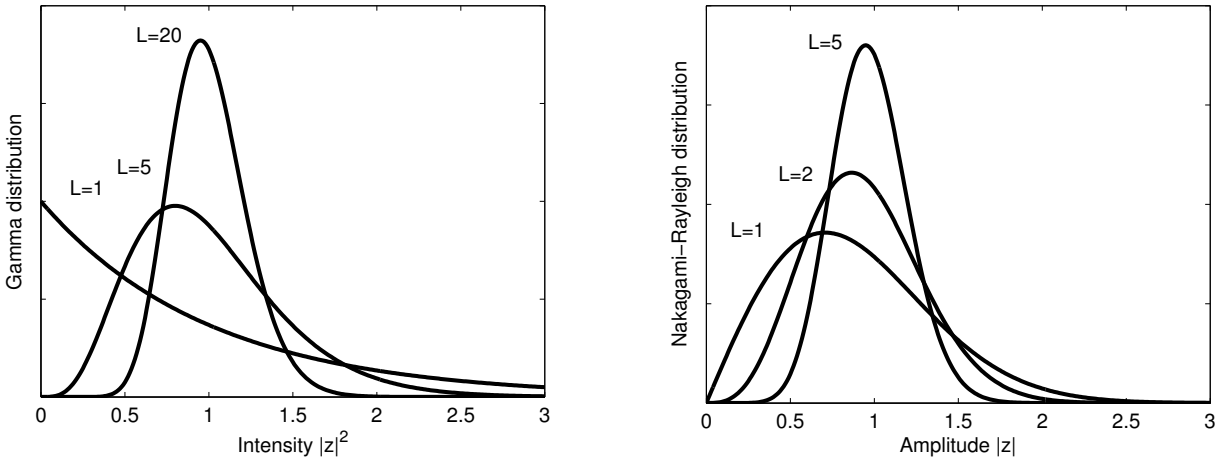
6.2.3 Multi-look SAR images

Multi-look SAR images result from the computation of the $K \times K$ sample covariance matrix of L scattering vectors $\boldsymbol{k}_i$ extracted from a spatial neighborhood:

$$\boldsymbol{C} = \frac{1}{L}\sum_{i=1}^{L} \boldsymbol{k}_i \boldsymbol{k}_i^\dagger, \tag{6.21}$$

where L is the equivalent number of looks. The matrix $\boldsymbol{C}$ provides the multi-look intensity image, the multi-look phase image and the empirical coherence image given by:

$$I_k = A_k^2 = \frac{1}{L}\sum_{i=1}^{L} |z_{i,k}|^2 \qquad \text{and} \qquad d_{k,l}e^{j\phi_{k,l}} = \frac{\sum_{i=1}^{L} z_{i,k}z_{i,l}^*}{\sqrt{\sum_{i=1}^{L}|z_{i,k}|^2 \sum_{i=1}^{L}|z_{i,l}|^2}}\,. \tag{6.22}$$



(a) (b)

Figure 6.9: Two distributions modeling *speckle*. (a) The gamma distribution describes the *speckle* observed in intensity images. (b) The Nakagami-Rayleigh distribution describes the *speckle* observed in amplitude images. Both distributions have a heavy right-tail explaining the highly bright pixels that are present in such images. Their shape becomes Gaussian when the equivalent number of looks L increases.

Note that under the ergodicity assumption (see Sec. 2.2.1), the empirical matrix $\boldsymbol{C}$, the intensity I_k, the phase difference $\phi_{k,l}$ and the empirical coherences $d_{k,l}$ are the respective sample estimates of the covariance matrix $\boldsymbol{\Sigma}$, the reflectivity σ_k^2, the true phase difference β_k and the coherence $D_{k,l}$. We will show in Sec. 6.3 that these sample estimates correspond to the maximum likelihood estimates.

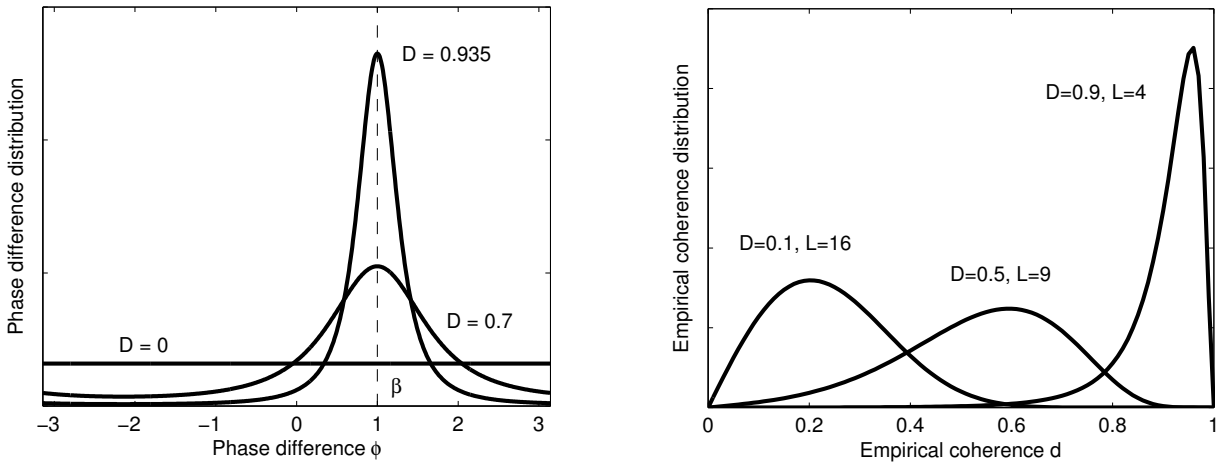
When $L \geq K$, the distribution of the multi-look data is described by a complex Wishart distribution given by:

$$p(\boldsymbol{C}|\boldsymbol{\Sigma}, L) = \frac{L^{LK}|\boldsymbol{C}|^{L-K}}{\Gamma_K(L)|\boldsymbol{\Sigma}|^L} \exp\left(-L \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{C})\right) \tag{6.23}$$

where $\operatorname{tr}(\cdot)$ is the matrix trace. The equivalent number of looks L acts as the shape parameter of the Wishart distribution.

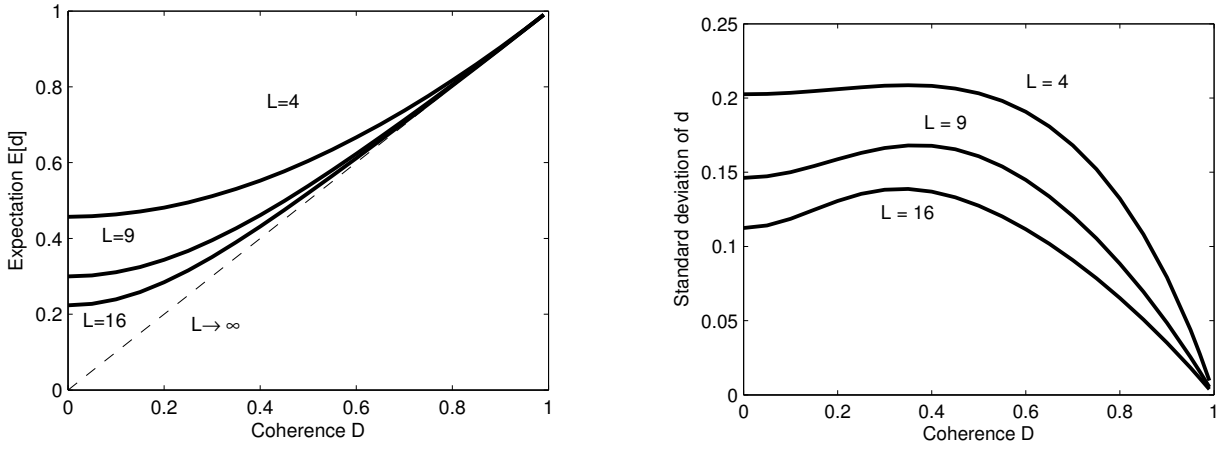
When $K = 1$, the matrix $\boldsymbol{C}$ simplifies to the multi-look intensity $I = A^2$ and $\boldsymbol{\Sigma}$ simplifies to the reflectivity σ^2. As a consequence, Eq. (6.23) simplifies to the gamma distribution with a scale parameter σ^2 and a shape parameter L (see Sec. 3.1.5). The square root of the intensity, i.e., the amplitude, follows a Nakagami-Rayleigh distribution (see Sec. 3.1.5). Figure 6.9 illustrates these distributions according to the equivalent number of looks L.

When $L < K$, the complex covariance matrix $\boldsymbol{C}$ is singular and Eq. (6.23) is no longer defined. Such a matrix cannot be modeled by a pdf defined on the cone of semi-positive hermitian matrices. The matrix $\boldsymbol{C}$ is said to have a degenerate distribution since we cannot have a complete description of the joint statistics of its elements. Its elements can only be described term by term by a pdf, referred to as an incomplete distribution of $\boldsymbol{C}$. A common situation is the single-look empirical covariance matrix defined by $\boldsymbol{C} = \boldsymbol{k}\boldsymbol{k}^\dagger$ (Eq. (6.21) with $L = 1$). It provides a matrix representation of the SLC data without loss of resolution and without loss of meaningful information. This matrix is singular and always provides a maximal empirical coherence $d = 1$.



(a) (b)

Figure 6.10: Two distributions modeling the errors in measuring phase differences. (a) The phase has a symmetric distribution defined on $]-\pi,\pi]$. It is uniform when the coherence D is zero and becomes narrow around the true phase β when the coherence D or the equivalent number of looks L increases. (b) The empirical coherence d has a distribution defined on $[0,1]$. Its shape is all the more sharper when the equivalent number of looks L is large. Its location depends on the true coherence D but its mean and mode are neither one nor the other centered on D.



(a) (b)

Figure 6.11: Two first-order statistics of the empirical coherence. (a) The expectation $\mathbb{E}[d]$ of the empirical coherence. The coherence d overestimates the true coherence D especially for small values of D. This bias decreases when the equivalent number of looks L increases. (b) The variance of the empirical coherence d. The variance non-linearly evolves with the true coherence D and starts to decrease from a given value. When the coherence $D = 1$, the empirical coherence is also equal to $d = 1$ and its variance is null.

K	L	Modality	Model
1	$\{1\}$	Single-look SAR	Exponential distribution
	$[2,\infty[$	Multi-look SAR	Gamma distribution
2	$\{1\}$	Single-look InSAR	Circular complex Gaussian
2	$[2,\infty[$	Multi-look InSAR	Complex Wishart
3	$\{1\}$	Single-look PolSAR	Circular complex Gaussian
	$\{2\}$	Two-look PolSAR	Degenerate distribution
	$[3,\infty[$	Multi-look Full PolSAR	Complex Wishart
6	$\{1\}$	Single-look PolInSAR	Circular complex Gaussian
	$[2,5]$	Multi-look PolInSAR	Degenerate distribution
	$[6,\infty[$	Multi-look PolInSAR	Complex Wishart

Table 6.1: Statistical models of different modalities of SAR images

6.2.4 Incomplete statistics of SAR images

The phase difference ϕ follows a distribution defined on $]-\pi,\pi]$ given by (Just and Bamler, 1994; Lee et al., 1994):

$$p(\phi|\beta,D,L) = \frac{\Gamma(L+1/2)(1-D^2)^L\Delta}{2\sqrt{\pi}\Gamma(L)(1-\Delta^2)^{L+1/2}} + \frac{(1-D^2)^L}{2\pi}{}_2F_1(L,1;1/2;\Delta^2)$$
$$\text{where} \qquad \Delta = D\cos(\phi-\beta) \tag{6.24}$$

and ${}_2F_1$ is the hypergeometric function. When $L=1$, Eq. (6.24) boils down to the following expression:

$$p(\phi|\beta,D) = \frac{1-D^2}{2\pi}\frac{1}{1-\Delta^2}\left(1+\frac{D\arccos(-\Delta)}{\sqrt{1-\Delta^2}}\right) . \tag{6.25}$$

As illustrated in Fig. 6.10, this distribution is symmetric, 2π periodic and centered on β (when $D>0$). Its scale depends only on the coherence D: the distribution is uniform when $D=0$ and tends towards a Dirac centered on β when $D\to 1$. The fluctuations are additive modulo 2π, i.e., the noisy phase difference ϕ can be decomposed as $\phi=\beta+\varepsilon\ [\mathrm{mod}\ 2\pi]$ where ε is a random variable centered on 0 and independent on β. Bamler and Hartl (1998) show that when $L=1$:

$$\mathbb{E}[\phi] = \beta\ , \tag{6.26}$$
$$\mathrm{Var}[\phi] = \frac{\pi^2}{3} - \pi\arcsin(D) + \arcsin^2(D) + \frac{\mathrm{Li}_2(D)}{2} \tag{6.27}$$

where Li_2 is Euler's dilogarithm. The variance of the phase difference depends on the coherence D: the noise is then *heteroscedastic*. In particular, when $D=0.935$, the standard deviation is close to 0.5 and when $D=0.7$, the standard deviation is close to 1. The SAR images are often composed of several regions of constant coherence D. The interferometric phase appears then as *homoscedastic* inside each of these regions.

The distribution of the empirical coherence $d\in[0,1]$ depends only on the true coherence D and the number of looks L and is given by (Touzi and Lopes, 1996):

$$p(d|D,L) = 2(L-1)(1-D^2)^L d(1-d^2)^{L-2}{}_2F_1(L,L;1;D^2d^2)\ . \tag{6.28}$$

As mentioned before, single-look complex images always provide a maximal empirical coherence $d=1$. Equation (6.28) is then defined as soon as $L>1$. This distribution is given in Fig. 6.10 with three different sets of parameters. This distribution is non-symmetric, its shape varies with D and, unlike previous distributions, its mode is not reached for $d=D$. Touzi and Lopes (1996); Touzi et al. (1999) deduced from Eq. (6.28) the moments of order k of the empirical coherence d:

$$m_k = (1-D^2)^L\frac{\Gamma(L)\Gamma(1+k/2)}{\Gamma(L+k/2)}{}_3F_2(1+k/2,L,L;L+k/2,1;D^2) \tag{6.29}$$

where ${}_3F_2$ is the generalized hypergeometric function. The expectation and the variance of d are then respectively given by:

$$\mathbb{E}[d] = m_1 \tag{6.30}$$

$$\mathrm{Var}[d] = m_2 - m_1^2 \,. \tag{6.31}$$

Figure 6.11 illustrates how the expectation and variance of the empirical coherence d vary as a function of the true coherence D and the equivalent number of looks L. It appears clearly that the empirical coherence d overestimates the coherence for low values of D and tends to the true coherence D when the number of looks L becomes large. The variance of d is also highly dependent on D. For low values of D and L, the empirical coherence d suffers from both bias and *heteroscedastic* noise.

Table 6.1 gives a non-exhaustive overview of the statistical models used in SAR imagery.

6.3 Estimation of the SAR parameters

6.3.1 Maximum likelihood estimation from i.i.d. samples

Assume that we have M independent and identically distributed (i.i.d.) K-dimensional scattering vectors $\boldsymbol{k}_i, i = 1, \ldots, M$, the maximum likelihood estimate (MLE) is derived from Eq. (6.17) as:

$$\hat{\boldsymbol{\Sigma}} = \arg\max_{\boldsymbol{\Sigma}} \sum_{i=1}^{M} \log p(\boldsymbol{k}_i|\boldsymbol{\Sigma}) = \frac{1}{M} \sum_{i=1}^{M} \boldsymbol{k}_i \boldsymbol{k}_i^\dagger \,. \tag{6.32}$$

By term identifications, it results that:

$$\hat{\sigma}_k^2 = \frac{1}{M} \sum_{i=1}^{M} |z_{k,i}|^2 \qquad \text{and} \qquad \hat{D}_{k,l} e^{j\hat{\beta}_{k,l}} \frac{\sum_{i=1}^{M} z_{k,i} z_{l,i}^*}{\sqrt{\sum_{i=1}^{M} |z_{k,i}|^2 \sum_{i=1}^{M} |z_{l,i}|^2}} \,. \tag{6.33}$$

The MLE boils down to the sample estimate of the complex covariance matrix. Note that multi-look images are then built from the application of the MLE in a local neighborhood with the underlying ergodicity assumption (following Sec. 3.3). The statistics of these estimates can be deduced from the ones of the multi-look SAR images.

Assume now that we have M i.i.d. observed empirical complex covariance matrices $\boldsymbol{C}_i$ built with the same equivalent number of looks L, the MLE is derived from Eq. (6.23) as:

$$\hat{\boldsymbol{\Sigma}} = \arg\max_{\boldsymbol{\Sigma}} \sum_{i=1}^{M} \log p(\boldsymbol{C}_i|\boldsymbol{\Sigma}) = \frac{1}{M} \sum_{i=1}^{M} \boldsymbol{C}_i \,. \tag{6.34}$$

As soon as $LM \geq K$, the estimated complex covariance matrix $\hat{\boldsymbol{\Sigma}}$ follows a Wishart distribution with a shape parameter LM.

The MLE is consistent and efficient meaning that it is asymptotically unbiased with respect to the number of samples M and its variance reaches the Cramer-Rao bound. For low values of M, the estimates of the reflectivities $\hat{\sigma}_k^2$ and the phase differences $\hat{\beta}_{k,l}$ are also unbiased. However, the estimated coherence $\hat{D}_{k,l}$ tends to be overestimated. Touzi et al. (1999) suggest inverting the functional link f between the biased estimate and the true coherence given by Eq. (6.29):

$$f(d, M) = (1 - D^2)^M \frac{\Gamma(M)\Gamma(1 + 1/2)}{\Gamma(M + 1/2)} {}_3F_2(1 + 1/2, M, M; M + 1/2, 1; D^2) \,. \tag{6.35}$$

The function f cannot be inverted in closed-form. In (Touzi et al., 1999), the authors numerically compute the values $f^{-1}(d; M)$ for several values of d and M.

6.3.2 The case of SAR interferometry

When an interferometric pair is considered, it seems reasonable to consider equal the (true) reflectivities of each pair of corresponding pixels, i.e., $\sigma_1^2 = \sigma_2^2 = \sigma^2$. The covariance matrix simplifies as follows:

$$\mathbf{\Sigma} = \sigma^2 \begin{pmatrix} 1 & De^{j\beta} \\ De^{-j\beta} & 1 \end{pmatrix} . \tag{6.36}$$

This hypothesis is naturally verified in regions with good coherence. By reducing the number of degrees of freedom (from 4 to 3 unknowns), the estimation variance is improved. Denoising techniques must trade-off variance reduction and resolution preservation. As the sample size is restricted by resolution preservation considerations, it is desirable to reduce the variance with such a hypothesis. Seymour and Cumming (1994) derive the MLE under this interferometric assumption. It leads to:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{M} |z_{1,i}|^2 + \sum_{i=1}^{M} |z_{2,i}|^2}{2M} \qquad \text{and} \qquad \hat{D}e^{j\hat{\beta}} = \frac{2\sum_{i=1}^{M} z_{1,i} z_{2,i}^*}{\sum_{i=1}^{M} |z_{1,i}|^2 + \sum_{i=1}^{M} |z_{2,i}|^2} . \tag{6.37}$$

Equation (6.37) defines the same estimator of the phase as the sample estimator. The coherence estimator differs in the denominator. In this case, $\hat{\mathbf{\Sigma}}$ no longer follows a Wishart distribution. When the assumption $\sigma_1^2 = \sigma_2^2$ holds, Seymour and Cumming (1994) show that their estimator is more efficient than the classical sample estimate. For instance, an estimate of the coherence can be obtained without averaging in a local neighborhood whereas the sample estimator requires averaging at least two pixels to define a coherence. The single-look empirical coherence depends only on the amplitudes and is defined by $\hat{D} = \frac{2|z_1||z_2|}{|z_1|^2+|z_2|^2}$. Note that this definition of the empirical coherence is like the generalized likelihood ratio between two observed intensity values (see Chap. 4). As well as the generalized likelihood ratio, the coherence can be interpreted in this case as a measure of the assumption that $\sigma_1^2 = \sigma_2^2$.

6.3.3 Estimation with mixed populations: the LLMMSE approach

In order to limit the resolution loss of multi-looking, Lee (1980) suggests taking into account that samples in a local neighborhood can arise from different populations. The idea, generalized in Lee et al. (1999) for polarimetric data, is to combine at each pixel position x the multi-look empirical covariance matrix $\boldsymbol{C}$ with the single-look empirical covariance matrix $\boldsymbol{k}\boldsymbol{k}^\dagger$:

$$\hat{\mathbf{\Sigma}} = (1-\alpha)\boldsymbol{C} + \alpha \boldsymbol{k}\boldsymbol{k}^\dagger \tag{6.38}$$

where α depends on the spatial configuration. Ideally, α should be null in homogeneous areas and equal to one in heterogeneous areas (e.g., around edges). The authors suggest finding the parameter α by minimizing locally the mean square error (MSE). This filter referred to as the local linear minimum mean square estimator (LLMMSE) is an aggregation-based filter as presented in Sec. 2.6.

To deal with multi-dimensional SAR data, the MSE is expressed on the mean reflectivity $\sigma^2 = \mathrm{tr}(\mathbf{\Sigma})$ also called the span of the matrix $\mathbf{\Sigma}$. The risk associated to the estimate $\hat{\mathbf{\Sigma}}$ is then given by:

$$R[\hat{\mathbf{\Sigma}}(x)] = \mathbb{E}\left[\left(\mathrm{tr}(\hat{\mathbf{\Sigma}}) - \mathrm{tr}(\mathbf{\Sigma})\right)^2\right] = \mathbb{E}\left[\left((1-\alpha(x))I_L + \alpha I - \sigma^2\right)^2\right] \tag{6.39}$$

where $I_L = \mathrm{tr}(\boldsymbol{C})$ is the multi-look span intensity, and $I = \mathrm{tr}(\boldsymbol{k}\boldsymbol{k}^\dagger)$ is the single-look span intensity. The minimization of Eq. (6.39) under reasonable assumptions leads to the following weight definition (Lee, 1980):

$$\alpha = \max\left(0, \frac{\mathrm{Var}[I_L] - I_L^2\eta^2}{\mathrm{Var}[I_L](1+\eta^2)}\right) \tag{6.40}$$

where $\mathrm{Var}[I_L]$ is the local estimation of the span variance calculated in the local neighborhood and η^2 is a fixed parameter linked to the variance of the normalized *speckle* in the span intensity, i.e., $\mathrm{Var}[I/\sigma^2]$. Under the ergodicity assumption, the estimated variance $\mathrm{Var}[I_L]$ should be approximately $\sigma^2\eta^2$ and I_L be σ^2. The resulting weight α should then be close to zero. When the samples in the local neighborhood arise from different populations, the estimated variance $\mathrm{Var}[I_L]$ should increase and the weight α tends to one. When the channels are independent, the normalized *speckle* has a variance $\eta^2 = 1/LK$. However, due to the correlation between channels and also the correlation of *speckle* in the local neighborhood the value η^2 should be reduced (see Lee et al., 1999).

To improve the estimation quality and prevent resolution loss, rather than considering a fixed sliding window, this approach can be used with a spatially-adaptive window as described in Sec. 2.3.2 and proposed in (Lee, 1981; Lee et al., 1999; Vasile et al., 2006). Otherwise one can select a neighborhood that prevent mixing different populations. Following this idea, we suggest in the next section using non-local neighborhoods.

6.4 Non-local estimation of the SAR parameters

6.4.1 Weighted maximum likelihood estimation

Rather than combining *neighboring* pixels (i.e., spatially close), we suggest using non-local neighborhoods to estimate properly the covariance matrix $\boldsymbol{\Sigma}$ and avoid mixing different populations. As we have suggested in Sec. 5.1, the extension of the non-local means of Buades et al. (2005) should rely on the weighted maximum likelihood estimation (WMLE). From Eq. (6.17), the WMLE can be derived as the weighted average of the empirical covariance matrices:

$$\hat{\boldsymbol{\Sigma}}(x) = \arg\max_{\boldsymbol{\Sigma}} \sum_{x'} w(x,x') \log p(\boldsymbol{C}(x')|\boldsymbol{\Sigma}) = \frac{\sum_{x'} w(x,x')\boldsymbol{C}(x')}{\sum_{x'} w(x,x')} \tag{6.41}$$

where x' is a pixel located in a search window around x and $w(x,x') > 0$ is the data-driven weight depending on the statistical similarity between pixels x and x'. If we further assume equal reflectivities (for instance in interferometric conditions), the WMLE would be given by substituting the sums by weighted sums in Eq. (6.37).

Eq. (6.41) provides unbiased estimates as soon as the weights $w(x,x')$ select enough candidate pixels x' i.i.d. with the pixel of interest x. As we have seen in Chap. 5, the determining of weights can be seen as a detection problem: two pixels can be combined if they are *in match*. Given the low signal-to-noise ratio of SAR images, this statistical test is performed by comparing together the two patches P_x and $P_{x'}$ centered respectively on x and x'. In this framework, the weight can be considered as the membership value that $\boldsymbol{C}(x')$ belongs to the set of random variables i.i.d. with $\boldsymbol{C}(x)$.

6.4.2 Weight derivation: multi-channel patch similarity

Patch similarity between empirical covariance matrices

Following the general methodology of Chap. 5, we define patch similarity as a measure of how likely the two patches could be considered as two noisy realizations of the same noiseless patch.

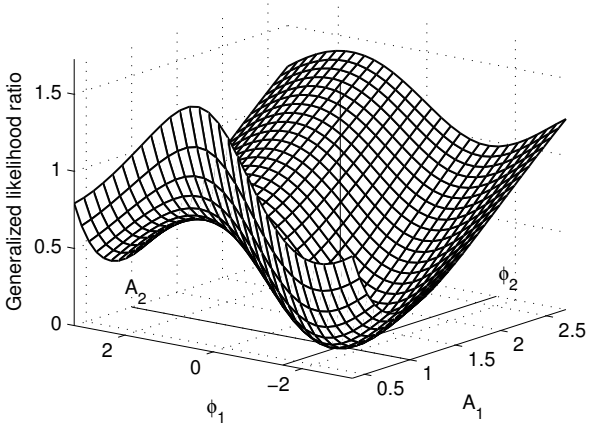


Figure 6.12: Negative logarithm of GLR between two matrices $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ differing only in the amplitude A_1 (resp. A_2) and the phase difference ϕ_1 (resp. ϕ_2) (see text). The curve shows the evolution of GLR with respect to A_1 and ϕ_1 for given values of A_2 and ϕ_2.

The evaluation of the similarity between two noisy covariance matrices $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ can then be rephrased as the following hypothesis test (i.e., a parameter test):

$$\mathscr{H}_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 \equiv \boldsymbol{\Sigma}_{12} \qquad \text{(null hypothesis)}, \tag{6.42}$$

$$\mathscr{H}_1 : \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2 \qquad \text{(alternative hypothesis)}. \tag{6.43}$$

According to the Neyman-Pearson theorem, the optimal criterion for the hypothesis test is given by the likelihood ratio (Kay, 1998):

$$\mathcal{L}(\boldsymbol{C}_1, \boldsymbol{C}_2) = \frac{p(\boldsymbol{C}_1, \boldsymbol{C}_2 | \boldsymbol{\Sigma}_{12}, \mathscr{H}_0)}{p(\boldsymbol{C}_1, \boldsymbol{C}_2 | \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \mathscr{H}_1)}. \tag{6.44}$$

The application of the likelihood ratio test requires the knowledge of the matrices $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ and $\boldsymbol{\Sigma}_{12}$ (the noise-free patches) which are, in practice, unavailable. The problem is thus a *composite hypothesis problem*. The generalized likelihood ratio (GLR) replaces these unknown matrices by their maximum likelihood estimates. Given the complex Wishart distribution of SAR images (see Eq. (6.23)), GLR is given by:

$$\mathcal{L}_G(\boldsymbol{C}_1, \boldsymbol{C}_2) = \frac{\sup_{\boldsymbol{\Sigma}_{12}} p(\boldsymbol{C}_1, \boldsymbol{C}_2 | \mathscr{H}_0, \boldsymbol{\Sigma}_{12})}{\sup_{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2} p(\boldsymbol{C}_1, \boldsymbol{C}_2 | \mathscr{H}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)} = \left(\frac{2^K \sqrt{|\boldsymbol{C}_1||\boldsymbol{C}_2|}}{|\boldsymbol{C}_1 + \boldsymbol{C}_2|} \right)^{2L}. \tag{6.45}$$

GLR provides a similarity measure between $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$. Note that this similarity measure has been proposed recently and independently in (Chen et al., 2011) for the comparison of polarimetric SAR data. Given the independence assumption (i.e., noise is considered uncorrelated), patch similarity can be computed pixelwise. Following the model in Sec. 5.4.1, the weights between the two patches P_x and $P_{x'}$ are then defined as:

$$w(x, x') = \varphi\left[-\log \mathcal{L}_G(\boldsymbol{C}(\mathrm{P}_x), \boldsymbol{C}(\mathrm{P}_{x'}) \right] \tag{6.46}$$

$$= \varphi\left[2L \sum_{\tau \in \mathrm{P}} \log \left(\frac{|\boldsymbol{C}(x+\tau) + \boldsymbol{C}(x'+\tau)|}{\sqrt{|\boldsymbol{C}(x+\tau)||\boldsymbol{C}(x'+\tau)|}} \right) - 2LK|P| \log 2 \right] \tag{6.47}$$

where φ is the kernel decay function as defined in Sec. 2.5.1. Note that when $K = 1$, the Wishart distribution boils down to the gamma distribution, and Eq. (6.46) corresponds to the

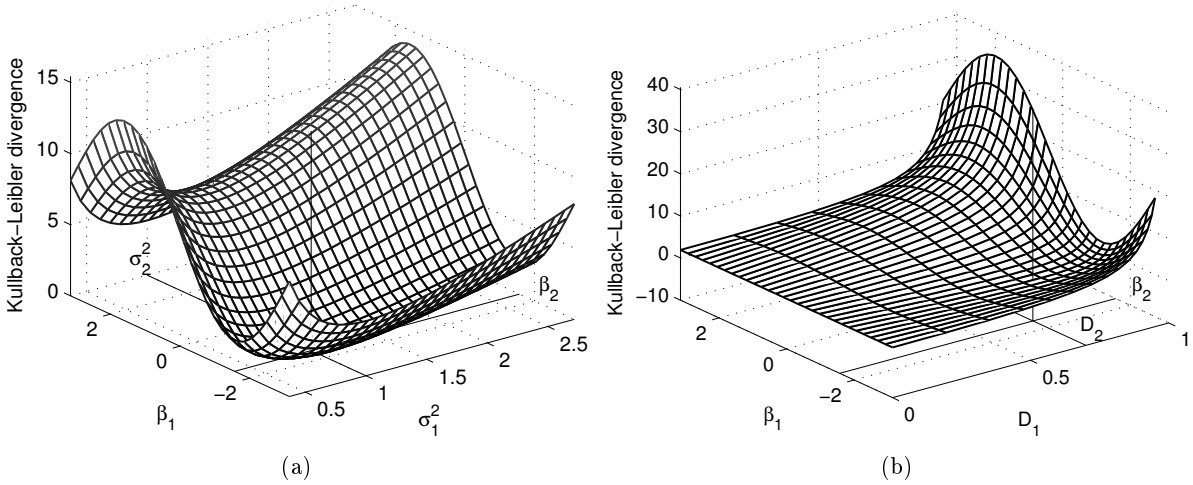


Figure 6.13: Symmetrical Kullback-Leibler divergence between $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ by varying (a) σ_1^2 and β_1 for the given values of σ_2^2, β_2 with $D_1 = D_2 = 0.7$, and (b) D_1 and β_1 for the given values of β_2, D_2 with $\sigma_1^2 = \sigma_2^2$.

expression given in Eq. (5.17) derived in the case of gamma noise.

Figure 6.12 represents the negative logarithm of GLR between two matrices $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$. The matrices $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are built from the same set of scattering vectors $\boldsymbol{k}$ except to modifications: the amplitude of the first channel is A_1 for $\boldsymbol{C}_1$ (resp. A_2 for $\boldsymbol{C}_2$), and the phase difference between the two first channels is ϕ_1 for $\boldsymbol{C}_1$ (resp. ϕ_2 for $\boldsymbol{C}_2$). The curve shows the evolution of GLR with respect to A_1 and ϕ_1 for given values of A_2 and ϕ_2. The criterion is minimum when observed data are identical: $A_1 = A_2$ and $\phi_1 = \phi_2$. Moreover, this criterion manages well with the phase wrapping without creating discontinuities when ϕ_1 jumps from $-\pi$ to π due to wrapping. For a given value of A_1, the criterion is minimum when ϕ_1 and ϕ_2 are in-phase, and maximum when they are out-of-phase.

Patch similarity between speckle-free covariance matrices

In an oracle setting, we have shown in Sec. 5.4.2 that weights can instead be defined by comparing the noise-free patches extracted from the unknown image $\mathbf{\Sigma}$. The symmetrical Kullback-Leibler divergence between patches extracted from $\mathbf{\Sigma}$ provides a statistical test for Eq. (6.42) based on the comparison of noise-free patches. It aims at selecting a maximum of almost i.i.d. samples in $\boldsymbol{C}$ in order to reach an optimal trade-off between bias and noise reduction. The weights in the WMLE are defined as:

$$\begin{aligned} w(x,x') &= \varphi\left[\mathcal{D}_{KL}(\mathbf{\Sigma}(\mathrm{P}_x), \mathbf{\Sigma}(\mathrm{P}_{x'}))\right] \\ &= \varphi\left[L \sum_{\tau \in \mathrm{P}} \left(\mathrm{tr}\left(\mathbf{\Sigma}^{-1}(x+\tau)\mathbf{\Sigma}(x'+\tau) + \mathbf{\Sigma}^{-1}(x'+\tau)\mathbf{\Sigma}(x+\tau)\right) - 2K\right)\right] . \end{aligned} \tag{6.48}$$

Note that when $K = 1$, the Wishart distribution boils down to the gamma distribution, and Eq. (6.48) corresponds to the expression given in Eq. (5.47) derived in the case of gamma noise.

Figure 6.13 represents the symmetrical Kullback-Leibler divergence between $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$. In 6.13(a), the variations are given with respect to the values of σ_1^2 and β_1, for given values of σ_2^2 and β_2 with $D_1 = D_2 = 0.7$. In 6.13(b), the variations are given with respect to the values of β_1

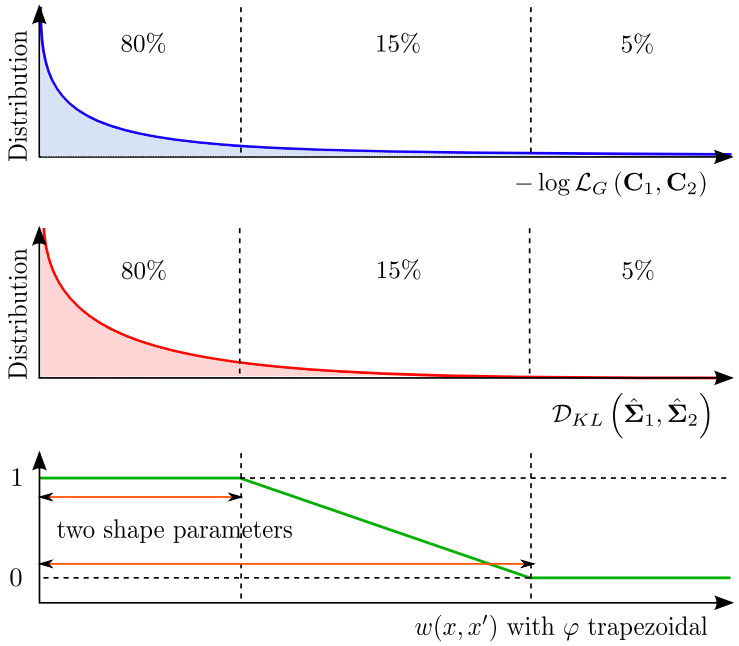


Figure 6.14: (top) Distribution of GLR when the two compared noisy patches share a common noise-free patch. (center) Distribution of the symmetrical Kullback-Leibler divergence when the two compared pre-filtered patches share a common noise-free patch. (bottom) According to this probability of false alarm, we can define the shape of a trapezoidal kernel function to control the amount of smoothing.

and D_1, for given values of β_2 and D_2 with $\sigma_1^2 = \sigma_2^2$. The criterion decreases when all parameters at pixel $_1$ get closer to the parameters at pixel $_2$ and becomes null when the parameters are equal. Moreover, this criterion manages well with the phase wrapping, without creating discontinuities when β_1 moves from $-\pi$ to π. For a given value of σ_1^2, the criterion is minimum when β_1 and β_2 are in-phase and maximum when they are out-of-phase. Note the satisfying behavior of the similarity criterion: the better the coherence (i.e., closer to 1) the larger the phase similarity, since phases are then more reliable.

Joint similarity driven by patches of empirical and speckle-free covariance matrices

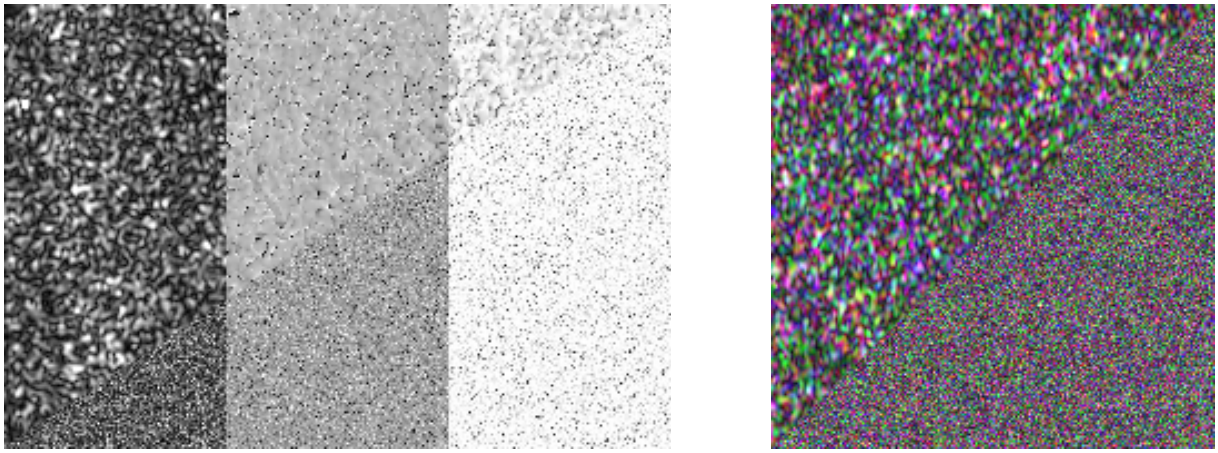
Because of the high level of noise in SAR images, weights provided by the similarity between noisy patches should be refined iteratively by using the similarity between previous estimated patches (see. Sec. 5.4.3). Theoretically, this refinement can be interpreted as the Bayesian test for Eq. (6.42) which decomposes using Bayes' rule as:

$$\frac{p(\mathscr{H}_0|\boldsymbol{C}_1, \boldsymbol{C}_2)}{p(\mathscr{H}_1|\boldsymbol{C}_1, \boldsymbol{C}_2)} = \underbrace{\frac{p(\boldsymbol{C}_1, \boldsymbol{C}_2|\mathscr{H}_0)}{p(\boldsymbol{C}_1, \boldsymbol{C}_2|\mathscr{H}_1)}}_{\text{Likelihood ratio}} \times \underbrace{\frac{p(\mathscr{H}_0)}{p(\mathscr{H}_1)}}_{\text{Extra information}} . \tag{6.49}$$

This refinement can be achieved by taking a convex combination of both criteria leading to the following definition:

$$w(x, x') = \varphi\left[(1-\lambda) f\left(-\log \mathcal{L}_G(\boldsymbol{C}(\mathrm{P}_x), \boldsymbol{C}(\mathrm{P}_{x'}))\right) + \lambda g\left(\mathcal{D}_{KL}(\hat{\mathbf{\Sigma}}(\mathrm{P}_x), \hat{\mathbf{\Sigma}}(\mathrm{P}_{x'}))\right)\right] \tag{6.50}$$

where the parameter $\lambda \in [0, 1]$ controls the confidence we have in the pre-estimate and f and g are two increasing affine transforms chosen such that both criteria answer with the same dynamic.



(a) (b)

Figure 6.15: (a) The amplitude, noisy phase and single-look empirical coherence in (top-left corner) an homogeneous area extracted from an interferometric pair of SAR images and (bottom-right corner) a simulated homogeneous area in interferometric conditions. (b) Pauli-based visualization of (top-left corner) an homogeneous area extracted from a polarimetric SAR image and (bottom-right corner) a simulated homogeneous area in polarimetric conditions. In both examples, true SAR data present a high spatially correlated *speckle* compared to the uncorrelated simulated ones.

The scale of the functions f and g should be chosen by controlling the desired amount of noise reduction. Assume the kernel function φ is a trapezoidal kernel defined as:

$$\varphi(t) = \begin{cases} 1 & \text{if } t \leq 1 \\ 2-t & \text{if } 1 < t \leq 2 \\ 0 & \text{otherwise} \end{cases} . \tag{6.51}$$

We want to control the level of noise reduction, i.e., the percentage of samples in an homogeneous area having a weight equal to 1, a non-null weight or a null weight. To reach good noise reduction in homogeneous areas and avoid blurring around edges, a reasonable choice illustrated in Fig. 6.14 is to select 80% of samples with a weight equal to one, 15% with a non-null weight and 5% with a null weight. To obtain such statistical selection and according to Eq. (6.51), the similarity criterion should have an answer t such that:

$$\mathbb{P}(t < 1 | \mathbf{\Sigma}_{12}, \mathscr{H}_0) = 0.80 \tag{6.52}$$

$$\mathbb{P}(t < 2 | \mathbf{\Sigma}_{12}, \mathscr{H}_0) = 0.95 \tag{6.53}$$

whatever the underlying unknown matrix $\mathbf{\Sigma}_{12}$.

The scale of the function f then relies on the probabilities of false-alarms associated to GLR, or equivalently to the distribution of GLR under the assumption $\mathcal{H}_0$. When $K = L = |P| = 1$, the distribution of the answer t of GLR under $\mathcal{H}_0$ is given by:

$$p(t|\mathscr{H}_0) = \frac{e^{t/2}}{\sqrt{e^t - 1}} \left[\left(\frac{e^{t/2} - \sqrt{e^t - 1}}{1 + (e^{t/2} - \sqrt{e^t - 1})^2} \right)^2 + \left(\frac{e^{t/2} + \sqrt{e^t - 1}}{1 + (e^{t/2} + \sqrt{e^t - 1})^2} \right)^2 \right] \tag{6.54}$$

which is, up to a change of variable, a Fisher distribution. Our experiments show that this distribution is well approximated by $\chi^2_1 \left(\frac{t}{2 - 2\log 2} \right)$ where χ^2_1 is the chi-square distribution with

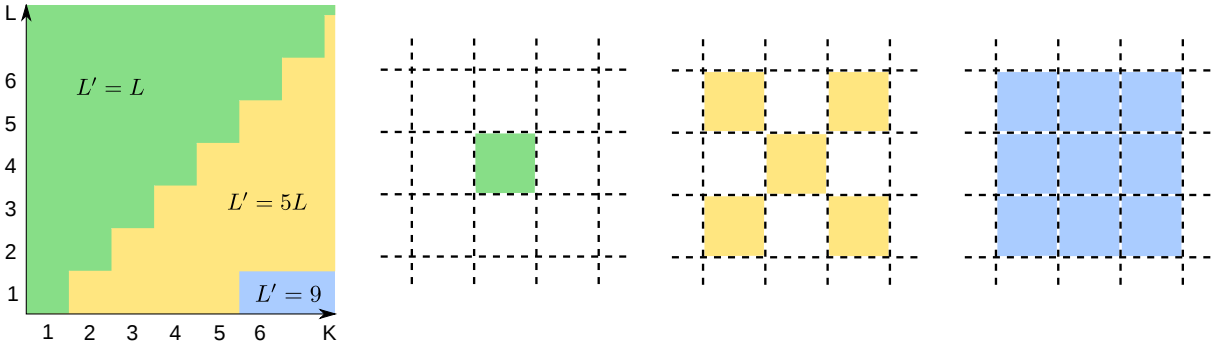


Figure 6.16: Illustration of the selected samples that are used to reach an equivalent number of looks $L' \geq K$. When $L \geq K$ only the pixel of interest is selected leading to $L' = L$. When $K/5 \leq L \leq K$, the 5 neighbor pixels on the diagonal directions are selected. When $L < K/5$, the 9 neighbor pixels are selected.

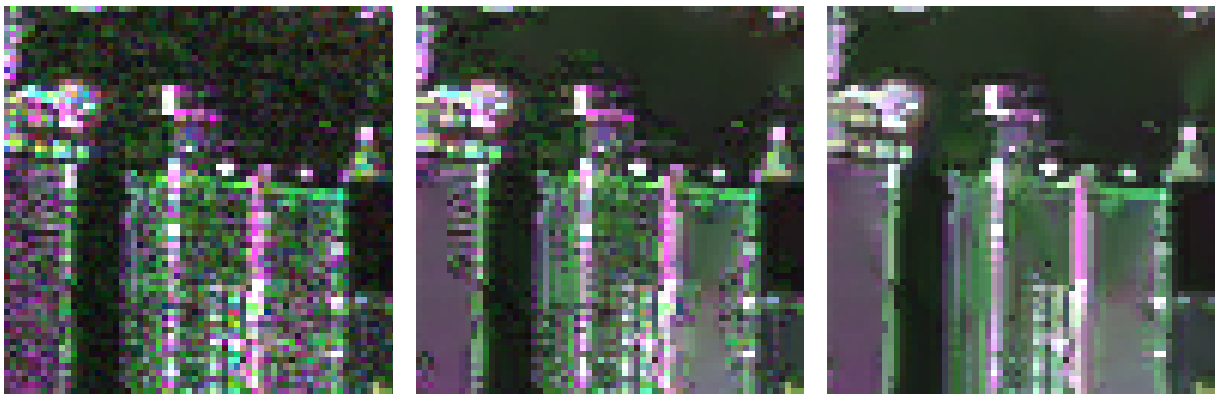
1 degree of freedom. This distribution is then independent on $\mathbf{\Sigma}_{12}$ meaning that the use of GLR leads to an estimator with *stationary relative variance* estimator: the same level of noise reduction can be maintained everywhere in the image (this supports the conclusion in Sec. 3.8). When $K > 1$ and the levels of coherence in $\mathbf{\Sigma}_{12}$ vary spatially in the image, we cannot ensure such a *stationarity*. However, we have seen in Chap. 4 that GLR asymptotically has a constant false-alarm rate (CFAR) with respect to the patch size $|P|$. Under reasonable patch sizes, we can always find a scale for the function f such that the above probabilities (almost) hold true for any value of $\mathbf{\Sigma}_{12}$.

In practice, it is frequently observed that the *speckle* in SAR images present a spatial correlation which is not taken into account by the model considered in Sec. 6.2.1 since SAR images are often upsampled (see Fig. 6.15). When *speckle* is spatially correlated, two patches extracted in the same homogeneous area will appear dissimilar since they will present different patterns due to the blur applied on the *speckle*. Equation (6.54) can no longer be used to select the scale of f which should be increased to reach the same level of false alarm (i.e., the level of noise reduction) at the expense of the probability of detection (i.e., the level of blurring effect). A first solution is to downsample the data so that the *speckle* appears as uncorrelated. In this case Eq. (6.54) can be used successfully at the expense of a loss of resolution. That was the solution in (Deledalle et al., 2009b, 2011a). Another solution is to choose another scale for the function f from studying empirical statistics of the answers of GLR computed on two homogeneous areas manually extracted from the image to process. Note that in this case, when the independence assumption is violated, trapezoidal kernels appear to be significantly more robust than exponential kernels (as used for instance in (Chen et al., 2011)).

The scale of the function g can be chosen in the same vein from the empirical statistics of the answers of the symmetrical Kullback-Leibler divergence computed on two homogeneous areas extracted from the pre-estimated image at hand.

6.4.3 The case of degenerate distributions

We have mentioned in Section 6.2 that when the equivalent number of looks is smaller than the data dimension, i.e., $L < K$, the distribution of $\boldsymbol{C}$ is degenerate. As a consequence, our definition of weights based on likelihood cannot be used: when $\boldsymbol{C}$ is singular the expression of GLR in Eq. (6.45) is undefined. Our attempts to regularize $\boldsymbol{C}$ using diagonal loading methods did not provide satisfying results. In (Deledalle et al., 2010d), we had suggested to cancel off-diagonal elements to ensure $\boldsymbol{C}$ to be diagonal. Good performances were already obtained, even



(a) Noisy image (b) $L_{min} = K = 3$ (c) $L_{min} = 9$

Figure 6.17: (a) A PolSAR image. (b) Illustration of the *noise halo* spread during iterations due to the *rare patch effect*. (c) The redistribution of weights to enforce a minimum of noise reduction limits this undesirable effect.

if theoretically this allows GLR to put in correspondence matrices with different polarimetric signatures.

Another solution consists in comparing two sets of L' observations $\mathcal{S}_1$ and $\mathcal{S}_2$ instead of only two observations $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$. Assuming that all matrices in $\mathcal{S}_1$ (resp. $\mathcal{S}_2$) share a common underlying covariance $\boldsymbol{\Sigma}_1$ (resp. $\boldsymbol{\Sigma}_2$), the GLR is given by:

$$\mathcal{L}_G(\mathcal{S}_1, \mathcal{S}_2) = \left(\frac{2^K \sqrt{|\boldsymbol{C}'_1||\boldsymbol{C}'_2|}}{|\boldsymbol{C}'_1 + \boldsymbol{C}'_2|} \right)^{2L'} \tag{6.55}$$

where $\boldsymbol{C}'_1$ (resp. $\boldsymbol{C}'_2$) denotes the sample mean performed on $\mathcal{S}_1$ (resp. $\mathcal{S}_2$). In other words, GLR relies on a multi-looked image ensuring a minimum equivalent number of looks $L' \geq K$. Note that this multi-looking is only performed to select the suitable samples to combine while the WMLE in Eq. (6.41) is still expressed on single-look data. We denote $\boldsymbol{C}'$ as the image of adherence since it drives GLR to select the suitable noisy observations. Bright targets yield a higher spatial correlation in the horizontal and vertical directions than in diagonal directions. For this reason, we suggest using samples located in the diagonal directions. Figure 6.16 indicates the selected samples and the resulting number of looks L' according to L and K.

6.4.4 Enforcing a minimum amount of smoothing

In an image, some patches are (almost) unique (i.e., not found elsewhere inside the search window). The direct application of the algorithm would produce highly noisy estimates for the central value of these patches since the weighted maximum likelihood estimate would be computed over too few samples. This problem is referred to as the *rare patch effect*. In Chap. 5 this was not an issue for the denoising of the intensity of SAR images. With multi-dimensional SAR data this results in two important problems:

1. The pre-estimate $\hat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\Sigma}$ can be singular if the number of combined samples is smaller than the data dimension K. In the proposed iterative scheme, we have assumed that the substitution of $\boldsymbol{\Sigma}$ by the pre-estimate $\hat{\boldsymbol{\Sigma}}$ in Eq. (6.23) allows us to define the Wishart distribution of $\boldsymbol{C}$. As a consequence, when $\hat{\boldsymbol{\Sigma}}$ is singular, the weights based on the Kullback-Leibler divergence in Eq. (6.48) would be undefined since they require the inversion of

$\hat{\mathbf{\Sigma}}$. The pre-estimate $\hat{\mathbf{\Sigma}}$ should result from the combination of at least K samples to be considered as a proper estimate of $\mathbf{\Sigma}$ and then to be used to properly refine the weights.

2. When considering an iterative joint estimation of K co-registered channels, the high variance of the estimator for "rare" patches leads to a decrease in the similarity between pre-estimated patches with the iterations. At the algorithm end, the resulting denoised image contains regions of high residual variance. Since the lack of repetitive patterns is all the more important around high contrasted edges and bright targets, this high residual noise is referred to as *noise halo*. It is then desirable to enforce a minimum amount of smoothing (i.e. variance reduction) to consider $\hat{\mathbf{\Sigma}}$ as a proper estimate of $\mathbf{\Sigma}$.

To guarantee a minimum amount of smoothing, and therefore limit the variance of the estimation, we propose to estimate the equivalent number of looks of the denoised pixels. Due to our non-local (data driven) approach, the equivalent number of looks varies from one pixel to another. It depends on the number of similar patches found in the search window, and can be approximated, for each pixel x, by:

$$\hat{L}(x) = \frac{\left(\sum_{x'} w(x,x')\right)^2}{\sum_{x'} w(x,x')^2} \tag{6.56}$$

according to the variance reduction of a weighted average (see Sec. 3.6). To enforce a minimum amount of smoothing, we suggest redefining the weights $w(x,x')$ in the cases where the equivalent number of looks $\hat{L}(x)$ falls below a given threshold L_{min}. One option is to redistribute equally the weights of the L_{min} most similar patches whenever $\hat{L}(x) < L_{min}$. "Rare" patches often contain a bright scatterer. To prevent biasing the estimation, we propose restricting the selection of the L_{min} patches to those whose central value is not too bright nor too dark compared to that of the reference patch, following the ideas of (Lee, 1983; Yaroslavsky, 1985; Tomasi and Manduchi, 1998). The correction of the weights can be performed as follows:

- Compute $\hat{L}(x)$ for each pixel x,
- If $\hat{L}(x) < L_{min}$, redistribute the L_{min} highest weights:
  - Create a vector $\boldsymbol{w}$ containing all the weights $w(x,x')$ such that:
$$\mathrm{tr}(\boldsymbol{C}(x))/4 < \mathrm{tr}\ \boldsymbol{C}(x') < 4\ \mathrm{tr}(\boldsymbol{C}(x))\ , \tag{6.57}$$
  - Sort the vector $\boldsymbol{w}$ in descending order,
  - Redistribute equally the weights of the L_{min} most similar pixels:
$$\boldsymbol{w}_k \leftarrow \frac{1}{L_{min}} \sum_{l=1}^{L_{min}} \boldsymbol{w}_l \quad \forall k \in 1..L_{min}\ . \tag{6.58}$$

Note that Eq. (6.57) is the selection rule associated to GLR (see Sec. 3.8).

Fig. 6.17 illustrates on a PolSAR image the *rare patch effect* and two solutions obtained by ensuring a minimum equivalent number of looks of 3 and 9. For the smallest minimum amount of smoothing $L_{min} = K = 3$, the *noise halo* spreads during the iterations. When $L_{min} = 9$, our redistribution rule limits this undesirable effect. We will see in Chap. 7 that the problem of the *rare patch effect* can be solved using patches with variable shape and size. This approach has not been considered here for two reasons. First, it increases the computing time. Due to the large size of SAR images, SAR filters must be as fast as possible to be applied on large databases. The choice of the size and the shape of patches is based on an aggregation-based filter which requires an unbiased estimator of the risk. We have seen in Sec. 3.6 that to our knowledge there is no suitable risk estimator for data damaged by a gamma distribution and therefore neither for a Wishart distribution.

Algorithm Non Local SAR (NL-SAR)

Input: noisy image $\boldsymbol{C}$, image of adherence $\boldsymbol{C}'$ and pre-estimate $\hat{\boldsymbol{\Sigma}}^{i-1}$
Output: estimate $\hat{\boldsymbol{\Sigma}}^{i}$
for all pixel positions x in the image **do**
  Initialize $\hat{\boldsymbol{\Sigma}}^{i}(x)$ and N to zero
  for all pixel positions x' in $W(x)$ **do**
    $t \leftarrow 0$
    for all translation vectors $\tau \in \mathrm{P}$ **do**
      Compute $-\log \mathcal{L}_G(\boldsymbol{C}'(x+\tau), \boldsymbol{C}'(x'+\tau))$ ▷ use Eq. (6.46)
      Compute $\mathcal{D}_{KL}(\hat{\boldsymbol{\Sigma}}^{i-1}(x+\tau), \hat{\boldsymbol{\Sigma}}^{i-1}(x'+\tau))$ ▷ use Eq. (6.48)
      $t \leftarrow t + (1-\lambda) f\left(-\log \mathcal{L}_G(\boldsymbol{C}'(x+\tau), \boldsymbol{C}'(x'+\tau))\right)$
        $+\lambda g\left(\mathcal{D}_{KL}(\hat{\boldsymbol{\Sigma}}^{i-1}(x+\tau), \hat{\boldsymbol{\Sigma}}^{i-1}(x'+\tau))\right)$
    end for
    $\hat{\boldsymbol{\Sigma}}^{i}(x) \leftarrow \hat{\boldsymbol{\Sigma}}^{i}(x) + \varphi(t)\boldsymbol{C}(x')$
    $N \leftarrow N + \varphi(t)$
  end for
  Enforce a minimum noise reduction ▷ see Sec. 6.4.4
  $\hat{\boldsymbol{\Sigma}}^{i}(x) \leftarrow \hat{\boldsymbol{\Sigma}}^{i}(x)/N$
end for
return $\hat{\boldsymbol{\Sigma}}^{i}$

Figure 6.18: Pseudo-code of the non local InSAR algorithm. The procedure has to be repeated iteratively. At iteration i the pre-estimated covariance matrix $\hat{\boldsymbol{\Sigma}}^{i-1}$ is used to refine the estimates. In practice, the first pre-estimates can be chosen as a constant image, and four iterations are enough to reach a good estimate.

6.4.5 Algorithm and implementation

This section describes the whole procedure used in NL-SAR. At each site x, the pixels x' present in the search window W_x are inspected sequentially to produce a weight by comparing two surrounding patches P_x and $\mathrm{P}_{x'}$. For each corresponding pixels $x+\tau$ and $x'+\tau$ in P_x and $\mathrm{P}_{x'}$, the similarity is computed by comparing the values of adherence $\boldsymbol{C}'(x+\tau)$ and $\boldsymbol{C}'(x'+\tau)$ using Eq. (6.46) and Eq. (6.55) and the pre-estimated parameters $\hat{\boldsymbol{\Sigma}}(x+\tau)$ and $\hat{\boldsymbol{\Sigma}}(x+\tau)$ using Eq. (6.48). These similarities are combined to produce the weights $w(x,x')$ (see Eq. (6.50)). In practice, the logarithm of the weights is computed to limit numerical errors. Once all weights are obtained for each pixel x, the minimum noise reduction procedure is performed (Section 6.4.4) before computing the new parameters $\hat{\boldsymbol{\Sigma}}$ (6.41). The procedure is performed iteratively. At the end of the iteration $i-1$, the estimated parameters provide the pre-estimated matrix $\hat{\boldsymbol{\Sigma}}^{i-1}$ used at iteration i. The procedure is repeated until there is no more change between two consecutive estimates. In practice, the first pre-estimates can be chosen as the identity matrix, i.e., with unit reflectivity and null coherence. According to Eq. (6.48), this is equivalent to performing the first iteration of NL-SAR with weights based only on the likelihood term. Finally, four iterations are performed to reach good estimates.

The pseudo-code of NL-SAR is given in Fig. 6.18 and the global scheme is illustrated in Fig. 6.19. The algorithm complexity is $O(|\Omega||\mathrm{W}||\mathrm{P}|)$ where $|\Omega|$, $|\mathrm{W}|$ and $|\mathrm{P}|$ are respectively the image size, the search window size and the similarity patch-size. Several optimizations of

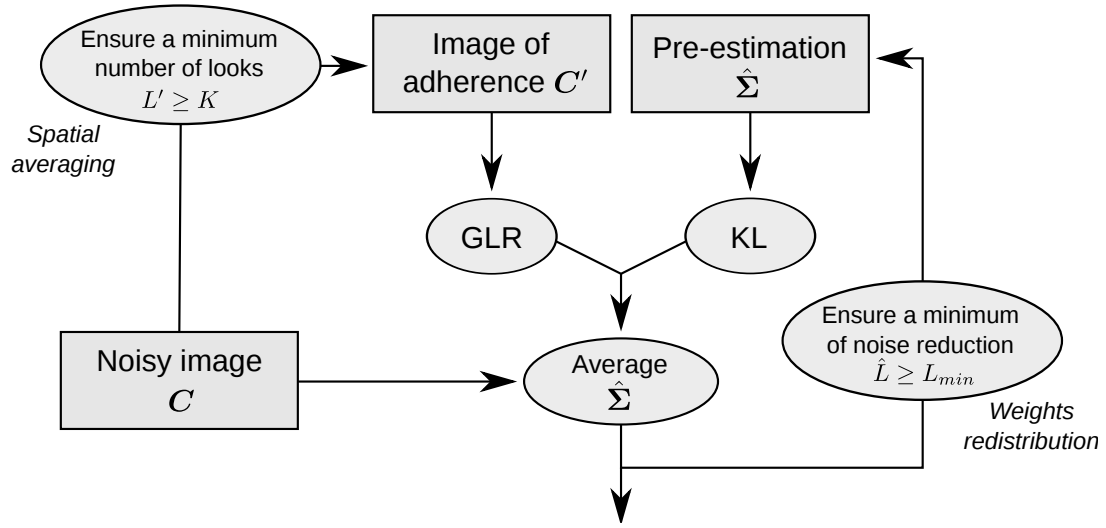


Figure 6.19: Scheme of the NL-SAR procedure. The estimated covariance matrix $\mathbf{\Sigma}$ is iteratively refined by combining noisy observation selected according to the similarities evaluated on the image of adherence and a pre-filtered image. Ensuring an equivalent number of looks $L' > K$ in the adherence image and a minimum of noise reduction $\hat{L} > L_{min}$ in the pre-estimate is necessary to evaluate similarity and reduce the *rare patch effect*.

the non-local means have been proposed in (Buades et al., 2005; Coupe et al., 2006; Goossens et al., 2008). We have extended the solution proposed by Darbon et al. (2008) for the NL-SAR algorithm with a time complexity given by $O(4|\Omega||W|)$. Finally, with a C implementation, the computation time of our method is of about 10 seconds per iteration for a polarimetric image of size $|\Omega| = 512 \times 512$ and windows of size $|\mathrm{W}| = 15 \times 15$ and $|\mathrm{P}| = 5 \times 5$ using an Intel Core 2 Duo 3.00GHz. Thanks to a parallel implementation of Darbon's optimization using OpenMP, NL-SAR can perform one iteration in about 30 seconds using 16 Intel Core 2 Duo Xeon 2.27GHz on a $|\Omega| = 4096 \times 4096$ polarimetric image. Our implementation is released under public license. The technical documentation of NL-SAR is given in Appendix B.

6.5 Experiments and results

6.5.1 Description and setting of the compared methods

In our experiments, comparisons will be performed with the classical boxcar filter on a 7×7 sliding window, the refined Lee estimator (Lee et al., 2003) on a 7×7 sliding window, and the IDAN filter with an adaptive neighborhood of maximum size 50 pixels (Vasile et al., 2006). For these three filters, we use the implementation provided by the PolSARPro project of ESA/IETR. These filters will be compared to a non-iterative version of NL-SAR using weights based only on the comparisons of noisy data. The iterative NL-SAR filter will be applied with a search window of size $|\mathrm{W}| = 15 \times 15$ and a similarity window of size $|\mathrm{P}| = 5 \times 5$. A minimum noise reduction of level $L_{min} = 9$ will be maintained. We use 4 iterations of the iterative NL-SAR filter to reach satisfying estimates.

6.5.2 Comparisons on simulated SAR data

Given the true images of reflectivities, phase differences and coherences, K single-look complex (SLC) images can be generated according to the complex circular Gaussian distribution model

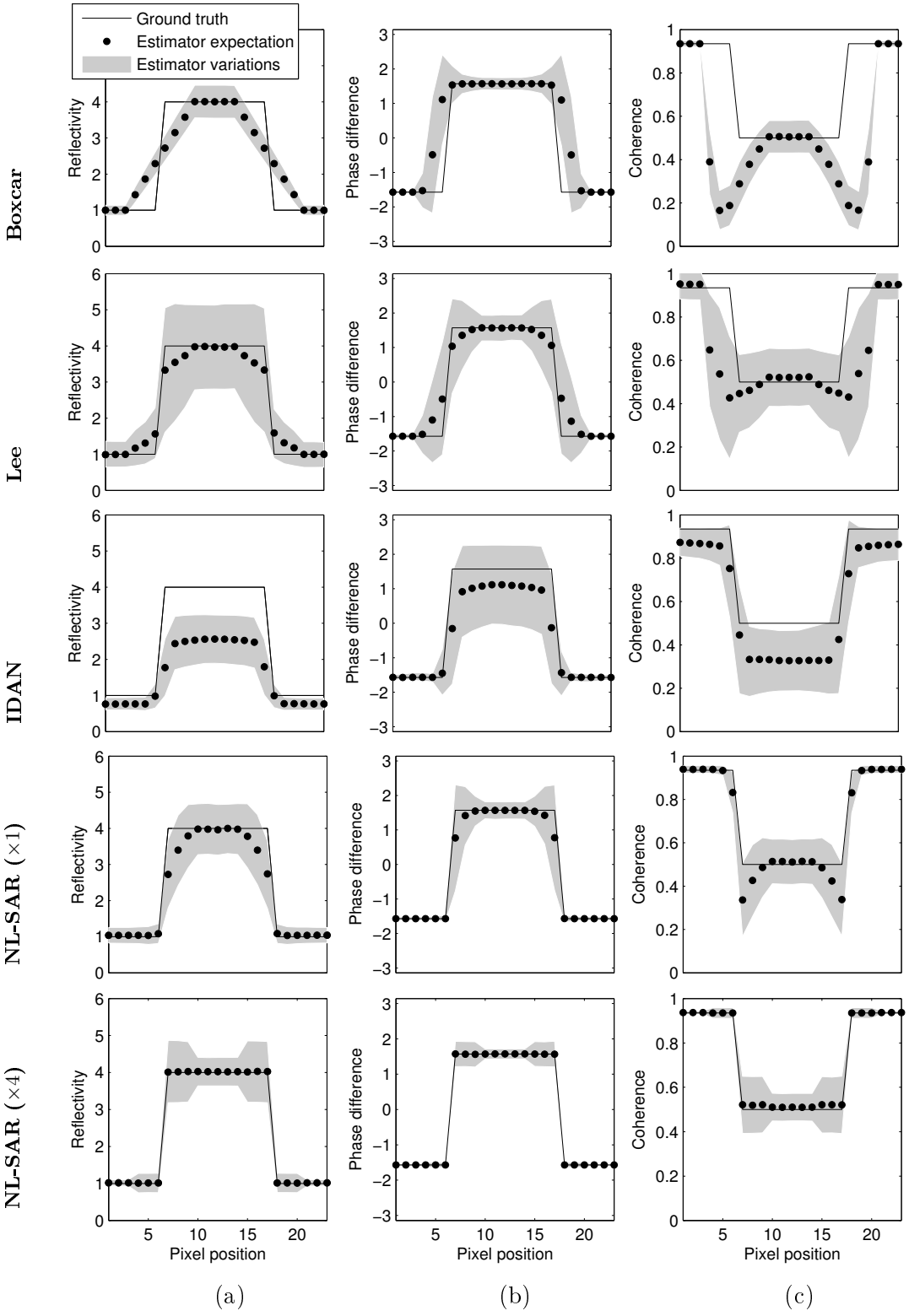


Figure 6.20: (a) Reflectivity, (b) phase difference and (c) coherence of a simulated rectangular function. The statistical answers are given from top to bottom by the boxcar estimator, Lee's estimator, the IDAN estimator, the non-iterative NL-SAR estimator and the (iterative) NL-SAR estimator.

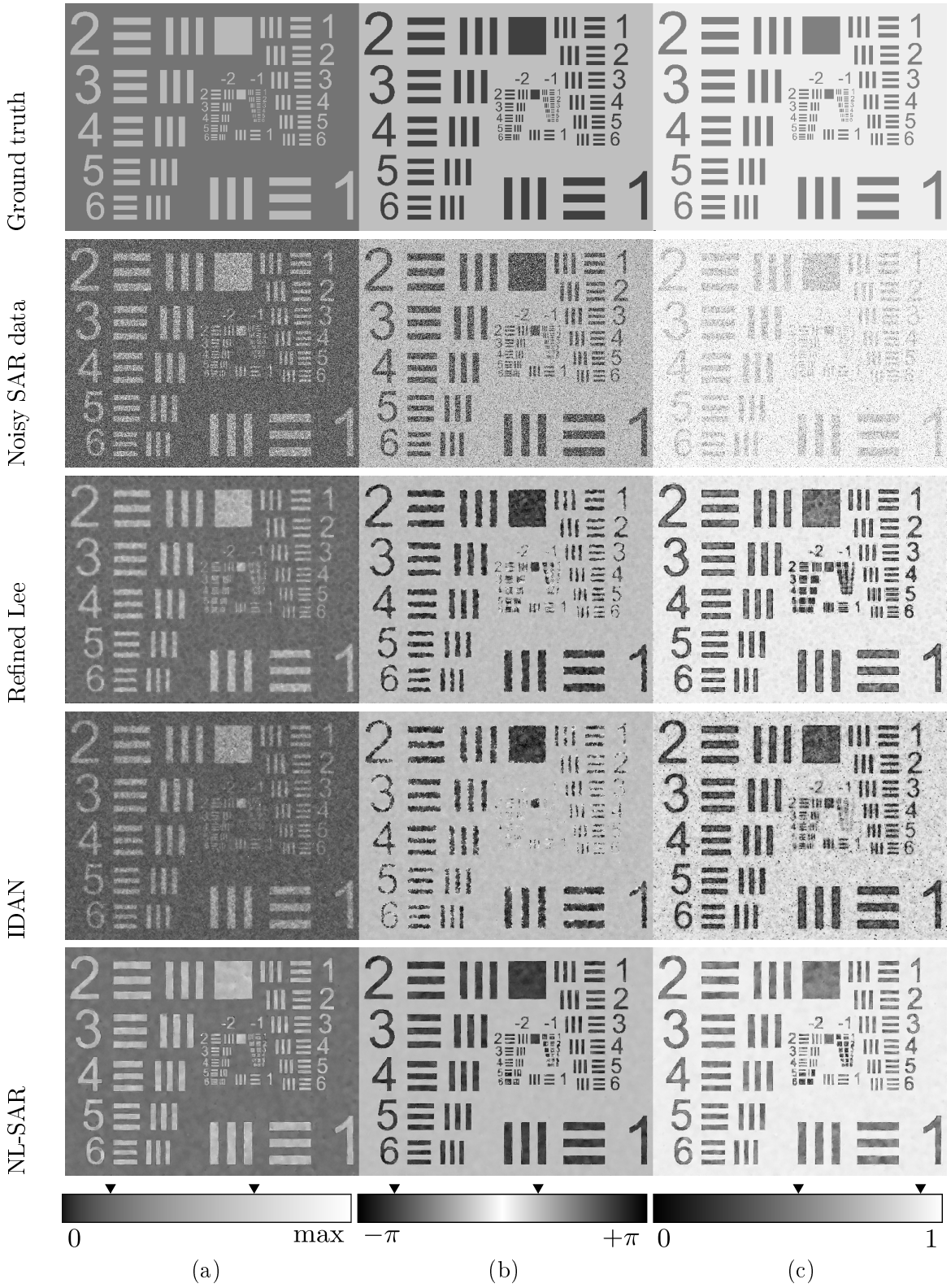


Figure 6.21: (a) Reflectivity, (b) phase difference and (c) coherence of a resolution test pattern obtained from top to bottom by the ground truth, the SLC images (maximum likelihood estimator of (Seymour and Cumming, 1994)), the refined Lee estimator (Lee et al., 2003), the IDAN estimator (Vasile et al., 2006) and the NL-SAR estimator. A colorbar of the range value is shown for each channel with pointers to indicate the true underlying values.

given in Eq. (6.16). The simulation procedure is given in (Lee et al., 2003), as follows:

- Compute a matrix $\boldsymbol{L}$ such that $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^*$. For example, in the interferometric case, the lower triangle matrix $\boldsymbol{L}$ in the Cholesky decomposition is a good candidate:

$$\boldsymbol{L} = \sqrt{R} \begin{pmatrix} 1 & 0 \\ De^{-j\beta} & \sqrt{1-D^2} \end{pmatrix}, \tag{6.59}$$

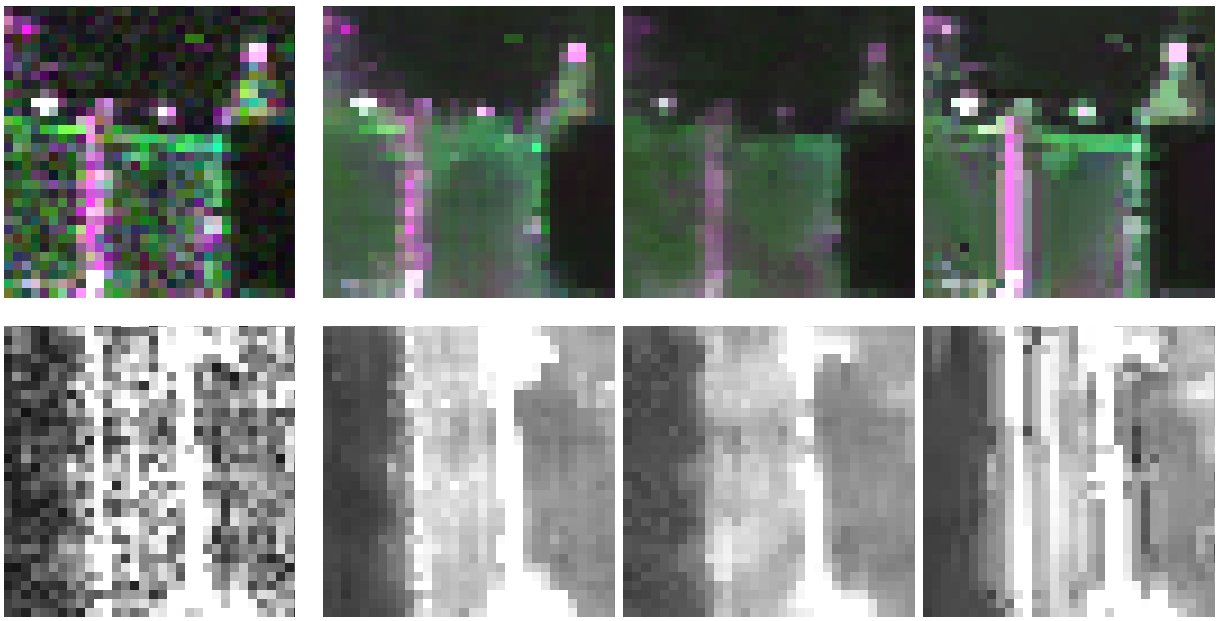
- Generate a vector $\boldsymbol{k}_0$ of K independent complex random variables with independent real and imaginary parts following a Gaussian distribution of variance $\sigma^2/2$,
- Finally, the scattering complex vector $\boldsymbol{k}$ is given by multiplying the matrix $\boldsymbol{L}$ by $\boldsymbol{k}_0$

$$\boldsymbol{k} = \boldsymbol{L}\boldsymbol{k}_0. \tag{6.60}$$

Gallager (2008) shows that any circular complex random vector can be generated by this process, and this process can only generate circular complex random vectors. There is equivalence between this generative model and Eq. (6.16). Next, multi-look data can be simulated by computing the empirical covariance matrix from L independent scattering vectors $\boldsymbol{k}$ generated as described above.

Figure 6.20 shows the statistical answer of the five estimators on a cut through a line of width 10. The statistics have been measured on denoised images of over 10 000 noisy generated images. The ground truth, the mean and an interval of variation (about 70% of the estimates) are represented on the graphics for the three estimated components. We can notice that the boxcar filter is unbiased with a low variance in homogeneous areas but presents a strong spatial bias around the edges of the rectangular function. This spatial bias produces large underestimates of the coherence around edges which is denoted in (Lee et al., 2003) as the dark ring effect. The refined Lee estimator presents less spatial bias but has a higher variance. This is due in part to the edge-aligned windows containing less samples to reduce the variance, but also, to the window selection process which presents high variations. IDAN provides a good restoration of the edges but unfortunately a bias is introduced even in homogeneous areas. This is due to the selection rule used during the growing region which tends to lower reflectivity and coherence values (see Chap. 3). Moreover, the bias increases on the line since the adaptive neighborhood selects samples out of the line. As a result the variance is bigger than for the boxcar filter even if there are as many values to estimate the cross-correlation. We assume this phenomenon could be reduced by using a more suitable similarity criterion to define the region growing. NL-SAR provides the best bias-variance trade-off. Indeed, compared to the boxcar filter, the refined Lee estimator and IDAN, (iterative) NL-SAR is neither biased in homogeneous areas nor around edges. Moreover, its variance is equivalent to the one of the boxcar filter in homogeneous areas. NL-SAR has a larger variance around edges than in homogeneous areas since these regions present less redundant patterns. The non-iterative NL-SAR provides a trade-off between the boxcar filter and the iterative NL-SAR.

Figure 6.21 presents the obtained estimated images for two generated single-look complex images representing a 600×464 resolution test pattern. On the original resolution test pattern, the contrasts between the lowest value and highest value, for all channels, are the same as on the line cut of Figure 6.20. The images obtained with the NL-SAR estimator seem to be well smoothed with a better edge and shape preservation. The images obtained by the boxcar and the IDAN estimators are more noisy than the images obtained by the NL-SAR filter (the remaining variance is larger). Moreover, the boxcar estimator blurs the edges resulting in a loss of resolution



(a) Noisy image (b) Refined Lee (c) IDAN (d) NL-SAR

Figure 6.22: Visual evaluation of the target preservation from (a) a PolSAR and a SAR image using (b) the refined Lee filter, (c) the IDAN filter and (d) our NL-SAR filter.

and large underestimations of the coherence around edges. The IDAN filter preserves the shapes but the noise variance remains essentially large in the coherence image, and small details are essentially lost in the phase difference image. Finally, our NL-SAR estimator seems to work efficiently by preserving small structures with satisfying noise reduction.

6.5.3 Overview of results on different SAR data

This section presents an overview of results obtained on real co-registered SLC SAR images. The SAR images are assumed to follow Goodman's model presented in Sec. 6.2.1.

The experiments are performed on three data sets:

- an interferometric pair of images acquired simultaneously (mono-pass) over a single building of complex shape in Toulouse (France) by RAMSES (aerial sensor). They are in X-band (wavelengths ranging from 2 to 5 cm) with a resolution under one meter in azimuth and slant range.

- a polarimetric image acquired over a wide area of San Francisco (USA) by AIRSAR (aerial sensor). It is in L-band (wavelengths ranging from 15 to 30 cm) with a resolution of about ten meters in azimuth and slant range.

- a polarimetric image acquired over a small urban area of Kaufbeuren (Germany) by F-SAR (aerial sensor). It is in S-band (wavelengths ranging from 7 to 15 cm) with a resolution of 0.5 meter/pixel in azimuth and 0.64 meter/pixel in range.

Figure 6.22 compares the performance of denoising approaches to preserve the bright scatterers numerous in SAR images. The IDAN filter tends to oversmooth and diffuse the bright scatterers of the image resulting in a blur effect. While the refined Lee filter preserves well the bright scatterers, the NL-SAR filter can also restore targets of lower amplitudes, preserve edges

and significantly reduce the level of noise. In both examples, on an image of amplitude and a polarimetric image, our NL-SAR filter seems to produce the best trade-off in terms of target preservation and noise reduction. Note that NL-SAR preserves well the three bright lines on the left of the building whereas the two other filters blur them. This attests the efficiency of the patch-based approach: the three lines act as a rail on which the similarity patch slides in order to combine all pixels parallel to the bright lines.

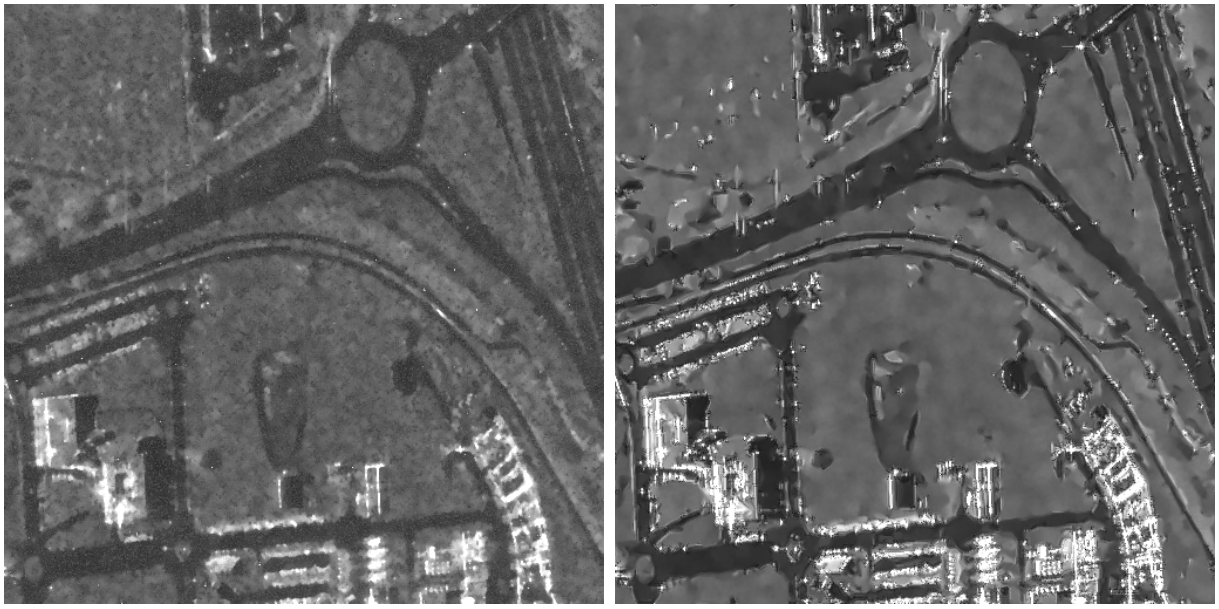
Figures 6.23, 6.24 and 6.25 present the obtained results for the different filters to estimate the interferometric parameters. Figures 6.26 and 6.27 present the obtained estimates of the polarimetric parameters for the different denoising filters. For all these images, the range is on the horizontal axis and the azimuth on the vertical axis. The results obtained with our NL-SAR estimator seem to be well smoothed with a better edge and shape preservation than other filters. As mentioned in Chap. 3, IDAN underestimates the intensities whereas the refined Lee filter and the NL-SAR filter preserve this information. As predicted by Fig. 6.20, the coherence is slightly overestimated by the refined Lee filter and underestimated by the IDAN filter. The NL-SAR provides a lower coherence since it is able to select a larger number of i.i.d. samples. With NL-SAR, the *speckle* effect is strongly reduced and the spatial resolution seems to be well preserved: buildings, streets and homogeneous areas are well restored.

Figure 6.28 and 6.29 shows the entropy/anisoptropy/alpha angle decomposition obtained from the three estimators on the two polarimetric data used above. In both cases, the parameters extracted from the NL-SAR estimate appear smoother with better edge preservation. The entropy, usually overestimated by the two other filters, is lower using NL-SAR.

6.6 Conclusion

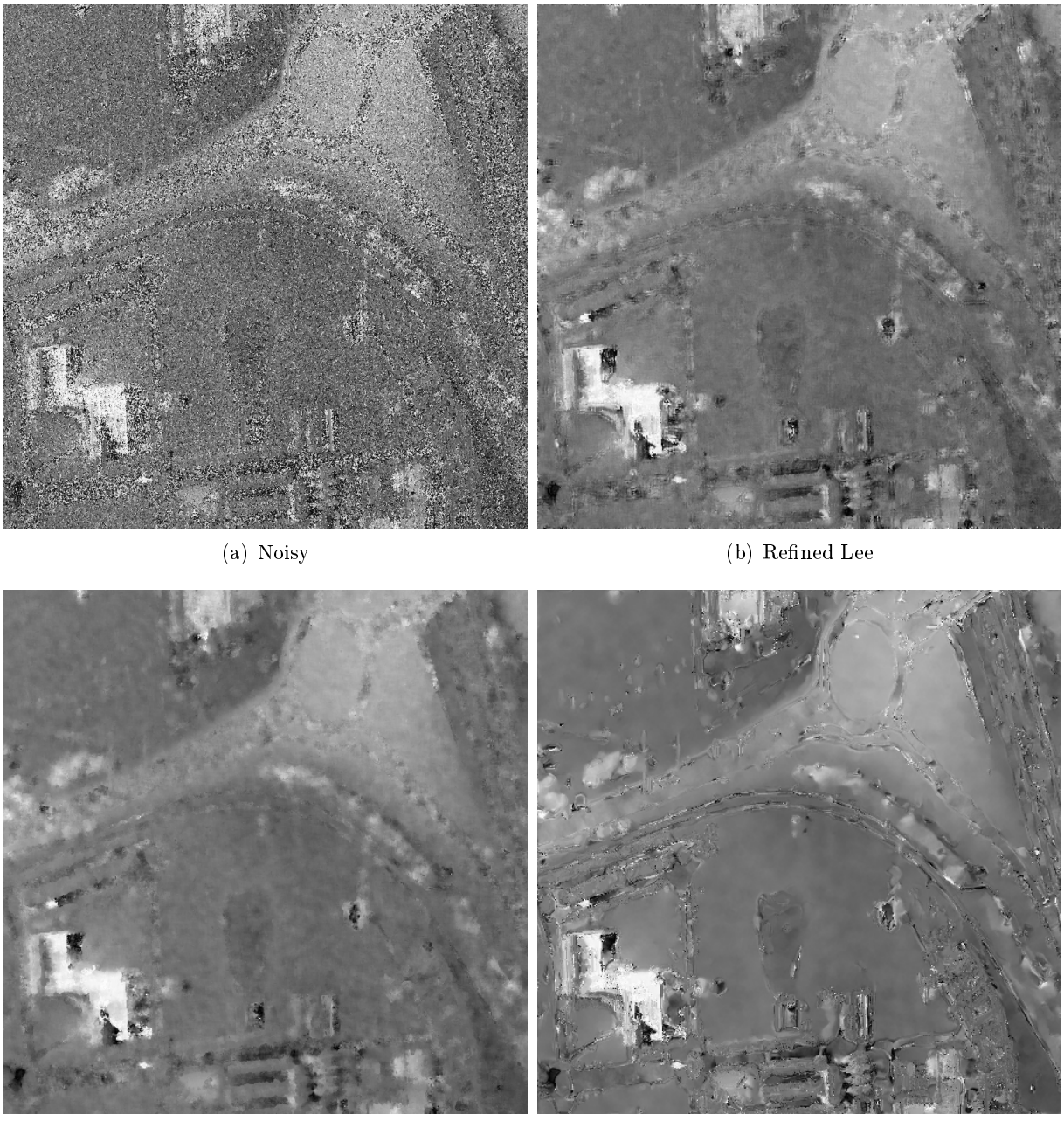
A new approach was proposed for the estimation of the SAR parameters. This method is based on the non-local means filter (Buades et al., 2005) whose originality rests on the weighted combination of pixel values which can be far apart. We apply the general iterative methodology proposed in Chap. 5 to select suitable pixels by evaluating patch-based similarity considering the noisy amplitudes, the complex cross-correlations and previous estimates. Finally, the reflectivity, the phase differences and the coherence between the different channels are jointly estimated. The proposed estimator out-performs *state-of-the-art* estimators in terms of both noise reduction and edge preservation. The noise, present in the input images, is well smoothed in the homogeneous regions and the object contours are well restored (preservation of the resolution). Moreover we can consider from our experiments that the reflectivity, the actual interferometric phase and the coherence are well recovered, without introducing strong undesired artifacts and with a good restoration of bright scatterers.

(a) Noisy (b) Refined Lee



(c) IDAN (d) NL-SAR

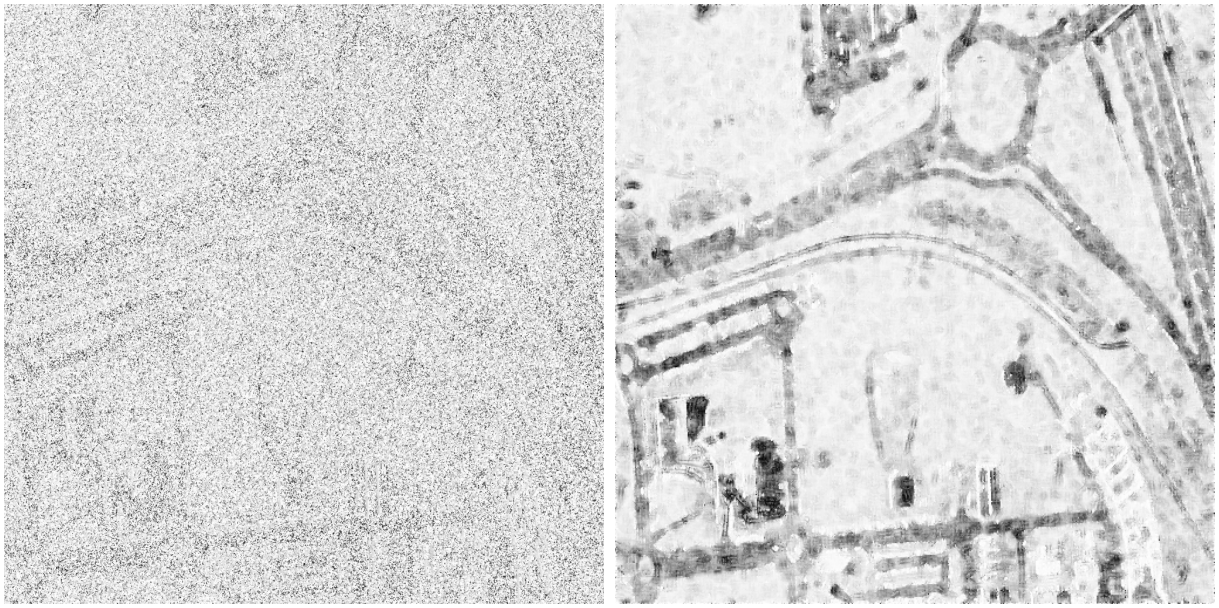
Figure 6.23: (a) The intensity of a SAR image of an urban area in Toulouse (France) sensed by RAMSES ©ONERA. The reflectivity images estimated jointly on a pair of interferometric SAR images using (b) the refined Lee estimator (Lee et al., 2003), (c) the IDAN estimator (Vasile et al., 2006) (d) and the NL-SAR estimator.



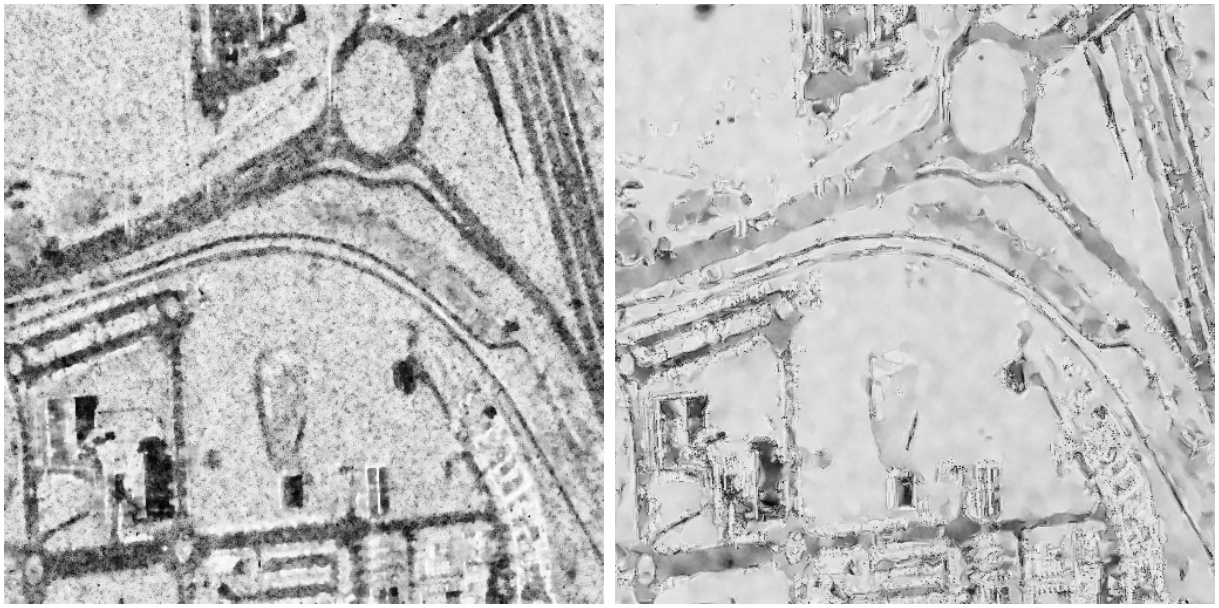
(a) Noisy (b) Refined Lee

(c) IDAN (d) NL-SAR

Figure 6.24: (a) The interferometric phase between a pair of SAR images of a urban area in Toulouse (France), sensed by RAMSES ©ONERA with a mono-pass. Interferometric phase images estimated jointly on the pair of interferometric SAR images using (b) the refined Lee estimator (Lee et al., 2003), (c) the IDAN estimator (Vasile et al., 2006) (d) and the NL-SAR estimator.

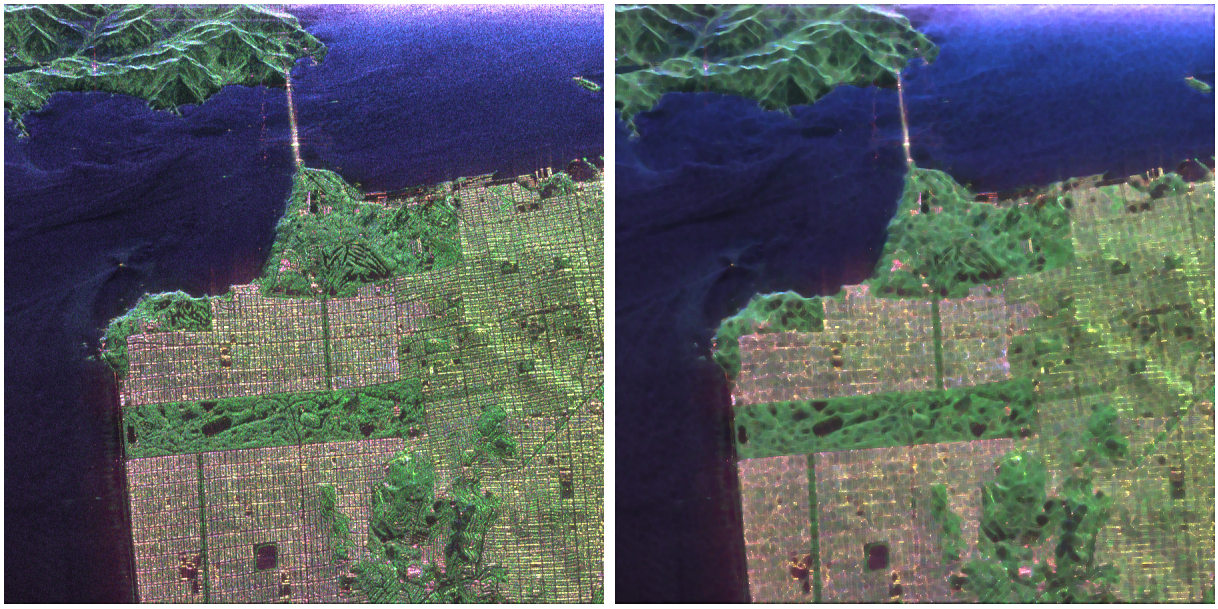


(a) Noisy (b) Refined Lee

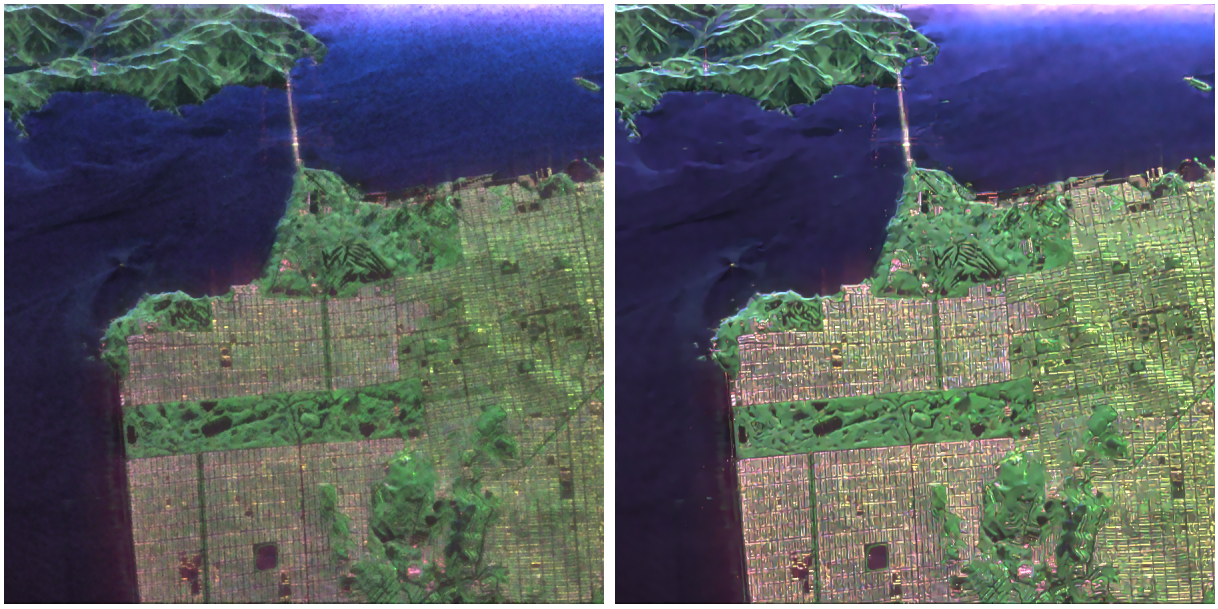


(c) IDAN (d) NL-SAR

Figure 6.25: (a) The coherence between a pair of SAR images of a urban area in Toulouse (France), sensed by RAMSES ©ONERA with a mono-pass (maximum likelihood estimation of Seymour and Cumming (1994)). Coherence images estimated jointly on the pair of interferometric SAR images (b) the refined Lee estimator (Lee et al., 2003), (c) the IDAN estimator (Vasile et al., 2006) (d) and the NL-SAR estimator.

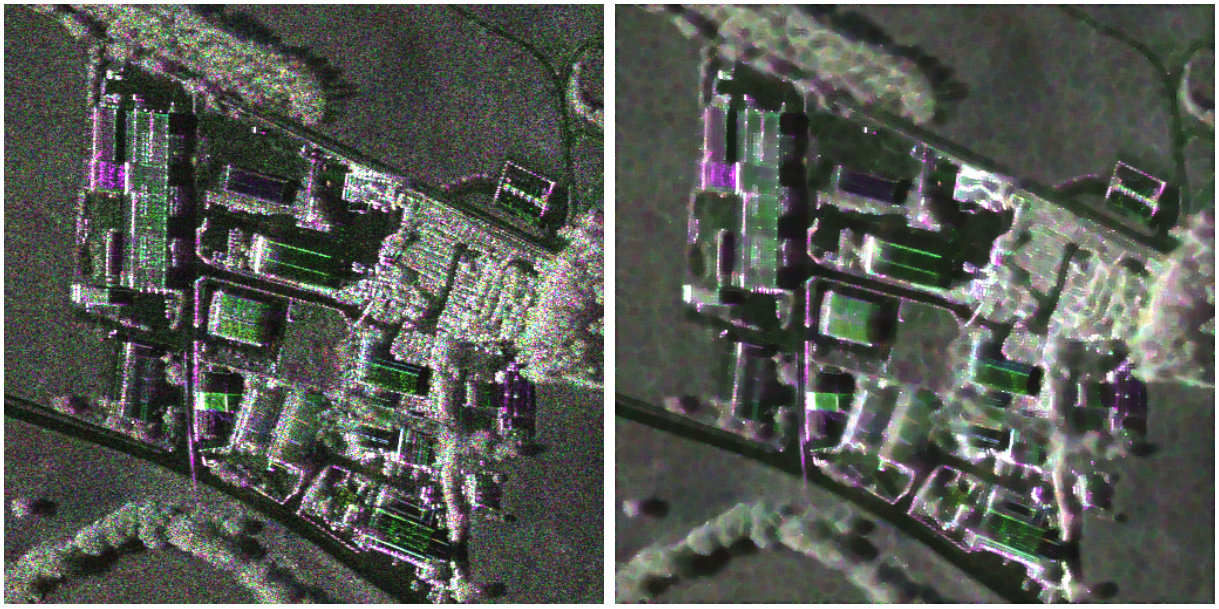


(a) Noisy (b) Refined Lee

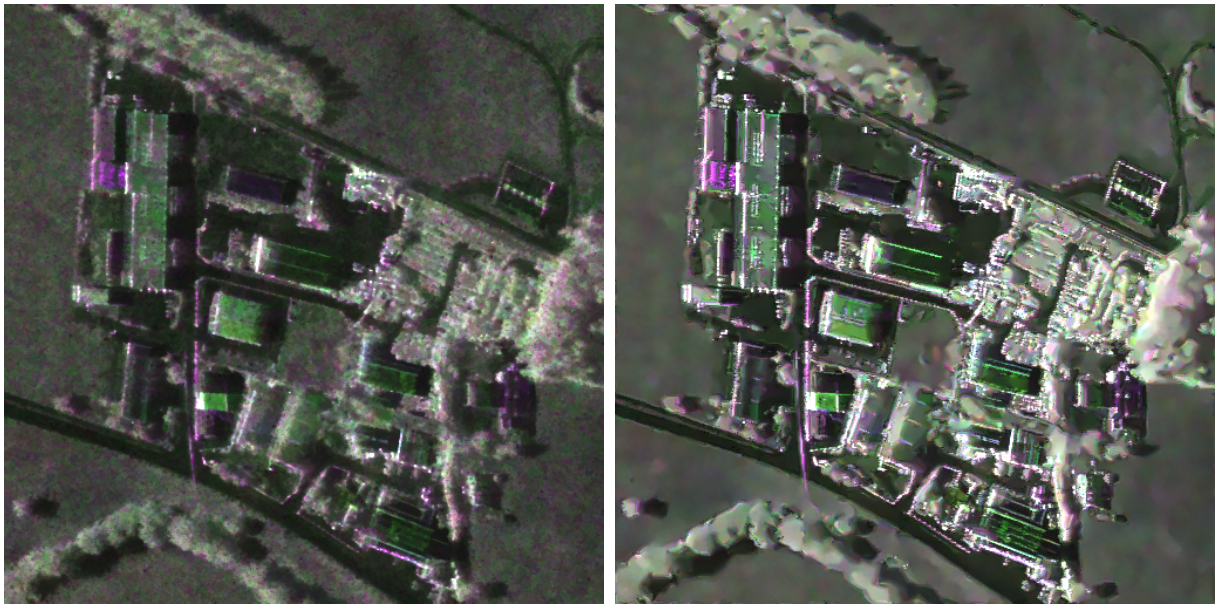


(c) IDAN (d) NL-SAR

Figure 6.26: (a) The polarimetric image of San Francisco (USA), sensed by L-band AIRSAR ©NASA-JPL-Caltech displayed using an RGB representation based on the Pauli basis. Polarimetric images estimated using (b) the refined Lee estimator (Lee et al., 2003), (c) the IDAN estimator (Vasile et al., 2006) (d) and the NL-SAR estimator.

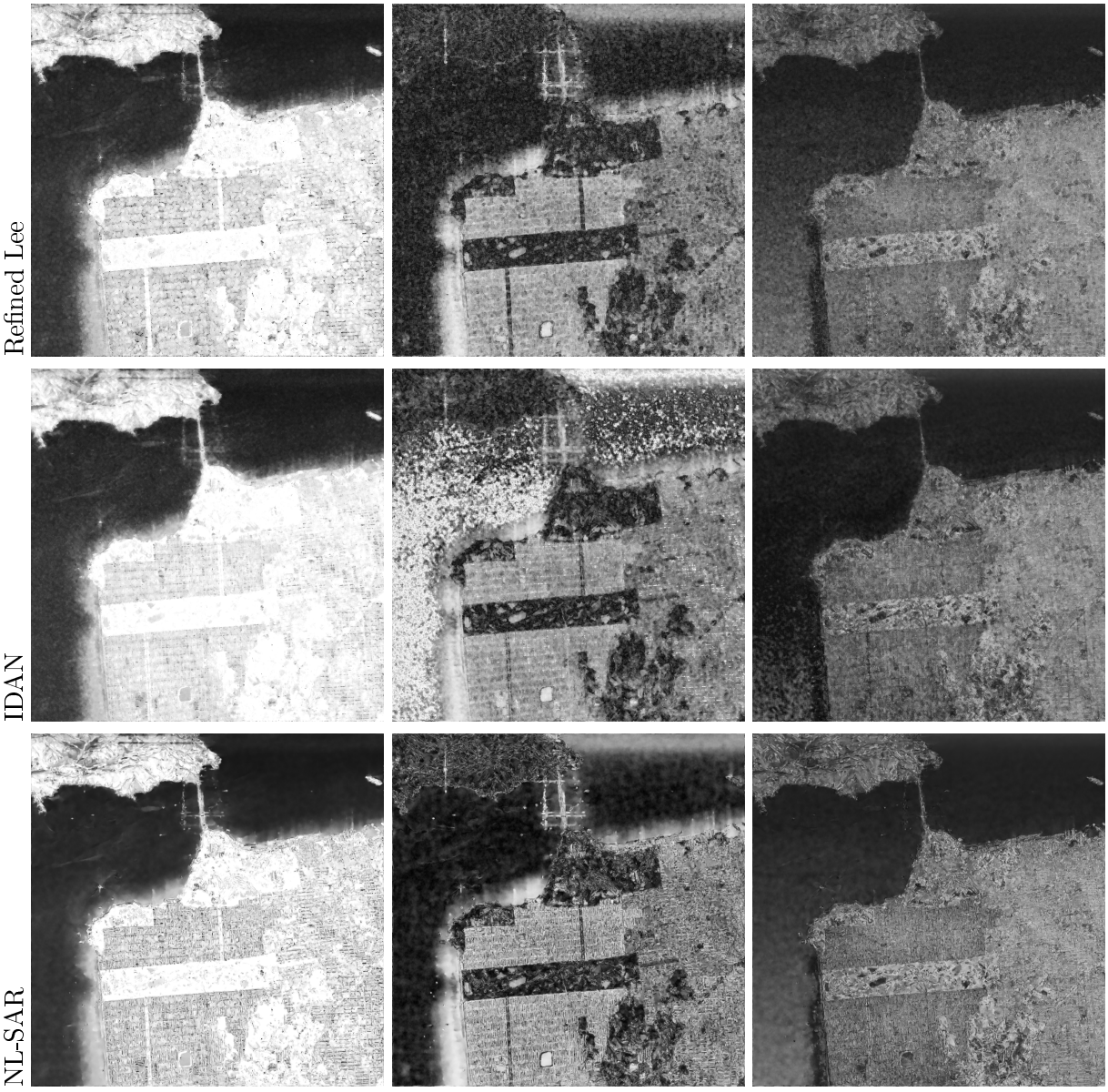


(a) Noisy (b) Refined Lee



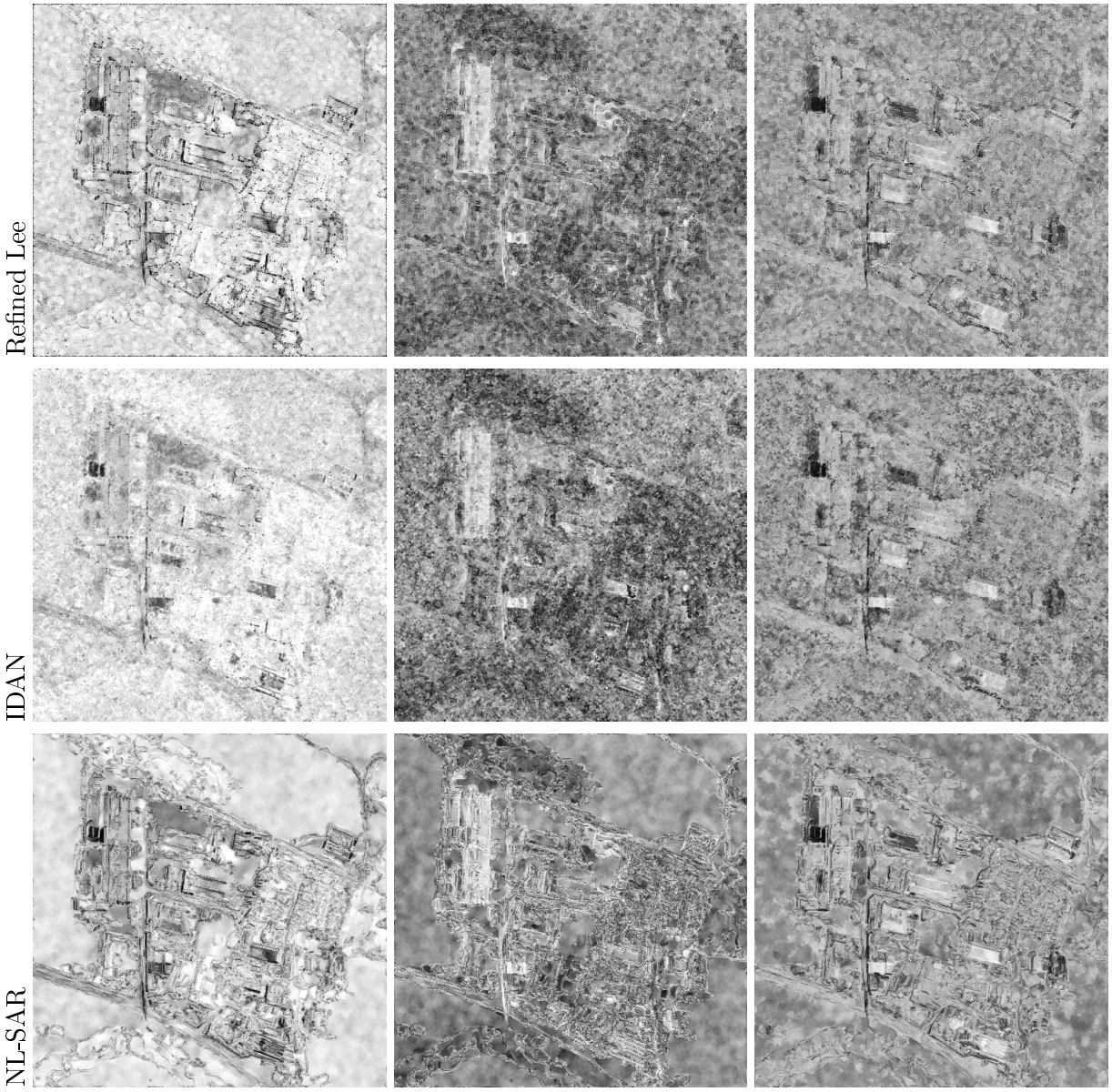
(c) IDAN (d) NL-SAR

Figure 6.27: (a) The polarimetric image of Kaufbeuren (Germany), sensed by S-band F-SAR ©DLR displayed using an RGB representation based on the Pauli basis. Polarimetric images estimated using (b) the refined Lee estimator (Lee et al., 2003), (c) the IDAN estimator (Vasile et al., 2006) (d) and the NL-SAR estimator.



(a) Entropy (b) Anisotropy (c) Alpha angle

Figure 6.28: Polarimetric information extracted from the polarimetric images of San Francisco (USA), L-band AIRSAR ©NASA-JPL-Caltech, obtained from top to bottom by the refined Lee estimator (Lee et al., 2003), the IDAN estimator (Vasile et al., 2006) and the NL-SAR estimator.



(a) Entropy (b) Anisotropy (c) Alpha angle

Figure 6.29: Polarimetric information extracted from the polarimetric images of Kaufbeuren (Germany), S-band F-SAR ©DLR, obtained from top to bottom by the refined Lee estimator (Lee et al., 2003), the IDAN estimator (Vasile et al., 2006) and the NL-SAR estimator.

Chapter 7

Spatially adaptive patches

This chapter relates a joint work with Vincent Duval and Joseph Salmon.

We propose in this chapter an extension of the non-local means (NL means) denoising algorithm where the usual square patches used to compare pixel neighborhoods are replaced by various shapes that can take advantage of the local geometry of the image. We provide a fast algorithm to compute the NL means with arbitrary shapes thanks to the Fast Fourier Transform. We then consider local combinations of the estimators associated with various shapes by using Stein's Unbiased Risk Estimate (SURE). Experimental results show that this algorithm improves the standard NL means performance and is close to *state-of-the-art* methods, both in terms of visual quality and numerical results. Moreover, common visual artifacts usually observed by denoising with NL means are reduced or suppressed thanks to our approach.

Patch-based methods are already quite efficient at dealing with smooth regions and textures. However, since they use patches with a fixed square shape and a fixed scale over the whole image, their performances may be limited for dealing with edges, mostly for edges with high contrast. Indeed, edges with high contrast present few redundancies and their denoised version suffer from a persistence of residual noise: this is named the *noise halo*. In order to overcome this drawback, more directional priors may be considered, using locally chosen scales and orientations of shapes. Few attempts have been made to use several patch sizes (see Mairal et al., 2008) for learning with patches or (Salmon and Strozecki, 2010) for the NL means.

As far as we know, the only work trying to handle variable shapes rather than simple square has recently been proposed by Dabov et al. (2009) as a way to improve the BM3D algorithm. The authors propose to adapt the shapes used by the algorithm: they locally select a shape by applying Lepski's method, and then perform the same steps of the BM3D algorithm with these shapes rather than with common square blocks (i.e., patches).

Recently, spatial-adaptive methods for selecting the parameters of NL means have been proposed in (Doré and Cheriet, 2009; Duval et al., 2011). Both methods locally select the parameters which minimize a local estimate of the risk (i.e., the Mean Square Error, MSE) by considering respectively Stein's Unbiased Risk Estimate (SURE) (Stein, 1973, 1981) or the C_p criterion (Mallows, 1973). The use of SURE for NL means was originally proposed in order to select the best bandwidth parameter (Van De Ville and Kocher, 2009). Applications of SURE emerged for choosing the smoothing parameter in families of linear estimates (Li, 1985) such as for model selection, ridge regression, smoothing splines, etc. It was then widely used in the wavelet community after the introduction of the SURE-Shrink algorithm (Donoho and Johnstone, 1995). Solo (1996) gave a general form of SURE for an estimator defined as a minimizer of a regular energy, especially for least square regression regularized by a Sobolev norm or the Total Variation. More recently, linear combinations of estimates based on SURE were considered (Blu and Luisier, 2007) instead of the selection of a single one. Moreover, Ramani et al. (2008) have described a

Figure 7.1: Illustration of the *noise halo* appearing around edges with high contrast on images denoised by NL means. The input noisy images was corrupted version of the noise-free images presented on Fig. 7.6 damaged by AWGN with standard deviation $\sigma = 20$. Noise halo arises from an abrupt lack of redundancy around edges sometimes referred as the *rare patch effect*.

Monte Carlo approach to evaluate SURE when a closed-form expression is not available or too computer-intensive. The proposed approach is in the same vein as (Van De Ville and Kocher, 2009; Duval et al., 2011) and proposes to locally select or aggregate the best shapes.

Our contributions— The aim of this chapter is to highlight the potential benefit of replacing the simple square patches with more general shapes, in the classical NL means filter. To this end, we propose a fast algorithm, based on the fast Fourier transform, which allows to compute the solution of the NL means for arbitrary patch shapes. Then we explain how to combine the estimators associated with each shape in a suitable way. We select or combine locally the shape-based estimates by measuring the performance of their associated denoisers with SURE. We coin such type of algorithms Non-Local Means with Shape-Adaptive Patches (NLM-SAP).

The main advantage of using adaptive patch shapes in the context of the NL means is to reduce the *noise halo* produced around edges with high contrast. Our method is an improvement of the NL means taking into account the anisotropy of natural images. It is all the more relevant when the images to denoise present edges with high contrast for which the classical NL means fails. The difference with (Dabov et al., 2009) is that we use a previously fixed family of shapes instead of learning them while processing the image.

Organization of the chapter— We introduce in Sec. 7.1 a more general framework using general shapes instead of square patches to measure the similarity between pixels. This leads in Sec. 7.2 to the natural problem of locally selecting or combining the best shapes in our NLM-SAP algorithm. Section 7.3 illustrates numerically, and above all visually, the gain in aggregating various shape-based estimates in a proper manner.

7.1 From patches to shapes: beyond the *rare patch effect*

In practice, we have seen that the original algorithm suffers from a *noise halo* around edges, due to an abrupt lack of redundancy of the image. This phenomenon is sometimes referred to as the *rare patch effect* (Fig. 7.1). Statistically, it leads to an NL means estimator with large variance around edges. Several solutions have already been proposed to handle this drawback (Salmon and Strozecki, 2010; Duval et al., 2011; Louchet and Moisan, 2011). Here, we generalize these approaches by considering general shapes rather than simple square patches (cf. Fig. 7.2).

In order to deal with patches of arbitrary shapes, we reformulate the way the distance between two pixels is measured in terms of patches. The weighted Euclidean distance used in the NL

means (see Sec. 2.5) can be generalized for any patch shape by using the following expression:

$$\hat{u}(x) = \frac{\sum_{x'} w(x,x')v(x')}{\sum_{x'} w(x,x')} \tag{7.1}$$

$$\text{where} \quad w(x,x') = \varphi\left(\frac{d_S^2(v(\mathrm{P}_{x'}), v(\mathrm{P}_x))}{2|P|h^2}\right) \tag{7.2}$$

$$\text{and} \quad d_S^2(v(\mathrm{P}_{x'}), v(\mathrm{P}_x)) = \sum_{\tau\in\Omega} S(\tau)\left(v(x+\tau) - v(x'+\tau)\right)^2, \tag{7.3}$$

where S encodes the shape of a patch P. With this notation we can easily rewrite the original NL means with a simple S by choosing:

$$S(\tau) = \begin{cases} 1, & \text{if } \|\tau\|_\infty \leq \frac{p-1}{2}, \\ \\ 0, & \text{otherwise,} \end{cases} \tag{7.4}$$

where p is the half-width of the square patch.

One of our contributions is to provide an efficient algorithm, based on the Fast Fourier Transform (FFT), to compute the distances in Eq. (7.3). Our implementation is independent of the shape, and can thus be applied with different shapes. As soon as we consider the use of anisotropic shapes, and not just squares or disks centered on the pixel of interest, two questions emerge. The first one is how to choose the collection of shapes to consider. The second issue is to propose a way to combine the estimators provided by each shape.

The collection of shapes should be composed of more than one shape to locally take into account the geometrical properties of natural images. Consider for instance the use of a single vertically elongated patch. This could be interesting to handle vertical features, but eventually we would not optimally deal with horizontal details with only one oriented shape. So, the collection should be diversified and numerous enough to identify directional features (see Fig. 7.8 for a visual illustration). At the same time, it should remain small enough so that the algorithm is not computationally intensive.

7.1.1 Fast algorithm to handle shapes

In this section, we present a fast way to compute the NL means weights for general shapes, based on the 2D-FFT. It is inspired from works initiated in (Wang et al., 2006) and (Darbon et al., 2008) to speed up the NL means algorithm. However, contrary to these approaches, ours can deal with non-square and/or non-binary patches, i.e., with general shapes S. Like them, our method is independent of the shape size. Let us also mention that other fast implementations of the NL means have been proposed in (Mahmoudi and Sapiro, 2005; Bilcu and Vehvilainen, 2008; Pang et al., 2009): such methods use a pre-selection of the patches based either on statistical tests or gradients comparisons. We should however emphasize that the final estimates with those approaches are approximate solutions of the original NL means. Our method does not rely on such tricks and computes the exact NL means in the case of a square shape in Eq. (7.3).

Wang et al. (2006) and Darbon et al. (2008) propose to compute the Euclidean distances using “Summed Area Tables" introduced by Crow (1984) (also called “Integral Images" by Viola and Jones (2001)). This allows them to reduce the computational cost of the NL means from $O(|W|\cdot|\Omega|\cdot|\mathrm{P}|)$ to $O(|W|\cdot|\Omega|)$, where $|W|$ is the number of pixels in the search window, $|\Omega|$ is the image domain size and $|\mathrm{P}|$ is the patch size (we refer to (Darbon et al., 2008) for more details). To compute these integral images, the authors change the original algorithm by swapping the two “for” loops: instead of considering all the shifts for each pixel, they consider all the pixels for each shift.

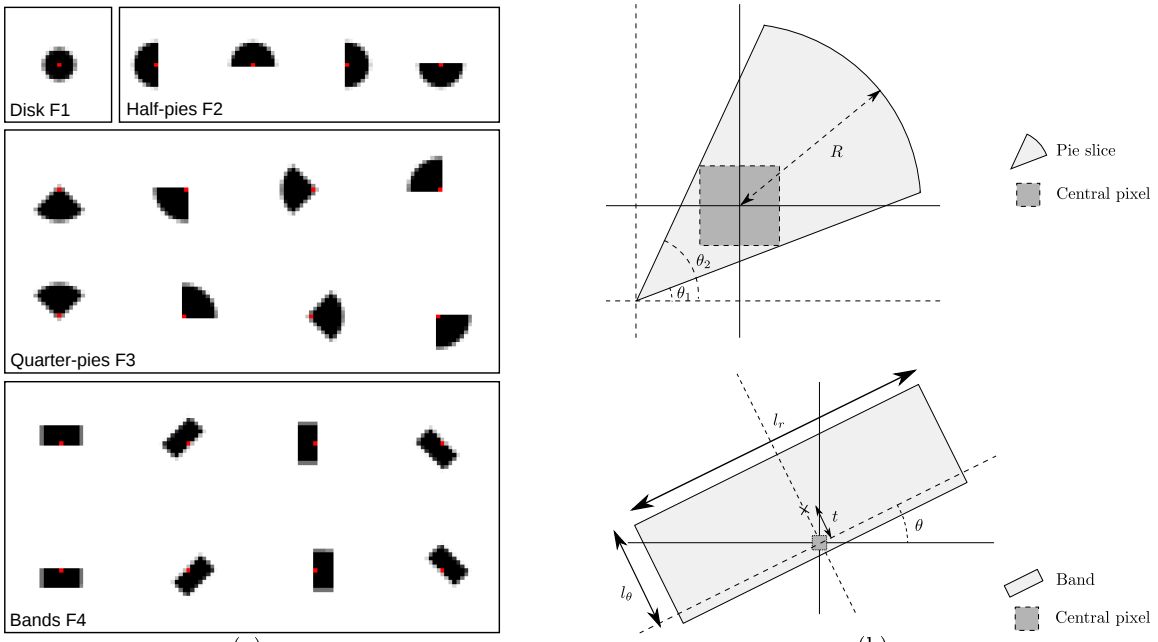


Figure 7.2: (a) Examples of shapes considered. The "central" pixel is shown in red, dark pixels illustrate high weights. Shapes are grouped in four categories: F1. the isotropic *disk* family, F2. the *half-pies* family, F3. the *quarter-pies* family and F4. the *bands* family. (b) Parametrization of the *pie* slices and *bands*.

We use basically the same swapping trick. Notice that Eq. (7.3) can be reformulated for any translation parameter δ (i.e., taking $x' = x + \delta$) as a discrete convolution:

$$d_S^2(v(\mathrm{P}_x), v(\mathrm{P}_{x+\delta})) = \sum_{\tau \in \Omega} S(\tau) \left(v(x+\tau) - v(x+\delta+\tau)\right)^2 = (\check{S} \star \Delta_\delta)(x)\,, \tag{7.5}$$

where $\check{S}(\tau) = S(-\tau)$, $\Delta_\delta(x) = (v(x) - v(x+\delta))^2$ and $\star$ is the convolution operator. This term can be interpreted as the correlation between the shape S and the square difference of the observed image and the δ-shifted version. The convolution $\check{S} \star \Delta_\delta$ can be computed quickly thanks to the following relation:

$$\check{S} \star \Delta_\delta = \mathcal{F}^{-1}(\mathcal{F}(\check{S})\mathcal{F}(\Delta_\delta)) = \mathcal{F}^{-1}(\mathcal{F}(S)^*\mathcal{F}(\Delta_\delta))\,, \tag{7.6}$$

where $\mathcal{F}$ is the 2D discrete Fourier transform (2D-FFT) and $\mathcal{F}^{-1}$ is its inverse transform. According to Eq. (7.6), and given a translation δ, we only need to perform one term by term multiplication in Fourier domain and two 2D-FFT (note that $\mathcal{F}(S)$ can be computed off-line). The repetition of this procedure for every translation δ covering the search window, leads to an algorithm (whose pseudo-code is detailed in Fig. 7.3) with a complexity of $O(|W| \cdot |\Omega| \cdot \log(|\Omega|))$. A similar algorithm using recursive filters was recently described by Condat (2010).

7.1.2 Families of shapes

The main purpose of this chapter is to show that the use of different shapes allows to reduce the *rare patch effect*. This point of view is a generalization of the NL means based on square patches with the reprojection studied by Salmon and Strozecki (2010), since each translated patch can be regarded as a decentered shape. Here, h is fixed and the challenge is to find shapes with enough similar candidates in the search window to reduce the noise.

Algorithm 2D-FFT NL means for an arbitrary shape

Inputs: noisy image $\boldsymbol{v}$, 2D-FFT of the shape $\mathcal{F}(S)$
Parameters: search window W, bandwidth h
Output: estimated image $\hat{\boldsymbol{u}}$

Initialize accumulator images A and B to zero
for all shift vector δ in the search window W **do**
Compute the square difference image Δ_δ

$$\Delta_\delta(x) := (v(x) - v(x+\delta))^2 \text{ for all pixels } x \text{ in } \Omega$$

Compute the 2D-FFT $\mathcal{F}(\Delta_\delta)$
Perform the convolution of Δ_δ by the shape $\check{S}$

$$d_S^2(v(\mathrm{P}.), v(\mathrm{P}._{+\delta})) \leftarrow \left(\mathcal{F}^{-1}\left(\mathcal{F}(S)^* \mathcal{F}(\Delta_\delta)\right)\right)(\cdot)$$

▷ $O(|\Omega| \cdot \log|\Omega|)$ operations using 2D-FFT

for all pixels x in Ω **do**
Compute the weights

$$w(x, x+\delta) = \varphi\left(\frac{d_S^2(v(\mathrm{P}_x), v(\mathrm{P}_{x+\delta}))}{2h^2}\right)$$

Update the accumulators

$$A(x) \leftarrow A(x) + w(x, x+\delta) v(x+\delta)$$
$$B(x) \leftarrow B(x) + w(x, x+\delta)$$

end for
end for
Final (normalized) estimator $\hat{u}(x) = \frac{A(x)}{B(x)}$ for all pixel x
Note: the central pixel ($\delta = 0$) is treated as a special case

▷ see Sec. 7.3 for details

Figure 7.3: NL means pseudo-code for an arbitrary patch shape S. Pre-computations (based on 2D-FFT) of distances between shapes from the noisy image and shapes from its shifted version leads to a smaller complexity of $O(|W| \cdot |\Omega| \cdot \log|\Omega|)$, independent of the shape S.

We now present several types of families we have considered. The first collections consist of classical squares and disks shapes. Then, we propose more directional shapes such as *pie slices* and *bands* displayed in Fig. 7.2.

Squares: To begin with, we apply our framework to the most commonly used shapes, i.e., the square shapes of odd length (so the squares have centers we can consider). For instance, choosing:

$$S(\tau) = \begin{cases} 1, & \text{if } \|\tau\|_\infty \leq \frac{p-1}{2}, \\ 0, & \text{otherwise,} \end{cases} \tag{7.7}$$

leads to the classical (simplified) NL means definition with square patches of size $p \times p$ and distance between patches measured by the Euclidean norm.

Gaussian: The original, but less common choice, is to set:

$$S(\tau) = \begin{cases} \exp(-(\tau_1^2 + \tau_2^2)/2a^2), & \text{if } \|\tau\|_\infty \leq \frac{p-1}{2}, \\ 0, & \text{otherwise.} \end{cases} \tag{7.8}$$

Equation (7.8) means that the norm $\|\cdot\|_{2,a}$ is used to measure the distance between patches. This limits the influence of square patches corners and leads to a more isotropic comparison between patches.

Disks: Disk shapes are defined in the same way, using the Euclidean norm instead:

$$S(\tau) = \begin{cases} 1, & \text{if } \|\tau\|_2 \leq \frac{p-1}{2}, \\ 0, & \text{otherwise.} \end{cases} \tag{7.9}$$

A non-binary version may also be defined for pixels crossed by the boundary.

Pie slices: We study a family of shapes, denoted as "pie", whose elements are defined with three parameters: two angles and a radius. These shapes represent a portion of a disk delimited by two lines and surrounding the discrete central pixel.

Bands: This family of shapes is simply composed of rectangles, potentially rotated and decentered with respect to the pixel of interest.

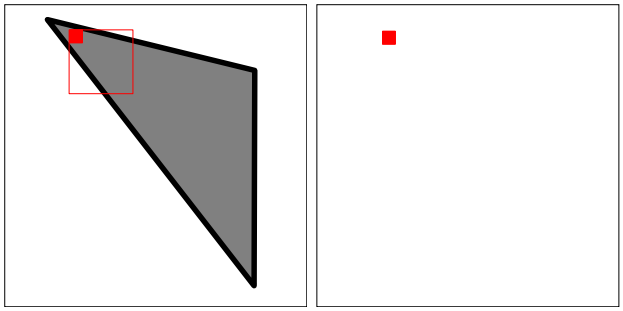
7.1.3 Connection with previous work

One of our main concern is to address the *rare patch effect* of the NL means algorithm. Different methods have been designed to limit this drawback or to improve the NL means in terms of quality or speed. The first attempt was proposed to speed-up the algorithm. The idea is to denoise patch by patch rather than pixel by pixel. Taking into account patches overlaps, a fast implementation of the NL means is reached by using a sub-sampled grid of pixels (Buades et al., 2005; Kervrann et al., 2007). Quality improvement can also be obtained by properly using overlapping patches. Indeed, we get |P| estimates for each pixel (where |P| is the number of elements in a patch). Some authors (Buades et al., 2005; Kervrann and Boulanger, 2006) simply propose to uniformly average those |P| estimates while a weighted average is performed in (Salmon and Strozecki, 2010). In our framework, these blockwise approaches are equivalent to use |P|, possibly decentered, square shapes (cf. Fig. 7.4).

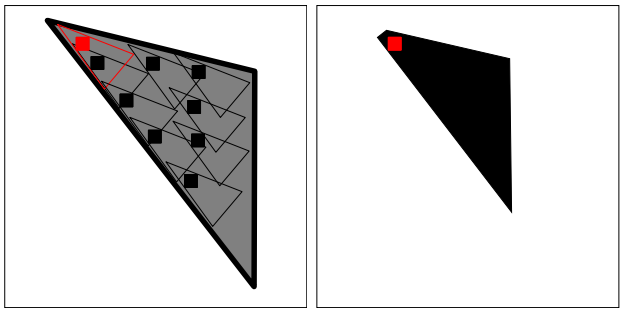
Other methods have been introduced to reduce the *noise halo*. Louchet and Moisan (2011) use a total variation-based pre-filtering of the image and set locally its parameter so that the NL means find enough similar patches. Duval et al. (2011) aim to select locally the bandwidth parameter h using SURE (introduced in details in Sec. 7.2.1) to select enough patches according to a bias-variance trade-off.

7.2 Aggregation of shape-based estimates

In this section, we investigate a way to aggregate the NL means estimators based on different shapes of "patches". We have extended the standard square shape to other shapes such as *disks*,



(a) Square patch and its neighborhood



(b) Suitable shape and its neighborhood

Figure 7.4: Examples of neighborhood associated with a square patch or a suitable shape. If the patch is square (a) fewer similar patches candidates are found than with the shape (b). The pixel of interest is in red and the selected pixels obtained by the two methods to denoise the red pixel are in black.

pies or *bands* (see Sec. 7.1.2). Thus, the new goal in this context is to determine how to locally take the most of each proposed denoiser.

For any pixel x in the image, we have built a collection of K pre-estimates $\overline{u}_1(x), \cdots, \overline{u}_K(x)$ based on the different shapes. We first suggest using the weighted variance minimization (WAV) presented in Sec. 2.6.1. The application of the WAV methodology of Salmon and Strozecki (2010) to general shapes define the weights as inversely proportional to the (approximate) variance of the corresponding estimator. In the context of the NL means, this approximate variance can be obtained in closed-form in the same way as in (Kervrann and Boulanger, 2006), assuming that the coefficients $w(x, x')$ in Eq. (7.1) can be treated as deterministic. Measuring the performance of the estimators in term of variance is well justified since the halo effect results in the high variance of our estimators around the edges (see Salmon and Strozecki, 2010). However, it tends to over-smooth the edges and the thin details since it does not consider the bias of each estimator.

7.2.1 SURE and the exponential weighted aggregation

A way to take the bias into account is to consider a risk estimate rather than the variance to locally attribute more weight to the estimators with small risks. We suggest to estimate this risk by using the Stein unbiased risk estimator (SURE) as presented in Sec. 2.6.4. Van De Ville and Kocher (2009) give a closed-form expression of SURE for the NL means. They aim at selecting globally the best bandwidth for a given image. Here, our approach is different, despite the use of the same tool. Indeed, our choice of shape is done locally (i.e., for each pixel), since it is very unlikely that a single shape should be optimal for a whole natural

image. Our method is closer to the one proposed by Duval et al. (2011). The authors rely on SURE to locally determine the parameters: the bandwidth h and the patch size p. They have shown that a local choice of h reduces the visual artifacts, especially the *rare patch effect*. Let us now rephrase Stein's Lemma (Stein, 1973, 1981) in our NLM-SAP framework.

SURE provides an estimate of the risk for the K shape-based denoised values $\overline{u}_k(x)$ at each pixel x defined by:

$$\hat{R}[\overline{u}_k(x)] = (\overline{u}_k(x) - v(x))^2 + 2\sigma^2 \frac{\partial \overline{u}_k(x)}{\partial v(x)} - \sigma^2 \,. \tag{7.10}$$

As seen in Sec. 2.6.4, Eq. (7.10) is unbiased: $\mathbb{E}(\hat{R}[\overline{u}_k(x)]) = \mathbb{E}|\overline{u}_k(x) - u(x)|^2$. The main contribution of (Van De Ville and Kocher, 2009; Duval et al., 2011) is that they give a closed-form expression of $\frac{\partial \overline{u}_k(x)}{\partial v(x)}$ for NL means. Indeed, thanks to Eq. (7.1), its expression can be recast in the following form:

$$\frac{\partial \overline{u}_k(x)}{\partial v(x)} = \frac{\varphi(0)}{C_x} + \frac{1}{C_x}\sum_{x'} v(x') \frac{\partial w(x,x')}{\partial v(x)} - \left(\frac{1}{C_x}\sum_{x'} v(x') w(x,x')\right)\left(\frac{1}{C_x}\sum_{x''} \frac{\partial w(x,x'')}{\partial v(x)}\right) \tag{7.11}$$

where $C_x = \sum_{x'} w(x,x')$ is a normalization constant. In our NLM-SAP framework, our shape-based norm defined in Eq. (7.3) leads to the following expression of the derivative of the weights $w(x,x')$:

$$\frac{\partial w(x,x')}{\partial v(x)} = \frac{1}{h^2}\varphi'\left(\frac{d_S^2(v(\mathrm{P}_x), v(\mathrm{P}_{x'}))}{2h^2}\right)\left(S\left(0\right)\left[v(x) - v(x')\right] + S\left(x - x'\right)\left[v(x) - v(2x - x')\right]\right) \tag{7.12}$$

where S encodes the shape of our k-th shape-based estimator. Finally, combining equations (7.10), (7.11) and (7.12) leads to an unbiased estimate of the risk of our NLM-SAP denoiser.

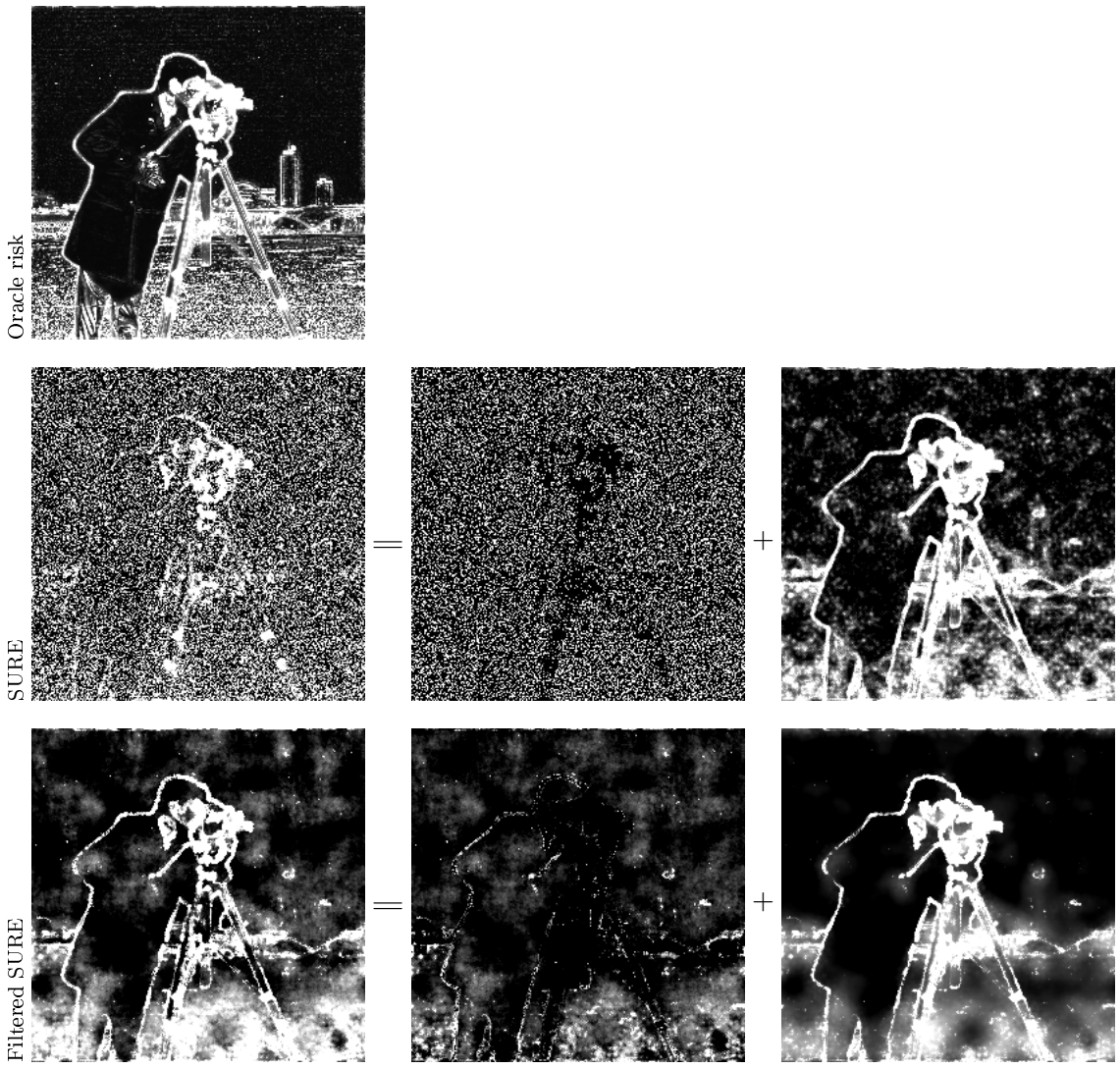
Given the risk associated to each estimator, we suggest to use the exponential weighted aggregation (EWA) presented in Sec. 2.6.5. In the same spirit as the WAV approach, it considers that it might be better to combine several estimates rather than just selecting one. The aggregating of the estimators is performed by a weighted average with weights based on the confidence attributed to each estimator and measured in term of the risk:

$$\hat{u}(x) = \sum_{k=1}^{K} \beta_k v_k(x) \,, \tag{7.13}$$

$$\text{with} \qquad \beta_k = \frac{\exp(-\hat{R}[\overline{u}_k(x)]/T)}{\sum_{l=1}^{K} \exp(-\hat{R}[\overline{u}_l(x)](x)/T)} \,.$$

The temperature parameter $T > 0$ is a smoothing parameter, that controls the confidence attributed to the risk estimates. If $T \to \infty$, then the EWA is simply the average of the pre-estimate. Conversely, when $T \to 0$, then EWA selects the pre-estimate minimizing the risk as discussed before. Most theoretical work about EWA (see Leung and Barron, 2006; Dalalyan and Tsybakov, 2008) recommend a large temperature parameter $T = 4\sigma^2$ under a few assumptions (like independence) on the estimators $\overline{u}_1(x), \cdots, \overline{u}_K(x)$. In practice, since assumptions on the estimators family may not be satisfied, a smaller value is used, such as $T = 0.4\sigma^2$.

As discussed in Sec. 2.6.4, a remaining problem with SURE is that it has an extremely large local variance. To take a local decision for each pixel x is difficult since this estimator has large oscillations (see for instance Fig. 7.5). In the next paragraph, we present how to regularize the risk maps, i.e., the "images" $\hat{r}_k = \hat{R}[\overline{\boldsymbol{u}}_k]$.



(a) Risk (b) Method noise (c) Divergence

Figure 7.5: Maps of the risk associated to a circular shape: (first line) the oracle risk map, (second line) the SURE map without regularization and (third line) the SURE map with Yaroslavsky regularization. Second and third rows correspond to the decomposition (7.10).

7.2.2 Regularizing the risk maps

In practice, the estimation of the risk given by (7.10) is too noisy to guide a local choice of the shape (see Fig. 7.5). To make it more robust, it is necessary to locally regularize the risk maps, so as to approximate at each pixel the expectations used in Eq. (7.10): our aim is to find estimates $\hat{r}_k(x)$ close to the true risks $\mathbb{E}[(\overline{u}_k(x) - u(x))^2]$. These true risks will be referred to as the oracle risks since in our simulations we will compute them using the true image.

In (Duval et al., 2011), the convolution of the risk map is an efficient way to estimate the local risk in view of setting h since on both sides of an edge a large value of h should be used. Here, the anisotropy of the shapes implies that on one side of an edge the risk may be low whereas it may be very high on the other side. Since the convolution diffuses the risks on both sides of the edges, any comparison of the risks associated with each shape becomes unstable.

In order to average the risks on each side of edges separately, we have adopted a variant of

(a) *Cameraman* (b) *City* (c) *Lake* (d) *Windmill*

Figure 7.6: Chosen 256×256 noise-free images for our experiments, from left to right: *cameraman*, *city*, *lake* and *windmill*. These images present edges with high contrast for which the classical NL means suffer from the *rare patch effect*.

	Cameraman	City	Lake	Windmill
Noisy input image	22.13/0.400	22.13/0.567	22.13/0.456	22.13/0.385
Patch shape area of 12.5 px^2	**29.59**/0.822	28.11/0.873	28.68/0.849	30.91/0.879
Patch shape area of 25 px^2	29.38/0.828	27.94/0.880	28.46/0.855	30.72/0.895
Patch shape area of 50 px^2	29.06/0.825	27.59/0.879	28.33/0.857	30.35/0.899
Combination of these three scales	29.58/**0.844**	**28.32**/**0.897**	**28.93**/**0.878**	**31.08**/**0.912**

Table 7.1: Gain by using multi-scale isotropic shapes in terms of PSNR and SSIM values (PSNR/SSIM). Circular patch shapes respectively with area of 12.5, 25 and 50 px^2 are used and their results are compared to the one obtained when using a combination of these three scales of patches.

the Yaroslavsky filter (Yaroslavsky, 1985). Up to a constant, the estimator of the risk (7.10) can be decomposed in two terms: the square of the method noise (Buades et al., 2005), and a divergence term $D_k(x) = 2\sigma^2 \frac{\partial \hat{u}_k(x)}{\partial v(x)}$. This divergence term has little variance compared to the noisy image and to the method noise (see Fig. 7.5) and contains all the needed information to guide the averaging process: it is uniformly high in the regions where the halo effect is likely to appear, whereas it is low in smooth regions.

As a consequence, better results are obtained by guiding the Yaroslavsky filter with the self-similarity of the divergence instead of the risk itself:

$$\hat{r}_k(x) = \frac{1}{C(x)} \sum_{x'} \mathbb{1}_{\{|D_k(x) - D_k(x')| \leq \kappa\}} \, \hat{r}_k(x'), \tag{7.14}$$

where the sum is taken over a small neighborhood of x, $C(x)$ is a normalizing constant and κ is a bandwidth parameter. Fig. 7.5 shows that this regularization procedure provides smooth risk maps, following edges of the underlying noise-free image, better than without regularization. We have displayed the oracle risk map and the estimated risk map provided by SURE and the Yaroslavsky regularization. For illustration purpose, we also show the decomposition of SURE as the sum of the square of the method noise and the divergence term.

Other attempts to regularize the risk map were performed (median filter, variants of Perona-Malik diffusion and NL means). Yaroslavsky regularization provides the best trade-off between computing time, visual and numerical results, and we have thus adopted this approach.

7.3 Numerical and Visual Results

This section presents quantitative and qualitative results obtained on four images synthetically corrupted by AWGN. Unless otherwise specified, the corrupted images are obtained from four 256×256 noise-free images presented on Fig. 7.6: the famous *cameraman* image and *city*, *lake* and *windmill*[1]. These images are particularly interesting in the study of our proposed Non-Local Means with Adaptive Patch Shapes (NLM-SAP) since they present edges with high contrast for which the classical NL means suffer from the *rare patch effect* (see Fig. 7.1).

In all the experiments, **unless otherwise specified**, the NLM-SAP is used with the following default parameters:

- search window: width $\ell = 11$ px,
- shape family: 15 shapes from families F1 and F2 on Fig. 7.2.a with shape areas of 12.5, 25 and 50 px^2,
- aggregation: EWA with $T = 0.4\sigma^2$,
- risk regularization: Yaroslavsky regularization with search window of size 11×11 px^2 and κ is proportional to the estimated standard deviation of the divergence map.

As soon as we consider shapes of different areas, the parameter h has to adapt to the sizes of the shapes. Since we use a trapezoidal kernel, we set the bandwidth parameter to $h^2 = 2\sqrt{8\sigma^4/|S|}$ as suggested in (Buades et al., 2009) where $|S|$ is the equivalent size of the shape $|S| = \frac{(\sum S(\tau))^2}{\sum S(\tau)^2}$.

The main limitation in computing time is due to the number K of shapes required by our NLM-SAP algorithm. We need to perform K times an NL means like algorithm, i.e., one for each shape. Thanks to our FFT acceleration, the computing time required for one shape, whatever the shape, is of about $2s$ for a 256×256 image with a Matlab implementation on an Intel Pentium 64-bit, 3.00 GHz. By comparison, the naive Matlab implementation of NL means takes about $100s$, for square patches of area 7×7 px^2. The computation of one local SURE map, using Yaroslavsky filtering, takes about $0.2s$ per shape. Finally, NLM-SAP[2] using 15 shapes leads to a computing time of about $32s$ which is still less than the naive Matlab implementation of NL means.

Visual results are given to assess the denoising qualities relative to the different settings of NLM-SAP and to compare NLM-SAP with other denoising approaches (see Fig. 7.12). Numerical criteria support our claims: the Peak Signal to Noise Ratio (PSNR) and the Structural SIMilarity (SSIM) defined by Wang et al. (2004) (see Sec. 2.1.3).

7.3.1 Behavior of NLM-SAP

In this section, we study the behavior of NLM-SAP according to some parameters such as the type of family, the type of aggregation and the type of risk regularization. Each noisy image is corrupted by AWGN with standard deviation $\sigma = 20$.

Table 7.1 and Figure 7.7 illustrate the gain of performance to use multi-scale patch shapes instead of using only one fixed size. In this experiment, we consider three circular shapes of areas: 12.5, 25 and 50 px^2. Comparatively, for the original version of NL means, the authors suggest to use square patches of fixed size $7 \times 7 = 49$ px^2. Surprisingly, using the smallest shapes provides always the best PSNR. It means that the bias and the *noise halos* introduced by using large patches are actually more penalizing than the remaining noise left by the use of small patches.

[1] three sub-images extracted from Laurent Condat's database: http://www.greyc.ensicaen.fr/~lcondat/

[2] our Matlab implementation is available online

(a) Area 12.5 px^2 (b) Area 50 px^2 (c) Combination

Figure 7.7: Results obtained with circular shapes of different scales (12.5 px^2, 50 px^2, combination of 12.5, 25 and 50 px^2). The smallest patch size provide best PSNR but has more artifacts in smooth regions, whereas larger ones suffer from the noise halo. Combination of scales limits those two issues.

	Cameraman	City	Lake	Windmill
Noisy input image	22.13/0.400	22.13/0.567	22.13/0.385	22.13/0.456
Disk shapes (family F1)	29.58/**0.844**	28.32/**0.897**	28.93/**0.878**	31.08/0.912
Half-pie shapes (family F2)	**29.72**/0.843	28.48/0.896	29.00/0.877	31.29/0.912
Quarter-pie shapes (family F3)	29.64/0.842	28.27/0.891	28.89/0.875	31.24/0.912
Band shapes (family F4)	29.72/0.841	28.45/0.894	28.98/0.875	31.36/0.912
Combination of F1 and F2	29.74/**0.844**	**28.53**/**0.897**	**29.04**/**0.878**	31.32/**0.913**
Combination of F1, F2, F3 and F4	**29.75**/0.842	28.49/0.895	29.02/0.876	**31.40**/**0.913**

Table 7.2: Gain by using anisotropic or mixture of isotropic and anisotropic shapes in terms of PSNR and SSIM values (PSNR/SSIM). The studied patch shapes are the isotropic *disks*, the *half-pies*, the *quarter-pies*, the *bands* and some combination of them (see Fig. 7.2.a).

	Cameraman	City	Lake	Windmill
Noisy input image	22.13/0.400	22.13/0.567	22.13/0.385	22.13/0.456
Weighted Avg. based on Var. (WAV)	29.64/0.841	28.15/0.887	28.69/0.868	31.10/0.910
Exp. Weighted Average (EWA)	**29.74**/**0.844**	**28.53**/**0.897**	**29.04**/**0.878**	**31.32**/**0.912**

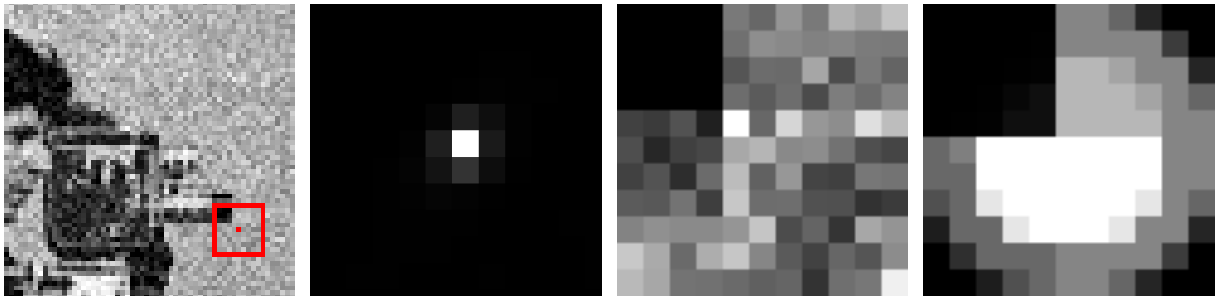
Table 7.3: Comparisons of different aggregation procedures in terms of PSNR and SSIM values (PSNR/SSIM). The compared aggregation types are WAV and EWA.

The aggregation of these three scales of shapes with our NLM-SAP methods improves slightly the PSNR. Visually speaking, using only small isotropic patches already decreases the *rare patch effect* while using too large isotropic shapes produces a strong *noise halo* effect. However, the level of noise is much more decreased by using large shapes than small ones in homogeneous areas. Finally, combining different scales of isotropic patch shapes leads to a diminution of both the level of noise and the halo effect. This is well reflected in Tab. 7.1 by the gain in term of the SSIM criterion which provides quality measurements closer to our perception system. We will see in the following that the results can be further improved by considering both multi-scale and anisotropic patch shapes.

Table 7.2 gives numerical results obtained by using different families. The compared families are the ones presented on Fig. 7.2.a, i.e., the *disks*, the *half-pies*, the *quarter-pies* and the *bands*. Combination of these families are also studied. Our experiments show that most suitable shape

Figure 7.8: Eight denoised images obtained for different oriented patch shapes. The proposed final aggregate is in the center. Each denoiser provides good performance in a specific target direction but suffers from *noise halos* in the other directions. The final (central) aggregate takes advantage of every oriented-denoiser to provide high quality restored edges. The patch shape used is indicated in white.

families, both in terms of PSNR and SSIM, have to contain isotropic shapes, directional shapes and various scales of shapes. Increasing the number of shapes does not necessarily improve the denoising quality. Using 15 shapes from families F1 and F2 with the three different scales, seems to be a good trade-off between computing time and denoising quality. Figure 7.8 illustrates why using directional shapes is important to reduce the *rare patch effect*. Indeed, each oriented patch shape enables the restoration of edges in the target direction but is inappropriate in the other directions. Then a fine aggregation of them leads to high quality restoration of edges in all directions. Figure 7.9 displays weight maps induced by using patches with only one fixed square shape (i.e., NL means) compared to patches with adaptive scales and orientations (i.e., NLM-SAP). For NL means, all the weights are concentrated around the target pixel: this is the *rare patch effect*. For NLM-SAP, the weights are more spread, and other similar pixels are detected thanks to multi-scale and anisotropic patch shapes. It is clear that the limitation of the



(a) Target pixel (b) Neighborhood for NLM (c) Neigh. for NLM-SAP (d) Avg. shape

Figure 7.9: Illustration of the diminution of the *rare patch effect*. (a) The noisy image with an highlighted target pixel and its neighborhood. (b) The associated maps of weights obtained by using only square patches of fixed size (i.e., NL means). (c) The associated maps of weights with multi-scale and anisotropic patch shapes (i.e., NLM-SAP with F1 and F2). (d) Weighted average of the shapes combined by NLM-SAP.

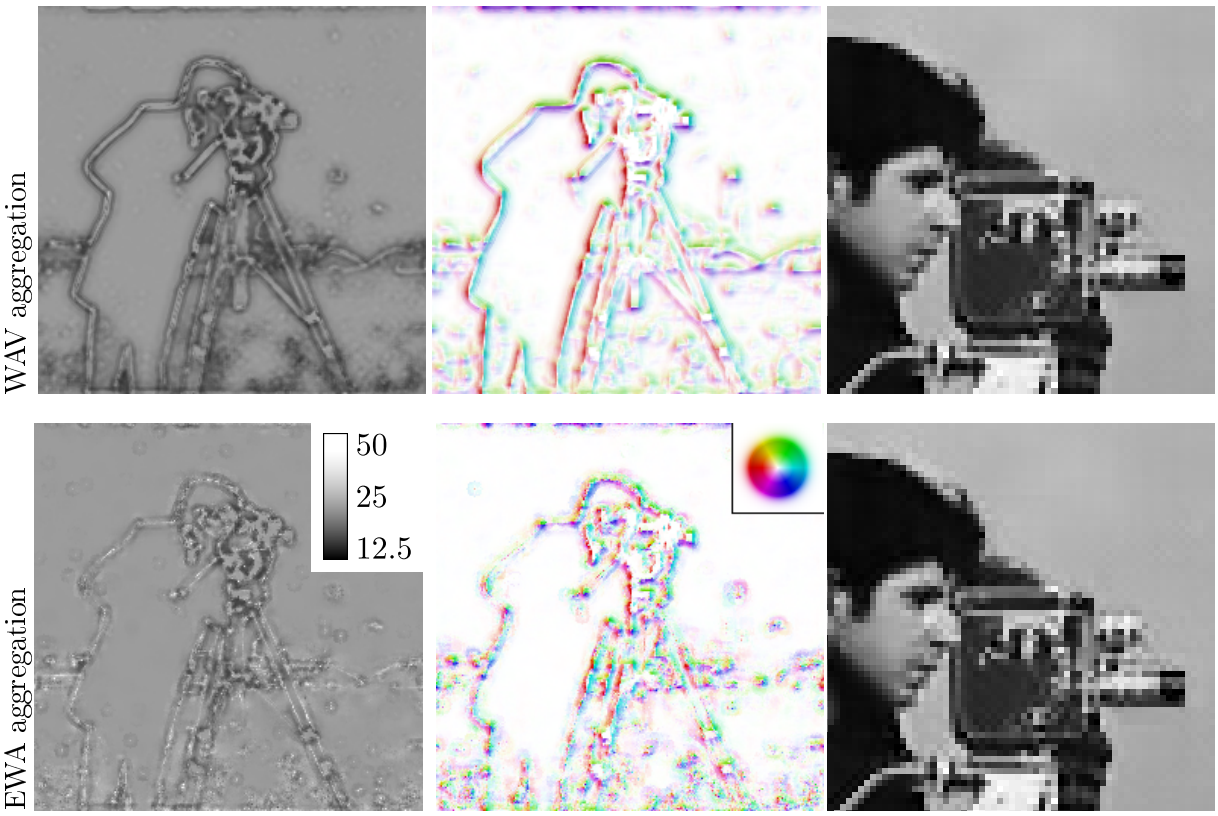
	Cameraman	City	Lake	Windmill
Noisy input image	22.13/0.400	22.13/0.567	22.13/0.456	22.13/0.385
Noisy risk maps (SURE maps)	29.13/0.817	27.41/0.865	28.38/0.846	30.15/0.872
Convolved risk maps	29.71/**0.845**	28.49/0.898	**29.13**/0.881	31.26/**0.912**
Yaroslavsky risk maps	**29.74**/0.844	**28.53**/**0.897**	29.04/0.878	**31.32**/**0.912**
True risk maps	32.09/0.880	32.31/0.938	32.27/0.922	34.43/0.935

Table 7.4: Comparisons of regularization procedures of the risk maps in terms of PSNR and SSIM values (PSNR/SSIM). The compared regularization procedures are the ones using the noisy risk maps directly (i.e., SURE maps), the convolved risk maps and the risk maps obtained by Yaroslavsky regularization.

rare patch effect around edges with high contrast leads to a good reduction of the *noise halo*.

Table 7.3 presents the numerical performance associated with the WAV and the EWA aggregation procedures. As expected, EWA provides best results, in terms of PSNR and SSIM, since compared to the other three it combines estimates with the best bias-variance trade-off. The local behaviors of NLM-SAP for WAV and EWA are presented on Fig. 7.10. The average areas and the average orientations of the selected shapes are given for the *cameraman* image. It summarizes for all pixels the information of the average shape as given in Fig. 7.9.d. In smooth regions, anisotropic shapes are not necessarily worse than isotropic ones (like disks or squares). In fact all shapes with the same size should perform equally: weights provided by WAV and EWA are close to uniform distributions. The selected patch shapes clearly adapt to the local scale and orientation of the image geometry. The chosen sizes of the shapes are smaller around edges and textured areas than in homogeneous areas. The chosen orientation follows the orthogonal orientation of the shapes which is consistent with the remarks given in Sec. 7.1.

Finally, we have studied the influence of the regularization of the risk maps on the aggregation results. Three methodologies are compared: aggregation using the noisy risk maps (i.e., SURE maps), the convolved risk maps (using a disk kernel of radius 4) and the risk maps obtained by Yaroslavsky regularization. Table 7.4 gives the corresponding numerical performance and Fig. 7.11 illustrates the behavior of each type of risk map regularization. The risk maps based on convolution and Yaroslavky of SURE maps provide the best results in terms of PSNR and SSIM. However, the choice of the local sizes and orientations of the patch shapes is more relevant in the maps obtained by Yaroslavsky regularization, in terms of scale adaptivity, feature directions and spatial coherency. Using Yaroslavsky filtering, the NLM-SAP acts as expected by selecting



(a) Average areas (b) Average orientations (c) Final aggregation

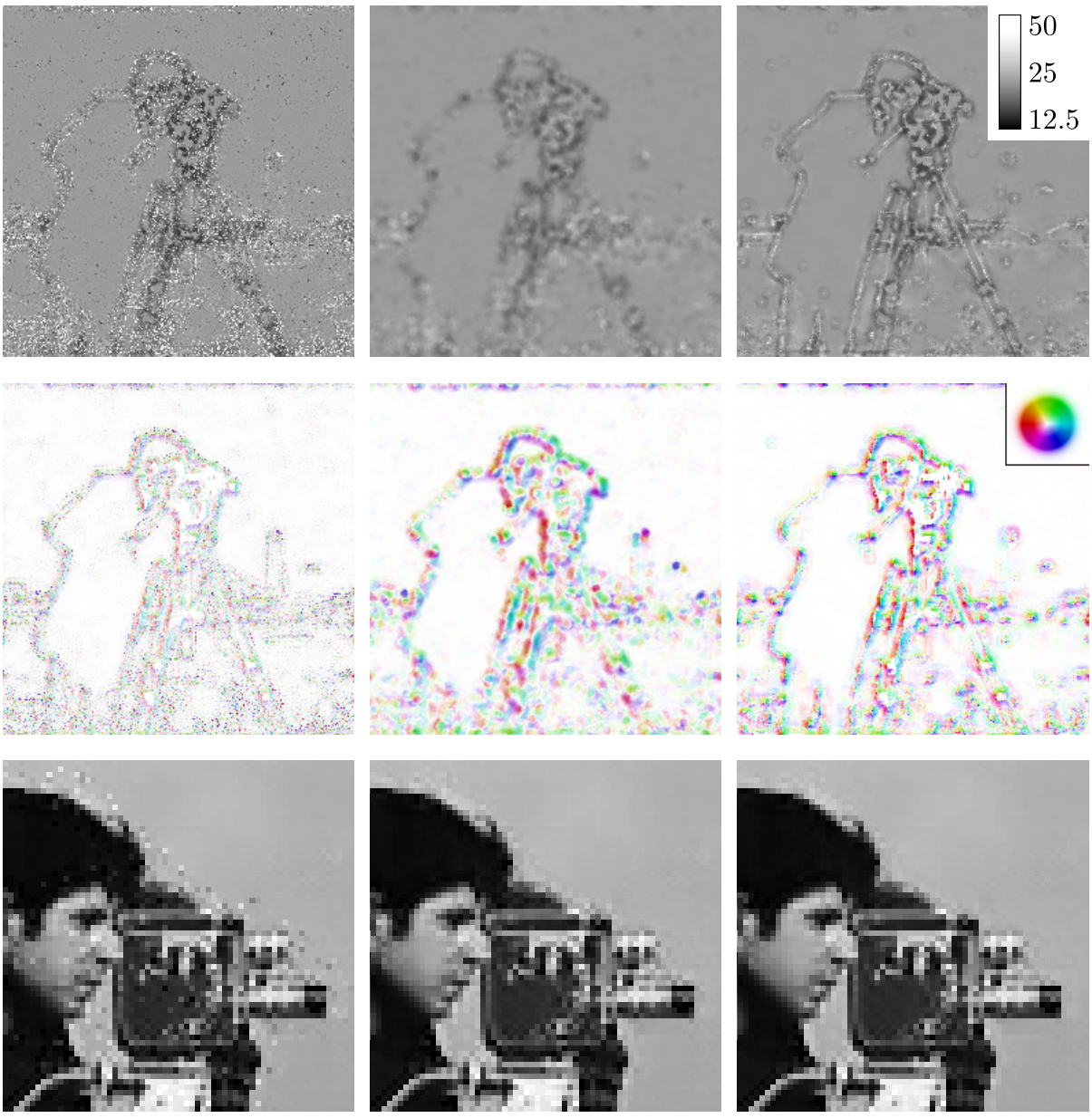
Figure 7.10: (a) Average areas and (b) average orientations of selected shapes for different aggregation procedures on a noisy realization of the *cameraman* image. (c) The final aggegate using (top) the WAV aggregation and (bottom) the EWA aggregations. The average areas and the average orientations are represented using colors whose legends are given on the top right corners.

big sizes of shapes, even around edges, since the shape orientations have been chosen properly to reduce the *rare patch effect*. By comparison, the convolution forces the size of shapes to be small around edges since it cannot select properly the suitable orientations. This slight differences of behaviors can be noticed around the camera (cf. Fig. 7.11). Other regularization strategies have been investigated (median filter, NL means,anisotropic diffusion) but we have not found striking improvements. However, Tab. 7.4 shows that there is still a gap of numerical performance between regularizations of the risk maps and an "oracle risk map" defined as the image of local square errors associated to each shape: $(\overline{u}_k(x) - u(x))^2$.

7.3.2 Comparisons with *state-of-the-art* methods

In this section, the proposed NLM-SAP approach is compared to *state-of-the-art* denoising methods on a large dataset of standard images at different noise levels σ. Comparisons have been performed with the classical (pixelwise) NL means, the blockwise NL means with and without WAV reprojection (with square patches of a single scale) the pixelwise NL means using SURE-based adaptive bandwidth selection (Duval et al., 2011), a refinement of the NL means by Goossens *et al.* (Goossens et al., 2008), the Block-Matching and 3D filtering (BM3D) denoiser (Dabov et al., 2007), and our proposed NLM-SAP approach.

Table 7.5 shows that NLM-SAP brings a gain of PSNR of about 1 dB compared to the



(a) Noisy risk (b) Convolved risk (c) Yaroslavsky

Figure 7.11: (top) Average areas and (middle line) average orientations of selected shapes for different risk maps on a noisy realization of the *cameraman* image. (bottom) Corresponding results focused on the cameraman's neck. From left to right, results using the noisy risk maps (i.e., SURE maps), the convolved risk maps and the risk maps obtained by Yaroslavsky regularization. The average areas and the average orientations are represented using colors whose legends are given on the top right corners.

classical NL means (for $\sigma \leq 20$). The SSIM is also usually increased. The BM3D approach leads to better numerical results than all Non Local based approaches. Figure 7.12 gives the visual results. While the blockwise NL means and the refinement of the NL means (Goossens et al., 2008) illustrate the *rare patch effect* by the presence of *noise halos*, BM3D and NLM-SAP have reduced a lot this phenomenon. Our NLM-SAP provides smooth results with accurate details, such as the cameraman's head, the house windows, the windmill blades, the tree-trunk and the

	NL means	Block-NLM	WAV NLM	Duval *et al.*	INLM	BM3D	NLM-SAP
	$\sigma = 10$						
barbara	32.23/0.969	32.61/0.971	32.89/0.970	33.85/0.970	34.29/0.973	**34.90/0.977**	33.69/0.970
boat	32.00/0.956	32.41/0.958	32.61/0.955	32.77/0.955	33.21/0.962	**33.85/0.967**	32.99/0.953
bridge	29.08/0.884	29.23/0.889	29.40/0.891	29.70/0.887	30.46/0.904	**30.66/0.906**	30.03/0.896
cameraman	32.13/0.909	32.47/0.913	32.85/0.921	33.11/0.920	33.52/0.926	**34.05/0.930**	33.50/0.923
city	30.60/0.922	31.00/0.926	31.49/0.941	32.07/0.948	32.01/0.943	**33.14/0.955**	32.73/0.952
couple	31.99/0.952	32.41/0.954	32.67/0.952	32.81/0.955	33.25/0.959	**33.93/0.967**	33.07/0.948
fingerprint	28.77/0.988	28.83/0.988	28.95/0.988	30.66/0.986	32.14/0.990	**32.41/0.990**	30.44/0.988
flinstones	30.33/0.976	30.71/0.978	31.07/0.977	31.50/0.977	31.68/0.978	**32.40/0.980**	31.85/0.978
hill	30.32/0.859	30.66/0.869	30.96/0.871	30.93/0.863	31.43/0.877	**31.85/0.883**	31.49/0.871
lake	31.64/0.919	32.10/0.924	32.42/0.936	32.56/0.936	32.95/0.940	**33.62/0.949**	33.22/0.943
lena	34.08/0.962	34.47/0.964	34.65/0.962	34.81/0.961	35.34/0.965	**35.79/0.969**	35.00/0.959
man	32.14/0.953	32.53/0.955	32.75/0.951	32.87/0.951	33.34/0.958	**33.90/0.963**	33.20/0.949
mandril	30.11/0.954	30.26/0.955	30.39/0.952	31.29/0.950	32.73/0.960	**33.09/0.966**	31.11/0.948
windmill	33.00/0.938	33.48/0.941	34.06/0.953	35.05/0.957	34.62/0.958	**35.81/0.966**	35.24/0.958
	$\sigma = 20$						
barbara	29.87/0.936	30.30/0.939	30.31/0.937	30.62/0.939	30.95/0.946	**31.76/0.953**	30.41/0.930
boat	29.29/0.892	29.63/0.893	29.55/0.886	29.59/0.897	29.92/0.902	**30.81/0.927**	29.67/0.877
bridge	25.68/0.739	26.11/0.756	26.17/0.743	25.89/0.738	26.20/0.761	**26.76/0.775**	26.24/0.728
cameraman	28.59/0.823	29.01/0.831	29.23/0.838	29.58/0.856	29.49/0.852	**30.34/0.871**	29.74/0.844
city	26.58/0.863	27.07/0.868	27.30/0.877	27.85/0.893	28.00/0.893	**29.06/0.912**	28.53/0.897
couple	29.03/0.892	29.42/0.895	29.41/0.889	29.25/0.893	29.82/0.903	**30.67/0.927**	29.37/0.877
fingerprint	26.48/0.958	26.94/0.960	27.16/0.958	27.20/0.957	27.75/0.965	**28.80/0.972**	27.45/0.951
flinstones	27.19/0.958	27.70/0.959	28.03/0.955	28.92/0.961	28.54/0.962	**29.55/0.966**	29.04/0.960
hill	27.50/0.733	27.86/0.745	27.78/0.735	27.62/0.741	27.99/0.756	**28.51/0.779**	27.83/0.724
lake	27.78/0.854	28.23/0.861	28.18/0.865	28.48/0.872	28.82/0.881	**29.38/0.894**	29.04/0.878
lena	31.61/0.926	31.99/0.928	31.95/0.924	31.67/0.922	32.37/0.932	**32.98/0.940**	31.92/0.918
man	29.34/0.886	29.64/0.887	29.54/0.880	29.55/0.889	29.81/0.895	**30.52/0.915**	29.62/0.872
mandril	27.02/0.869	27.31/0.872	27.35/0.864	27.33/0.867	28.22/0.884	**29.04/0.910**	27.45/0.846
windmill	29.36/0.883	29.94/0.890	30.18/0.901	31.44/0.917	30.85/0.917	**32.06/0.935**	31.32/0.912
	$\sigma = 40$						
barbara	26.65/0.855	26.78/0.858	26.67/0.856	26.86/0.866	27.49/0.879	**28.04/0.896**	26.04/0.831
boat	26.26/0.781	26.38/0.784	26.21/0.779	26.27/0.789	26.74/0.800	**27.64/0.848**	26.06/0.762
bridge	23.01/0.529	23.14/0.540	23.02/0.533	23.16/0.568	23.15/0.552	**23.98/0.615**	22.91/0.511
cameraman	25.89/0.707	26.11/0.724	25.88/0.724	26.29/0.768	26.62/0.766	**27.26/0.801**	26.28/0.749
city	23.50/0.733	23.75/0.746	23.42/0.736	23.77/0.766	24.04/0.779	**25.25/0.829**	23.70/0.747
couple	25.73/0.769	25.87/0.772	25.75/0.767	25.67/0.775	26.33/0.792	**27.43/0.849**	25.39/0.737
fingerprint	23.45/0.855	23.78/0.859	23.72/0.856	23.93/0.885	24.08/0.878	**25.27/0.926**	23.07/0.817
flinstones	24.76/0.903	24.86/0.898	24.63/0.890	25.29/0.915	25.35/0.916	**26.07/0.933**	25.02/0.888
hill	24.57/0.555	24.73/0.568	24.67/0.564	24.63/0.583	24.97/0.590	**25.87/0.659**	24.43/0.540
lake	24.71/0.715	24.88/0.732	24.67/0.727	24.99/0.759	25.25/0.774	**25.86/0.805**	24.79/0.742
lena	28.31/0.846	28.51/0.850	28.42/0.848	28.37/0.853	29.14/0.868	**29.77/0.883**	28.22/0.839
man	26.42/0.779	26.52/0.781	26.45/0.778	26.46/0.785	26.83/0.792	**27.57/0.833**	26.30/0.761
mandril	23.75/0.678	23.83/0.677	23.74/0.671	23.98/0.707	24.16/0.705	**25.22/0.789**	23.39/0.633
windmill	26.41/0.769	26.59/0.787	26.36/0.789	27.46/0.837	27.23/0.839	**28.04/0.878**	26.65/0.815

Table 7.5: Comparisons of denoising approaches for different degradation levels in terms of PSNR and SSIM values (PSNR/SSIM). The compared methods are the classical (pixelwise) NL means (Buades et al., 2005), the blockwise NL means for square patches, the blockwise NL means using WAV reprojection for square patches, the pixelwise NL means using SURE-based adaptive bandwidth selection (Duval et al., 2011), the Improved NL means (INLM) (Goossens et al., 2008) the BM3D denoiser (Dabov et al., 2007), and our proposed NLM-SAP approach.

car. Visually, the quality of images obtained with NLM-SAP challenges those obtained with BM3D.

Figure 7.12: Comparisons of the visual denoising performance of the proposed NLM-SAP approach and other *state-of-the-art* methodologies ($\sigma = 20$). From top to bottom, the input noisy images, the results obtained by the classical (pixelwise) NL means (Buades et al., 2005), the blockwise NL means for square patches the BM3D denoiser (Dabov et al., 2007), and our proposed NLM-SAP approach.

7.3.3 Conclusions on the experiments

The best results were obtained by using the NLM-SAP with the default parameters given in the previous section. Let us also mention that another version of the algorithm could be appealing to users particularly interested in fast and simple implementation. It consists in replacing the EWA aggregation with a WAV aggregation using only three scales of isotropic shapes. The simple scheme is then 5 times faster than the more elaborated version using 15 shapes at the price of a slight decrease of the PSNR. Note also that the WAV aggregation scheme is parameter-free.

7.4 Conclusion

In this chapter, we have addressed the problem of the *rare patch effect* arising in the NL means procedure and responsible for the noisy halos created around edges with high contrast. The proposed solution consists of substituting the square patches of fixed size by spatially adaptive patch shapes. A fast implementation of NL means, based on FFT calculations, has been proposed in this context to handle any kind of patch shape with arbitrary scale. Thanks to this acceleration, different estimates are obtained by using different patch shapes, typically one isotropic patch shape and four edge oriented patch shapes, all of them with three different scales. We have extended SURE-based approaches to aggregate properly these different shape-based estimates in a spatially adaptive way. To get an efficient locally adaptive filter, we have shown that the SURE-based risk maps require to be regularized and that Yaroslavsky regularization can be used to this purpose. Simulations have shown that exponentially weighted aggregation based on the regularized risk maps of the different shape-based estimates could lead to both numerical and visual improvements (the *noise halo* is suppressed around edges). Visually, our method challenges other NL means improvements we have considered in our comparisons. It is still out-performed by BM3D in terms of PSNR and SSIM.

Future work could be to reduce computation time of the algorithm by choosing more convenient shapes or parallel implementations. Another extension might be to apply our framework to more general kinds of noise, by adapting results given in Chap. 5.

Chapter 8

Conclusion and perspectives

Conclusion

There is an ever increasing interest for high resolution radar images. However, strong *speckle* that corrupts these images make their interpretation very difficult. *Speckle* differs significantly from additive Gaussian noise and thus requires adapted denoising methods. The starting point of this thesis has been to bridge the gap between cutting edge denoising methods restricted to Gaussian noise and SAR despeckling techniques. To best preserve the high resolution of SAR images, adaptive selection of neighborhoods must be designed. Given the strong noise level, the use of patches is necessary for robust selection. This raised the question of how to compare noisy patches.

Patch similarity: In the light of detection theory, we studied several similarity criteria that have been proposed in the literature to deal with the problem of patch comparisons in non-Gaussian noise conditions. We introduced a new similarity criterion based on the generalized likelihood ratio test. The properties and the performance on a task-based evaluation were in favor of the proposed criterion.

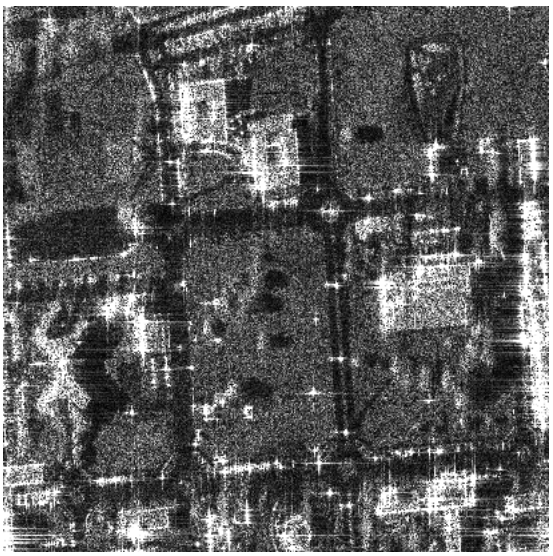
Data driven selection: When noise departs from the Gaussian distribution, particular attention should be paid to the bias introduced by the selection procedure. The selection of pixels with similar values can discard samples of interest which biases the subsequent maximum likelihood estimation. The generalized likelihood ratio provides a selection-rule leading to an unbiased estimator with an equal variance reduction in the homogeneous areas: the dynamic and contrast of objects are preserved.

In the heterogeneous areas, an oracle study shows that the estimation should rather introduce a small bias. This bias resulting in blur effects can be controled efficiently thanks to the good properties of the generalized likelihood ratio. The smoothing parameters can be chosen to maintain the same smoothing effect in the homogeneous areas (independently of the underlying signal) and minimize the blur in the heterogeneous areas: the edges, textures and punctual targets are preserved.

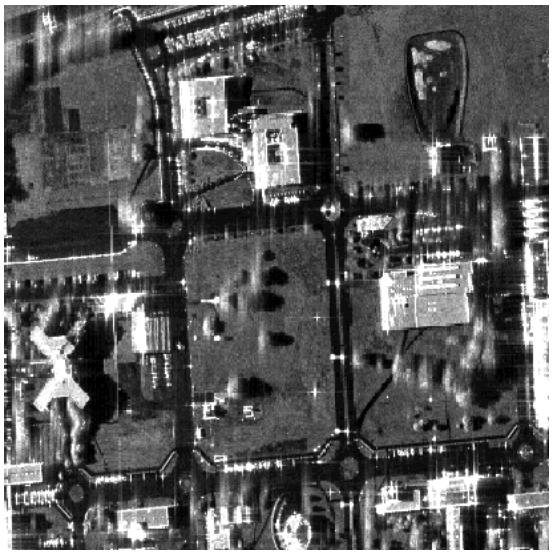
Two-step or iterative selection: When the signal intensity is low compared to the noise intensity, it is important to drive the selection according to two images: the noisy one and a pre-filtered one (possibly obtained iteratively). An oracle study shows that the Kullback-Leibler divergence can be used efficiently to refine the detection performance of the generalized likelihood ratio test. When an estimate of the mean square error is available, one should use it to automatically choose the trade-off between the fidelities to both images.

Non-local SAR (NL-SAR): The proposed methodology has retained a lot of attention for the denoising of amplitude SAR images. Our results have made the cover of the IEEE GRSL Newsletter

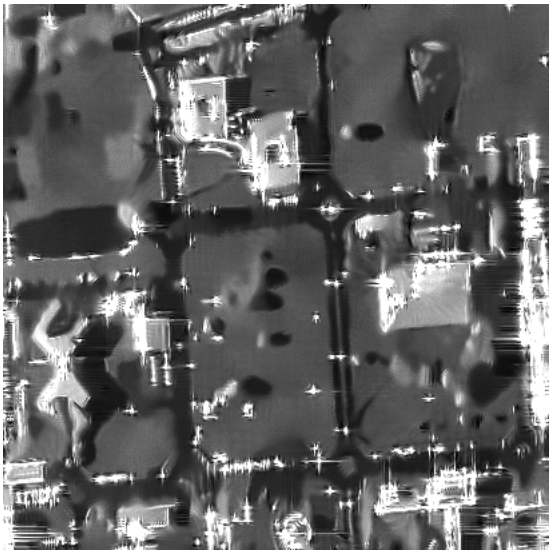
(a) IEEE GRSL Newsletter of March 2011



(b) Single-look image



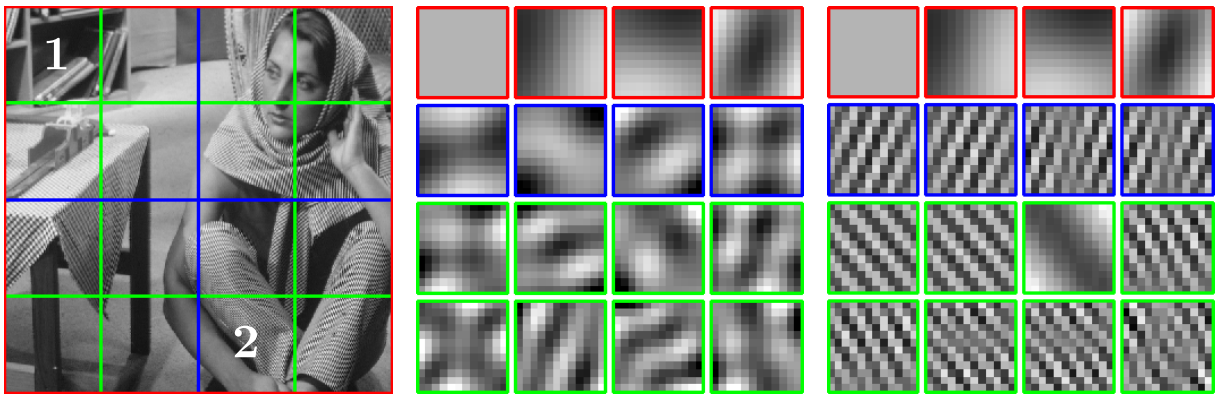
(c) 100-look SAR image



(d) Result of NL-SAR

Figure 8.1: (a) The cover of the IEEE GRSL Newsletter of March 2011 illustrating the performance of NL-SAR by comparing the images in (b), (c) and (d). (b) A single-look image of resolution 1m × 1m (acquired by ONERA, multi-looked by CNES, ©ONERA CNES). (c) A 100-look image obtained by multi-looking a very high resolution image of the same urban area. This image can be considered as a ground truth. (d) The denoised version of the single-look image.

of March 2011 (see Fig. 8.1). This illustration provides an interesting validation of our denoising technique. It compares our denoising result obtained from a decimated single-look noisy image (1m ground resolution) with a 100-look image (i.e., at each 1m × 1m pixel, 100 values from a 10cm×10cm image are averaged).The 100-look image can be considered as a ground truth, in the sense that the remaining speckle noise is extremely low. Note however that, due to anisotropy of some targets, there are some differences which are not due to speckle only. The figure illustrates that fine details are well preserved by our approach. For amplitude SAR images, our method-



(a) Quadtree decomposition (b) 16 first axes in part 1 (c) 16 first axes in part 2

Figure 8.2: An image and its 16 first principal patches obtained over two stacks extracted respectively in two different leaves of the quadtree decomposition. Here, the four main patches are kept at each node of the quadtree and three level of decomposition is used. The resulting dictionaries seem to describe more and more local features.

ology has led to advancements in *speckle* reduction and inspired other methods (Parrilli et al., 2010; Teuber and Lang, 2011; Feng et al., 2011; Zhong et al., 2011). For interferometric and polarimetric SAR images, our filter can be considered as the current *state-of-the-art* method. In this thesis, it has been applied to the latest aerial radar system of the DLR, F-SAR, and its potential has been illustrated on such high-resolution SAR data.

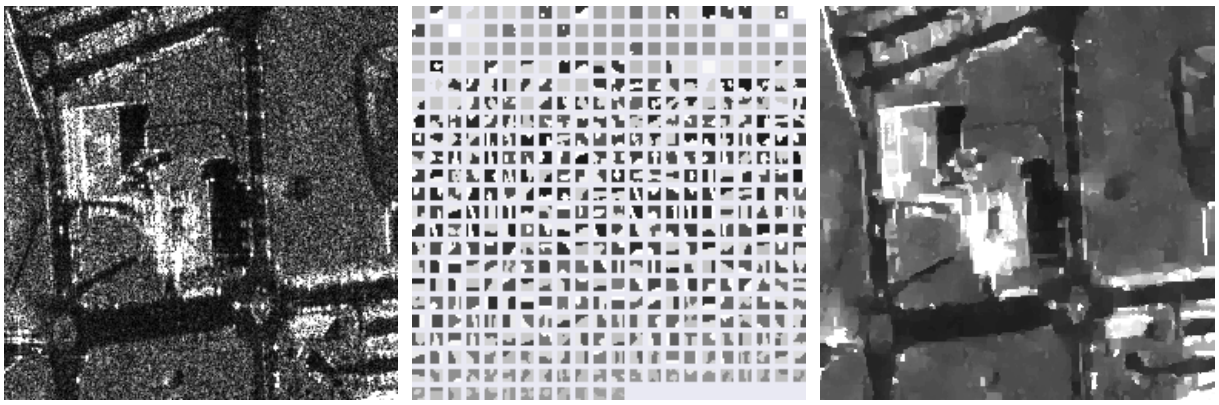
Poisson NL means: The same methodology has led to appealing results in photon-limited imagery such as fluorescence microscopy and astronomy. I have been awarded for this work the best student paper award at ICIP'2010. This approach has also been mentioned as the *state-of-the-art* technique in (Lee et al., 2011) even if I consider that the BM3D and the SAFIR approaches with optimal inverse Anscombe transform still provides at least as good results (Dabov et al., 2007; Boulanger et al., 2008; Mäkitalo and Foi, 2011).

NL means with shape adaptive patches (NLM-SAP): Concurrently, we also proposed a spatial adaptation of the scale and the orientation of patches to take into account the local geometry of images. This was a joint work with Vincent Duval and Joseph Salmon. This technique relied on a fast algorithm to compute the solution of the NL means for arbitrary shapes of patches. The availability of many estimates of the NL means for different shapes pointed us to the problem of aggregation. We proposed to use a local estimation of the Stein unbiased risk estimator to locally combine the best scales and shapes of patches. The resulting filter has led to results of good quality challenging other NL means improvements.

Remaining issues and perspectives

Rare features: towards the use of dictionary

When using denoising techniques, one should care when their inherent assumptions are violated. We have seen that on images with high contrast, such as in high dynamic range images, the presence of "rare" features defeats the patch redundancy assumption. The use of patches with adaptive sizes and shapes allowed us to exploit the redundancy of directional and multi-scale features. However, there can still be unique features even when considering a geometrical adaptive definition of the redundancy. I remember a passionate discussion with Vincent Duval and



(a) Noisy image (b) Dictionary (c) Filtered image

Figure 8.3: (a) A SLC SAR image (acquired by RAMSES ©ONERA CNES), (b) a dictionary learned on the noisy image and (c) the filtered image reconstructed by using the learned dictionary.

Joseph Salmon on what denoising techniques should do in this case. Should we enforce reducing the noise at the cost of a bias, or, should we leave some noisy parts in the image? Of course it depends on the application at hand.

In the case of polarimetric SAR images, a punctual target (often a bright scatterer) may very well have no replica in the search window, especially if it has a non-common polarimetric signature. It is vain to restore such a feature by exploiting the image content thus one could leave this target unchanged. However to analyze its polarimetric signature, one should find at least three replicas (to prevent singularity issues). We then suggested enforcing a minimum of noise reduction to properly estimate the polarimetric information but if there is no replica we will necessarily bias such punctual targets. Again, what is preferable?

If features have no replica in the image content, one could instead search in an extra database. A simple idea could be to exploit a large collection of images of the same nature. Of course, to prevent high running time, we should search in a dictionary providing a summary of relevant patterns and including "rare" features. The construction of such dictionary is still an open problem that has been briefly reviewed in this thesis. Recently, Jospeh Salmon, Arnak Dalalyan and I have designed a denoising technique based on a dictionary obtained by an adaptive principal component analysis that we presented at BMVC'2011 (Fig. 8.2 extracted from Deledalle et al., 2011d, gives an illustration of the idea). Following the spirit of this thesis and the recent improvements in image modeling, one could consider the learning of a dictionary with multi-scale and anisotropic features and/or under non-Gaussian noise conditions. Beyond the problem of denoising, the learning of crude low-level features in polarimetric imagery could be of particular interest. An early work following this idea has already provided appealing results on an amplitude SAR image (see Fig. 8.3).

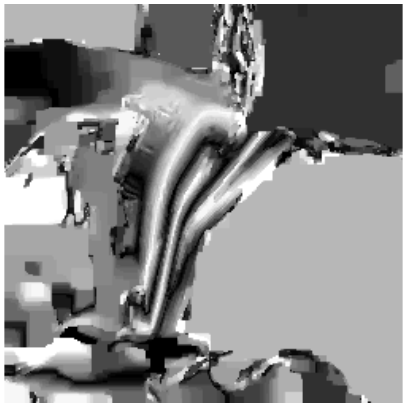
Testing patch equality vs patch similarity

In this thesis we considered robust similarity criteria evaluating that two noisy patches share the same noise-free patch under non-Gaussian conditions. Our hypothesis test is expressed as an equality between the underlying noise-free patches.

When two different images are available, e.g., in change detection, flickering reduction, stereovision or image registration, a change of illumination may have occured between the two acquisitions. One should rather be interested in finding if two noisy patches are identical up to an affine

(a) Amplitude, phase and empirical coherence in SAR interferometry

(b) Joint non-local estimation without regularization of the interferometric phase



(c) Joint non-local estimation with regularization of the interferometric phase

Figure 8.4: (a) From left to right, the amplitude, the phase difference and the empirical coherence of an interferometric pair of two SLC SAR images of an hypdrolic dam in Serre-Ponçon sensed by TerraSAR-X (image courtesy of Astrium). (b) The non-local joint estimation of the reflectivity, the true phase difference and the true coherence. (c) The non-local joint estimation with regularization of the interferometric phase.

transform. The same probleme arises in dictionary learning from noisy data. One wants to learn the geometry of atomic features rather than their intensities. Our definition of similarity using the generalized likelihood ratio could be applied for such applications by adapting the definition of the hypothesis test (i.e., introducing additional nuisance parameters).

Under the Gaussian noise assumption, the square difference between noisy values estimates the square difference between noise-free values. This no longer holds true with the generalized likelihood ratio. We have seen that the performance of our methodology are all the more relevant for strong noises, whereas the square differences are challenging with lower noise levels. We believe that the similarity criterion should approach the square difference at low noise levels. Our definition of similarity using the generalized likelihood ratio could be derived in this sense by testing the proximity of noise-free patches rather than their equality.

Noise reduction versus regularity: towards a joint approach

Selection-based filters reduce the local variance while avoiding the introduction of bias. Such filters produce an image with smaller fluctuations, hence the name of "noise reduction". Some image analysis techniques or image interprets should expect the homogeneous areas to appear at a constant level without fluctuations. For instance, in a 3D reconstruction obtained from a pair of interferometric SAR images, one expects that the roof of a building is flat or planar. Also, in non-coherent regions (for instance shadows or the vegetation) the interferometric phase is non-informative and one should instead regularize/extrapolate the phase based on coherent areas (for instance the ground or man-made structures).

Several works defined the regularity in the patch domain with a data-fidelity term in the spatial domain (for instance using non-local graphs). In an early work, we obtained promising results by considering a spatial regularity of the interferometric phase with a data-fidelity term defined in the patch domain (see Fig. 8.4). This model could be extended to phase unwrapping.

The limit of joint filtering

In this thesis, we considered jointly estimating the different channels of multi-variate images. In SAR polarimetry, the joint estimation is preferable to preserve the polarimetric characteristics and prevent cross-talk between channels. We have also chosen to drive our filter by considering the joint information brought by all channels while it is usually driven only by the intensity. In interferometric SAR filtering, the denoising is then driven by the phases and the amplitudes, preventing from mixing regions of different heights with the same reflectivity and vice-versa. However, during my stay at DLR, Andreas Reigber pointed out that the filtering of the amplitude and the phase would be more efficient if performed independently. Indeed, we could reduce the variance more in both channels by independently mixing the amplitudes of the same reflectivity and the noisy phases of the same height. But at the same time, we would lose the robustness of the similarity brought by the joint comparison of the different channels. It sounds that a compromise must be made.

Publication list

Papers in refereed journals

Deledalle, C.-A., Denis, L., and Tupin, F. (2012). How to compare noisy patches? Patch similarity beyond Gaussian noise. *Accepted in International Journal of Computer Vision*.

Deledalle, C.-A., Duval, V., and Salmon, J. (2011f). Non-local Methods with Shape-Adaptive Patches (NLM-SAP). *Journal of Mathematical Imaging and Vision*, pages 1–18.

Deledalle, C.-A., Denis, L., and Tupin, F. (2011b). NL-InSAR : Non-Local Interferogram Estimation. *IEEE Trans. Geosci. Remote Sens.*, 49(4):1441–1452.

Deledalle, C.-A., Denis, L., and Tupin, F. (2009b). Iterative Weighted Maximum Likelihood Denoising with Probabilistic Patch-Based Weights. *IEEE Trans. Image Process.*, 18(12):2661–2672.

International conference papers

Salmon, J., Deledalle, C.-A., Willett, R., Harmany, Z. (2012). Poisson noise reduction with non-local PCA. Accepted in *IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE.

Deledalle, C.-A., Tupin, F., and Denis, L. (2011e). Patch similarity under non-gaussian noise. In *IEEE Int. Conf. Image Process. (ICIP)*. IEEE.

Deledalle, C.-A., Salmon, J., and Dalalyan, A. (2011d). Image denoising with patch based PCA: local versus global. In *British Machine Vision Conference (BMVC)*.

Lucas, A., Deledalle, C.-A., Aharonson, O.and Hayes, A., and Team, T. C. R. (2011b). Improved Landform Interpretation on Titan from SAR Image Processing. In *Titan Surface Workshop*.

Cao, F., Deledalle, C.-A., Nicolas, J.-M., Tupin, F., Denis, L., Ferro-Famil, L., Pottier, E., and Lopéz-Martínez, C. (2011). Influence of speckle filtering of polarimetric SAR data on different classification methods. In *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*. IEEE.

Hachicha, S., Deledalle, C.-A., Chaabane, F., and Tupin, F. (2011). Multi-temporal sar classification according to change detection operators. In *Analysis of Multi-temporal Remote Sensing Images (Multi-Temp)*, pages 133–136. IEEE.

Lucas, A., Aharonson, O., Deledalle, C.-A., Hayes, A., and Team, T. C. R. (2011a). Improved SAR Processing and Landform Interpretation on Titan. In *Titan Science Meeting*.

Deledalle, C.-A., Duval, V., and Salmon, J. (2011c). Anisotropic non-local means with spatially adaptive patch shapes. In *Scale Space and Variational Methods in Computer Vision (SSVM)*.

Deledalle, C.-A., Tupin, F., and Denis, L. (2010c). Poisson NL means: Unsupervised non local means for Poisson noise. In *IEEE Int. Conf. Image Process. (ICIP)*, pages 801–804. IEEE.

Deledalle, C.-A., Tupin, F., and Denis, L. (2010d). Polarimetric SAR estimation based on non-local means. In *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, pages 2515–2518. IEEE.

Deledalle, C.-A., Tupin, F., and Denis, L. (2010b). A non-local approach for SAR and interferometric SAR denoising. In *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, pages 714–717. IEEE.

Deledalle, C.-A., Nicolas, J., Tupin, F., Denis, L., Fallourd, R., and Trouvé, E. (2010a). Glacier monitoring: correlation versus texture tracking. In *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, pages 513–516. IEEE.

Benzid, S., Deledalle, C.-A., Abdelfattah, R., Chaabane, F., and Tupin, F. (2010). Change detection in a multitemporal series of radar images. In *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, pages 1473–1476. IEEE.

French conferences

Salmon, J., Deledalle, C.-A., and Duval, V. (2011). Méthodes non-locales et formes de patchs adaptatives. In *Journées de Statistique*.

Deledalle, C.-A. (2011). Débruitage non-local : Adapation au type de bruit et aux structures de l'image. In *Journées de la SMAI*.

Deledalle, C.-A., Denis, L., and Tupin, F. (2009a). Débruitage Non-Local Itératif fondé sur un Critère de Similarité Probabiliste. In *GRETSI*.

Appendix A

Patch-similarity: closed-form expressions and sketch of proofs

This document is a supplementary material of the paper *How to compare noisy patches? Patch similarity beyond Gaussian noise* (Deledalle et al., 2012). The derivation of the closed-form expressions of seven similarity criteria for Gaussian, gamma, Poisson and Cauchy noise is given in a first part. Proof sketches of some useful properties for similarity criteria described in (Deledalle et al., 2012), are given in a second part.

A.1 Derivation of closed-form expressions of similarity criteria

We derive in this section the closed-form expression of the 7 different similarity criteria between patches $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ considered in (Deledalle et al., 2012):

- $\mathcal{G}$, the usual similarity criterion based on square differences: $\mathcal{G}(\boldsymbol{v}_1, \boldsymbol{v}_2) = \exp\left(-\|\boldsymbol{v}_1 - \boldsymbol{v}_2\|_2^2/h\right)$,
- $\mathcal{S}$, based on variance stabilizing transform $\boldsymbol{s}$: $\mathcal{S}(\boldsymbol{v}_1, \boldsymbol{v}_2) = \mathcal{G}(\boldsymbol{s}(\boldsymbol{v}_1), \boldsymbol{s}(\boldsymbol{v}_2))$,
- $\mathcal{L}_B$, the Bayesian likelihood ratio: $\mathcal{L}_B(\boldsymbol{v}_1, \boldsymbol{v}_2) = \frac{\int p(\boldsymbol{v}_1|\boldsymbol{u}_{12}=\boldsymbol{t})p(\boldsymbol{v}_2|\boldsymbol{u}_{12}=\boldsymbol{t})p(\boldsymbol{u}_{12}=\boldsymbol{t})\,\mathrm{d}\boldsymbol{t}}{\int p(\boldsymbol{v}_1|\boldsymbol{u}_1=\boldsymbol{t}_1)p(\boldsymbol{u}_1=\boldsymbol{t}_1)\mathrm{d}\boldsymbol{t}_1 \int p(\boldsymbol{v}_2|\boldsymbol{u}_2=\boldsymbol{t}_2)p(\boldsymbol{u}_2=\boldsymbol{t}_2)\mathrm{d}\boldsymbol{t}_2}$,
- $\mathcal{L}_G$, the generalized likelihood ratio: $\mathcal{L}_G(\boldsymbol{v}_1, \boldsymbol{v}_2) = \frac{\sup_{\boldsymbol{t}} p(\boldsymbol{v}_1,\boldsymbol{v}_2|\boldsymbol{u}_{12}=\boldsymbol{t},\mathscr{H}_0)}{\sup_{\boldsymbol{t}_1,\boldsymbol{t}_2} p(\boldsymbol{v}_1,\boldsymbol{v}_2|\boldsymbol{u}_1=\boldsymbol{t}_1,\boldsymbol{u}_2=\boldsymbol{t}_2,\mathscr{H}_1)}$,
- $\mathcal{Q}_B$, the Bayesian joint likelihood: $\mathcal{Q}_B(\boldsymbol{v}_1, \boldsymbol{v}_2) = \int p(\boldsymbol{v}_1|\boldsymbol{u}_1 = \boldsymbol{t})\, p(\boldsymbol{v}_2|\boldsymbol{u}_2 = \boldsymbol{t})\, p(\boldsymbol{u}_{12} = \boldsymbol{t})\, \mathrm{d}\boldsymbol{t}$,
- $\mathcal{Q}_G$, the maximum joint likelihood: $\mathcal{Q}_G(\boldsymbol{v}_1, \boldsymbol{v}_2) = p(\boldsymbol{v}_1|\boldsymbol{u}_1 = \hat{\boldsymbol{t}}_{12})\, p(\boldsymbol{v}_2|\boldsymbol{u}_2 = \hat{\boldsymbol{t}}_{12})$,
- $\mathcal{K}_B$, the mutual information kernel: $\mathcal{K}_B(\boldsymbol{v}_1, \boldsymbol{v}_2) = \mathcal{Q}_B(\boldsymbol{v}_1, \boldsymbol{v}_2)/\sqrt{\mathcal{Q}_B(\boldsymbol{v}_1, \boldsymbol{v}_1)\mathcal{Q}_B(\boldsymbol{v}_2, \boldsymbol{v}_2)}$.

where, $\boldsymbol{v}$ denotes the available (i.e., noisy) data, while $\boldsymbol{u}$ are the parameters of interest that are to be recovered.

We consider uncorrelated noise, so that patch similarity is the product over the patch of similarity between pixels. We study first Gaussian noise, then Gamma noise, Poisson noise, and finally Cauchy-distributed noise.

A.1.1 Gaussian noise case

Given $\sigma \in \mathbb{R}_*^+$, a Gaussian random variable V follows the probability density function (pdf):

$$p(v|u) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(v-u)^2}{2\sigma^2}\right] , \tag{A.1}$$

with expectation $\mathbb{E}[V] = u$ and variance $\mathrm{Var}[V] = \sigma^2$. Gaussian fluctuations are additive, it is straightforward to show that V can be decomposed as $u + N$ with N a zero mean Gaussian random variable.

Fisher information

Fisher information associated with a Gaussian pdf is given by:

$$\mathcal{I}(u) \triangleq \mathbb{E}_V\left[\left(\frac{\partial}{\partial u}\log p(v|u)\right)^2\right] = \int \left(\frac{\partial}{\partial u}\log p(v|u)\right)^2 p(v|u)\, \mathrm{d}v \tag{A.2}$$

$$= \int \left(\frac{v-u}{\sigma^2}\right)^2 \frac{e^{-\frac{(v-u)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}\, \mathrm{d}v = \underbrace{\frac{1}{\sigma^4}\int (v-u)^2 \frac{e^{-\frac{(v-u)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}\, \mathrm{d}v = \frac{\sigma^2}{\sigma^4}}_{\text{by definition of variance}} = \frac{1}{\sigma^2} . \tag{A.3}$$

Jeffreys' prior

Jeffreys' prior follows from Fisher information:

$$p(u) \triangleq \sqrt{|\mathcal{I}(u)|} = \frac{1}{\sigma} . \tag{A.4}$$

Bayesian joint likelihood

With Jeffreys' prior, we can derive the Bayesian joint likelihood as follows:

$$\mathcal{Q}_B(v_1, v_2) = \int p(v_1|u_1 = t)p(v_2|u_2 = t)p(u_{12} = t)\mathrm{d}t = \int \left(\frac{e^{-\frac{(v_1-t)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}\right)\left(\frac{e^{-\frac{(v_2-t)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}\right)\left(\frac{1}{\sigma}\right)\mathrm{d}t \tag{A.5}$$

$$= \underbrace{\frac{1}{2\pi\sigma^3}\int e^{-\frac{(v_1-t)^2}{2\sigma^2}} e^{-\frac{(v_2-t)^2}{2\sigma^2}}\mathrm{d}t = \frac{e^{-\frac{(v_1-v_2)^2}{4\sigma^2}}}{2\pi\sigma^3}}_{\text{by convolution of two Gaussian functions}} . \tag{A.6}$$

Bayesian likelihood ratio

Let $\mathcal{D}_B$ be the denominator term appearing in the Bayesian likelihood ratio and expressed as:

$$\mathcal{D}_B(v) = \int p(v|u = t)p(u = t)\mathrm{d}t = \int \left(\frac{e^{-\frac{(v-t)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}\right)\left(\frac{1}{\sigma}\right)\mathrm{d}t = \frac{1}{\sigma} . \tag{A.7}$$

Using the expression of $\mathcal{Q}_B(v_1, v_2)$ and $\mathcal{D}_B(v)$, it results that the Bayesian likelihood ratio is given by:

$$\mathcal{L}_B(v_1, v_2) = \frac{\mathcal{Q}_B(v_1, v_2)}{\mathcal{D}_B(v_1)\mathcal{D}_B(v_2)} = \frac{\frac{e^{-\frac{(v_1-v_2)^2}{4\sigma^2}}}{2\pi\sigma^3}}{\frac{1}{\sigma}\frac{1}{\sigma}} = \frac{e^{-\frac{(v_1-v_2)^2}{4\sigma^2}}}{2\pi\sigma} . \tag{A.8}$$

Mutual information kernel

Using the expression of $\mathcal{Q}_B(v_1, v_2)$ and $\mathcal{Q}_B(v, v)$, it comes that the mutual information kernel is:

$$\mathcal{K}_B(v_1, v_2) = \frac{\mathcal{Q}_B(v_1, v_2)}{\sqrt{\mathcal{Q}_B(v_1, v_1)\mathcal{Q}_B(v_2, v_2)}} = \frac{\frac{e^{-\frac{(v_1-v_2)^2}{4\sigma^2}}}{2\pi\sigma^3}}{\sqrt{\frac{e^0}{2\pi\sigma^3}\frac{e^0}{2\pi\sigma^3}}} = e^{-\frac{(v_1-v_2)^2}{4\sigma^2}} \ . \tag{A.9}$$

Maximum joint likelihood

The priorless extension of $\mathcal{Q}_B(v_1, v_2)$, i.e. the maximum joint likelihood is obtained as follows:

$$\mathcal{Q}_G(v_1, v_2) = \sup_t p(v_1|u_1 = t)p(v_2|u_2 = t) = \left(\frac{e^{-\frac{(v_1-\frac{v_1+v_2}{2})^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}\right)\left(\frac{e^{-\frac{(v_2-\frac{v_1+v_2}{2})^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}\right) \tag{A.10}$$

$$= \frac{1}{2\pi\sigma^2}e^{-\frac{(v_1-v_2)^2}{8\sigma^2}}e^{-\frac{(v_-v_2)^2}{8\sigma^2}} = \frac{e^{-\frac{(v_1-v_2)^2}{4\sigma^2}}}{2\pi\sigma^2} \ . \tag{A.11}$$

since under Gaussian noise the maximum likelihood estimator (MLE) is the mean.

Generalized likelihood ratio

Let $\mathcal{D}_G$ be the denominator term appearing in the generalized likelihood ratio and expressed as:

$$\mathcal{D}_G(v) = \sup_t p(v|u = t) = \frac{e^0}{\sqrt{2\pi}\sigma} = \frac{1}{\sqrt{2\pi}\sigma} \ . \tag{A.12}$$

Using the expression of $\mathcal{Q}_G(v_1, v_2)$ and $\mathcal{D}_G(v)$, it results that the generalized likelihood ratio is given by:

$$\mathcal{L}_G(v_1, v_2) = \frac{\mathcal{Q}_G(v_1, v_2)}{\mathcal{D}_G(v_1)\mathcal{D}_G(v_2)} = \frac{\frac{e^{-\frac{(v_1-v_2)^2}{4\sigma^2}}}{2\pi\sigma^2}}{\frac{1}{\sqrt{2\pi}\sigma}\frac{1}{\sqrt{2\pi}\sigma}} = e^{-\frac{(v_1-v_2)^2}{4\sigma^2}} \ . \tag{A.13}$$

A.1.2 Gamma noise case

Given the positive integer $L \in \mathbb{N}^*$, a Gamma random variable V can be described by the following pdf:

$$p(v|u) = \frac{L^L v^{L-1} e^{-\frac{Lv}{u}}}{\Gamma(L)u^L}. \tag{A.14}$$

Its expectation is $\mathbb{E}[V] = u$ and variance $\mathrm{Var}[V] = \frac{u^2}{L}$. The relation $\mathrm{Var}[V] \propto \mathbb{E}[V]^2$ indicates a multiplicative behaviour. Indeed, it is straightforward to show that V can be decomposed as $u \times S$ with S a Gamma random variable of parameter $u_S = 1$.

Fisher information

Fisher information associated with a Gamma pdf is given by:

$$\mathcal{I}(u) = E\left[\left(\frac{\partial}{\partial u}\log p(v|u)\right)^2 |u\right] = \int \left(\frac{Lv}{u^2} - \frac{L}{u}\right)^2 \frac{L^L v^{L-1} e^{-\frac{Lv}{u}}}{\Gamma(L)u^L}\mathrm{d}v \tag{A.15}$$

$$= \underbrace{\frac{L^2}{u^4}\int (v-u)^2 \frac{L^L v^{L-1} e^{-\frac{Lv}{u}}}{\Gamma(L)u^L}\mathrm{d}v}_{\text{by definition of variance}} = \frac{L^2}{u^4}\frac{u^2}{L} = \frac{L}{u^2}\,. \tag{A.16}$$

Jeffreys' prior

Fisher information allows to define Jeffreys' prior as:

$$p(u) \triangleq \sqrt{|\mathcal{I}(u)|} = \frac{\sqrt{L}}{u} \tag{A.17}$$

Bayesian joint likelihood

With Jeffreys prior, we can derive the Bayesian joint likelihood as follows:

$$\mathcal{Q}_B(v_1, v_2) = \int p(v_1|t_1 = t)p(v_2|t_2 = t)p(t_{12} = t)\mathrm{d}t = \int \left(\frac{L^L v_1^{L-1} e^{-\frac{Lv_1}{t}}}{\Gamma(L)t^L}\right)\left(\frac{L^L v_2^{L-1} e^{-\frac{Lv_2}{t}}}{\Gamma(L)t^L}\right)\left(\frac{\sqrt{L}}{t}\right)\mathrm{d}t \tag{A.18}$$

$$= \frac{L^{2L+1/2} v_1^{L-1} v_2^{L-1}}{\Gamma(L)^2}\int \frac{e^{-\frac{L(v_1+v_2)}{t}}}{t^{2L+1}}\mathrm{d}t = \frac{L^{2L+1/2} v_1^{L-1} v_2^{L-1}}{\Gamma(L)^2}\frac{\Gamma(2L)}{(L(v_1+v_2))^{2L}} \tag{A.19}$$

$$= \frac{\sqrt{L}\Gamma(2L)}{\Gamma(L)^2}\left(\frac{1}{v_1 v_2}\left(\frac{v_1 v_2}{(v_1+v_2)^2}\right)^L\right) \tag{A.20}$$

by using

$$\int \frac{e^{-\frac{A}{t}}}{t^N}\mathrm{d}t = \frac{\Gamma(N-1)}{A^{N-1}}\,. \tag{A.21}$$

Bayesian likelihood ratio

Let $\mathcal{D}_B$ be the denominator term appearing in the Bayesian likelihood ratio and expressed as:

$$\mathcal{D}_B(v) = \int p(v|u = t)p(u = t)\mathrm{d}t = \int \left(\frac{L^L v^{L-1} e^{-\frac{Lv}{t}}}{\Gamma(L)t^L}\right)\left(\frac{\sqrt{L}}{t}\right)\mathrm{d}t \tag{A.22}$$

$$= \frac{L^{L+1/2} v^{L-1}}{\Gamma(L)}\int \frac{e^{-\frac{Lv}{t}}}{t^{L+1}}\mathrm{d}t = \frac{L^{L+1/2} v^{L-1}}{\Gamma(L)}\frac{\Gamma(L)}{(Lv)^L} = \frac{\sqrt{L}}{v}\,. \tag{A.23}$$

Using the expression of $\mathcal{Q}_B(v_1, v_2)$ and $\mathcal{D}_B(v)$, it comes that the Bayesian likelihood ratio is given by:

$$\mathcal{L}_B = \frac{\mathcal{Q}_B(v_1, v_2)}{\mathcal{D}_B(v_1)\mathcal{D}_B(v_2)} = \frac{\frac{\sqrt{L}\Gamma(2L)}{\Gamma(L)^2}\frac{v_1^{L-1}v_2^{L-1}}{(v_1+v_2)^{2L}}}{\frac{\sqrt{L}}{v}\frac{\sqrt{L}}{v}} = \frac{\Gamma(2L)}{\sqrt{L}\Gamma(L)^2}\left(\frac{v_1 v_2}{(v_1+v_2)^2}\right)^L\,. \tag{A.24}$$

Mutual information kernel

Using the expression of $\mathcal{Q}_B(v_1, v_2)$ and $\mathcal{Q}_B(v, v)$, it results that the mutual information kernel is given by:

$$\mathcal{K}_B(v_1, v_2) = \frac{\mathcal{Q}_B(v_1, v_2)}{\sqrt{\mathcal{Q}_B(v_1, v_1)\mathcal{Q}_B(v_2, v_2)}} = \frac{\frac{\sqrt{L}\Gamma(2L)}{\Gamma(L)^2}\frac{v_1^{L-1}v_2^{L-1}}{(v_1+v_2)^{2L}}}{\sqrt{\frac{\sqrt{L}\Gamma(2L)}{\Gamma(L)^2}\frac{v_1^{2L-2}}{(2v_1)^{2L}}\frac{\sqrt{L}\Gamma(2L)}{\Gamma(L)^2}\frac{v_2^{2L-2}}{(2v_2)^{2L}}}} = 2^{2L}\left(\frac{v_1 v_2}{(v_1+v_2)^2}\right)^L . \tag{A.25}$$

Maximal joint likelihood

The priorless extension of $\mathcal{Q}_B(v_1, v_2)$, i.e. the maximum joint likelihood is obtained as follows:

$$\mathcal{Q}_G(v_1, v_2) = \int \sup_t p(v_1|t_1 = t)p(v_2|t_2 = t)\mathrm{d}t = \left(\frac{2^L L^L v_1^{L-1} e^{-\frac{2Lv_1}{v_1+v_2}}}{\Gamma(L)(v_1+v_2)^L}\right)\left(\frac{2^L L^L v_2^{L-1} e^{-\frac{2Lv_2}{v_1+v_2}}}{\Gamma(L)(v_1+v_2)^L}\right) \tag{A.26}$$

$$= \frac{2^{2L}L^{2L}v_1^{L-1}v_2^{L-1}e^{-2L}}{\Gamma(L)^2(v_1+v_2)^{2L}} = \frac{2^{2L}L^{2L}e^{-2L}}{\Gamma(L)^2}\left(\frac{1}{v_1 v_2}\left(\frac{v_1 v_2}{(v_1+v_2)^2}\right)^L\right) . \tag{A.27}$$

since under Gamma noise the MLE is the mean.

Generalized likelihood ratio

Let $\mathcal{D}_G$ be the denominator term appearing in the generalized likelihood ratio and expressed as:

$$\mathcal{D}_G(v) = \sup_t p(v|u = t) = \frac{L^L e^{-L}}{\Gamma(L)v} . \tag{A.28}$$

Using the expression of $\mathcal{Q}_G(v_1, v_2)$ and $\mathcal{D}_G(v)$, it results that the generalized likelihood ratio is given by:

$$\mathcal{L}_G = \frac{\mathcal{Q}_G(v_1, v_2)}{\mathcal{D}_G(v_1)\mathcal{D}_G(v_2)} = \frac{\frac{2^{2L}L^{2L}e^{-2L}}{\Gamma(L)}\frac{v_1^{L-1}v_2^{L-1}}{(v_1+v_2)^{2L}}}{\frac{L^L e^{-L}}{\Gamma(L)v}\frac{L^L e^{-L}}{\Gamma(L)v}} = 2^{2L}\left(\frac{v_1 v_2}{(v_1+v_2)^2}\right)^L . \tag{A.29}$$

Variance stabilization criterion

Variance stabilization of Gamma random values can be performed using a log transform:

$$\boldsymbol{s}(V) = \log V \Rightarrow \mathrm{Var}[\boldsymbol{s}(V)] = \mathrm{Var}[\log V] = \Psi(1, L) \tag{A.30}$$

where $\Psi(1, L)$ is the first-order Polygamma function of L (e.g. Xie et al., 2002b). The resulting similarity criterion is then given by:

$$\mathcal{S}(v_1, v_2) = \exp\left[-\left(\log\frac{v_1}{v_2}\right)^2\right] . \tag{A.31}$$

A.1.3 Poisson noise case

A Poisson random variable V can be described by the following pdf:

$$p(v|u) = \frac{u^v e^{-u}}{v!} \,. \tag{A.32}$$

Its expectation is $\mathbb{E}[v] = u$ and variance $\mathrm{Var}[V] = u$. Note that the relation $\mathrm{Var}[V] = \mathbb{E}[v]$ is non-homogeneous, which is challenging, since, as a consequence, V cannot be related to u through additive or multiplicative decomposition.

Fisher information

Fisher information associated with a Poissonian pdf is given by:

$$\mathcal{I}(u) = E\left[\left(\frac{\partial}{\partial u}\log p(v|u)\right)^2 |u\right] = \int \left(\frac{\partial}{\partial u}\log p(v|u)\right)^2 p(v|u)\mathrm{d}v \tag{A.33}$$

$$= \int \left(\frac{v}{u} - 1\right)^2 \frac{u^v e^{-u}}{v!}\mathrm{d}v = \underbrace{\frac{1}{u^2}\int (v-u)^2 \frac{u^v e^{-u}}{v!}\mathrm{d}v}_{\text{by definition of variance}} = \frac{u}{u^2} = \frac{1}{u} \,. \tag{A.34}$$

Jeffreys' prior

The corresponding Jeffreys' prior is:

$$p(u) \triangleq \sqrt{|\mathcal{I}(u)|} = \frac{1}{\sqrt{u}} \,. \tag{A.35}$$

Bayesian joint likelihood

With Jeffreys' prior, we can derive the Bayesian joint likelihood as follow:

$$\mathcal{Q}_B(v_1, v_2) = \int p(v_1|u_1 = t)p(v_2|u_2 = t)p(u_{12} = t)\mathrm{d}t = \int \left(\frac{t^{v_1}e^{-t}}{v_1!}\right)\left(\frac{t^{v_2}e^{-t}}{v_2!}\right)\left(\frac{1}{\sqrt{t}}\right)\mathrm{d}t \tag{A.36}$$

$$= \frac{1}{v_1! v_2!}\int t^{v_1+v_2-1/2}e^{-2t}\mathrm{d}t = \frac{1}{\sqrt{2}}\frac{\Gamma(v1+v2+1/2)}{2^{v1+v2}v_1!v_2!} \tag{A.37}$$

by using

$$\int t^N e^{-At}\mathrm{d}t = \int \frac{e^{-\frac{A}{t}}}{t^{N+2}}\mathrm{d}t = \frac{\Gamma(N+1)}{A^{N+1}} \,. \tag{A.38}$$

Bayesian likelihood ratio

Let $\mathcal{D}_B$ be the denominator term appearing in the Bayesian likelihood ratio and expressed as:

$$\mathcal{D}_B(v) = \int p(v|u = t)p(u = t)\mathrm{d}t = \int \left(\frac{t^v e^{-t}}{v!}\right)\left(\frac{1}{\sqrt{t}}\right)\mathrm{d}t = \frac{1}{v!}\int t^{v-1/2}e^{-t}\mathrm{d}t \tag{A.39}$$

$$= \frac{\Gamma(v+1/2)}{v!} \,. \tag{A.40}$$

Using the expression of $\mathcal{Q}_B(v_1, v_2)$ and $\mathcal{D}_B(v)$, it results that the Bayesian likelihood ratio is given by:

$$\mathcal{L}_B = \frac{\mathcal{Q}_B(v_1, v_2)}{\mathcal{D}_B(v_1)\mathcal{D}_B(v_2)} = \frac{\frac{1}{\sqrt{2}}\frac{\Gamma(v_1+v_2+1/2)}{2^{v_1+v_2}v_1!v_2!}}{\frac{\Gamma(v_1+1/2)}{v_1!}\frac{\Gamma(v_2+1/2)}{v_2!}} = \frac{1}{\sqrt{2}}\frac{\Gamma(v_1+v_2+1/2)}{2^{v_1+v_2}\Gamma(v_1+1/2)\Gamma(v_2+1/2)} \,. \tag{A.41}$$

Mutual information kernel

Using the expression of $\mathcal{Q}_B(v_1, v_2)$ and $\mathcal{Q}_B(v, v)$, the mutual information kernel can be written as:

$$\mathcal{K}_B(v_1, v_2) = \frac{\mathcal{Q}_B(v_1, v_2)}{\sqrt{\mathcal{Q}_B(v_1, v_1)\mathcal{Q}_B(v_2, v_2)}} = \frac{\frac{1}{\sqrt{2}}\frac{\Gamma(v1+v2+1/2)}{2^{v1+v2}v_1!v_2!}}{\sqrt{\frac{1}{\sqrt{2}}\frac{\Gamma(2v_1+1/2)}{2^{2v_1}v_1!^2}\frac{1}{\sqrt{2}}\frac{\Gamma(2v_2+1/2)}{2^{2v_2}v_2!^2}}} = \frac{\Gamma(v_1+v_2+1/2)}{\sqrt{\Gamma(2v_1+1/2)\Gamma(2v_2+1/2)}} \,. \tag{A.42}$$

Maximal joint likelihood

The priorless extension of $\mathcal{Q}_B(v_1, v_2)$, i.e. the maximum joint likelihood is obtained as follows:

$$\mathcal{Q}_G(v_1, v_2) = \int \sup_t p(v_1|t_1 = t)p(v_2|t_2 = t)\mathrm{d}t = \left(\frac{\frac{(v_1+v_2)^{v_1}}{2^{v_1}}e^{-\frac{v_1+v_2}{2}}}{v_1!}\right)\left(\frac{\frac{(v_1+v_2)^{v_2}}{2^{v_2}}e^{-\frac{v_1+v_2}{2}}}{v_2!}\right) \tag{A.43}$$

$$= \frac{(v_1+v_2)^{v_1+v_2}}{(2e)^{v_1+v_2}v_1!v_2!} \,. \tag{A.44}$$

since once again, the MLE for Poisson noise is the mean.

Generalized likelihood ratio

Let $\mathcal{D}_G$ be the denominator term appearing in the generalized likelihood ratio and expressed as:

$$\mathcal{D}_G(v) = \sup_t p(v|u = t) = \frac{v^x e^{-v}}{v!} \,. \tag{A.45}$$

Using the expression of $\mathcal{Q}_G(v_1, v_2)$ and $\mathcal{D}_G(v)$, it comes that the generalized likelihood ratio is:

$$\mathcal{L}_G = \frac{\mathcal{Q}_G(v_1, v_2)}{\mathcal{D}_G(v_1)\mathcal{D}_G(v_2)} = \frac{\frac{(v_1+v_2)^{v_1+v_2}}{(2e)^{v_1+v_2}v_1!v_2!}}{\frac{v_1^{v_1}e^{-v_1}}{v_1!}\frac{v_2^{v_2}e^{-v_2}}{v_2!}} = \frac{(v_1+v_2)^{v_1+v_2}}{2^{v_1+v_2}v_1^{v_1}v_2^{v_2}} \,. \tag{A.46}$$

Variance stabilization criterion

Approximated variance stabilization of Poisson random values can be performed using Anscombe transform:

$$\boldsymbol{s}(V) = 2\sqrt{V + \frac{3}{8}} \Rightarrow (u \gg 0 \Rightarrow \mathrm{Var}[\boldsymbol{s}(V)] = 1) \,. \tag{A.47}$$

The resulting similarity criterion is then given by:

$$\mathcal{S}(v_1, v_2) = \exp\left(-4\left(\sqrt{v_1 + \frac{3}{8}} - \sqrt{v_2 + \frac{3}{8}}\right)^2\right) . \tag{A.48}$$

A.1.4 Cauchy noise case

A Cauchy random variable V can be described by the following pdf:

$$p(v|u) = \frac{1}{\pi\gamma\left[1+\left(\frac{v-u}{\gamma}\right)^2\right]} . \tag{A.49}$$

where u is the mode and γ is a shape parameter. Cauchy fluctuations are additive, it is straightforward to show that V can be decomposed as $u+N$ with N a Cauchy random variable with a mode in 0 and the scale parameter γ. The particularity of Cauchy random variables is that their expectation and variance do not exist. A consequence is that the sample mean and the sample variance do not converge wrt the number of observations. Surprisingly, all criteria are still defined in this case, except the variance stabilization criterion since we have not found a transformation g such as $g(V)$ has a finite and constant variance whatever u.

Fisher information

Fisher information associated with a Cauchy pdf is given by:

$$\mathcal{I}(u) = E\left[\left(\frac{\partial}{\partial u}\log p(v|u)\right)^2 |u\right] = \int \left(\frac{\partial}{\partial u}\log p(v|u)\right)^2 p(v|u)\mathrm{d}v \tag{A.50}$$

$$= \int \left(\frac{2(v-u)}{\gamma^2\left[1+\left(\frac{v-u}{\gamma}\right)^2\right]}\right)^2 \frac{1}{\pi\gamma\left[1+\left(\frac{v-u}{\gamma}\right)^2\right]}\mathrm{d}v = \frac{1}{2\gamma^2} . \qquad \text{(derived with Maple)} \tag{A.51}$$

Jeffreys' prior

Fisher information gives Jeffreys' prior as:

$$p(u) \triangleq \sqrt{|\mathcal{I}(u)|} = \frac{1}{\sqrt{2}\gamma} . \tag{A.52}$$

Bayesian joint likelihood

With Jeffreys' prior, we can derive the Bayesian joint likelihood as follows:

$$\mathcal{Q}_B(v_1,v_2) = \int p(v_1|u_1=t)p(v_2|u_2=t)p(u_{12}=t)\mathrm{d}t \tag{A.53}$$

$$= \int \left(\frac{1}{\pi\gamma\left[1+\left(\frac{v_1-t}{\gamma}\right)^2\right]}\right)\left(\frac{1}{\pi\gamma\left[1+\left(\frac{v_2-t}{\gamma}\right)^2\right]}\right)\left(\frac{1}{\sqrt{2}\gamma}\right)\mathrm{d}t \tag{A.54}$$

$$= \frac{\sqrt{2}}{\pi\gamma^2\left[4+\left(\frac{v_1-v_2}{\gamma}\right)^2\right]} . \tag{A.55}$$

Bayesian likelihood ratio

Let $\mathcal{D}_B$ be the denominator term appearing in the Bayesian likelihood ratio and expressed as:

$$\mathcal{D}_B(v) = \int p(v|u=t)p(u=t)\mathrm{d}t = \int \left(\frac{1}{\pi\gamma\left[1+\left(\frac{v-t}{\gamma}\right)^2\right]} \right) \left(\frac{1}{\sqrt{2}\gamma} \right) \mathrm{d}t = \frac{1}{\sqrt{2}\gamma} \,. \tag{A.56}$$

Using the expression of $\mathcal{Q}_B(v_1, v_2)$ and $\mathcal{D}_B(v)$, it results that the Bayesian likelihood ratio is given by:

$$\mathcal{L}_B = \frac{\mathcal{Q}_B(v_1, v_2)}{\mathcal{D}_B(v_1)\mathcal{D}_B(v_2)} = \frac{\frac{\sqrt{2}}{\pi\gamma^2\left[4+\left(\frac{v_1-v_2}{\gamma}\right)\right]}}{\frac{1}{\sqrt{2}\gamma}\frac{1}{\sqrt{2}\gamma}} = \frac{2\sqrt{2}}{\pi\left[4+\left(\frac{v_1-v_2}{\gamma}\right)^2\right]} \,. \tag{A.57}$$

Mutual information kernel

Using the expression of $\mathcal{Q}_B(v_1, v_2)$ and $\mathcal{Q}_B(v, v)$, it results that the mutual information kernel is given by:

$$\mathcal{K}_B(v_1, v_2) = \frac{\mathcal{Q}_B(v_1, v_2)}{\sqrt{\mathcal{Q}_B(v_1, v_1)\mathcal{Q}_B(v_2, v_2)}} = \frac{\frac{\sqrt{2}}{\pi\gamma^2\left[4+\left(\frac{v_1-v_2}{\gamma}\right)^2\right]}}{\sqrt{\frac{\sqrt{2}}{\pi\gamma^2\left[4+\left(\frac{v_1-v_1}{\gamma}\right)^2\right]}\frac{\sqrt{2}}{\pi\gamma^2\left[4+\left(\frac{v_2-v_2}{\gamma}\right)^2\right]}}} = \frac{1}{1+\left(\frac{v_1-v_2}{2\gamma}\right)^2} \tag{A.58}$$

Maximal joint likelihood

The priorless extension of $\mathcal{Q}_B(v_1, v_2)$, i.e. the maximum joint likelihood is obtained as follows:

$$\mathcal{Q}_G(v_1, v_2) = \int \sup_t p(v_1|t_1=t)p(v_2|t_2=t)\mathrm{d}t = \frac{1}{\pi\gamma\left[1+\left(\frac{v_1-\frac{v_1+v_2}{2}}{\gamma}\right)^2\right]}\frac{1}{\pi\gamma\left[1+\left(\frac{v_1-\frac{v_1+v_2}{2}}{\gamma}\right)^2\right]} \tag{A.59}$$

$$= \frac{1}{\pi^2\gamma^2\left[1+\left(\frac{v_1-v_2}{2\gamma}\right)^2\right]^2} \tag{A.60}$$

$$\tag{A.61}$$

since for a dataset of one or two elements the mean is the MLE (note that it is no longer the case for larger datasets).

Generalized likelihood ratio

Let $\mathcal{D}_G$ be the denominator term appearing in the generalized likelihood ratio and expressed as:

$$\mathcal{D}_G(v) = \sup_t p(v|u=t) = \frac{1}{\pi\gamma\left[1+\left(\frac{v-v}{\gamma}\right)^2\right]} = \frac{1}{\pi\gamma} \,. \tag{A.62}$$

Using the expression of $\mathcal{Q}_G(v_1, v_2)$ and $\mathcal{D}_G(v)$, it results that the generalized likelihood ratio is given by:

$$\mathcal{L}_G = \frac{\mathcal{Q}_G(v_1, v_2)}{\mathcal{D}_G(v_1)\mathcal{D}_G(v_2)} = \frac{\frac{1}{\pi^2\gamma^2\left[1+\left(\frac{v_1-v_2}{2\gamma}\right)^2\right]^2}}{\frac{1}{\pi\gamma}\frac{1}{\pi\gamma}} = \frac{1}{\left[1+\left(\frac{v_1-v_2}{2\gamma}\right)^2\right]^2} \,. \tag{A.63}$$

Variance stabilization criterion

Cauchy random variables have neither expectation nor variance. Our attempts to transform Cauchy r.v. into random variables with constant variance did not succeed.

A.2 Proof sketches for similarity criteria properties

A.2.1 Bayesian joint likelihood

× **Max. self-similarity:** Assume V is Gamma distributed with $L = 1$ and $v_1 = 2v_2$:

$$\mathcal{Q}_B(v_1, v_2) = \frac{1}{(v_1+v_2)^2} = \frac{1}{9v_2^2} > \frac{1}{16v_2^2} = \frac{1}{(2v_2+2v_2)^2} = \frac{1}{(v_1+v_1)^2} = \mathcal{Q}_B(v_1, v_1) \tag{A.64}$$

which breaks the property of max. self-similarity.

× **Eq. self-similarity:** Assume V is Gamma distributed with $L = 1$ and $v_1 = 2v_2$:

$$\mathcal{Q}_B(v_1, v_1) = \frac{1}{(v_1+v_1)^2} = \frac{1}{(2v_2+2v_2)^2} = \frac{1}{16v_2^2} < \frac{1}{4v_2^2} = \frac{1}{(v_2+v_2)^2} = \mathcal{Q}_B(v_2, v_2) \tag{A.65}$$

which breaks the property of eq. self-similarity.

× **Id. of indiscernible:** It requires the eq. self-similarity property.

× **Invariance:** Assume V is Gamma distributed with $L = 1$ and consider $V' = \sqrt{V}$, i.e., the mapping function $g(.) = \sqrt{.}$, then:

$$\mathcal{Q}_{BV_1,V_2}(v_1, v_2) = \frac{1}{(v_1+v_1)^2} \tag{A.66}$$

$$\mathcal{Q}_{BV_1',V_2'}(\sqrt{v_1}, \sqrt{v_2}) = \int p(\sqrt{v_1}|u_1 = t)p(\sqrt{v_2}|u_2 = t)p(u_{12} = t)\mathrm{d}t \tag{A.67}$$

$$= \left|\frac{\mathrm{d}\sqrt{v_1}}{\mathrm{d}v_1}\right|^{-1}\left|\frac{\mathrm{d}\sqrt{v_2}}{\mathrm{d}v_2}\right|^{-1}\int p(v_1|u_1 = t)p(v_2|u_2 = t)p(u_{12} = t)\mathrm{d}t \tag{A.68}$$

$$= 4\sqrt{v_1 v_2}\mathcal{Q}_{BV_1,V_2}(v_1, v_2) \,. \tag{A.69}$$

The equality does not hold for any value $v_1 > 0$ or $v_2 > 0$.

× **Asymp. CFAR:** The closed-from expression of $\mathcal{Q}_B$ obtained for Gamma distribution is clearly not asymptotically CFAR, since the expectation of the similarity criterion is inversely proportional to the underlying parameters.

× **Asymp. UMPI:** $\mathcal{L}_G$ being UMPI, it defeats $\mathcal{S}$.

A.2.2 Maximum joint likelihood

Since $\mathcal{Q}_G$ corresponds to $\mathcal{Q}_B$ in the Gamma case, we can use the same counter-examples as above.

A.2.3 Bayesian likelihood ratio

× **Max. self-similarity:** Assume V to take values in $\{v_1, v_2, v_3\}$ and $u \in \{a, b, c\}$. Assume the distribution of V to be defined by:

$$p(v_1|a) = 5/8 \qquad p(v_1|b) = 2/8 \qquad p(v_1|c) = 1/8 \tag{A.70}$$
$$p(v_2|a) = 2/8 \qquad p(v_2|b) = 4/8 \qquad p(v_2|c) = 3/8 \tag{A.71}$$
$$p(v_3|a) = 1/8 \qquad p(v_3|b) = 2/8 \qquad p(v_3|c) = 4/8\,. \tag{A.72}$$

Note that the observations are statistically identifiable through their likelihood and their MLE. Assume $p(u)$ to be described by

$$p(u = a) = 0/2 \tag{A.73}$$
$$p(u = b) = 1/2 \tag{A.74}$$
$$p(u = c) = 1/2\,. \tag{A.75}$$

The self Bayesian likelihood ratio for v_2 is given by

$$\mathcal{L}_B(v_2, v_2) = \frac{p(v_2|a)^2 p(a) + p(v_2|b)^2 p(b) + p(v_2|c)^2 p(c)}{(p(v_2|a)p(a) + p(v_2|b)p(b) + p(v_2|c)p(c))^2} = \frac{\frac{2\times2\times0}{8\times8\times2} + \frac{4\times4\times1}{8\times8\times2} + \frac{3\times3\times1}{8\times8\times2}}{(\frac{2\times0}{8\times2} + \frac{4\times1}{8\times2} + \frac{3\times1}{8\times2})^2} = \frac{50}{49}\,. \tag{A.76}$$

The Bayesian likelihood ratio between v_1 and v_2 is given by

$$\mathcal{L}_B(v_1, v_2) = \frac{(p(v_1|a)p(v_2|a)p(a) + p(v_1|b)p(v_2|b)p(b) + p(v_1|c)p(v_2|c)p(c))}{(p(v_1|a)p(a) + p(v_1|b)p(b) + p(v_1|c)p(c))(p(v_2|a)p(a) + p(v_2|b)p(b) + p(v_2|c)p(c))} \tag{A.77}$$

$$= \frac{\frac{5\times2\times0}{8\times8\times2} + \frac{2\times4\times1}{8\times8\times2} + \frac{1\times3\times1}{8\times8\times2}}{(\frac{5\times0}{8\times2} + \frac{2\times1}{8\times2} + \frac{1\times1}{8\times2})(\frac{2\times0}{8\times2} + \frac{4\times1}{8\times2} + \frac{3\times1}{8\times2})} = \frac{22}{21}\,. \tag{A.78}$$

Since $50/49 < 22/21$ then $\mathcal{L}_B(v_2, v_2) < \mathcal{L}_B(v_1, v_2)$. The max. self-similarity does not hold.

Open question: what are the sufficient and necessary conditions on the likelihood p to ensure the max. self similarity of $\mathcal{L}_B$?

× **Eq. self-similarity:** Consider the case of Poisson noise, the eq. self similarity is given by:

$$\mathcal{L}_B(v, v) = \frac{1}{\sqrt{2}} \frac{\Gamma(v + v + 1/2)}{2^{v+v}\Gamma(v + 1/2)\Gamma(v + 1/2)} = \frac{1}{\sqrt{2}} \frac{\Gamma(2v + 1/2)}{2^{2v}\Gamma(v + 1/2)^2} \tag{A.79}$$

which depends, as illustrated on Figure A.1, on the value of v.

× **Id. of indiscernible:** It requires the eq. self-similarity property.

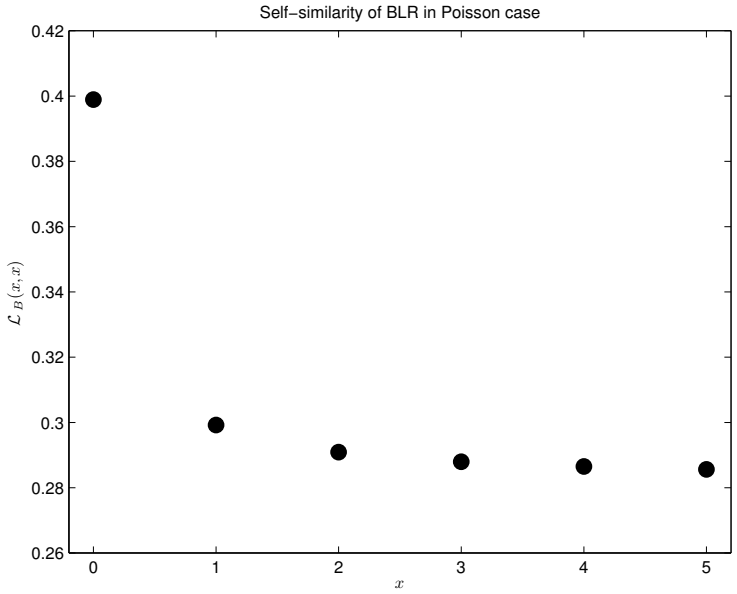


Figure A.1: Self Bayesian likelihood ratio $\mathcal{L}_B(v,x)$ with respect to the value v in the case of Poisson noise.

√ **Invariance:** Let g be an invertible and differentiable mapping function of the rv V to V', then:

$$\mathcal{L}_{BV_1',V_2'}(g(v_1),g(v_2)) = \frac{\int p(g(v_1)|u_{12}=t)p(g(v_2)|u_{12}=t)p(u_{12}=t)\,\mathrm{d}t}{\int p(g(v_1)|u_1=t)p(u_1=t)\,\mathrm{d}t \int p(g(v_2)|u_2=t)p(u_2=t)\,\mathrm{d}t} \tag{A.80}$$

$$= \frac{\left|\frac{\mathrm{d}g(v_1)}{\mathrm{d}v_1}\right|^{-1}\left|\frac{\mathrm{d}g(v_2)}{\mathrm{d}v_2}\right|^{-1}\int p(v_1|u_{12}=t)p(v_2|u_{12}=t)p(u_{12}=t)\,\mathrm{d}t}{\left|\frac{\mathrm{d}g(v_1)}{\mathrm{d}v_1}\right|^{-1}\left|\frac{\mathrm{d}g(v_2)}{\mathrm{d}v_2}\right|^{-1}\int p(v_1|u_1=t)p(u_1=t)\,\mathrm{d}t \int p(v_2|u_2=t)p(u_2=t)\,\mathrm{d}t} \tag{A.81}$$

$$= \mathcal{L}_{BV_1,V_2}(v_1,v_2) \tag{A.82}$$

The Bayesian likelihood ratio fulfils the invariance property.

× **Asymp. CFAR:** We can always choose a prior on the underlying parameters, favouring the similarity for a range of underlying parameters, implying that $\mathcal{L}_B$ would not be CFAR.

× **Asymp. UMPI:** $\mathcal{L}_G$ being UMPI, it defeats $\mathcal{S}$.

A.2.4 Generalized likelihood ratio

√ **Eq. self-similarity:** The self generalized likelihood ratio is always equal to one:

$$\mathcal{L}_G(v,v) = \frac{\sup_t p(v|u=t)^2}{\left(\sup_t p(v|u=t)\right)^2} = 1 \tag{A.83}$$

since the superior bound is reached at the same value(s) t for $p(v|u=t)$ and $p(v|u=t)^2$.

√ **Max. self-similarity:** The superior bound of a product is always inferior to the product of the superior bounds, then:

$$\mathcal{L}_G(v_1, v_2) = \frac{\sup_t p(v_1|u_1 = t)p(v_2|u_2 = t)}{\sup_t p(v_1|u_1 = t) \sup_t p(v_2|u_2 = t)} \leq 1 \,. \tag{A.84}$$

√ [1] **Id. of indiscernible:** Assume the observations are statistically identifiable through their MLE. Let two observations $v_1 \neq v_2$. Let $\hat{t}_1$ and $\hat{t}_2$ be respectively the maximum likelihood estimates of v_1 and v_2, and $\hat{t}_{12}$ be the maximum likelihood estimator of $\{v_1, v_2\}$. Since $v_1 \neq v_2$ and observations are statistically identifiable through their MLE, $\hat{t}_1 \neq \hat{t}_2$. Since the MLE is unique, then, either:

$$p(v_1|u_1 = \hat{t}_1) \geq p(v_1|u_1 = \hat{t}_{12}) > 0 \tag{A.85}$$
$$p(v_2|u_2 = \hat{t}_2) > p(v_2|u_2 = \hat{t}_{12}) > 0 \tag{A.86}$$

or

$$p(v_1|u_1 = \hat{t}_1) > p(v_1|u_1 = \hat{t}_{12}) > 0 \tag{A.87}$$
$$p(v_2|u_2 = \hat{t}_2) \geq p(v_2|u_2 = \hat{t}_{12}) > 0 \tag{A.88}$$

Then, in any case, $p(v_1|u_1 = \hat{t}_1)p(v_2|u_2 = \hat{t}_1) > p(v_1|u_1 = \hat{t}_{12})p(v_2|u_2 = \hat{t}_{12})$, i.e., $v_1 \neq v_2 \Rightarrow \mathcal{L}_G(v_1, v_2) < 1$.

√ **Invariance:** Let g be an invertible and differentiable mapping function of the rv V to V', then:

$$\mathcal{L}_{GV_1',V_2'}(g(v_1), g(v_2)) = \frac{\sup_t p(g(v_1)|u_1 = t)p(g(v_2)|u_2 = t)}{\sup_t p(g(v_1)|u_1 = t) \sup_t p(g(v_2)|u_2 = t)} \tag{A.89}$$
$$= \frac{\left|\frac{\mathrm{d}g(v_1)}{\mathrm{d}v_1}\right|^{-1}\left|\frac{\mathrm{d}g(v_2)}{\mathrm{d}v_2}\right|^{-1} \sup_t p(v_1|u_1 = t)p(v_2|u_2 = t)}{\left|\frac{\mathrm{d}g(v_1)}{\mathrm{d}v_1}\right|^{-1}\left|\frac{\mathrm{d}g(v_2)}{\mathrm{d}v_2}\right|^{-1} \sup_t p(v_1|u_1 = t) \sup_t p(v_2|u_2 = t)} \tag{A.90}$$
$$= \mathcal{L}_{GV_1,V_2}(v_1, v_2) \tag{A.91}$$

The generalized likelihood ratio fulfils the invariance property (see also Kay and Gabriel, 2003)).

√ **Asymp. CFAR:** According to (Kay, 1998).

√ **Asymp. UMPI:** Due to its convergence to the likelihood ratio $\mathcal{L}$, which is Neyman-Pearson optimal, $\mathcal{L}_G$ is UMPI (Lehmann, 1959).

A.2.5 Mutual information kernel

√ **Eq. self-similarity:** The self mutual information kernel is always equal to one:

$$\mathcal{K}_B(v, v) = \frac{\mathcal{Q}_B(v, v)}{\sqrt{\mathcal{Q}_B(v, v)\mathcal{Q}_B(v, v)}} = 1 \,. \tag{A.92}$$

√ **Max. self-similarity:** This property derived directly from the Cauchy-Schwartz inequality.

[1] Holds true under the assumption that the observations are statistically identifiable through their MLE.

✓ [2] **Id. of indiscernible:** Assume the observations are statistically identifiable through their likelihood. See u as a random variable with distribution $p(u)$. Let $P_1 = p(v_1|u)$ and $P_2 = p(v_2|u)$ be the two r.v. resulting of the evaluation of the likelihood of the r.v. u. We can rewrite the mutual information kernel as the correlation between P_1 and P_2:

$$\mathcal{K}_B(v, v) = \frac{\mathbb{E}[P_1 P_2]}{\sqrt{\mathbb{E}[P_1^2]\mathbb{E}[P_2^2]}} \tag{A.93}$$

We get that the mutual information is maximal if the correlation between P_1 and P_2 is equal to one:

$$\mathcal{K}_B(v, v) = 1 \Rightarrow \quad \frac{\mathbb{E}[P_1 P_2]}{\sqrt{\mathbb{E}[P_1^2]\mathbb{E}[P_2^2]}} = 1 \tag{A.94}$$

i.e., for all u, $p(v_1|u) = a\, p(v_2|u)$ with $a > 0$ since a pdf is a positive function. Under normalization constraint and since the observations are statistically identifiable through their likelihood, $v_1 = v_2$.

✓ **Invariance:** Let g be an invertible and differentiable mapping function of the rv V to V', then:

$$\mathcal{Q}_{GV_1', V_2'}(g(v_1), g(v_2)) = \frac{\int p(g(v_1)|u_1 = t)p(g(v_2)|u_2 = t)p(u_{12} = t)\mathrm{d}t}{\sqrt{\int p(g(v_1)|u_1 = t)^2 p(u_{12} = t)\mathrm{d}t \int p(g(v_2)|u_2 = t)^2 p(u_{12} = t)\mathrm{d}t}} \tag{A.95}$$

$$= \frac{\left|\frac{\mathrm{d}g(v_1)}{\mathrm{d}v_1}\right|^{-1}\left|\frac{\mathrm{d}g(v_2)}{\mathrm{d}v_2}\right|^{-1}\int p(v_1|u_1 = t)p(v_2|u_2 = t)p(u_{12} = t)\mathrm{d}t}{\left|\frac{\mathrm{d}g(v_1)}{\mathrm{d}v_1}\right|^{-1}\left|\frac{\mathrm{d}g(v_2)}{\mathrm{d}v_2}\right|^{-1}\sqrt{\int p(v_1|u_1 = t)^2 p(u_{12} = t)\mathrm{d}t \int p(v_2|u_2 = t)^2 p(u_{12} = t)\mathrm{d}t}} \tag{A.96}$$

$$= \mathcal{Q}_{GV_1, V_2}(v_1, v_2) \tag{A.97}$$

The mutual information kernel fulfils the invariance property.

× **Asymp. CFAR:** We can always choose a prior on the underlying parameters, favouring the similarity for a range of underlying parameters, implying that $\mathcal{Q}_G$ would not be CFAR.

× **Asymp. UMPI:** $\mathcal{L}_G$ being UMPI, it defeats $\mathcal{S}$.

A.2.6 Variance stabilization criterion

It is important to note that all properties below require that a variance stabilizer $\boldsymbol{s}$ exists.

✓ **Eq. self-similarity:** Thanks to the Gaussian kernel, the self similarity of $\mathcal{S}$ is always equal to one:

$$\mathcal{S}(v, v) = \exp\left(\frac{\|s(v) - s(v)\|_2^2}{h}\right) = 1\,. \tag{A.98}$$

[2] Holds true under the assumption that the observations are statistically identifiable through their likelihood.

√ **Max. self-similarity:** This property follows from the property of the Euclidean distance:

$$\begin{aligned}
& \|s(v_1) - s(v_2)\|_2^2 \geq 0 && \text{(A.99)}\\
\Leftrightarrow\quad & \|s(v_1) - s(v_2)\|_2^2 \geq \|s(v_1) - s(v_1)\|_2^2 && \text{(A.100)}\\
\Leftrightarrow\quad & \exp\left(-\frac{\|s(v_1) - s(v_2)\|_2^2}{h}\right) \leq \exp\left(-\frac{\|s(v_1) - s(v_1)\|_2^2}{h}\right) && \text{(A.101)}\\
\Leftrightarrow\quad & \mathcal{S}(v_1, v_2) \leq \mathcal{S}(v_1, v_1)\ . && \text{(A.102)}
\end{aligned}$$

√ **Id. of indiscernible:** This property is obtained as follows:

$$\begin{aligned}
& \mathcal{S}(v_1, v_2) = 1 && \text{(A.103)}\\
\Rightarrow\quad & \exp\left(-\frac{\|s(v_1) - s(v_2)\|_2^2}{h}\right) = 1 && \text{(A.104)}\\
\Rightarrow\quad & \|s(v_1) - s(v_2)\|_2^2 = 0 && \text{(A.105)}\\
\Rightarrow\quad & s(v_1) = s(v_2) && \text{(A.106)}\\
\Rightarrow\quad & v_1 = v_2 \quad \text{since } s \text{ is invertible}\ . && \text{(A.107)}
\end{aligned}$$

√ **Invariance:** If s stabilizes the variance of V then $s \circ g^{-1}$ stabilizes the variance of $g(V)$. Hence:

$$\mathcal{S}_{V_1', V_2'}(g(V_1), g(V_2)) = \mathcal{N}((s \circ g^{-1})(g(V_1)), (s \circ g^{-1})(g(V_2))) = \mathcal{N}(s(V_1), s(V_2)) = \mathcal{S}_{V_1, V_2}(V_1, V_2)\ . \tag{A.108}$$

√ **Asymp. CFAR:** If s stabilizes the variance of V, and given that $\mathbb{E}[\|s(V) - s(V)\|_2^2] = 2\mathrm{Var}[s(V)]$, then $\mathcal{S}$ is asymptotically CFAR.

× **Asymp. UMPI:** $\mathcal{L}_G$ being UMPI, it defeats $\mathcal{S}$.

Appendix B

NL-SAR: an open-source software for speckle reduction

NL-SAR is originally designed to denoise multi-modalities of SAR images with non-local filtering. However, NL-SAR is also a flexible suite of tools to manipulate SAR images. There are 3 ways to interract with NL-SAR: in command line, with IDL and with Matlab.

Two other interfaces should be available soon: using dynamic or static library, and using PolSARPro.

So far, the command line version is the most stable one while others can crash, for instance, if you do not provide the good input in arguments. Feel free to fix such bugs or contribute to NL-SAR as you wish under the term of the license (see Section B.1).

B.1 License

This software is a computer program whose purpose is to provide a suite of tools to manipulate SAR images.

This software is governed by the CeCILL license under French law and abiding by the rules of distribution of free software. You can use, modify and/ or redistribute the software under the terms of the CeCILL license as circulated by CEA, CNRS and INRIA at the following URL "http://www.cecill.info".

As a counterpart to the access to the source code and rights to copy, modify and redistribute granted by the license, users are provided only with a limited warranty and the software's author, the holder of the economic rights, and the successive licensors have only limited liability.

In this respect, the user's attention is drawn to the risks associated with loading, using, modifying and/or developing or reproducing the software by the user in light of its specific status of free software, that may mean that it is complicated to manipulate, and that also therefore means that it is reserved for developers and experienced professionals having in-depth computer knowledge. Users are therefore encouraged to load and test the software's suitability as regards their requirements in conditions enabling the security of their systems and/or data to be ensured and, more generally, to use and operate it in the same conditions as regards security.

The fact that you are presently reading this means that you have had knowledge of the CeCILL license and that you accept its terms.

B.2 Installation

B.2.1 Dependencies

First check the following dependencies:

software	feature	status
gcc	to compile the project	required
gcc version > 4.2	to enable parallelization with OpenMP	optional
lapack	to enable non-local filtering with covariance matrices higher than 3 x 3	optional
blas	to enable non-local filtering with covariance matrices higher than 3 x 3	optional
gsl	to enable non-Wishart distribution [exerimental]	optional
gslcblas	to enable non-Wishart distribution [exerimental]	optional
fftw3	to enable non-local filtering with fft implementation and car filters	optional
fftw3f	to enable non-local filtering with fft implementation and car filters	optional
idl	to enable IDL interface	optional
matlab	to enable Matlab interface	optional
pdflatex	to create the documentation	optional

The above pieces of software have to be present in your environement variable `PATH` (for binaries) or `LD_LIBRARY_PATH` (for libraries) otherwise their associated feature will be disabled.

Once you have checked your dependencies, you can compile and install NL-SAR in two ways: as a super user or as a non super user.

B.2.2 Installation for super users

Fist configure and compile NL-SAR by typing in a shell prompt:

```
> ./configure
> make
> sudo make install
```

This will install NL-SAR's in `/usr/local/`, NL-SAR's IDL interface `<IDL_PATH>/lib/nlsar/` and NL-SAR's Matlab interface in `<MATLAB_PATH>/toolbox/nlsar/`.

B.2.3 Installation for non super users

Fist configure and compile NL-SAR by typing in a shell prompt:

```
> ./configure --prefix=<PREFIX> --prefix-idl=<IDL_PATH> --prefix-matlab=<MATLAB_PATH>
> make
> make install
```

This will install NL-SAR's in `<PREFIX>`, NL-SAR's IDL interface `<IDL_PATH>/lib/nlsar/` and NL-SAR's Matlab interface in `<MATLAB_PATH>/toolbox/nlsar/`.

You will need to update your environement paths variables. Make sure you are placed in the NL-SAR's directory and type the followings:

```
> echo export PATH=<PREFIX>/bin:'$PATH' >> $HOME/.bashrc
> echo export LD_LIBRARY_PATH=<PREFIX>/lib:'$LD_LIBRARY_PATH' >> $HOME/.bashrc
> source $HOME/.bashrc
```

B.2.4 Update IDL environment

If you are using IDL, update the IDL environment as follows:

```
> echo "PREF_SET, 'IDL_PATH', '<IDL_PATH>/lib/nlsar/:<IDL_DEFAULT>', /COMMIT" | idl
> echo export LD_LIBRARY_PATH=<IDL_PATH>/lib/nlsar/:'$LD_LIBRARY_PATH' >> $HOME/.bashrc
> source $HOME/.bashrc
```

B.2.5 Update Matlab environment

If you using Matlab, update the Matlab environment as follows:

```
> echo export MATLABPATH=<PREFIX_MATLAB>/toolbox/nlsar/:'$MATLABPATH' >> $HOME/.bashrc
> source $HOME/.bashrc
```

B.3 Images formats and input/output commands

Supported formats:

- RAT format,
- PolSARPro format.
- XIMA format (read only)

Note that NL-SAR deals only with images of intensity or of covariance matrices. Other inputs will not produce what you want. If you have amplitude or complex images, use the program `sarjoin` which build an intensity image or an image of covariance matrices from amplitude or complex images (see Section B.3.7).

B.3.1 RAT format

- A RAT file is assumed to be an image of complex covariance matrices. A RAT file containing vectorial data will produce an error message. Only the arrays of the following types are implemented so far:
  - float (`var` = 4)
  - float complex (`var` = 6)
  - double complex (`var` = 9)

  RAT files with other types will produce an error message. Fell free to contact me to extend to other modalities.

B.3.2 PolSARPro format

A PolSARPRo data is a directory containing binary files (with extensions `.bin`) and a `config.txt` file, You can find more details about this format there: http://earth.eo.esa.int/polsarpro/Manuals/PolSARpro_DataFormat.pdf So far, only complex covariance matrix data formats are implemented.

B.3.3 XIMA format

An XIMA data is a binary file without header wich comes with a file with same name and extension .dim. You can find more details about this format there: http://perso.telecom-paristech.fr/~nicolas/XIMA/index.html. So far, only cxf (complex float) data formats are implemented.

B.3.4 Reading information

- From command line:

```
> sarinfo infile
dimensions:
    M = 512
    N = 256
    D = 3
```

- From IDL:

```
> PRINT, sarinfo('infile')
          512         256           3
```

- From Matlab:

```
> [M, N, D] = sarinfo('../test/mire3.rat')
M =
     512
N =
     256
D =
       3
```

B.3.5 Reading data

The following commands import a SAR image from disk to memory

- From IDL:

```
> sarimage = sarread('infile')
```

Look at the matrix dimensions:

```
> PRINT, size(sarimage, /DIMENSIONS)
       3           3         256         512
```

- From Matlab:

```
> sarimage = sarread('infile');
```

Look at the matrix dimensions:

```
> size(sarimage)
ans =
     3     3   256   512
```

Note that a command line version would be meaningless.

B.3.6 Writing data

The following commands export a SAR image from memory to disk.

- From IDL:

```
> sarwrite, sarimage, 'outfile'
```

- From Matlab:

```
> sarwrite(sarimage, 'outfile');
```

Note that a command line version would be meaningless.

B.3.7 Join

The following commands creates an intensity image or a covariance matrix from an amplitude image or a complex image:

- From command line:

```
> sarjoin infile1 [infile2 ... infileN] outfile
```

Note that it is the only command of NL-SAR which deals with single look complex data as input. If you have only one file in input, this function basically computes the intensity image from an amplitude or complex image.

B.3.8 Conversion

The following example converts a RAT file to PolSARPro format:

- From command line

```
> sarconvert infile.rat outdir
```

- From IDL:

```
> sarimage = sarread('infile.rat')
> sarwrite, sarimage, 'outdir'
```

- From Matlab:

```
> sarimage = sarread('infile.rat');
> sarwrite(sarimage, 'outdir');
```

B.3.9 Extraction

The following commands extract a subarea from position (x, y) to position $(x + width - 1, y + height - 1)$ with a decimation step:

- From command line

```
> sarextract infile outfile x y width height step
```

- From IDL:

```
> sarout = sarin[*, *, y:(y+height-1):step, x:(x+width-1):step]
```

- From Matlab:

```
> sarout = sarin(:, :, y + (1:step:height)), x + (1:step:width));
```

B.3.10 RGB export

- From command line:

```
> sar2png infile rgbexport.png [alpha]
```

where `alpha` is an optional parameter to enhance contrast (default 3)

- From IDL:

```
> rgbexport = sar2rgb(sarimage [, alpha])
```

The argument `alpha` is the same as for the command line version.
Look at the matrix dimensions:

```
> PRINT, SIZE(sarimage, /DIMENSIONS)
       3           3         256         512
> PRINT, SIZE(rgbexport, /DIMENSIONS)
       3         256         512
```

- From Matlab:

```
> rgbexport = sar2rgb(sarimage [, alpha]);
```

The argument `alpha` is the same as for the command line version.
Look at the matrix dimension:

```
> size(sarimage)
ans =
     3      3    256    512
> size(rgbexport)
ans =
   512    256      3
```

The storing convension for the RGB image is reversed compared to our usual convention to ensure compatibility with the Matlab Image Toolbox.

B.3.11 Viewer

- From command line:

```
> sarshow infile [alpha]
```

The argument `alpha` is the same as for the RGB export.
The first time, you yill probably have the following message:

```
Please set your environment variable SARFOO_VIEWER
```

You need to define an environment variable `SARFOO_VIEWER` pointing to your favourite image viewer. For instance, if you like Eye Of Gnome, type the following:

```
> echo export SARFOO_VIEWER=eog >> $HOME/.bashrc
> source $HOME/.bashrc
```

Or, if you prefer Konqueror

```
> echo export SARFOO_VIEWER=konqueror >> $HOME/.bashrc
> source $HOME/.bashrc
```

- Form IDL:

```
> sarshow, sarimage [, alpha]
```

The argument `alpha` is the same as for the RGB export.

- Form Matlab:

```
> sarshow(sarimage [, alpha]);
```

The argument `alpha` is the same as for the RGB export.

B.4 Image filtering

B.4.1 Boxcar filter

Description

Estimate the complex covariance matrice as:

$$\hat{\mathbf{\Sigma}}(x) = \sum_{x' \in Box_{hW}(x)} \boldsymbol{C}(x') \tag{B.1}$$

where $Box_{hW}(x)$ is a square search window centered on x of width $(2hW + 1)$, we call hW the half-width.

Howto

- From command line:

```
> sarboxcar infile outfile [hW]
```

where `hW` is the half-width of the box (default 1).

- From IDL:

```
> sarout = sarboxcar(sarin [, hW])
```

The arguments are the same as for the command line version.

- From Matlab:

```
> sarout = sarboxcar(sarin [, hW]);
```

The arguments are the same as for the command line version.

B.4.2 Diskcar filter

Same as the box car filter, just replace $Box_{hW}(x)$ by $Disk_{hW}(x)$, i.e. a disk centered on x with half-width (i.e. radius) of hW.

B.4.3 Iterative Non-local Means filter

Description

Estimate iteratively $\mathbf{\Sigma}(x)$ using:

$$\hat{\mathbf{\Sigma}}^t(x) = \begin{cases} \dfrac{\sum\limits_{x' \in Disk_{hW}(x)} w(x,x')\boldsymbol{C}(x')}{\sum\limits_{x' \in Disk_{hW}(x))} w(x,x')} & \text{if } \hat{L}_{res} \geq L_{\min} \\ \sum_{x' \in Best_{L_{\min}}(x)} \boldsymbol{C}(x') & \text{otherwise} \end{cases} \tag{B.2}$$

where $\hat{L}_{res}$ is an estimate of the resulting equivalent number of looks:

$$\hat{L}_{res} = \frac{\left(\sum\limits_{x' \in Disk_{hW}(x)} w(x,x') \right)^2}{\sum\limits_{x' \in Disk_{hW}(x)} w(x,x')^2} \tag{B.3}$$

$Best_{Lmin}(x)$ refers to the L_{min} candidates x' having the highest weights $w(x,x')$. The non-local filter computes a weighted average of candidates covariance matrices and forces the resulting equivalent number of looks to be higher than a threshold $L_{\min}$.

The weights are defined as:

$$w(x,x') = \psi \Bigg[(1-\lambda) \times \phi_{GLR} \left(\sum_{\tau \in Box_{hP}(0)} -\log GLR(C'(x+\tau), C'(x'+\tau)) \right) + \tag{B.4}$$

$$\lambda \times \phi_{KL} \left(\sum_{\tau \in Box_{hP}(0)} KL(\hat{\Sigma}^{t-1}(x+\tau), \hat{\Sigma}^{t-1}(x'+\tau)) \right) \Bigg] \tag{B.5}$$

where ϕ_{GLR} (resp. ϕ_{KL}) is a linear function scaling the values $q_{1,GLR}$ and $q_{2,GLR}$ (resp. the values $q_{1,KL}$ and $q_{2,KL}$) to the values 1 and 2:

$$\phi_X(x) = \frac{x + q_{2,X} - 2q_{1,X}}{q_{2,X} - q_{1,X}} \tag{B.6}$$

and, ψ is a trapezoidal kernel defined as:

$$\psi(x) = \begin{cases} 1 & \text{if } x \leq 1 \\ 2 - x & \text{if } 1 < x \leq 2 \\ 0 & \text{otherwise} \end{cases} \tag{B.7}$$

Finally, we can list 10 parameters inherent to the method:

- model: the model used to derive GLR and KL (so far, it is only “wishart”)

- T: the number of iterations
- hW: half-width of the circular search window
- hP: half-width of the square patches
- $L_{\min}$: the minimum equivalent number of looks to ensure
- λ: the trade-off between your confidence on GLR versus KL. (set 0 if you do not trust KL or 1 if you do not trust GLR, otherwise 0.5 sounds good).
- $q_{1,GLR}$, $q_{2,GLR}$: two quantiles of the distribution of GLR in homogeneous noisy areas If you choose $q_{1,GLR}$ as the 0.8-quantile and $q_{2,GLR}$ as the 0.95-quantile, it means that in homogeneous area you will average at least 80% of the candidates, reject 1% of the candidates and attribute a weight betwwen 0 and 1 for the remaining candidates.
- $q_{1,KL}$, $q_{2,KL}$: two quantiles of the distribution of KL in pre-filtered homogeneous noisy areas (typically choose $q_{1,KL}$ as the 0.8-quantile and $q_{2,KL}$ as the 0.95-quantile).

And a last parameter inherent to the algorithm used for calculating (B.2):

- naive: the naive version of NL means (complexity $O(NMD^2(2hW+1)^2(2hP+1)^2)$)
- issd: the version of Darbon et al. using integrate sum square difference (complexity $O(NMD^2(2hW+1)^2)$)

  NB: this implementation is protected by a U.S. patent. For this reason, this implementation is not included by default in the NL-SAR package. If the patent does not hold in your country, feel free to download this implementation online.
- fft: the naive version of NL means (complexity $O(NMD^2(2hW+1)^2\log_2(NMD^2))$)

Howto

- From command line:

```
> sarnl filein fileout L [verbose implem model T hW hP Lmin \
                          lambda q1glr q2glr q1kl q2kl]
```

  where `L` is the equivalent number of looks of the input noisy image. If `verbose` = 1, steps and progressing bars are displayed on the standard output. A description of the other arguments is given in the previous section. Type `sarnl` to see the default values of the optional parameters.
- From IDL:

```
> sarout = sarnl(sarin, L [,verbose ,implem, model, T, hW, hP, Lmin, $
                           lambda, q1glr, q2glr, q1kl, q2kl])
```

  The arguments are the same as for the command line version.
- From Matlab:

```
> sarout = sarnl(sarin, L [,verbose ,implem, model, T, hW, hP, Lmin, ...
                           lambda, q1glr, q2glr, q1kl, q2kl]);
```

  The arguments are the same as for the command line version.

Note that the verbose mode for IDL and Matlab might not work properly.

B.4.4 Semi-supervised INLM filter

Description

The iterative non-local filter lies on 10 parameters whose 4 parameters depends on the distribution of GLR and KL on respectively an homogenous noisy area and on a prefiltered homogeneous area. Instead of providing in input the values of the quantiles (that you need to estimate) you can provide also a noisy area and ask to estimate for you the desired alpha-quantiles on this area. Note that if you have a large set of data but sensed in the same condition you need to extract only one noisy are for the whole dataset: the semi supervision is very lazy.

Howto

- From command line:

```
> sarnlcal filein filenoise fileout L [verbose implem model T hW hP Lmin \
                                       lambda alpha-q1 alpha-q2]
```

  If `verbose` = 1, steps and progressing bars are displayed on the standard output. A description of the other arguments is given in the previous sections. Type `sarnlcal` to see the default values of the optional parameters.

- From IDL:

```
> sarout = sarnlcal(sarin, sarnoise, L [,verbose ,implem, model, T, hW, hP, Lmin, $
                                        lambda, alpha-q1, alpha-q2])
```

  The arguments are the same as for the command line version.

- From Matlab:

```
> sarout = sarnlcal(sarin, sarnoise, L [,verbose ,implem, model, T, hW, hP, Lmin, ...
                                        lambda, alpha-q1, alpha-q2]);
```

  The arguments are the same as for the command line version.

B.5 Examples

B.5.1 InSAR image without orbital components

The following example has been run on a SLC RAMSES image of resolution about 1×1 m^2 with computer with 16 Intel Core 2 Duo Xeon CPU 2.27GHz. With high resolution data and correlated spekcle noise (due to zero padding and windowing in the Fourier domain) large windows need to be used. We choose a search window of radius 10 pixels. It means that the maximum noise reduction is of about 314 looks, and patches of width 9 (half-width $hP = 4$). A minimum equivalent number of 9 looks is required. We have chosen the 0.80 and 0.95-quantile, implying that in average the resulting equivalent number of looks in homogeneous noisy area would be around 250 looks.

```
> ls
filein_hh.cxf   filein_hh.dim   filein_bb.cxf   filein_bb.dim
> sarjoin filein_hh.cxf filein_bb.cxf filein_2x2.rat
> ls
filein_2x2.rat   filein_hh.cxf   filein_hh.dim   filein_bb.cxf   filein_bb.dim
> sarinfo filein_2x2
dimensions:
    M = 512
    N = 512
```

```
    D = 2
> sarshow filein_2x2.rat 1.5
```

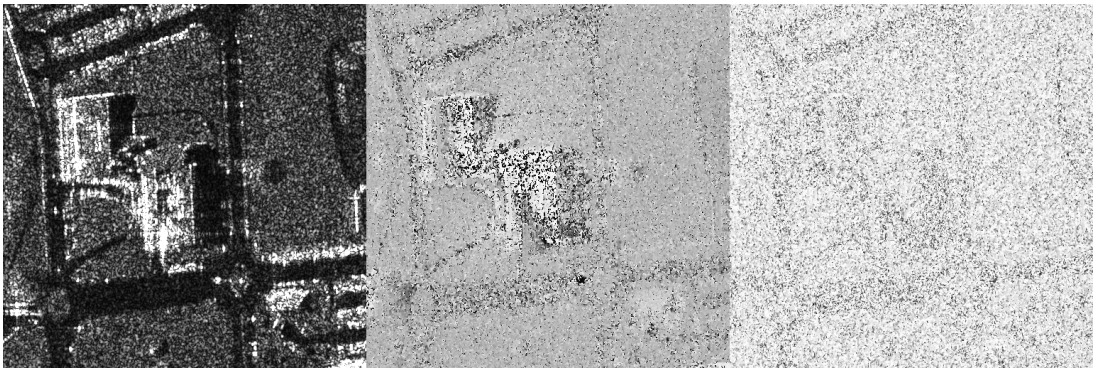


Figure B.1: The 512×512 image

```
> sarextract filein_2x2.rat rat noise.rat 0 340 200 100
> ls
filein_2x2.rat   filein_hh.cxf   filein_hh.dim   filein_bb.cxf   filein_bb.dim   noise.rat
> sarshow noise.rat
```

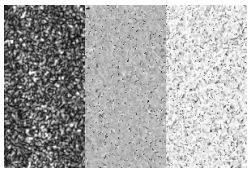


Figure B.2: A noisy area extracted from the input noisy data

```
> time sarnlcal filein_2x2.rat noise.rat fileout_1it.rat 1 1 issd wishart 1 10 4 9 \
                                                          0.5 0.80 0.95
Estimation of GLR quantiles
        0.80-quantile =         1.5287
        0.95-quantile =         1.70116
Estimation of KL quantiles
|================================================| 100%
|================================================| 100%
        0.80-quantile =         0.106519
        0.95-quantile =         0.270247
Process to the 1 iterations
|================================================| 100%
|================================================| 100%

real    0m3.303s
user    0m50.515s
sys     0m4.584s
> ls
filein_2x2.rat   filein_hh.cxf   filein_hh.dim
filein_bb.cxf   filein_bb.dim   fileout_1it.rat
> sarshow fileout_1it.rat 1.5
```

```
> time sarnlcal filein_2x2.rat noise.rat fileout_1it.rat 1 1 issd wishart 4 10 4 9 \
                                                          0.5 0.80 0.95
```

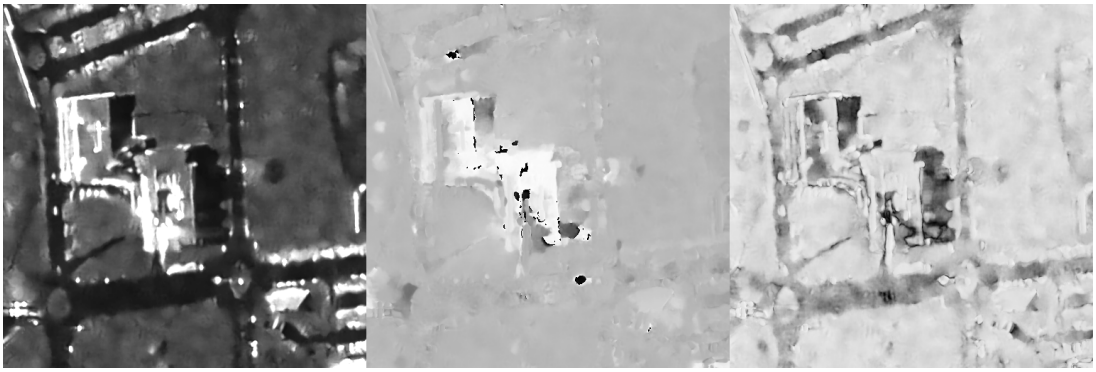


Figure B.3: Result after 1 iteration

```
Estimation of GLR quantiles
        0.80-quantile =         1.5287
        0.95-quantile =         1.70116
Estimation of KL quantiles
|================================================| 100%
|================================================| 100%
        0.80-quantile =         0.106519
        0.95-quantile =         0.270247
Process to the 1 iterations
|================================================| 100%
|================================================| 100%
|================================================| 100%
|================================================| 100%
|================================================| 100%
|================================================| 100%
|================================================| 100%
|================================================| 100%

real    0m27.648s
user    8m28.092s
sys     1m58.539s
> ls
filein_2x2.rat   filein_hh.cxf   filein_hh.dim
filein_bb.cxf   filein_bb.dim   fileout_1it.rat
> sarshow fileout_1it.rat 1.5
```

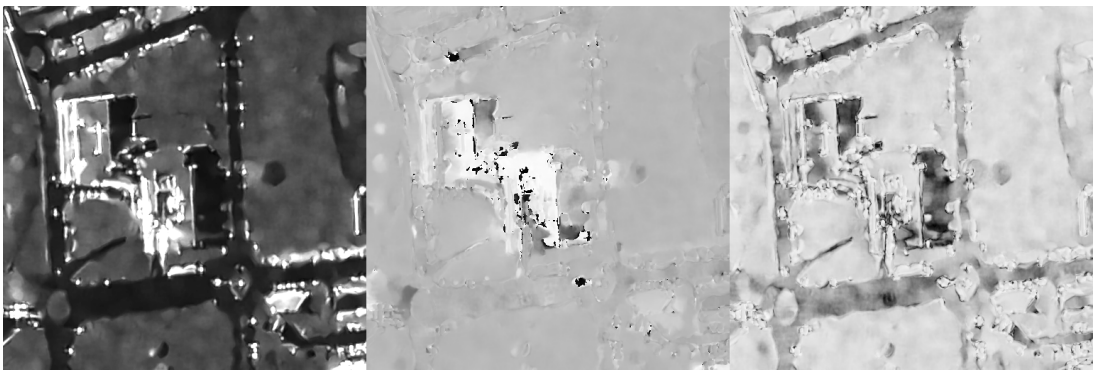


Figure B.4: Result after 4 iterations

B.5.2 PolSAR image with high resolution

The following example has been run on a SLC X-Band F-SAR image of resolution 20×20 cm^2 with computer with 16 Intel Core 2 Duo Xeon CPU 2.27GHz. With high resolution data and correlated spekcle noise (due to zero padding and windowing in the Fourier domain) large windows need to be used. We choose a search window of radius 10 pixels. It means that the maximum noise reduction is of about 314 looks, and patches of width 9 (half-width $hP = 4$). A minimum equivalent number of 9 looks is required. We have chosen the 0.80 and 0.95-quantile, implying that in average the resulting equivalent number of looks in homogeneous noisy area would be around 250 looks.

```
> ls
filein_Xhh.rat   filein_Xhv.rat   filein_Xvv.rat
> sarjoin filein_Xhh.rat filein_Xvv.rat filein_Xhv.rat filein_3x3.rat
> ls
filein_3x3.rat   filein_Xhh.rat   filein_Xhv.rat   filein_Xvv.rat
> sarinfo filein_3x3
dimensions:
    M = 4096
    N = 4096
    D = 3
> sarshow filein_3x3.rat
```

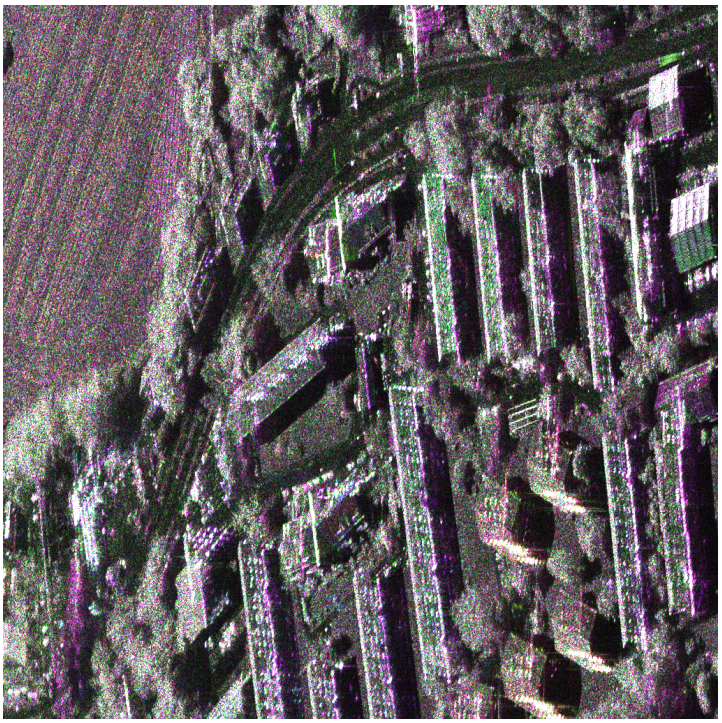


Figure B.5: Zoom on 1/8 of the 4096×4096 image

```
> sarextract filein_3x3.rat rat noise.rat 0 2048 256 256
> ls
filein_3x3.rat   filein_Xhh.rat   filein_Xhv.rat   filein_Xvv.rat   noise.rat
> sarshow noise.rat
```

```
> time sarnlcal filein_3x3.rat noise.rat fileout_1it.rat 1 1 issd wishart 1 10 4 9 \
                                                    0.5 0.80 0.95
```

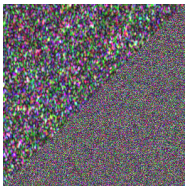


Figure B.6: A noisy area extracted (and rescale) from the input noisy data

```
Estimation of GLR quantiles
        0.80-quantile =         3.77725
        0.95-quantile =         4.05229
Estimation of KL quantiles
|================================================| 100%
|================================================| 100%
        0.80-quantile =         0.087599
        0.95-quantile =         0.189808
Process to the 1 iterations
|================================================| 100%
|================================================| 100%

real    2m58.006s
user    35m47.210s
sys     3m44.646s
> ls
filein_3x3.rat   filein_Xhh.rat   filein_Xhv.rat   filein_Xvv.rat   fileout_1it.rat
noise.rat
> sarshow fileout_1it.rat
```

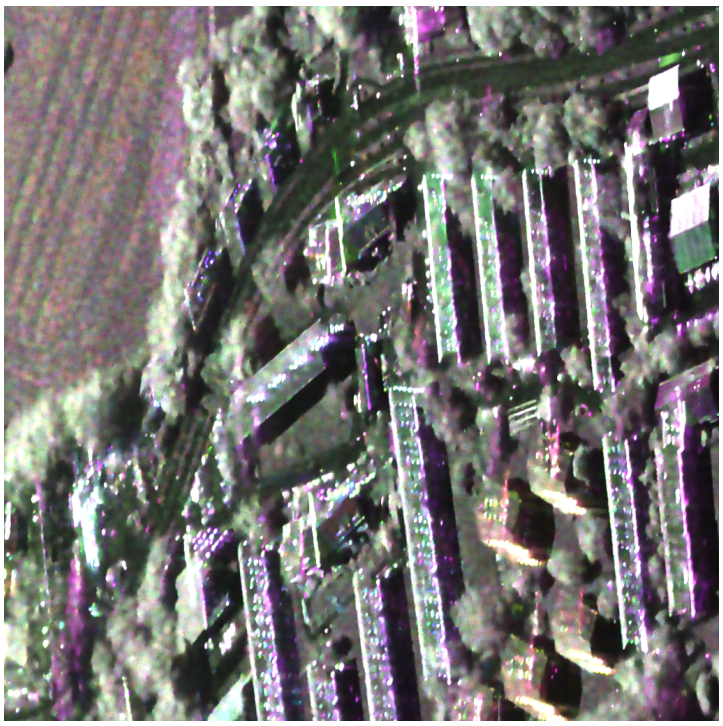


Figure B.7: Result after 1 iteration

```
> time sarnlcal filein_3x3.rat noise.rat fileout_4it.rat 1 1 issd wishart 4 10 4 9 \
                                                        0.5 0.80 0.95
Estimation of GLR quantiles
        0.80-quantile =         3.77725
        0.95-quantile =         4.05229
Estimation of KL quantiles
|================================================| 100%
|================================================| 100%
        0.80-quantile =         0.087599
        0.95-quantile =         0.189808
Process to the 4 iterations
|================================================| 100%
|================================================| 100%
|================================================| 100%
|================================================| 100%
|================================================| 100%
|================================================| 100%
|================================================| 100%
|================================================| 100%

real    13m31.742s
user    212m43.354s
sys     32m29.862s
> ls
filein_3x3.rat   filein_Xhh.rat   filein_Xhv.rat   filein_Xvv.rat   fileout_1it.rat
fileout_4it.rat   noise.rat
> sarshow fileout_4it.rat
```

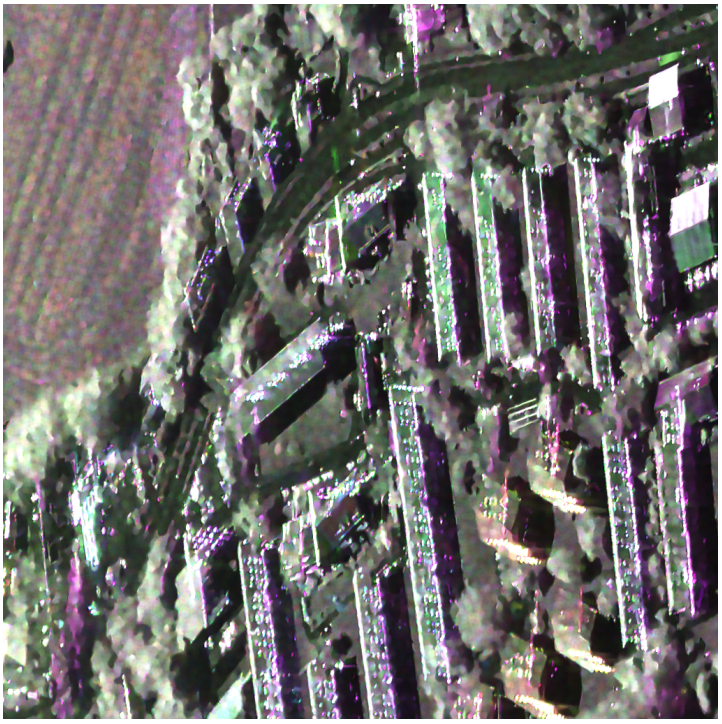


Figure B.8: Result after 4 iterations

Bibliography

Abramovich, F., Sapatinas, T., and Silverman, B. (1998). Wavelet thresholding via a bayesian approach. *Journal of the Royal Statistical Society. Series B (Methodological)*, 60(4):725–749.

Achim, A., Tsakalides, P., and Bezerianos, A. (2003). SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling. *IEEE Trans. Geosci. Remote Sens.*, 41(8):1773–1784.

Aharon, M., Elad, M., and Bruckstein, A. (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.*, 54(11):4311–4322.

Aja-Fernández, S. and Alberola-López, C. (2006). On the estimation of the coefficient of variation for anisotropic diffusion speckle filtering. *IEEE Trans. Image Process.*, 15(9):2694–2701.

Alliney, S. (1992). Digital filters as absolute norm regularizers. *IEEE Trans. Signal Process.*, 40(6):1548–1562.

Alter, F., Matsushita, Y., and Tang, X. (2006). An intensity similarity measure in low-light conditions. *Lecture Notes in Computer Science*, 3954:267–280.

Argenti, F. and Alparone, L. (2002). Speckle removal from SAR images in the undecimated wavelet domain. *IEEE Trans. Geosci. Remote Sens.*, 40(11):2363–2374.

Argenti, F., Bianchi, T., and Alparone, L. (2006). Multiresolution MAP despeckling of SAR images based on locally adaptive generalized Gaussian pdf modeling. *IEEE Trans. Image Process.*, 15(11):3385–3399.

Aubert, G. and Aujol, J. (2008). A variational approach to remove multiplicative noise. *SIAM Journal Applied Math.*, 68(4):925–946.

Awate, S. and Whitaker, R. (2006). Unsupervised, information-theoretic, adaptive image filtering for image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(3):364–376.

Awate, S. P. and Whitaker, R. T. (2005). Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering. In *IEEE Comput. Vis. and Pattern Recognition (CVPR)*, pages 44–51. IEEE.

Azzabou, N., Paragios, N., and Guichard, F. (2007a). Image denoising based on adapted dictionary computation. In *IEEE Int. Conf. Image Process. (ICIP)*, pages 109–112. IEEE.

Azzabou, N., Paragios, N., and Guichard, F. (2007b). Uniform and textured regions separation in natural images towards MPM adaptive denoising. *Lecture Notes in Computer Science*, 4485:418–429.

Bamler, R. and Hartl, P. (1998). Synthetic aperture radar interferometry. *Inverse problems*, 14:R1.

Bardsley, J. and Luttman, A. (2009). Total variation-penalized poisson likelihood estimation for ill-posed problems. *Advances in Computational Mathematics*, 31(1):35–59.

Baxter, J. (1995). Learning internal representations. In *Conference on Computational Learning Theory (COLT)*, pages 311–320. ACM.

Baxter, J. and Bartlett, P. (1998). The canonical distortion measure in feature space and 1-nn classification. In *Advances in Neural Inf. Process. Syst. (NIPS)*, volume 10, pages 245–250. MIT Press.

Benazza-Benyahia, A. and Pesquet, J. (2005). Building robust wavelet estimators for multicomponent images using Stein's principle. *IEEE Trans. Image Process.*, 14(11):1814–1830.

Benoît, L., Mairal, J., Bach, F., and Ponce, J. (2011). Sparse image representation with epitomes. In *IEEE Comput. Vis. and Pattern Recognition (CVPR)*, pages 2913–2920. IEEE.

Benzid, S., Deledalle, C.-A., Abdelfattah, R., Chaabane, F., and Tupin, F. (2010). Change detection in a multitemporal series of radar images. In *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, pages 1473–1476. IEEE.

Bhuiyan, M., Ahmad, M., and Swamy, M. (2007). Spatially adaptive wavelet-based method using the Cauchy prior for denoising the SAR images. *IEEE Trans. Circuits Syst. Video Technol.*, 17(4):500–507.

Bianchi, T., Argenti, F., and Alparone, L. (2008). Segmentation-Based MAP Despeckling of SAR Images in the Undecimated Wavelet Domain. *IEEE Trans. Geosci. Remote Sens.*, 46(9):2728–2742.

Bilcu, R. C. and Vehvilainen, M. (2008). Combined non-local averaging and intersection of confidence intervals for image de-noising. In *IEEE Int. Conf. Image Process. (ICIP)*, pages 1736–1739. IEEE.

Bioucas-Dias, J. and Figueiredo, M. (2010). Multiplicative noise removal using variable splitting and constrained optimization. *IEEE Trans. Image Process.*, 19(7):1720–1730.

Bioucas-Dias, J., Katkovnik, V., Astola, J., and Egiazarian, K. (2008). Absolute phase estimation: adaptive local denoising and global unwrapping. *Applied optics*, 47(29):5358–5369.

Bioucas-Dias, J. and Valadão, G. (2007). Phase unwrapping via graph cuts. *IEEE Trans. Image Process.*, 16(3):698–709.

Blu, T. and Luisier, F. (2007). The SURE-LET approach to image denoising. *IEEE Trans. Image Process.*, 16(11):2778–2786.

Bougleux, S., Elmoataz, A., and Melkemi, M. (2009). Local and nonlocal discrete regularization on weighted graphs for image and mesh processing. *International Journal of Computer Vision*, 84(2):220–236.

Boulanger, J., Kervrann, C., Bouthemy, P., Elbau, P., Sibarita, J., and Salamero, J. (2010). Patch-based nonlocal functional for denoising fluorescence microscopy image sequences. *IEEE Trans. Med. Imag.*, 29(2):442–454.

Boulanger, J., Sibarita, J., Kervrann, C., and Bouthemy, P. (2008). Non-parametric regression for patch-based fluorescence microscopy image sequence denoising. In *IEEE Int. Symposium on Biomedical Imaging (ISBI)*, pages 748–751. IEEE.

Boykov, Y., Veksler, O., and Zabih, R. (1998). Markov random fields with efficient approximations. In *IEEE Comput. Vis. and Pattern Recognition (CVPR)*, pages 648–655. IEEE.

Boykov, Y., Veksler, O., and Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1222–1239.

Bratsolis, E. and Sigelle, M. (2003). Fast SAR image restoration, segmentation, and detection of high-reflectance regions. *IEEE Trans. Geosci. Remote Sens.*, 41(12):2890–2899.

Brown, L., Cai, T., Zhang, R., Zhao, L., and Zhou, H. (2010). The root–unroot algorithm for density estimation as implemented via wavelet block thresholding. *Probability theory and related fields*, 146(3):401–433.

Brox, T., Kleinschmidt, O., and Cremers, D. (2008). Efficient Nonlocal Means for Denoising of Textural Patterns. *IEEE Trans. Image Process.*, 17(7):1083–1092.

Buades, A., Coll, B., and Morel, J. (2008). Nonlocal Image and Movie Denoising. *International Journal of Computer Vision*, 76(2):123–139.

Buades, A., Coll, B., and Morel, J.-M. (2005). A review of image denoising algorithms, with a new one. *SIAM Journal Multiscale Model. Simul.*, 4(2):490–530.

Buades, A., Coll, B., and Morel, J.-M. (2009). Non-local means denoising. *Image Processing on Line*.

Buccigrossi, R. and Simoncelli, E. (1999). Image compression via joint statistical characterization in the wavelet domain. *IEEE Trans. Image Process.*, 8(12):1688–1701.

Cameron, W. and Leung, L. (1990). Feature motivated polarization scattering matrix decomposition. In *IEEE Int. Radar Conf. (RADAR)*, pages 549–557. IEEE.

Candès, E. (1998). *Ridgelets: theory and applications*. PhD thesis, Stanford University.

Candès, E. and Donoho, D. (1999). Ridgelets: A key to higher-dimensional intermittency? *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 357(1760):2495.

Candes, E. and Donoho, D. (2000). *Curvelets: A Surprisingly Effective Nonadaptive Representation for Objects with Edges*. Vanderbilt University Press.

Cao, F., Deledalle, C.-A., Nicolas, J.-M., Tupin, F., Denis, L., Ferro-Famil, L., Pottier, E., and Lopéz-Martínez, C. (2011). Influence of speckle filtering of polarimetric SAR data on different classification methods. In *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*. IEEE.

Chan, T. and Esedoḡlu, S. (2005). Aspects of total variation regularized L1 function approximation. *SIAM Journal Applied Math.*, 65(5):1817–1837.

Chang, S., Yu, B., Vetterli, M., et al. (2000). Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans. Image Process.*, 9(9):1532–1546.

Chatterjee, P. and Milanfar, P. (2010). Is denoising dead? *IEEE Trans. Image Process.*, 19(4):895–911.

Chatterjee, P. and Milanfar, P. (2011). Patch-based locally optimal denoising. In *IEEE Int. Conf. Image Process. (ICIP)*. IEEE.

Chen, J., Chen, Y., An, W., Cui, Y., and Yang, J. (2011). Nonlocal Filtering for Polarimetric SAR Data: A Pretest Approach. *IEEE Trans. Geosci. Remote Sens.*, 49(5):1744 –1754.

Chen, L. (1975). Poisson approximation for dependent trials. *The Annals of Probability*, 3(3):534–545.

Chen, S., Donoho, D., and Saunders, M. (1999). Atomic decomposition by basis pursuit. *SIAM Journal Sci. Comput.*, 20(1):33–61.

Cho, T., Avidan, S., and Freeman, W. (2009). The patch transform. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):1489–1501.

Cloude, S., Papathanassiou, K., and Pottier, E. (2001). Radar polarimetry and polarimetric interferometry. *Trans. Institute of Electronics, Information and Communication Engineers*, 84(12):1814–1822.

Cloude, S. and Pottier, E. (1995). Concept of polarization entropy in optical scattering. *Optical engineering*, 34(06):1599–1610.

Cloude, S. and Pottier, E. (1996). A review of target decomposition theorems in radar polarimetry. *IEEE Trans. Geosci. Remote Sens.*, 34(2):498–518.

Cloude, S. and Pottier, E. (1997). An entropy based classification scheme for land applications of polarimetric SAR. *IEEE Trans. Geosci. Remote Sens.*, 35(1):68–78.

Coifman, R. and Donoho, D. (1995). Translation-invariant de-noising. *Lecture Notes in Statistics*, pages 125–150.

Comaniciu, D., Ramesh, V., and Meer, P. (2003). Kernel-Based Object Tracking. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(5):564–577.

Condat, L. (2010). A Simple Trick to Speed Up and Improve the Non-Local Means. *Submitted, hal-00512801*.

Coupé, P., Hellier, P., Kervrann, C., and Barillot, C. (2008). Bayesian non-local means-based speckle filtering. In *IEEE Int. Symposium on Biomedical Imaging (ISBI)*, pages 1291–1294. IEEE.

Coupe, P., Yger, P., and Barillot, C. (2006). Fast Non Local Means Denoising for 3D MR Images. *Lecture Notes in Computer Science*, 4191:33–40.

Criminisi, A., Pérez, P., and Toyama, K. (2004). Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Process.*, 13(9):1200–1212.

Crow, F. C. (1984). Summed-area tables for texture mapping. In *SIGGRAPH*, pages 207–212. ACM.

Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2008). A Nonlocal and Shape-Adaptive Transform-Domain Collaborative Filtering. In *Local and Non-Local Approximation (LNLA)*, pages 1–8. IEEE.

Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. O. (2007). Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16(8):2080–2095.

Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. O. (2009). BM3D image denoising with shape-adaptive principal component analysis. In *Proc. Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS'09)*.

Dalalyan, A. S. and Tsybakov, A. B. (2008). Aggregation by exponential weighting, sharp pac-bayesian bounds and sparsity. *Mach. Learn.*, 72(1):39–61.

Darbon, J., Cunha, A., Chan, T. F., Osher, S., and Jensen, G. J. (2008). Fast nonlocal filtering applied to electron cryomicroscopy. In *IEEE Int. Symposium on Biomedical Imaging (ISBI)*, pages 1331–1334. IEEE.

Darbon, J. and Sigelle, M. (2006). Image restoration with discrete constrained total variation part ii: Levelable functions, convex priors and non-convex cases. *Journal of Mathematical Imaging and Vision*, 26(3):277–291.

Daubechies, I. (1992). Ten lectures on wavelets. In *CBMS-NSF Regional Conference Series in Applied Mathematics*, volume 61. SIAM.

Davis, G., Mallat, S., and Avellaneda, M. (1997). Adaptive greedy approximations. *Constructive approximation*, 13(1):57–98.

Deledalle, C.-A., Denis, L., and Tupin, F. (2009a). Débruitage Non-Local Itératif fondé sur un Critère de Similarité Probabiliste. In *GRETSI*.

Deledalle, C.-A., Denis, L., and Tupin, F. (2009b). Iterative Weighted Maximum Likelihood Denoising with Probabilistic Patch-Based Weights. *IEEE Trans. Image Process.*, 18(12):2661–2672.

Deledalle, C.-A., Denis, L., and Tupin, F. (2011a). NL-InSAR : Non-Local Interferogram Estimation. *IEEE Trans. Geosci. Remote Sens.*, 49(4):1441–1452.

Deledalle, C.-A., Denis, L., and Tupin, F. (2012). How to compare noisy patches? Patch similarity beyond Gaussian noise. *Accepted in International Journal of Computer Vision*.

Deledalle, C.-A., Duval, V., and Salmon, J. (2011b). Anisotropic non-local means with spatially adaptive patch shapes. In *Scale Space and Variational Methods in Computer Vision (SSVM)*.

Deledalle, C.-A., Duval, V., and Salmon, J. (2011c). Non-local Methods with Shape-Adaptive Patches (NLM-SAP). *Journal of Mathematical Imaging and Vision*, pages 1–18.

Deledalle, C.-A., Nicolas, J., Tupin, F., Denis, L., Fallourd, R., and Trouvé, E. (2010a). Glacier monitoring: correlation versus texture tracking. In *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, pages 513–516. IEEE.

Deledalle, C.-A., Salmon, J., and Dalalyan, A. (2011d). Image denoising with patch based PCA: local versus global. In *British Machine Vision Conference (BMVC)*.

Deledalle, C.-A., Tupin, F., and Denis, L. (2010b). A non-local approach for SAR and interferometric SAR denoising. In *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, pages 714–717. IEEE.

Deledalle, C.-A., Tupin, F., and Denis, L. (2010c). Poisson NL means: Unsupervised non local means for Poisson noise. In *IEEE Int. Conf. Image Process. (ICIP)*, pages 801–804. IEEE.

Deledalle, C.-A., Tupin, F., and Denis, L. (2010d). Polarimetric SAR estimation based on non-local means. In *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, pages 2515–2518. IEEE.

Deledalle, C.-A., Tupin, F., and Denis, L. (2011e). Patch similarity under non-gaussian noise. In *IEEE Int. Conf. Image Process. (ICIP)*. IEEE.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.

Denis, L., Tupin, F., Darbon, J., and Sigelle, M. (2009). SAR image regularization with fast approximate discrete minimization. *IEEE Trans. Image Process.*, 18(7):1588–1600.

Do, M. and Vetterli, M. (2005). The contourlet transform: an efficient directional multiresolution image representation. *IEEE Trans. Image Process.*, 14(12):2091–2106.

Donoho, D. (2006). Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306.

Donoho, D. and Johnstone, I. (1995). Adapting to Unknown Smoothness Via Wavelet Shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224.

Donoho, D. and Johnstone, J. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455.

Doré, V. and Cheriet, M. (2009). Robust NL-means filter with optimal pixel-wise smoothing parameter for statistical image denoising. *IEEE Trans. Signal Process.*, 57(5):1703–1716.

Dubois, D. and Prade, H. (1988). *Possibility theory*. Plenum Press New York.

Durand, S., Fadili, J., and Nikolova, M. (2010). Multiplicative noise removal using L1 fidelity on frame coefficients. *Journal of Mathematical Imaging and Vision*, 36(3):201–226.

Duval, V., Aujol, J., and Gousseau, Y. (2009). The TVL1 Model: A Geometric Point of View. *SIAM Journal Multiscale Model. Simul.*, 8(1):154–189.

Duval, V., Aujol, J.-F., and Gousseau, Y. (2011). A bias-variance approach for the non-local means. *SIAM Journal Imaging Sci.*, 4:760–788.

Efros, A. and Freeman, W. (2001). Image quilting for texture synthesis and transfer. In *SIGGRAPH*, pages 341–346. ACM.

Efros, A. and Leung, T. (1999). Texture synthesis by non-parametric sampling. In *International Conference on Computer Vision (ICCV)*, volume 2, pages 1033–1038.

Elad, M. and Aharon, M. (2006). Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.*, 15(12):3736–3745.

Eldar, Y. (2009). Generalized sure for exponential families: Applications to regularization. *IEEE Trans. Signal Process.*, 57(2):471–481.

Elmoataz, A., Lezoray, O., and Bougleux, S. (2008). Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing. *IEEE Trans. Image Process.*, 17(7):1047–1060.

Fan, J., Farmen, M., and Gijbels, I. (1998). Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, 60(3):591–608.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*, volume 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Feng, H., Hou, B., and Gong, M. (2011). SAR Image Despeckling Based on Local Homogeneous-Region Segmentation by Using Pixel-Relativity Measurement. *IEEE Trans. Geosci. Remote Sens.*, 49(7):2724–2737.

Figueiredo, M. and Bioucas-Dias, J. (2009). Deconvolution of poissonian images using variable splitting and augmented lagrangian optimization. In *IEEE Statistical Signal Processing (SSP)*, pages 733–736. IEEE.

Freeman, A. and Durden, S. (1998). A three-component scattering model for polarimetric SAR data. *IEEE Trans. Geosci. Remote Sens.*, 36(3):963–973.

Freeman, W., Jones, T., and Pasztor, E. (2002). Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2):56–65.

Gallager, R. (2008). Circularly-symmetric gaussian random vectors. Technical report, Massachusetts Institute of Technology.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6:721–741.

Ghiglia, D. and Romero, L. (1989). Direct phase estimation from phase differences using fast elliptic partial differential equation solvers. *Optics Letters*, 14(20):1107–1109.

Gilboa, G. and Osher, S. (2007). Nonlocal linear image regularization and supervised segmentation. *SIAM Journal Multiscale Model. Simul.*, 6(2):595–630.

Goldstein, R., Zebker, H., and Werner, C. (1988). Satellite radar interferometry: Two-dimensional phase unwrapping. *Radio Science*, 23(4):713–720.

Goodman, J. (1976). Some fundamental properties of speckle. *Journal of the Optical Society of America*, 66(11):1145–1150.

Goodman, J. (1984). Statistical Properties of Laser Speckle Patterns. *Laser speckle and related phenomena*, 9:9–75.

Goodman, J. (2006). *Speckle phenomena in optics: theory and applications*. Roberts & Company Publishers.

Goodman, N. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *The Annals of Mathematical Statistics*, 34(1):152–177.

Goossens, B., Luong, H. Q., Pizurica, A., and Philips, W. (2008). An improved non-local denoising algorithm. In *Local and Non-Local Approximation (LNLA)*, pages 143–156.

Goudail, F., Réfrégier, P., and Delyon, G. (2004). Bhattacharyya distance as a contrast parameter for statistical processing of noisy optical images. *Journal of the Optical Society of America*, 21(7):1231–1240.

Graham, L. (1974). Synthetic interferometer radar for topographic mapping. *Proceedings of the IEEE*, 62(6):763–768.

Guleryuz, O. (2007). Weighted averaging for denoising with overcomplete dictionaries. *IEEE Trans. Image Process.*, 16(12):3020–3034.

Hachicha, S., Deledalle, C.-A., Chaabane, F., and Tupin, F. (2011). Multi-temporal sar classification according to change detection operators. In *Analysis of Multi-temporal Remote Sensing Images (Multi-Temp)*, pages 133–136. IEEE.

Hajnsek, I., Pottier, E., and Cloude, S. (2003). Inversion of surface parameters from polarimetric sar. *IEEE Trans. Geosci. Remote Sens.*, 41(4):727–744.

Hammersley, J. and Clifford, P. (1971). Markov fields on finite graphs and lattices. Unpublished.

Hanssen, R. (2001). *Radar interferometry*. Kluwer Academic.

Harmany, Z., Marcia, R., and Willett, R. (2009). Sparse poisson intensity reconstruction algorithms. In *IEEE Statistical Signal Processing (SSP)*, pages 634–637. IEEE.

Hartley, R. and Zisserman, A. (2000). *Multiple view geometry*, volume 642. Cambridge university press.

He, L. and Greenshields, I. (2009). A Non-Local Maximum Likelihood Estimation Method for Rician Noise Reduction in MR Images. *IEEE Trans. Med. Imag.*, 28(2):165–172.

Horn, B. and Schunck, B. (1981). Determining optical flow. *Artificial intelligence*, 17(1-3):185–203.

Huang, L., Xiao, L., and Wei, Z. (2010). Multiplicative noise removal via a novel variational model. *EURASIP Journal on Image and Video Processing*, 2010.

Huang, Y., Ng, M., and Wen, Y. (2009). A new total variation method for multiplicative noise removal. *SIAM Journal Imaging Sci.*, 2(1):20–40.

Hudson, H. (1974). *Empirical Bayes estimation*. PhD thesis, Stanford University. Dept. of Statistics.

Hudson, H. (1978). A natural identity for exponential families with applications in multiparameter estimation. *The Annals of Statistics*, 6(3):473–484.

Huynen, J. (1970). *Phenomenological theory of radar targets*. Drukkerij Bronder-Offset NV, Rotterdam, the Netherlands.

Ishikawa, H. (2003). Exact optimization for Markov random fields with convex priors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(10):1333–1336.

Jain, A. (1989). *Fundamentals of digital image processing*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA.

Just, D. and Bamler, R. (1994). Phase statistics of interferograms with applications to synthetic aperture radar. *Applied Optics*, 33(20):4361–4368.

Katkovnik, V. (1999). A new method for varying adaptive bandwidth selection. *IEEE Trans. Signal Process.*, 47(9):2567–2571.

Katkovnik, V., Astola, J., and Egiazarian, K. (2008). Phase local approximation (PhaseLa) technique for phase unwrap from noisy data. *IEEE Trans. Image Process.*, 17(6):833–846.

Katkovnik, V., Egiazarian, K., and Astola, J. (2002). Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule. *Journal of Mathematical Imaging and Vision*, 16(3):223–235.

Katkovnik, V., Foi, A., Egiazarian, K., and Astola, J. (2010). From local kernel to nonlocal multiple-model image denoising. *International Journal of Computer Vision*, 86(1):1–32.

Katkovnik, V., Foi, A., Egiazarian, K. O., and Astola, J. T. (2004). Directional varying scale approximations for anisotropic signal processing. In *EUSIPCO*, pages 101–104.

Katkovnik, V., Katkovnik, V., Egiazarian, K., and Astola, J. (2006). *Local approximation techniques in signal and image processing*. Society of Photo Optical.

Katkovnik, V. and Spokoiny, V. (2008). Spatially adaptive estimation via fitted local likelihood techniques. *IEEE Trans. Signal Process.*, 56(3):873–886.

Kay, S. (1998). *Fundamentals of statistical signal processing Volume 2: Detection theory*. Prentice Hall.

Kay, S. and Gabriel, J. (2003). An invariance property of the generalized likelihood ratio test. *IEEE Signal Process. Lett.*, 10(12):352–355.

Kendall, M. and Stuart, A. (1979). *The advanced theory of statistics. Vol. 2: Inference and relationship*. Charles Griffin and Co., Ltd., London.

Kervrann, C. and Boulanger, J. (2006). Optimal spatial adaptation for patch-based image denoising. *IEEE Trans. Image Process.*, 15(10):2866–2878.

Kervrann, C. and Boulanger, J. (2008). Local Adaptivity to Variable Smoothness for Exemplar-Based Image Regularization and Representation. *International Journal of Computer Vision*, 79(1):45–69.

Kervrann, C., Boulanger, J., and Coupé, P. (2007). Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. In *Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 520–532.

Kim, H. and Hero III, A. (2001). Comparison of GLR and invariant detectors under structured clutter covariance. *IEEE Trans. Image Process.*, 10(10):1509–1520.

Kindermann, S., Osher, S., and Jones, P. (2005). Deblurring and denoising of images by nonlocal functionals. *SIAM Journal Multiscale Model. Simul.*, 4(4):1091–1115.

Krissian, K., Westin, C., Kikinis, R., and Vosburgh, K. (2007). Oriented speckle reducing anisotropic diffusion. *IEEE Trans. Image Process.*, 16(5):1412–1424.

Krogager, E. (1990). New decomposition of the radar target scattering matrix. *Electronics Letters*, 26(18):1525–1527.

Kwatra, V., Schödl, A., Essa, I., Turk, G., and Bobick, A. (2003). Graphcut textures: image and video synthesis using graph cuts. *ACM Transactions on Graphics*, 22(3):277–286.

Lanteri, H. and Theys, C. (2005). Restoration of astrophysical images: the case of poisson data with additive gaussian noise. *EURASIP Journal on Applied Signal Processing*, 2005:2500–2513.

Le, T., Chartrand, R., and Asaki, T. (2007). A variational approach to reconstructing images corrupted by poisson noise. *Journal of Mathematical Imaging and Vision*, 27(3):257–263.

Le Pennec, E. and Mallat, S. (2005). Sparse geometric image representations with bandelets. *IEEE Trans. Image Process.*, 14(4):423–438.

Lee, C., Lee, C., and Kim, C.-S. (2011). MMSE nonlocal means denoising algorithm for Poisson noise removal. In *IEEE Int. Conf. Image Process. (ICIP)*. IEEE.

Lee, J. (1980). Digital image enhancement and noise filtering by use of local statistics. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2:165–168.

Lee, J. (1983). Digital image smoothing and the sigma filter. *Computer Vision, Graphics, and Image Processing*, 24(2):255–269.

Lee, J., Cloude, S., Papathanassiou, K., Grunes, M., and Woodhouse, I. (2003). Speckle filtering and coherence estimation of polarimetric SAR interferometry data for forest applications. *IEEE Trans. Geosci. Remote Sens.*, 41(10):2254–2263.

Lee, J., Grunes, M., and De Grandi, G. (1999). Polarimetric SAR speckle filtering and its implication for classification. *IEEE Trans. Geosci. Remote Sens.*, 37(5):2363–2373.

Lee, J., Hoppel, K., Mango, S., and Miller, A. (1994). Intensity and phase statistics of multilook polarimetric and interferometric sar imagery. *IEEE Trans. Geosci. Remote Sens.*, 32(5):1017–1028.

Lee, J., Wen, J., Ainsworth, T., Chen, K., and Chen, A. (2009). Improved sigma filter for speckle filtering of SAR imagery. *IEEE Trans. Geosci. Remote Sens.*, 47(1):202–213.

Lee, J.-S. (1981). Speckle analysis and smoothing of synthetic aperture radar images. *Computer Graphics and Image Processing*, 17(1):24–32.

Lee, T. (1996). Image representation using 2D Gabor wavelets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(10):959–971.

Lehmann, E. (1959). Optimum invariant tests. *The Annals of Mathematical Statistics*, 30(4):881–884.

Lepski, O. V., Mammen, E., and Spokoiny, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.*, 25(3):929–947.

Leung, G. and Barron, A. R. (2006). Information theory and mixing least-squares regressions. *IEEE Trans. Inf. Theory*, 52(8):3396–3410.

Li, K.-C. (1985). From Stein's unbiased risk estimates to the method of generalized cross validation. *Ann. Statist.*, 13(4):1352–1377.

Liang, L., Liu, C., Xu, Y., Guo, B., and Shum, H. (2001). Real-time texture synthesis by patch-based sampling. *ACM Transactions on Graphics*, 20(3):127–150.

Lin, Q., Vesecky, J., and Zebker, H. (1992). New approaches in interferometric SAR data processing. *IEEE Trans. Geosci. Remote Sens.*, 30(3):560–567.

Louchet, C. (2008). *Variational and Bayesian models for image denoising : from total variation towards non-local means*. PhD thesis, Université Paris Descartes.

Louchet, C. and Moisan, L. (2011). Total variation as a local filter. *SIAM Journal Imaging Sci.*, 4:651–694.

Lowe, D. (1992). Robust model-based motion tracking through the integration of search and estimation. *International Journal of Computer Vision*, 8(2):113–122.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.

Lucas, A., Aharonson, O., Deledalle, C.-A., Hayes, A., and Team, T. C. R. (2011a). Improved SAR Processing and Landform Interpretation on Titan. In *Titan Science Meeting*.

Lucas, A., Deledalle, C.-A., Aharonson, O.and Hayes, A., and Team, T. C. R. (2011b). Improved Landform Interpretation on Titan from SAR Image Processing. In *Titan Surface Workshop*.

Luisier, F., Vonesch, C., Blu, T., and Unser, M. (2010). Fast interscale wavelet denoising of Poisson-corrupted images. *Signal Processing*, 90(2):415–427.

Mahmoudi, M. and Sapiro, G. (2005). Fast image and video denoising via nonlocal means of similar neighborhoods. *IEEE Signal Process. Lett.*, 12(12):839–842.

Mairal, J., Bach, F., Ponce, J., and Sapiro, G. (2010). Online learning for matrix factorization and sparse coding. *The Journal of Machine Learning Research*, 11:19–60.

Mairal, J., Bach, F., Ponce, J., Sapiro, G., and Zisserman, A. (2009). Non-local sparse models for image restoration. *International Conference on Computer Vision (ICCV)*, pages 2272–2279.

Mairal, J., Sapiro, G., and Elad, M. (2008). Learning multiscale sparse representations for image and video restoration. *SIAM Journal Multiscale Model. Simul.*, 7(1):214–241.

Mäkitalo, M. and Foi, A. (2011). Optimal inversion of the Anscombe transformation in low-count Poisson image denoising. *IEEE Trans. Image Process.*, 20(1):99–109.

Mäkitalo, M., Foi, A., Fevralev, D., and Lukin, V. (2010). Denoising of single-look SAR images based on variance stabilization and nonlocal filters. In *Math. Meth. Electromagn. Th. (MMET)*, pages 1–4.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(7):674–693.

Mallat, S. (2009a). Geometrical grouplets. *Applied and Computational Harmonic Analysis*, 26(2):161–180.

Mallat, S. (2009b). *A wavelet tour of signal processing: the sparse way*. Elsevier/Academic Press (with contributions form Gabriel Peyré).

Mallat, S. and Peyré, G. (2007). A review of bandlet methods for geometrical image representation. *Numerical Algorithms*, 44(3):205–234.

Mallat, S. and Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41(12):3397–3415.

Mallows, C. L. (1973). Some comments on cp. *Technometrics*, 15(4):661–675.

Massonnet, D. and Rabaute, T. (1993). Radar interferometry: limits and potential. *IEEE Trans. Geosci. Remote Sens.*, 31(2):455–464.

Matsushita, Y. and Lin, S. (2007). A Probabilistic Intensity Similarity Measure based on Noise Distributions. In *IEEE Comput. Vis. and Pattern Recognition (CVPR)*, pages 1–8. IEEE.

McKinney, J., Webster, M., Webb, K., and Weiner, A. (2000). Characterization and imaging in optically scattering media by use of laser speckle and a variable-coherence source. *Optics Letters*, 25(1):4–6.

Mignotte, M. (2008). A non-local regularization strategy for image deconvolution. *Pattern Recognition Letters*, 29(16):2206–2212.

Milanfar, P. (2011). A tour of modern image filtering. *IEEE Signal Processing Magazine (to appear)*.

Minka, T. (1998). Bayesian Inference, Entropy, and the Multinomial Distribution. Technical report, Carnegie Mellon University.

Minka, T. (2000). Distance measures as prior probabilities. Technical report, Carnegie Mellon University.

Moreno, P., Ho, P., and Vasconcelos, N. (2004). A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications. *Advances in Neural Inf. Process. Syst. (NIPS)*, 16:1385–1392.

Mumford, D. and Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math*, 42(5):577–685.

Muresan, D. D. and Parks, T. W. (2003). Adaptive principal components and image denoising. In *IEEE Int. Conf. Image Process. (ICIP)*, pages 101–104. IEEE.

Nemirovski, A. S. (2000). *Topics in non-parametric statistics*, volume 1738 of *Lecture Notes in Math.* Springer.

Nikolova, M. (2000). Local strong homogeneity of a regularized estimator. *SIAM Journal on Applied Mathematics*, 61(2):633–658.

Nikolova, M. (2003). Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers. *SIAM Journal Numerical Anal.*, 40(3):965–994.

Olshausen, B. and Field, D. (1996). Natural image statistics and efficient coding. *Network: Computation in Neural Systems*, 7(2):333–339.

Orchard, J., Ebrahimi, M., and Wong, A. (2008). Efficient nonlocal-means denoising using the SVD. In *IEEE Int. Conf. Image Process. (ICIP)*, pages 1732–1735. IEEE.

Palubinskas, G., Makarau, A., and Tao, J. (2011). Fusion of optical and radar data for the extraction of higher quality information. In *TerraSAR-X Science Team Meeting*.

Pang, C., Au, O., Dai, J., Yang, W., and F.Zou (2009). A Fast NL-Means Method in Image Denoising based on the Similarity of Spatially Sampled Pixels. In *IEEE Multimedia Signal Processing (MMSP)*, pages 1–4. IEEE.

Park, J., Song, W., and Pearlman, W. (1999). Speckle filtering of SAR images based on adaptive windowing. *IEE Proceedings on Vision, Image and Signal Processing*, 146(4):191–197.

Parrilli, S., Poderico, M., Angelino, C., Scarpa, G., and Verdoliva, L. (2010). A nonlocal approach for SAR image denoising. In *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, pages 726–729. IEEE.

Pascazio, V. and Schirinzi, G. (2001). Estimation of terrain elevation by multifrequency interferometric wide band SAR data. *IEEE Signal Process. Lett.*, 8(1):7–9.

Pati, Y., Rezaiifar, R., and Krishnaprasad, P. (1993). Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Signals, Systems and Computers*, pages 40–44. IEEE.

Pelloux, S., Robillard, J., Ferrera, R., Bilbaut, A., Ojeda, C., Saks, V., Ovize, M., and Tourneur, Y. (2006). Non-beating HL-1 cells for confocal microscopy: application to mitochondrial functions during cardiac preconditioning. *Progress in biophysics and molecular biology*, 90(1-3):270–298.

Perona, P. and Malik, J. (1990). Scale space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639.

Pesquet, J., Krim, H., and Carfantan, H. (1996). Time-invariant orthonormal wavelet representations. *IEEE Trans. Signal Process.*, 44(8):1964–1970.

Peyré, G., Bougleux, S., and Cohen, L. (2008). Non-local regularization of inverse problems. In *European Conference on Computer Vision (ECCV)*, pages 57–68. Springer.

Polzehl, J. and Spokoiny, V. (2006a). Propagation-separation approach for local likelihood estimation. *Probability Theory and Related Fields*, 135(3):335–362.

Polzehl, J. and Spokoiny, V. G. (2000). Adaptive weights smoothing with applications to image restoration. *Journal of the Royal Statistical Society. Series B (Methodological)*, 62(2):335–354.

Polzehl, J. and Spokoiny, V. G. (2006b). Propagation-separation approach for local likelihood estimation. *Probab. Theory Related Fields*, 135(3):335–362.

Portilla, J., Strela, V., Wainwright, M., and Simoncelli, E. (2003). Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Trans. Image Process.*, 12(11):1338–1351.

Prati, C. and Rocca, F. (1990). Limits to the resolution of elevation maps from stereo SAR images. *International Journal of Remote Sensing*, 11(12):2215–2235.

Raginsky, M., Willett, R., Harmany, Z., and Marcia, R. (2010). Compressed sensing performance bounds under poisson noise. *IEEE Trans. Signal Process.*, 58(8):3990–4002.

Ramani, S., Blu, T., and Unser, M. (2008). Monte-Carlo SURE: a black-box optimization of regularization parameters for general denoising algorithms. *IEEE Trans. Image Process.*, 17(9):1540–1554.

Raphan, M. and Simoncelli, E. P. (2007). Learning to be Bayesian without supervision. In *Advances in Neural Inf. Process. Syst. (NIPS)*, volume 19, pages 1145–1152. MIT Press.

Rosen, P., Hensley, S., Joughin, I., Li, F., Madsen, S., Rodriguez, E., and Goldstein, R. (2000). Synthetic aperture radar interferometry. *Proceedings of the IEEE*, 88(3):333–382.

Rudin, L., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268.

Salmon, J. (2010). On two parameters for denoising with Non-Local Means. *IEEE Signal Process. Lett.*, 17(3):269–272.

Salmon, J., Deledalle, C.-A., and Duval, V. (2011). Méthodes non-locales et formes de patchs adaptatives. In *Journées de Statistique*.

Salmon, J. and Le Pennec, E. (2009). NL-Means and aggregation procedures. In *IEEE Int. Conf. Image Process. (ICIP)*, pages 2977–2980. IEEE.

Salmon, J. and Strozecki, Y. (2010). From patches to pixels in non-local methods: Weighted-Average reprojection. In *IEEE Int. Conf. Image Process. (ICIP)*, pages 1929–1932. IEEE.

Sarabandi, K. (1992). Derivation of phase statistics from the Mueller matrix. *Radio Science-Washington*, 27(5):553–560.

Scharf, L. and Demeure, C. (1991). *Statistical signal processing: detection, estimation, and time series analysis*, volume 148. Addison-Wesley Reading.

Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1):7–42.

Schmitt, M., Magnard, C., Brehm, T., and Stilla, U. (2011). Towards airborne single pass decimeter resolution sar interferometry over urban areas. *Lecture Notes in Computer Science*, 6952:197–208.

Scott, D. (1992). *Multivariate density estimation*, volume 139. Wiley Online Library.

Seeger, M. (2002). Covariance kernels from Bayesian generative models. In *Advances in Neural Inf. Process. Syst. (NIPS)*, volume 2, pages 905–912. MIT Press.

Seymour, M. and Cumming, I. (1994). Maximum likelihood estimation for SAR interferometry. In *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, pages 2272–2274.

Shen, J. (2003). On the foundations of vision modeling: I. Weber's law and Weberized TV restoration. *Physica D: Nonlinear Phenomena*, 175(3-4):241–251.

Shensa, M. (1992). The discrete wavelet transform: Wedding the a trous and Mallat algorithms. *IEEE Trans. Signal Process.*, 40(10):2464–2482.

Shi, J. and Osher, S. (2008). A nonlinear inverse scale space method for a convex multiplicative noise model. *SIAM Journal Imaging Sci.*, 1(3):294–321.

Silverman, B. (1986). *Density estimation for statistics and data analysis*, volume 26. Chapman & Hall/CRC.

Simoncelli, E. (1997). Statistical models for images: Compression, restoration and synthesis. In *Signals, Systems and Computers*, volume 1, pages 673–678. IEEE.

Simoncelli, E. (1999). Bayesian denoising of visual images in the wavelet domain. *Lecture Notes in Statistics*, pages 291–308.

Simoncelli, E. and Adelson, E. (1996). Noise removal via bayesian wavelet coring. In *IEEE Int. Conf. Image Process. (ICIP)*, pages 379–382. IEEE.

Smith, S. and Brady, J. (1997). SUSAN – A new approach to low level image processing. *International Journal of Computer Vision*, 23(1):45–78.

Solo, V. (1996). A SURE-fired way to choose smoothing parameters in ill-conditioned inverse problems. In *IEEE Int. Conf. Image Process. (ICIP)*, volume 3, pages 89–92. IEEE.

Sprague, R. (1972). Surface roughness measurement using white light speckle. *Applied Optics*, 11(12):2811–2816.

Starck, J., Candes, E., and Donoho, D. (2002). The curvelet transform for image denoising. *IEEE Trans. Image Process.*, 11(6):670–684.

Steidl, G. and Teuber, T. (2010). Removing multiplicative noise by Douglas-Rachford splitting methods. *Journal of Mathematical Imaging and Vision*, 36(2):168–184.

Stein, C. M. (1973). Estimation of the mean of a multivariate distribution. In *Asymptotic Statist.*

Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, 9(6):1135–1151.

Strong, D. and Chan, T. (2003). Edge-preserving and scale-dependent properties of total variation regularization. *Inverse problems*, 19(6):165–187.

Takeda, H., Farsiu, S., and Milanfar, P. (2007). Kernel regression for image processing and reconstruction. *IEEE Trans. Image Process.*, 16(2):349–366.

Tasdizen, T. (2008). Principal components for non-local means image denoising. In *IEEE Int. Conf. Image Process. (ICIP)*, pages 1728–1731. IEEE.

Teuber, T. and Lang, A. (2011). A new similarity measure for nonlocal filtering in the presence of multiplicative noise. *Preprint University of Kaiserslautern.*

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.

Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *International Conference on Computer Vision (ICCV)*, pages 839–846.

Touzi, R. (2004). Target scattering decomposition of one-look and multi-look SAR data using a new coherent scattering model: The TSVM. In *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, pages 2491–2494. IEEE.

Touzi, R. and Lopes, A. (1996). Statistics of the Stokes parameters and of the complex coherence parameters in one-look and multilook speckle fields. *IEEE Trans. Geosci. Remote Sens.*, 34(2):519–531.

Touzi, R., Lopes, A., Bruniquel, J., and Vachon, P. (1999). Coherence estimation for SAR imagery. *IEEE Trans. Geosci. Remote Sens.*, 37(1):135–149.

Tschumperlé, D. and Brun, L. (2011). Non-Local Regularization and Registration of Multi-Valued Images By PDE's and Variational Methods on Higher Dimensional Spaces. *Mathematical Image Processing*, 5:181–197.

Tschumperlé, D. and Deriche, R. (2005). Vector-valued image regularization with PDEs: A common framework for different applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(4):506–517.

Tsui, K. and Press, S. (1982). Simultaneous estimation of several Poisson parameters under K-normalized squared error loss. *The Annals of Statistics*, 10(1):93–100.

Tsybakov, A. (2008). *Introduction to nonparametric estimation.* Springer Verlag.

Tsybakov, A. B. (2003). Optimal rates of aggregation. In *Conference on Computational Learning Theory*, pages 303–313.

Van De Ville, D. and Kocher, M. (2009). SURE-based Non-Local Means. *IEEE Signal Process. Lett.*, 16(11):973–976.

Van De Ville, D. and Kocher, M. (2011). Non-local means with dimensionality reduction and SURE-based parameter selection. *IEEE Trans. Image Process.*, 9(20):2683–2690.

Varma, M. and Zisserman, A. (2003). Texture classification: Are filter banks necessary? In *IEEE Comput. Vis. and Pattern Recognition (CVPR)*, pages 691–698. IEEE.

Vasile, G., Trouvé, E., Lee, J., and Buzuloiu, V. (2006). Intensity-Driven Adaptive-Neighborhood Technique for Polarimetric and Interferometric SAR Parameters Estimation. *IEEE Trans. Geosci. Remote Sens.*, 44(6):1609–1621.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *IEEE Comput. Vis. and Pattern Recognition (CVPR)*, pages 511–518.

Wang, J., Guo, Y.-W., Ying, Y., Liu, Y.-L., and Peng, Q.-S. (2006). Fast non-local algorithm for image denoising. In *IEEE Int. Conf. Image Process. (ICIP)*, pages 1429–1432. IEEE.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Signal Process.*, 13(4):600–612.

Wasserman, L. (2007). *All of Nonparametric Statistics (Springer Texts in Statistics)*. Springer.

Wegmann, B. and Zetzsche, C. (1990). Statistical dependence between orientation filter outputs used in a human-vision-based image code. In *Proceedings of SPIE*, volume 1360, pages 909–923.

Willett, R., Harmany, Z., and Marcia, R. (2010). Poisson image reconstruction with total variation regularization. In *IEEE Int. Conf. Image Process. (ICIP)*, pages 4177–4180. IEEE.

Willett, R. and Nowak, R. (2004). Fast multiresolution photon-limited image reconstruction. In *IEEE Int. Symposium on Biomedical Imaging (ISBI)*, pages 1192–1195. IEEE.

Wu, F. (1982). The Potts model. *Reviews of modern physics*, 54(1):235–268.

Xiao, L., Huang, L., and Wei, Z. (2010). A weberized total variation regularization-based image multiplicative noise removal algorithm. *EURASIP Journal on Advances in Signal Processing*, 2010:1–15.

Xie, H., Pierce, L., and Ulaby, F. (2002a). SAR speckle reduction using wavelet denoising and Markov random field modeling. *IEEE Trans. Geosci. Remote Sens.*, 40(10):2196–2212.

Xie, H., Pierce, L., and Ulaby, F. (2002b). Statistical properties of logarithmically transformed speckle. *IEEE Trans. Geosci. Remote Sens.*, 40(3):721–727.

Yaroslavsky, L. P. (1985). *Digital picture processing*, volume 9 of *Springer Series in Information Sciences*. Springer-Verlag, Berlin.

Yianilos, P. (1995). Metric learning via normal mixtures. Technical report, NEC Research Institute, Princeton, NJ.

Yu, Y. and Acton, S. (2002). Speckle reducing anisotropic diffusion. *IEEE Trans. Image Process.*, 11(11):1260–1270.

Zadeh, L. (1965). Fuzzy sets. *Information and control*, 8(3):338–353.

Zhang, L., Dong, W., Zhang, D., and Shi, G. (2010a). Two-stage image denoising by principal component analysis with local pixel grouping. *Pattern Recogn.*, 43(4):1531–1549.

Zhang, X., Burger, M., Bresson, X., and Osher, S. (2010b). Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *SIAM Journal Imaging Sci.*, 3(3):253–276.

Zhang, X. and Luo, Z. (1999). A new time-scale adaptive denoising method based on wavelet shrinkage. In *IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 1629–1632.

Zhong, H., Li, Y., and Jiao, L. (2011). SAR Image Despeckling Using Bayesian Nonlocal Means Filter With Sigma Preselection. *IEEE Geosci. Remote Sens. Lett.*, 99(99):809–813.

Zitova, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000.