



**HAL**  
open science

# LE SQUELETTE DE L'UNIVERS: Un outil d'analyse topologique des grandes structures

Thierry Sousbie

► **To cite this version:**

Thierry Sousbie. LE SQUELETTE DE L'UNIVERS: Un outil d'analyse topologique des grandes structures. Cosmologie et astrophysique extra-galactique [astro-ph.CO]. Ecole normale supérieure de lyon - ENS LYON, 2006. Français. NNT: . tel-00162559

**HAL Id: tel-00162559**

**<https://theses.hal.science/tel-00162559v1>**

Submitted on 13 Jul 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N°d'ordre : 394

N° attribué par la bibliothèque : 06ENSL0 394

## **THESE**

*en vue d'obtenir le grade de*

**Docteur de l'Ecole Normale Supérieure de Lyon**

**spécialité : Sc.de la Terre et Univers**

**Laboratoire de C.R.A.L.**

*école doctorale de : Physique et Astrophysique de Lyon*

*présentée et soutenue publiquement le 14/12/2006*

*par Monsieur Thierry SOUSBIE-*

---

**Titre :**

***Le squelette de l'Univers:un nouvel outil d'analyse topologique des grandes structures.***

---

*Directeurs de thèse : Hélène COURTOIS et Christophe PICHON*

*Après avis de : Monsieur Pierre SALATI, Membre/Rapporteur*

*Monsieur Joseph SILK, Membre/Rapporteur*

*Devant la Commission d'examen formée de :*

*Madame Hélène COURTOIS, Membre*

*Monsieur Bruno GUIDERDONI, Membre*

*Monsieur Simon PRUNET, Membre*

*Monsieur Pierre SALATI, Membre/Rapporteur*

*Monsieur Joseph SILK, Membre/Rapporteur*

*Hélène Courtois*  
*Président*  
*Simon Prunet*  
*Pierre Salati*  
*Joseph Silk*



## Résumé

La distribution de la matière dans l'Univers est supposée homogène et isotrope à très grande échelle mais l'observation de la distribution des galaxies lors de grandes campagnes de recensements comme le SDSS nous montre un véritable réseau d'amas et de filaments sur des échelles de plusieurs centaines de mégaparsecs.

De nombreuses méthodes ont été développées dans le but de caractériser cette distribution et nous nous proposons dans cette thèse de présenter l'adaptation en trois dimensions d'un nouvel outil : le squelette. Cette méthode vise à donner une définition mathématique claire des filaments ainsi qu'un algorithme numérique robuste permettant leur identification ainsi que le calcul de leurs propriétés.

Afin de pouvoir comparer les résultats obtenus à partir des simulations N-corps de matière noire aux observations, une nouvelle méthode, baptisée MoLUSC, spécialisée dans la création de catalogues virtuels de galaxies a aussi été élaborée. Elle se base sur les modèles semi-analytiques et est particulièrement efficace pour la fabrication de catalogues de grande taille simulant de manière suffisamment réaliste les propriétés galactiques.

Les utilisations de ces deux outils sont nombreuses et nous montrons par exemple qu'il est possible en mesurant la densité de longueur des filaments à une échelle donnée de contraindre la quantité de matière dans l'univers  $\Omega_m$ . Ces méthodes peuvent aussi être appliquées avec succès à la mesure statistique des propriétés du flux de matière noire le long des filaments, une mesure inédite. Nous présentons enfin de nombreuses applications possibles dont les résultats préliminaires sont très encourageants.



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Le modèle cosmologique standard . . . . .	2
1.1.1	Modèle de Friedmann-Lemaître . . . . .	2
1.1.2	Constante cosmologique et évolution de l'Univers . . . . .	4
1.1.3	Mesure des paramètres cosmologiques . . . . .	8
1.2	Croissance des structures . . . . .	16
1.2.1	Inflation et approche linéaire . . . . .	16
1.2.2	Évolution non linéaire : simulations numériques . . . . .	21
1.3	Les catalogues de galaxies . . . . .	28
<b>2</b>	<b>Catalogues virtuels de galaxies</b>	<b>38</b>
2.1	Modèles semi-analytiques pour la formation de galaxies . . . . .	38
2.1.1	GalICS . . . . .	38
2.1.2	Méthodes de biais . . . . .	42
2.2	MoLUSC . . . . .	42
2.2.1	Calcul des distributions de galaxies . . . . .	43
2.2.2	Construction de catalogues fictifs . . . . .	51
<b>3</b>	<b>Topologie des grandes structures</b>	<b>60</b>
3.1	Méthodes actuelles pour étudier la topologie . . . . .	62
3.1.1	Fonctions de corrélation . . . . .	63
3.1.2	Fonctionnelles de Minkowsky . . . . .	66
3.1.3	Arbre à recouvrement minimal . . . . .	70
3.2	Le squelette . . . . .	72
3.2.1	Vrai squelette . . . . .	73
3.2.2	Approximation locale . . . . .	77
3.3	Implémentation . . . . .	83
3.3.1	Calcul du champ de densité . . . . .	83
3.3.2	Résolution des équations . . . . .	90
3.3.3	Post-traitement . . . . .	103
<b>4</b>	<b>Mesurer l'univers avec le squelette 3D</b>	<b>116</b>
4.1	La longueur totale du squelette comme mesure des structures à grande échelle . . . . .	116
4.1.1	Sélection des galaxies dans le catalogue SDSS . . . . .	117

---

4.1.2	Création des catalogues virtuels . . . . .	118
4.1.3	Lissage et échantillonnage du champ de densité . . . . .	123
4.1.4	Calcul des squelettes . . . . .	125
4.1.5	Résultats . . . . .	126
4.2	Caractéristiques du flux de matière noire au voisinage du squelette . .	133
4.2.1	Méthode de calcul . . . . .	133
4.2.2	Mesure des moments le long du squelettes . . . . .	136
4.2.3	Influence comparée du squelette à différentes échelles . . . . .	138
4.2.4	Evolution avec le décalage spectral . . . . .	139
4.2.5	Moment cinétique des halos . . . . .	143
4.2.6	Propriétés des flux de matière dans les filaments . . . . .	143
4.3	Autres applications et perspectives . . . . .	147
4.3.1	Test de Alcock-Paczynski (A-P) . . . . .	147
4.3.2	Tests de gaussianité . . . . .	151
4.3.3	Section des filaments . . . . .	154
4.3.4	Mesures de distance entre deux squelettes . . . . .	156
<b>5</b>	<b>Conclusion</b>	<b>163</b>
	<b>Annexes</b>	<b>165</b>
<b>A</b>	<b>Tirage aléatoire selon une loi de probabilité donnée</b>	<b>165</b>
<b>B</b>	<b>Le squelette et l'Univers local</b>	<b>166</b>







# Remerciements

De nombreuses personnes m'ont aidé de différentes façons durant cette thèse ...

Hélène tout d'abord, même si ça n'était pas gagné au début. Merci de m'avoir laissé une telle liberté, de toujours m'avoir donné d'aussi bons conseils et aussi pour ta constante bonne humeur (et surtout de m'avoir permis de faire ces deux séjours à Hawaï). Merci aussi à tous ceux de l'IAP avec qui j'ai collaboré de près ou de loin et bien sûr tout particulièrement Christophe ainsi que Stéphane et Simon. Un grand merci à Julien aussi qui a toujours été là pour répondre à mes questions ...

J'ai aussi beaucoup apprécié mon séjour à Oxford qui fut très instructif, pour cela je remercie particulièrement Joe Silk de m'avoir accueilli ainsi que Greg Bryan auprès de qui j'aurai appris beaucoup de choses.

Merci aussi à tous les gens de l'observatoire, j'aurais passé trois années très agréables en votre compagnie.

Merci enfin à toute ma famille, mon frère, mes parents qui ont toujours été là pour m'aider ainsi que mes grands parents pour leur gentillesse. Merci aussi à tous mes amis et pour finir un gros bisou à Sandrine, merci d'être là :)



# Chapitre 1

## Introduction

Depuis l'observation des premières galaxies hors voie-lactée jusqu'aux récents recensements de leur distribution dans l'univers local qui ont commencé dans les années 80, on a successivement observé l'apparition de structures à des échelles de plus en plus grandes. Depuis CFA1 [14] et son "Great Wall" aux catalogues les plus récents tels que le 2dF [8] ou le SDSS [16], la précision et la taille grandissante de ces bases de données n'ont cessé de se montrer une source intarissable d'informations sur la formation et l'évolution de l'Univers. Aujourd'hui, elles nous permettent d'espérer trouver une réponse à certaines questions fondamentales de cosmologie. Parmi celles-ci, en voilà quelques unes que nous nous proposons d'aborder dans cette thèse :

- Comment se répartit la matière dans l'Univers ?
- La matière noire est censée constituer 85% du contenu en matière total de l'univers ... Que pouvons nous dire de sa distribution et de son influence sur l'évolution de celle des galaxies ?
- Quel modèle est-il le plus à même de rendre compte de l'évolution de l'univers ? Et surtout, comment contraindre ces modèles observationnellement ?
- Quelle est la nature du biais entre la distribution des galaxies et celle de la matière en général ? Comment le mesurer ?
- Les données actuelles permettent-elles de mettre en défaut le modèle de concordance ?

La possibilité d'apporter une réponse claire à ces questions dépend en grande partie de notre capacité à extraire des informations de ces catalogues et à les comparer à des prédictions théoriques. Par bonheur, il se trouve que nous sommes à une époque où la puissance de calcul numérique des ordinateurs ainsi que la qualité des modèles numériques atteignent un niveau tel qu'il nous est possible de simuler de vastes portions d'univers (de plusieurs centaines de mégaparsecs (Mpc) cube) tout en conservant une précision suffisante aux échelles plus modestes. En utilisant des programmes de simulation N-corps en arbre tels que GADGET2 [44], il est maintenant facile de simuler l'évolution de centaines de millions de particules traçant le champ de densité de la matière noire dans des boîtes pouvant aller de quelques mégaparsecs de côté jusqu'à plus d'un gigaparsec. Le principal défi réside donc dans la confrontation de ces immenses quantités de données aux observations.

## 1.1 Le modèle cosmologique standard

Nous allons ici présenter les bases du modèle cosmologique standard qui vise à décrire l'évolution dynamique de l'univers dans son ensemble. Le but en est simplement d'introduire les concepts fondamentaux et de présenter les paramètres libres et leur influence. Les ouvrages de Misner *et al.* [31], Weinberg [50], Peebles [36] et Padmanabhan [34] constituent des références en la matière et cette section en est fortement inspirée.

### 1.1.1 Modèle de Friedmann-Lemaître

Depuis Copernic, la cosmologie repose sur le principe central que nous ne sommes pas au centre de l'univers. L'approche moderne va plus loin et établit le principe cosmologique : non seulement nous ne sommes pas au centre de l'univers mais de plus il n'y a pas de centre. En d'autres termes, l'univers apparaît identique vu de n'importe quel endroit et sous n'importe quel angle : il est homogène et isotrope. Il est clair que cela n'est pas vrai à toutes les échelles, le système solaire par exemple n'est certainement pas homogène et isotrope. Nous supposons donc qu'il existe une échelle d'homogénéité au delà de laquelle le principe cosmologique est vrai, les inhomogénéités à plus petite échelle (mais suffisamment grande tout de même) pouvant éventuellement être traitées comme des perturbations (voir la section 1.2.1).

La relativité générale traduit le principe cosmologique en termes de symétries : un univers homogène et isotrope est invariant par rotation et translation. D'un point de vue mathématique, les conséquences de telles symétries se font ressentir sur un objet appelé métrique qui décrit la manière de mesurer les distances entre deux points. Dans un espace Euclidien par exemple, la distance  $ds$  entre deux points séparés par le vecteur  $d\mathbf{x} = (dx, dy, dz) = (dx^1, dx^2, dx^3)$  est :

$$ds^2 = dx^2 + dy^2 + dz^2. \quad (1.1)$$

En définissant la métrique  $\eta_{ij}$  comme la matrice permettant le passage des coordonnées aux distances, on obtient alors :

$$ds^2 = \sum_{i,j} \eta_{ij} dx^i dx^j = \eta_{ij} dx^i dx^j \quad (1.2)$$

où la convention de sommation implicite a été utilisée et où  $\eta_{ij} = \text{diag}(1, 1, 1)$ . La relativité restreinte nous apprend cependant que les variables de temps  $t$  et d'espace  $d\mathbf{x}$  doivent être considérées sur un pied d'égalité et que les distances doivent donc être mesurées dans l'espace temps de Minkowsky dont la métrique s'écrit  $\eta_{\mu\nu} = \text{diag}(c^2, -1, -1, -1)$ , la distance propre entre deux événements étant donnée par l'équation :

$$ds^2 = c^2 dt^2 - d\mathbf{x}^2, \quad (1.3)$$

avec  $c \approx 299,792,458$  m/s la vitesse de la lumière. Dans la suite, nous poserons  $c = 1$  pour simplifier les notations.

La théorie de la relativité générale propose de considérer la métrique comme un objet dynamique reliant les propriétés de l'espace à la distribution d'énergie supposée déformer l'espace temps. Son équation fondamentale s'écrit :

$$\left( G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \right) = 8\pi G T_{\mu\nu}. \quad (1.4)$$

Dans cette équation, le tenseur  $G_{\mu\nu}$  dépend uniquement de la métrique et de ses dérivées et décrit en quelques sortes la courbure de l'espace. La partie droite de l'équation relie quand à elle cette courbure au contenu en matière et énergie de l'univers décrit par le tenseur énergie-impulsion  $T_{\mu\nu}$  et la constante gravitationnelle  $G$ .

Le modèle de Friedmann-Lemaître regroupe le principe cosmologique et l'équation d'Einstein (1.4). Pour cela, il introduit dans l'équation (1.4) l'expression la plus générale possible de la métrique dans un espace-temps homogène et isotrope qui fut découverte par Robertson [41] et Walker [49] :

$$ds^2 \equiv g_{\mu\nu} dx^\mu dx^\nu = dt^2 - a^2(t) \left( \frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right). \quad (1.5)$$

Dans cette expression,  $a(t)$  est appelé facteur d'échelle et décrit l'expansion de l'univers à un temps  $t$ . Le terme  $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$  désigne quand à lui la métrique à deux dimensions de la surface d'une sphère. Enfin, le paramètre  $k$  est le facteur de courbure qui peut prendre une valeur positive pour un univers positivement courbé ou "fermé" (du type d'une sphère), nulle pour un univers plat, où négative pour un univers négativement courbé dit "ouvert" (du type d'une selle de cheval). Actuellement, les observations suggèrent que le facteur  $a(t)$  a toujours augmenté et même que cette augmentation tend à s'accélérer, de plus, la valeur de  $k$  est mesurée quasiment nulle. On peut alors montrer que sous l'hypothèse d'une métrique de Robertson-Walker, le tenseur de Ricci est diagonal et s'écrit donc :

$$\begin{aligned} R_{00} &= 3 \frac{\ddot{a}}{a} \\ R_{ij} &= g_{ij} \left( \frac{\ddot{a}}{a} + 2H^2 + 2\frac{k}{a^2} \right) \end{aligned} \quad (1.6)$$

où  $R_{00}$  est la composante temporelle et  $R_{ij}$  désigne les composantes spatiales du tenseur de Ricci. Dans cette expression la "constante" de Hubble  $H = \dot{a}/a$  a été introduite. Pour résoudre l'équation d'Einstein, il ne reste alors plus qu'à définir le tenseur énergie-impulsion que nous supposons être de type gaz parfait :

$$T^{\mu\nu} = (P + \rho) U^\mu U^\nu - P g^{\mu\nu} \quad (1.7)$$

où  $\rho$  désigne la densité d'énergie,  $P$  la pression et le vecteur  $U^\mu = dx^\mu/ds$  représente la quadri-vitesse du contenu de l'Univers. Dans le repère comobile où ce fluide est

au repos, on aura donc  $T^{\mu\nu} = \text{diag}(\rho, P, P, P)$ .

Il devient alors possible de résoudre les équations (1.4) afin d'obtenir les équations de Friedmann. La partie temporelle tout d'abord donne :

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \quad (1.8)$$

et la partie spatiale :

$$\frac{\ddot{a}}{\dot{a}} = \frac{-4\pi G}{3}(\rho + 3P). \quad (1.9)$$

De la même manière, il est possible d'exprimer la conservation du tenseur énergie impulsion  $D_\mu T^{\mu\nu} = 0$  sous la forme :

$$\frac{d}{dt}(\rho a^3) = -P \frac{da^3}{dt}. \quad (1.10)$$

Cette dernière équation peut simplement être interprétée comme exprimant l'adiabaticité, c'est à dire du type  $dU = -PdV$  où  $dU$  est la variation d'énergie interne du gaz,  $P$  sa pression et  $dV$  son volume. Cet ensemble de trois équations décrit l'évolution dynamique de l'Univers où il convient cependant encore de définir la nature du contenu énergétique de l'Univers par le biais de l'équation d'état qui sera postulée du type :

$$P = \omega\rho. \quad (1.11)$$

Cette équation peut être facilement comprise en se représentant une boîte de volume  $V = a^3$  suivant l'expansion de l'univers. Si la boîte est remplie de matière ordinaire, sa densité d'énergie est proportionnelle à sa masse divisée par son volume :  $\rho \propto a^{-3}$ . En utilisant l'équation (1.10), il est alors facile de voir que  $P = 0$  et donc  $\omega = 0$ . Pour des particules relativistes comme des photons, on trouvera ainsi  $\omega = 1/3$  et pour le terme de courbure  $k$ ,  $\omega = -1/3$ .

### 1.1.2 Constante cosmologique et évolution de l'Univers

Quelle que soit la valeur de la constante de Hubble  $H$ , on peut remarquer dans l'équation (1.8) qu'il est possible de définir une densité critique pour laquelle l'univers est plat (i.e.  $k = 0$ ) :

$$\rho_c \equiv \frac{3H^2}{8\pi G}. \quad (1.12)$$

Il devient alors pratique d'exprimer la densité d'énergie en fonction de cette densité critique et l'on définit donc :

$$\Omega = \frac{\rho}{\rho_c} = \frac{8\pi G}{3H^2}\rho. \quad (1.13)$$

Ce choix est pratique car il permet de faire rapidement le lien avec la courbure : si  $\Omega = 1$ , la courbure est nulle, pour  $\Omega > 1$ , l'univers est fermé et pour  $\Omega < 1$  il est ouvert.

Imaginons maintenant que l'univers soit exclusivement constitué de matière baryonique, de densité d'énergie  $\rho_m$ . Il nous est alors possible de prédire son évolution en fonction de la valeur de  $\Omega_m = \rho_m/\rho_c$ , la fraction d'énergie disponible sous forme de matière (ici, on aura donc  $\Omega_m = \Omega$ ). Définissons alors :

$$\begin{aligned} H_0 &= \left(\frac{\dot{a}}{a}\right)_0 \\ \Omega_0 &= \frac{8\pi G\rho_m(t_0)}{3H_0^2} \\ 1+z &= \frac{a_0}{a}. \end{aligned} \tag{1.14}$$

où  $H_0$  désigne la valeur actuelle de la constante de Hubble,  $\Omega_0$  la quantité d'énergie actuellement sous forme de matière et  $z$  le décalage spectral vers le rouge (voir la section suivante). On obtient alors, selon la valeur de  $\Omega_0$  :

– Pour  $\Omega_0 = 1$  :

$$t = \frac{2}{3H_0(1+z)^{3/2}}. \tag{1.15}$$

– Pour  $\Omega_0 > 1$  :

$$t = \frac{\Omega_0}{2H_0(\Omega_0 - 1)^{3/2}} \left( \arccos \left[ \frac{2 + \Omega_0(z - 1)}{(1+z)\Omega_0} \right] - \frac{2\sqrt{(1+z\Omega_0)(\Omega_0 - 1)}}{(1+z)\Omega_0} \right). \tag{1.16}$$

– Pour  $0 \leq \Omega_0 < 1$  :

$$t = \frac{\Omega_0}{2H_0(1 - \Omega_0)^{3/2}} \left( -\cosh^{-1} \left[ \frac{2 + \Omega_0(z - 1)}{(1+z)\Omega_0} \right] + \frac{2\sqrt{(1+z\Omega_0)(1 - \Omega_0)}}{(1+z)\Omega_0} \right). \tag{1.17}$$

Dans ces trois cas, la valeur de  $a$  ayant été normalisée à sa valeur présente :  $a \equiv a/a_0$ . Comme l'a justement remarqué Einstein en son temps, aucune de ces solutions ne conduit à un univers statique (voir la figure 1.1 page 7). En effet, dans le premier cas, l'univers est plat et l'expansion est éternelle ( $a \propto t^{2/3}$ ). Dans le deuxième cas, l'univers est sphérique et devrait donc finir par s'effondrer sur lui-même, le facteur d'expansion  $a$  décrivant une sorte de cycloïde en fonction du temps. Enfin, dans le dernier cas, l'univers a une géométrie hyperbolique (de type selle de cheval) et l'expansion est aussi éternelle (mais  $a$  n'est pas constant asymptotiquement, contrairement au cas  $\Omega_m = 1$ ).

Au début du XX<sup>ème</sup> siècle, il semblait évident que l'univers devait être statique. Les implications du modèle décrit précédemment semblaient donc inacceptables et Einstein lui-même décida que quelque chose devait manquer. La solution qu'il trouva s'appelle la "constante cosmologique". Le terme  $G_{\mu\nu}$  étant construit de manière à être l'expression covariante la plus générale possible dépendant uniquement de la



métrique et de ses deux premières dérivées, l'ajout d'un terme constant  $\Lambda g_{\mu\nu}$  est tout à fait envisageable. L'équation (1.4) devient alors :

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi GT_{\mu\nu}. \quad (1.18)$$

où  $\Lambda$  désigne la constante cosmologique. Un autre point de vue consiste à considérer la constante cosmologique comme appartenant au terme de droite, ce qui revient à ajouter une constante au tenseur énergie impulsion  $T_{\mu\nu} \mapsto T_{\mu\nu} + \Lambda g_{\mu\nu}$  qui peut alors être interprétée comme l'énergie du vide (constante et ne dépendant aucunement de la distribution de matière). L'équation de conservation (1.10) reste alors inchangée ainsi que les équations locales mais ce n'est pas le cas de la cosmologie. Après cet ajout, les équations de Friedmann (1.8) et (1.9) ne sont donc pas modifiées mais l'expression de la densité d'énergie  $\rho = \rho_m + \rho_\Lambda$  doit simplement tenir compte de l'ajout de cette nouvelle composante :

$$\rho_\Lambda = \frac{\Lambda}{8\pi G}. \quad (1.19)$$

De plus, l'équation d'état reliant pression et énergie donne  $P_\Lambda = -\rho_\Lambda$  (i.e.  $\omega_\Lambda = -1$ ), c'est à dire que la constante cosmologique a l'étrange propriété d'exercer une pression négative. Il devient facile de construire un univers statique (c'est à dire où  $\ddot{a} = \dot{a} = 0$ ) en choisissant des valeurs de  $\Omega_m$  et  $\Omega_\Lambda = \rho_\Lambda/\rho_c$  adéquates :

$$\rho_m = \frac{k}{4\pi G a^2} \quad (1.20)$$

$$\rho_\Lambda = \frac{k}{8\pi G a^2}. \quad (1.21)$$

Il est assez amusant de constater que c'est à la même époque que Edwin Hubble commença à mesurer les vitesses de récession des galaxies environnantes pour s'apercevoir qu'elles s'éloignaient toutes de la terre à une vitesse d'autant plus grande qu'elles étaient éloignées. C'est la fameuse loi de Hubble :

$$v = H_0 d, \quad (1.22)$$

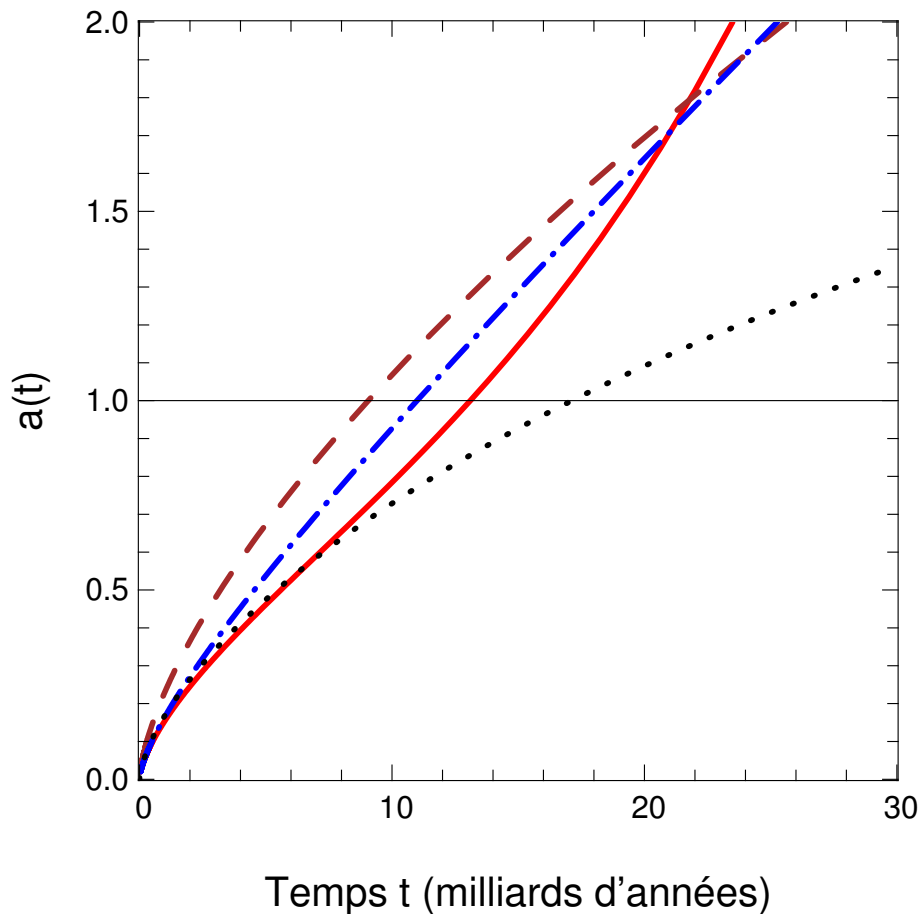
où la valeur de  $H_0$  est actuellement estimée à  $H_0 = 72 \pm 8$  km/s/Mpc [13]. La constante cosmologique n'étant plus nécessaire, elle fut abandonnée pour un temps pour mieux revenir vers la fin des années 90 avec les premières mesures effectuées sur les supernovæ [39] qui déterminèrent à partir d'un échantillon de 42 supernovæ que, dans le cas d'un espace plat ( $\Omega_k = 0$ ),

$$\Omega_m = 0.28 \pm 0.14 \quad (1.23)$$

$$\Omega_\Lambda = 1 - \Omega_m = 0.72 \pm 0.14. \quad (1.24)$$

En résolvant les équations de Friedmann avec constante cosmologique, il est possible de déterminer l'avenir d'un tel Univers, et l'on obtient dans le cas général où  $0 < \Omega_0 < 1$  et  $\Omega_{\Lambda 0} = 1 - \Omega_0$  :

$$t = \frac{2}{3H_0\sqrt{(1-\Omega_0)}} \sinh^{-1} \left( \frac{\sqrt{1/\Omega_0 - 1}}{(1+z)^{3/2}} \right). \quad (1.25)$$



**Fig. 1.1:** Les différentes évolutions du paramètre d'échelle  $a(t)$  selon la quantité de matière  $\Omega_m$  et de constante cosmologique  $\Omega_\Lambda$  dans l'univers. La courbe marron en tirets correspond à un univers fermé  $\Omega_m > 1$ ,  $\Omega_\Lambda = 0$  qui finira par s'effondrer sur lui-même, la courbe noire pointillée le cas d'un univers ouvert  $\Omega_m < 1$ ,  $\Omega_\Lambda = 0$  et la courbe bleue en tirets-points le cas d'un univers plat  $\Omega_m = 1$ ,  $\Omega_\Lambda = 0$ . La courbe continue rouge décrit l'évolution de  $a(t)$  pour le modèle actuel  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$  et le temps présent correspond à la valeur  $a(t_0) = 1$ .

En d'autres termes, pour les valeurs actuellement mesurées des paramètres cosmologiques, le facteur d'expansion  $a(t)$  est non seulement croissant, mais cette croissance est, de plus, accélérée. La figure 1.1 expose les évolutions possibles du paramètre d'échelle  $a(t)$  selon le modèle envisagé. Le temps correspondant à la valeur  $a(t) = 1$  donne l'âge actuel de l'Univers, ce qui exclu directement les modèles  $\Omega_\Lambda = 0$  avec  $\Omega_m \geq 1$  dans lesquels l'univers serait plus jeune que son contenu. Dans le modèle actuel, correspondant aux mesures des supernovæ, l'Univers est en train d'entrer dans une phase où son expansion s'accélère comme le montre la courbe pleine rouge.

Pour conclure cette section, notons cependant les deux principaux problèmes soulevés par la constante cosmologique. Son interprétation très pratique comme "énergie du vide" tout d'abord n'est pas si évidente que cela. En effet, en faisant la simple hypothèse que  $\Omega_\Lambda < 1$ , ce qui peut être considéré comme établi de nos jours,

l'ordre de grandeur de l'énergie de la constante cosmologique et celui prédit pour le vide  $\rho_v \approx m_{\text{Pl}}^4$  par la théorie quantique des champs (où  $m_{\text{Pl}} \approx 10^{19}$  GeV désigne la masse de Planck) sont totalement incompatibles :

$$\rho_\Lambda < 10^{-120} \rho_v. \quad (1.26)$$

À ce jour, aucune explication n'existe et il semble très étrange qu'un quelconque mécanisme naturel puisse supprimer les effets de l'énergie du vide juste dans les bonnes proportions pour qu'elle soit observée à une échelle énergétique de cent vingt ordres de grandeur plus petite. Ce problème est appelé sobrement "problème de la constante cosmologique" mais il en existe un second tout aussi gênant. Il est possible, à partir de l'équation de conservation (1.10), de prédire pour un modèle cosmologique donné l'évolution des différents composants si l'on suppose une équation d'état de type (1.11) avec  $\omega$  constant. On obtient alors :

$$\rho_i \propto a^{-3(1+\omega)} \quad (1.27)$$

où  $\rho_i$  désigne la densité d'énergie de chacune des composantes de l'univers (matière, constante cosmologique, photons ...). Le problème posé est appelé "problème de coïncidence" et provient du fait que nous vivons apparemment à une époque où les valeurs de  $\Omega_m$  et  $\Omega_\Lambda$  sont du même ordre de grandeur, ce qui constitue effectivement une grande coïncidence étant donné les évolutions très différentes suivies par ces deux densités : l'une est constante et l'autre varie comme l'inverse du cube du facteur d'expansion. Dans un univers sensé être éternel (il n'y a pas d'effondrement dans ce modèle), il est en effet étrange que nous vivions exactement pendant la période très courte où  $\Omega_m$  n'est pas négligeable devant  $\Omega_\Lambda$ . Ces deux problèmes semblent suggérer qu'il existe peut être un autre mécanisme que la constante cosmologique dans lequel les valeurs de  $\Omega_m$  et  $\Omega_\Lambda$  resteraient comparables et le modèle de quintessence par exemple propose d'expliquer cette énergie noire par un fluide modélisé par un champ scalaire avec une équation d'état (1.11) dépendante du temps ( $\omega = \omega(t)$ ).

### 1.1.3 Mesure des paramètres cosmologiques

Nous venons de voir dans la section précédente que l'évolution de l'univers est liée à la proportion de chacune de ses composantes. En définissant  $\Omega_i = \rho_i^0 / \rho_c^0$ , chaque valeur de  $i$  correspondant à un type de fluide différent (matière, constante cosmologique ou même baryon et rayonnement), il est possible d'exprimer la quantité d'énergie sous forme de courbure simplement :

$$\Omega_k = 1 - \Omega_\Lambda - \Omega_m = \frac{-k}{a_0^2 H_0^2} \quad (1.28)$$

avec  $H_0 = 100h$ . Dans cette expression, la quantité d'énergie sous forme de rayonnement  $\Omega_r$  est considérée comme négligeable (attention cependant au fait que cela n'a pas toujours été le cas, en effet  $\rho_r \propto a^{-4}$  alors que  $\rho_m \propto a^{-3}$ ). Les baryons sont quand à eux inclus dans le terme  $\Omega_m$ .

## Distances dans un univers de Friedmann-Robertson-Walker (FRW)

Il est donc extrêmement important de pouvoir mesurer les valeurs de  $\Omega_i$  aussi précisément que possible et le seul moyen existant est l'observation. Commençons donc par examiner les prédictions du modèle FRW sur les distances et le trajet de la lumière dans un univers en expansion. Etant donné que la principale observable à notre disposition est la lumière provenant des étoiles et des galaxies, nous allons tout d'abord définir le décalage spectral vers le rouge  $z$  qui décrit comment les spectres émis par les galaxies sont décalés du fait que l'univers s'est étendu entre le moment où la lumière a été émise  $t_e$  et celui où elle a été observée à  $t_o$ . Tout comme le reste de l'univers, la longueur d'onde  $\lambda$  d'un photon subit l'expansion, par conséquent, elle est proportionnelle à  $a(t)$  :

$$\lambda \propto a(t). \quad (1.29)$$

En d'autres termes,

$$\frac{\lambda_o}{\lambda_e} = a(t_o)/a(t_e) \equiv 1 + z, \quad (1.30)$$

où l'on vient de définir le décalage vers le rouge  $z$  (souvent appelé "redshift"). En posant  $a(t_o) = 1$ , il devient possible d'exprimer la distance à une source sous la forme de son décalage spectral :

$$z = \frac{1}{a} - 1. \quad (1.31)$$

Pour relier  $z$  à la distance, il existe plusieurs méthodes. L'une d'elle consiste à définir la distance de luminosité  $D_L$  par le rapport de la luminosité absolue  $L$  d'une source à sa luminosité apparente  $l$  :

$$D_L \equiv \sqrt{\frac{L}{4\pi l}}. \quad (1.32)$$

Dans un espace Euclidien, une onde sphérique émise se dilue avec le carré de la distance parcourue par les photons et  $D_L$  correspond à la distance usuelle. Dans l'univers en expansion, les choses sont cependant différentes. L'énergie émise pendant un temps  $\delta t_e$  est reçue en un temps  $\delta t_o = (1 + z)\delta t_e$  mais cette énergie est de plus divisée par un facteur  $1 + z$  du fait de l'augmentation de longueur d'onde des photons. Pour une source se trouvant à une distance  $D_e$  de l'observateur au moment de l'émission, on obtient donc :

$$l = \frac{L\delta t_e}{\delta t_o(1+z)} \frac{1}{4\pi D_e^2 (a_o/a_e)^2} = \frac{L}{4\pi D_e^2 (a_o/a_e)^2 (1+z)^2}, \quad (1.33)$$

soit :

$$D_L = D_e \frac{a_o}{a_e} (1+z). \quad (1.34)$$

A noter que  $D_L$  est différente de la "vraie" distance  $D_p = D_e(a_o/a_e)$  (que l'on aurait pu mesurer avec une règle suffisamment grande). La distance  $D_p$  est appelée distance

physique. Il est pratique de définir aussi une distance comobile  $D = D_p/a(t)$  qui a la propriété intéressante de rester constante pour deux points de vitesse relative nulle (il s'agit de la distance la plus couramment utilisée pour les simulations numériques). On obtient alors simplement :

$$D_L = D(1 + z). \quad (1.35)$$

Un autre moyen de mesurer une distance consiste simplement à utiliser une triangulation. La distance angulaire est définie comme le rapport de la taille physique d'un objet à sa taille angulaire observée et vaut :

$$D_A = \frac{D}{1 + z} = \frac{D_L}{(1 + z)^2}. \quad (1.36)$$

En réécrivant l'équation (1.8), il est aussi possible d'exprimer la distance en fonction du redshift ;

$$\frac{dz}{dt} = -H_0(1 + z)E(z) \quad (1.37)$$

avec

$$E(z) = \sqrt{\Omega_m(1 + z)^3 + \Omega_k(1 + z)^2 + \Omega_\Lambda}, \quad (1.38)$$

où l'on a utilisé l'équation (1.27). L'équation (1.37) se réécrit alors :

$$\frac{c}{H_0} \frac{dz}{E(z)} = (1 + z)cdt \quad (1.39)$$

où le membre de droite représente la distance comobile parcourue pendant un temps  $dt$  par un photon. On obtient, dans un espace plat :

$$D_{k=0} = \chi = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}. \quad (1.40)$$

Dans le cas général d'un espace courbe, la distance comobile mesurée peut être déduite du fait que les photons suivent des géodésiques (i.e.  $ds^2 = 0$ ). En utilisant la métrique de Robertson-Walker (1.5) et en fixant les coordonnées angulaires, l'équation des géodésiques s'écrit :

$$\frac{dt}{a(t)} = \frac{dr}{\sqrt{1 - kr^2}}. \quad (1.41)$$

En intégrant le membre de gauche entre le temps d'émission  $t_e$  et de réception  $t_o$  et celui de droite entre 0 et la distance parcourue  $D_{k=0}$ , l'expression de  $D$  en fonction de la courbure  $\Omega_k$  et de  $E(z)$  s'exprime comme :

$$D = f_k(\chi) = \begin{cases} \frac{c}{H_0\sqrt{\Omega_k}} \sinh\left(\frac{H_0\sqrt{\Omega_k}}{c} D_{k=0}\right) & \text{si } \Omega_k > 0 \\ D_{k=0} & \text{si } \Omega_k = 0 \\ \frac{c}{H_0\sqrt{-\Omega_k}} \sin\left(\frac{H_0\sqrt{-\Omega_k}}{c} D_{k=0}\right) & \text{si } \Omega_k < 0 \end{cases}.$$

## Mesures des paramètres cosmologiques avec les Supernovæ

La détermination des paramètres cosmologiques à partir de la mesure des courbes de luminosité des supernovæ se fait par le biais de la relation distance luminosité (1.32). On pense actuellement que les Supernovæ de type Ia sont des “chandelles standard”, c’est à dire que la quantité d’énergie lumineuse émise est proportionnelle à leur temps d’extinction, ce dernier étant très facile à mesurer. Connaissant la luminosité apparente et le décalage spectral, il devient alors possible de tracer la courbe  $D_L = f(z)$  et par conséquent de contraindre les valeurs des paramètres cosmologiques en utilisant l’équation (1.42) avec  $D_L = (1+z)D$ . En pratique,  $D_L$  est mesuré à partir des magnitudes apparentes  $m$  et absolues  $M$  en utilisant la formule :

$$m - M = 5 \log_{10} \frac{D_L}{1 \text{ pc}} - 5. \quad (1.42)$$

Il semble donc que la méthode des supernovæ soit excellente, plusieurs problèmes se posent cependant. Le suivi tout d’abord n’est pas évident, la durée de visibilité d’une explosion étant de l’ordre de trois semaines, de grandes campagnes monopolisant de grands télescopes sur plusieurs années sont nécessaires afin de pouvoir détecter un nombre significatif d’événements. Habituellement, ces détections se font en scrutant régulièrement une même zone du ciel afin de repérer les objets nouvellement apparus. En les suivant sur quelques jours, il est possible de déterminer, selon leur comportement, s’il s’agit de supernovæ ou non et le cas échéant de les spectrographier afin de s’assurer du résultat. Le suivi de la magnitude dès les premiers jours est crucial car il détermine la qualité de l’évaluation de la magnitude absolue. Un autre problème vient de la difficulté d’observer des supernovæ à grand décalage spectral. Ce sont certes des objets très brillants mais les durées des poses sont limitées par la nécessité de scruter les mêmes champs encore et encore. Comme le montre la figure 1.2, c’est à grand  $z$  que la distance de luminosité est la plus sensible aux variations des paramètres cosmologiques et les mesures proches sont donc très dégénérées.

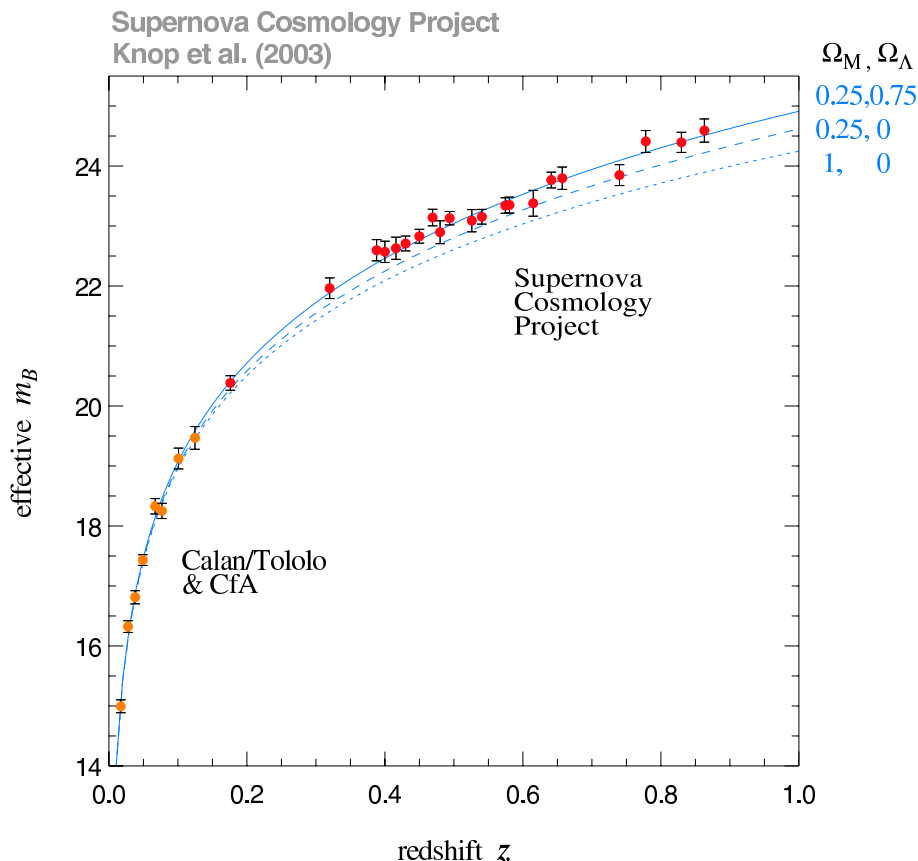
Les premiers résultats obtenus sont ceux du Supernovæ Cosmology Project [39] qui obtiennent avec 42 supernovæ une contrainte :

$$0.8\Omega_M - 0.6\Omega_\Lambda \approx -0.2 \pm 0.1. \quad (1.43)$$

et des données plus récentes incluant notamment 11 nouvelles supernovæ à grand redshift [23] prédisent dans l’hypothèse d’un univers plat (ce qui est déduit de l’étude du CMB) des valeurs de  $\Omega_m = 0.25 \pm 0.07$  et  $\Omega_\Lambda = 0.75 \pm 0.07$ , les marges d’erreur étant purement statistiques.

## Fond de rayonnement cosmologique

Il est facile de voir à partir de l’équation (1.27) que dans les premiers temps de l’univers le rayonnement constituait la majeure partie de l’énergie. Il se trouve que



**Fig. 1.2:** Diagramme de Hubble obtenu à partir de mesures de supernovæ [23]. Chacune des courbes bleues représente différents modèles cosmologiques mettant en évidence la forte dégénérescence à petit décalage spectral.

ce rayonnement devait aussi respecter le spectre de Planck, les diverses réactions chimiques étant à l'équilibre, et par conséquent  $\rho_r \propto T^4$ , soit :

$$\rho_r \propto a(t)^{-4} \propto T(t)^4. \quad (1.44)$$

Autrement dit, la température est inversement proportionnelle au facteur d'expansion.

Cette observation est en accord avec le modèle du Big Bang chaud dans lequel l'Univers est supposé extrêmement chaud et dense à son origine. Pour résumer rapidement, la température tendant vers l'infini à l'origine, le fluide cosmique n'était alors constitué que d'une soupe de particules élémentaires. La température diminuant avec l'expansion de l'univers, on estime que vers  $t \approx 10^{-6}$  sec (soit  $T \approx 10^{13}$  K), les premiers protons et neutrons purent se former et aux alentours de  $t \approx 3$  min ( $T \approx 10^{10}$  K) ceux-ci formèrent les premiers noyaux atomiques. Cette période est appelée nucléosynthèse primordiale et il se trouve que la prédiction théorique des quantités des divers éléments chimiques prédite par la théorie correspond parfaitement bien aux observations corroborant ainsi la théorie du Big-Bang (voir Riotto [40] par exemple). Il est intéressant aussi de noter que la quantité de baryons  $\Omega_b$  ob-

servable de nos jours peut être calculée à partir du rapport  $\eta$  du nombre de baryons au nombre de photons à l'époque, soit :

$$\Omega_b h^2 = 3.64 \cdot 10^{-3} (10^{10} \eta) \approx 0.02 \pm 0.002 \quad (1.45)$$

ce qui ne constitue qu'une très faible fraction de la matière observée ( $\Omega_m \approx 0.3$ ).

L'Univers étant toujours trop chaud pour que des atomes neutres et stables se forment avant  $T \approx 3000$  K ( $t \approx 300,000$  ans), les photons subissaient un énorme nombre de collisions avec les électrons libres par diffusion Thomson, rendant l'univers totalement opaque, matière et radiation étant encore liées. L'époque où se formèrent les premiers atomes est appelée recombinaison (pour une raison obscure, les électrons et les noyaux ne s'étant encore jamais combinés). Elle correspond de plus au moment où l'Univers devint transparent, photons et matière n'interagissant plus que marginalement. Le fond de rayonnement cosmologique (aussi appelé CMB pour "Cosmic Microwave Background") observé de nos jours correspond à cette étape.

Il est possible d'estimer le moment de la recombinaison par son décalage spectral  $z_r$  :

$$1 + z_r = \frac{a(t_0)}{a(t_r)} = \frac{T_r}{T_0} \approx 1100, \quad (1.46)$$

ce décalage spectral constituant le redshift maximal qu'il est possible d'observer, l'univers à des époques antérieures étant opaque. Il nous est donc aujourd'hui possible d'observer les photons ayant été "libérés" à  $t \approx 300,000$  ans et situés sur une sphère de rayon  $z_r$  autour de nous. Il s'agit d'une source d'information incroyable, les photons étant couplés à la matière baryonique avant la recombinaison, ceux-ci gardent l'empreinte de sa distribution à l'époque et peuvent donc nous fournir de précieux renseignements sur la formation des inhomogénéités dans la distribution de matière actuelle (sur la formation des halos de matière noire et des galaxies notamment).

Les premières observations ont été purement fortuites et c'est en voulant régler des problèmes d'interférence que Penzias et Wilson découvrirent les premières traces d'un rayonnement de type corps noir à une température de  $T \approx 3$  K [38]. Ce n'est qu'en 1990 que le satellite COBE commença à établir la première carte du rayonnement fossile sur le ciel complet [12]. Depuis, des expériences comme Boomerang [32], Maxima [19] et Archeops [3] ont affiné les mesures et la publication récente des résultats de la troisième année de cartographie du satellite WMAP [43] ont permis d'obtenir une très grande précision de mesure.

Globalement, le CMB peut être considéré comme homogène sur l'ensemble du ciel à une température de  $T \approx 2.728$ , les anisotropies ne dépassant pas  $\Delta T/T \approx 10^{-5}$ . Ce n'était pas quelque chose d'évident a priori. On peut considérer que des zones situées à plus de 1 degré d'écart n'ont jamais été en contact causal dans le modèle de Friedman-Lemaître (c'est à dire qu'elles sont situées à une distance plus grande



que ce que la lumière aurait pu parcourir). Partant de là, il est difficile d'expliquer comment des régions n'ayant jamais pu échanger d'information auraient précisément les mêmes caractéristiques et c'est ce qui a donné naissance à la théorie de l'inflation par exemple (voir la section 1.2.1). Les anisotropies, bien que faibles, sont ce qu'il y a de plus intéressant dans la mesure du CMB. Nous n'expliquerons pas ici la théorie complexe du CMB car ce n'est pas le propos de cette thèse mais on peut montrer que si l'on décompose ces anisotropies en harmoniques sphériques :

$$\delta T(\theta, \phi) = \sum_{l,m} a_{lm} Y_{lm}(\theta, \phi), \quad (1.47)$$

alors le tracé de la valeur des coefficients  $C_l = \langle |a_{lm}|^2 \rangle_m$  présente une série de pics dont la position et l'amplitude peuvent être reliés aux valeurs des paramètres cosmologiques (voir la figure 1.3). Ces pics donnent les corrélations existantes entre des zones séparées d'un angle constant sur le ciel et gardent les traces de la distribution et des interactions entre la matière et le rayonnement. L'ouverture angulaire  $\theta_H$  du premier pic situé à un angle de  $1^\circ$  par exemple s'écrit :

$$\theta_H \approx \frac{1^\circ}{\sqrt{\Omega_M \Omega_\Lambda}}. \quad (1.48)$$

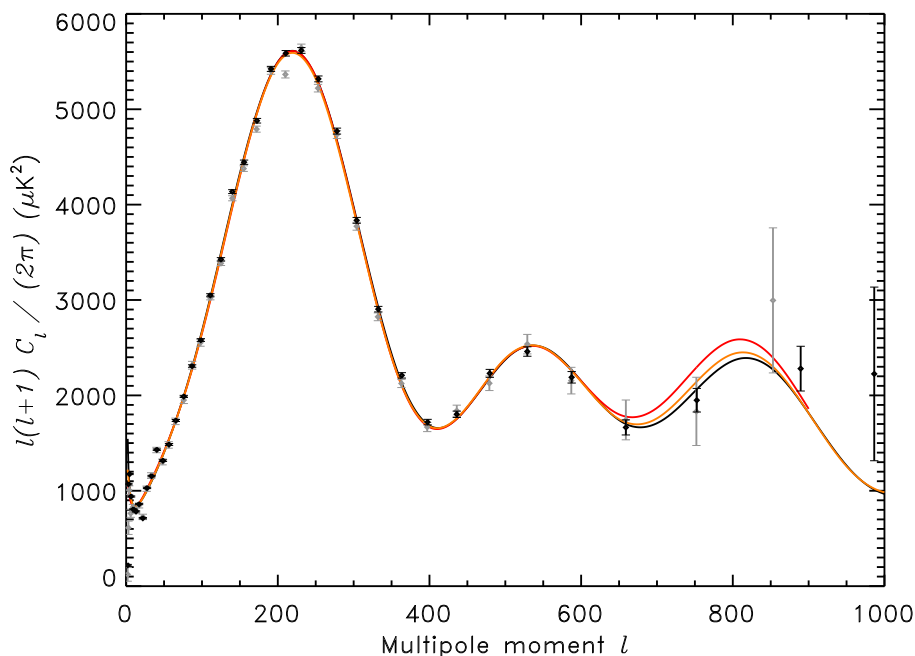
Les derniers résultats obtenus par Spergel [43] donnent ainsi :

$$\begin{aligned} h &= H_0/100 = 0.72 \pm 0.03 \\ \Omega_m &= 0.238 \pm 0.02 \\ \Omega_b &= 0.044 \pm 0.001 \\ \Omega_k &= -0.01 \pm 0.01. \end{aligned} \quad (1.49)$$

Ce sont des résultats d'une très grande précision mais ayant été publiés très récemment, il n'a pas été possible de les utiliser durant cette thèse. Les valeurs trouvées restent cependant proches des valeurs précédentes et l'impact de l'erreur commise sera de toute façon négligeable pour les calculs que nous effectuerons.

### Composition du contenu en matière, la matière noire

Nous venons de voir que d'après les mesures du CMB, la proportion de matière baryonique n'est que de  $\Omega_b \approx 0.04$  alors que la proportion de matière est de  $\Omega_m \approx 0.24$ . La question de la nature exacte de la masse manquante se pose donc. Il est clair que seule la matière baryonique peut actuellement être observée directement (car chauffée elle rayonne et émet des photons), la matière manquante et apparemment invisible est donc couramment appelée "matière noire". Si cette matière noire existe vraiment, elle constitue 85 % de toute la matière dans l'Univers et doit par conséquent avoir une forte influence, sur la formation des galaxies par exemple. Il se trouve que c'est le cas et c'est même la raison pour laquelle Zwicky a émis l'hypothèse de son existence en 1933 dans le but d'expliquer la dynamique du milieu inter-amas.



**Fig. 1.3:** *Spectre de puissance angulaire du CMB d'après les résultats de WMAP3 [43]. La ligne noire représente le meilleur ajustement du modèle aux données obtenues après la troisième année (points noirs) et les lignes oranges et points gris correspondent aux données de la première année. La présence des trois premiers pics de résonance est clairement détectée et c'est leur position et leur amplitude qui permettent l'estimation des paramètres cosmologiques.*

La mesure des courbes de rotation des galaxies est effectuée grâce aux raies d'émission dans les régions HII où bien à celles de l'hydrogène neutre. Ces deux raies ont l'avantage d'être présentes au delà de la partie visible (car chaude) de la galaxie et donc de permettre la mesure des champs de vitesses au delà de l'extension visible des galaxies. D'après la loi de Newton, l'accélération radiale  $\gamma_r$  s'écrit

$$\gamma_r = \frac{GM(r)}{r^2} = \frac{v(r)^2}{r} \quad (1.50)$$

où  $M(r)$  est la masse contenue dans la sphère de rayon  $r$ . Il est alors possible d'exprimer la vitesse comme :

$$v(r) = \sqrt{\frac{GM(r)}{r}}. \quad (1.51)$$

Les mesures montrant en général un profil de vitesse plat (voir McGaugh [28] par exemple), on peut en déduire que  $M(r) \propto r$  où  $\rho_m \propto r^{-2}$  ce qui est en contradiction avec les observations de la matière lumineuse bien plus concentrée. L'explication émise est finalement que, la matière s'étend bien plus loin que le bord visible des galaxies expliquant sous une forme non détectée : la matière noire.

Citons comme autres mesures confirmant l'existence de la matière noire la présence des amas galactiques où peuvent être regroupées des milliers de galaxies ou

encore les émissions X de ces amas qui permettent la détermination de leur masse et de leur fraction baryonique par analyse du rayonnement (voir McCammon [27] par exemple). Enfin, les effets de lentillage gravitationnel sont sans doute l'élément le plus convainquant. La relativité générale prédit une courbure des rayons lumineux par la présence d'un fort potentiel gravitationnel et c'est ce qui est observé au voisinage d'amas situés dans l'alignement d'une galaxie et de la Terre. Les distortions observées peuvent alors être reliées à la masse de l'objet défecteur et les mesures actuelles, qui utilisent le rayonnement X pour déterminer la fraction baryonique, confirment largement sa présence (voir Schneider [42] ou Tereno *et al.* [46] pour plus de détails sur la méthode et les résultats obtenus).

La matière noire n'étant observable qu'indirectement, se pose alors la question de la relation entre sa distribution et celle de la matière baryonique. Les catalogues de galaxies actuels (voir la section 1.3) recensent la position de centaines de milliers de galaxies dans l'univers local ( $z < \sim 0.3$ ). Les propriétés de ces distributions peuvent être étudiées par le biais du spectre de puissance linéaire des galaxies  $P_{gg}(k)$ , qui n'est autre que la transformée de Fourier de la fonction de corrélation à deux points (voir la section 3.1.1). Il reste alors nécessaire pour confronter les observations aux théories à relier  $P_{gg}(k)$  à la valeur de  $P_m(k)$ , le spectre de puissance linéaire de toute la matière (incluant la matière noire). La différence entre ces deux quantités est appelée le biais, une notion introduite par Kaiser [21] en 1987. Le biais peut être approximé par des modèles assez simples. L'idée consiste à supposer que les galaxies occupent seulement les pics en densité de la distribution de matière. Si l'on suppose alors l'indépendance d'échelle du biais (ce qui est justifiable à grande échelle) on obtient une simple relation linéaire du type :

$$P_{gg}(k) = bP_m(k), \quad (1.52)$$

valable pour  $k < k_{NL} \approx 0.3hMpc^{-1}$ . En pratique, la valeur du paramètre  $b$  est appelé le biais. la valeur de  $b$  est assez délicate à mesurer mais certaines méthodes utilisant les fonctions de corrélations à N-points permettent d'obtenir une valeur de  $b \approx 1.04 \pm 0.11$  (voir Durrer *et al.* [10] par exemple).

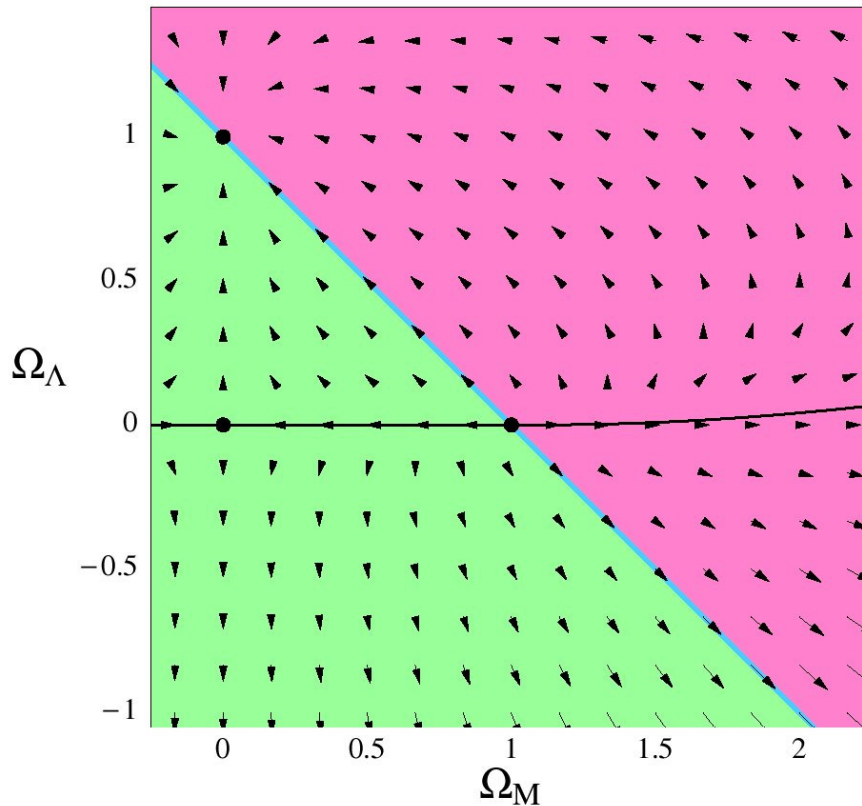
## 1.2 Croissance des structures

### 1.2.1 Inflation et approche linéaire

Dans les sections précédentes, nous avons décrit le modèle cosmologique standard sans toutefois jamais aborder la question des conditions initiales. L'étude de la nature de ces conditions initiales soulève au moins deux problèmes très gênants qui ont conduit à l'élaboration des modèles d'inflation.

#### Modèles d'inflation

Nous avons déjà présenté un des principaux inconvénients du modèle classique : le problème de coïncidence. Il est directement relié à un autre phénomène difficilement



**Fig. 1.4:** Evolutions possible de l'Univers dans l'espace des paramètres cosmologiques  $\{\Omega_\Lambda, \Omega_m\}$ . Le point  $(\Omega_\Lambda = 1, \Omega_m = 0)$  constitue l'évolution finale pour un grand nombre de valeurs initiales différentes des paramètres cosmologiques, ce qui constitue le problème de coïncidence (pourquoi  $\Omega_\Lambda \approx \Omega_m$  de nos jours?). Le problème de platitude tient au fait qu'il semblerait que nous nous situions sur la droite d'équation  $\Omega_\Lambda + \Omega_m = 1$ , mais dans ce cas, quel mécanisme pourrait expliquer le fait que l'Univers contienne exactement la bonne quantité d'énergie pour que la courbure soit précisément nulle? Figure extraite de Carroll [7].

explicable : la platitude de l'Univers. Comme le montre la figure 1.4, pour des conditions initiales quelconques, si l'on représente l'évolution de l'univers dans l'espace des paramètres cosmologiques, il existe un seul attracteur : le point  $(\Omega_\Lambda = 1, \Omega_m = 0)$ . il est remarquable de plus que, pour que la courbure de l'Univers soit nulle dans le présent, cela nécessite qu'elle l'ait toujours été, la droite d'équation  $\Omega_\Lambda + \Omega_m = 1$  n'étant absolument pas un attracteur. Dans ce cas comment se fait-il que nous vivions aujourd'hui dans un Univers apparemment plat à une très grande précision ?

Il existe aussi un autre problème encore plus gênant : celui de l'horizon. Un temps fini s'est écoulé depuis le Big-bang et les photons ont par conséquent parcouru une distance finie. Si l'on considère un photon se déplaçant selon une géodésique depuis le Big-bang, on peut utiliser l'équation (1.40) pour calculer la distance maximale qu'il a pu parcourir. Cette distance est notée  $r_{\text{hor}}$  pour distance à l'horizon car elle représente la distance au delà de laquelle l'Univers n'est plus observable. Dans un

univers de type Einstein-de Sitter ( $\Omega_m = 1$ ,  $\Omega_\Lambda = 0$ ), on obtient alors (en distance comobile) :

$$r_{\text{hor}} = \frac{c}{H_0} \int_{z_*}^{\infty} \frac{dz'}{(1+z')^{3/2}} = \frac{2c}{H_0} \sqrt{a_*}, \quad (1.53)$$

avec  $\sqrt{a_*} = 1/(1+z_*)$  l'époque à laquelle est mesurée la taille de l'horizon. Par conséquent, si l'on considère que l'époque à laquelle est observé le CMB est  $a_{\text{CMB}} = 1200$ , on obtient :

$$r_{\text{hor}} \approx 6.10^{-2} H_0^{-1}. \quad (1.54)$$

Dans tout les types d'univers plats, on peut montrer que ce résultat est une bonne approximation de la réalité et il est donc impossible que l'ensemble des zones observées du CMB aient un jour pu échanger la moindre information. Comment expliquer alors l'extraordinaire homogénéité mesurée ? La solution trouvée a pour nom l'inflation, un modèle proposé en 1981 par Guth [18].

L'idée générale est sans doute inspirée de la constante cosmologique : si l'Univers primordial subit une phase importante d'expansion accélérée, des zones très proches initialement peuvent se retrouver très éloignées par la suite et donc en apparence à des distances non causales. Cette inflation a de plus la plaisante propriété de rendre naturellement l'Univers extrêmement plat (ce qui est observé de nos jours) et les fluctuations du champ de densité peuvent alors être expliquées par les fluctuations quantiques, étirées sur de très grandes échelles. Une telle expansion, tout comme la constante cosmologique, peut être modélisée par un champ scalaire  $\phi$ . Considérons un modèle simple de Lagrangien :

$$\mathcal{L} = \frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi), \quad (1.55)$$

alors l'équation du mouvement s'écrit :

$$\ddot{\phi} + 3H\dot{\phi} + v'(\phi) = 0, \quad (1.56)$$

et l'énergie du champ :

$$\rho = \frac{1}{2} \dot{\phi}^2 + V(\phi), \quad (1.57)$$

et sa pression :

$$p = \frac{1}{2} \dot{\phi}^2 - V(\phi). \quad (1.58)$$

En supposant  $\dot{\phi} = 0$ , les équations de Friedmann donnent alors :

$$H^2 = \left( \frac{\dot{a}}{a} \right)^2 \propto V(\phi) = \text{const.} \quad (1.59)$$

Par conséquent,

$$a(t) \propto \exp(Ht), \quad (1.60)$$

et l'Univers s'étend à une vitesse exponentielle. En pratique, il suffit simplement que  $1/2\dot{\phi}^2 \ll V(\phi)$  pour obtenir ce type de comportement, ce type d'approximation

étant appelé “slow roll”. Il existe de nombreux modèles de potentiels répondant à ce critère, les références [25] et [26] en contenant une liste exhaustive.

Durant l’inflation, les fluctuations quantiques initiales, de très petite taille, se retrouvent étirées à très grande vitesse pour rapidement dépasser la taille de l’horizon  $r_{\text{hor}}$  et il est possible de prédire l’amplitude des fluctuations *scalaires* au moment où elles croisent l’horizon :

$$\frac{\delta\rho}{\rho} = \frac{H^2}{4\pi^{3/2}\dot{\phi}}. \quad (1.61)$$

Dans la plupart des cas, ce type de fluctuations conduit à un spectre de puissance du type “loi de puissance” :

$$P(k) \propto k^{n_s-1}. \quad (1.62)$$

Les mesures actuelles conduisent ainsi à un spectre de puissance initial quasiment invariant d’échelle (aux échelles suffisamment grandes) avec  $n_s = 0.951 \pm 0.015$  (voir Spergel [43]).

### Croissance linéaire des structures

La théorie de croissance linéaire des structures est connue depuis les années 70 et de nombreux ouvrages font référence en la matière, citons principalement [35] dans lequel tous les calculs sont très détaillés. Le lecteur intéressé est invité à se référer à cet ouvrage.

Partant de la distribution des particules dans l’espace des phases :

$$dN = f(\mathbf{x}, \mathbf{p}, t) d^3x d^3p, \quad (1.63)$$

il est possible d’écrire l’équation de Vlasov en utilisant le théorème de Liouville d’après lequel  $f$  doit être constant sur une géodésique. On obtient alors :

$$\frac{df}{d\tau} = 0 = \frac{df}{dt} + \frac{\mathbf{p}}{ma^2} \cdot \nabla f - m \nabla \phi \cdot \frac{d\mathbf{p}}{dt}, \quad (1.64)$$

avec  $\phi$  le potentiel gravitationnel exprimé dans un espace de type FRW (voir l’équation (1.91)) obéissant à l’équation de poisson (1.92). Cette équation peut être résolue en considérant les moments successifs de la fonction de distribution  $f(\mathbf{x}, \mathbf{p}, t)$  :

$$\rho(\mathbf{x}, t) \equiv \int d^3\mathbf{p} f(\mathbf{x}, \mathbf{p}, t), \quad (1.65)$$

$$\rho(\mathbf{x}, t) \mathbf{v}(\mathbf{x}, \tau) \equiv \int d^3\mathbf{p} \frac{\mathbf{p}}{am} f(\mathbf{x}, \mathbf{p}, t), \quad (1.66)$$

...

où  $\rho(\mathbf{x}, \tau)$  est le champ de densité et  $\mathbf{v}(\mathbf{x}, \tau)$  le champ de vitesse particulière. Les deux premiers moments permettent alors d’obtenir l’équation de continuité :

$$\frac{\partial\delta}{\partial t} + \frac{1}{a} \nabla \cdot (1 + \delta) \mathbf{v} = 0 \quad (1.67)$$

et, en utilisant (1.67), l'équation d'Euler :

$$\frac{\partial^2 \delta}{\partial t^2} + 2H \frac{\partial \delta}{\partial t} \frac{\nabla^2 \mathbf{p}}{a^2 \rho_0} + \frac{1}{a^2} \nabla \cdot (1 + \delta) \nabla \phi + \frac{1}{a^2} \frac{\partial^2}{\partial \alpha \partial \beta} [(1 + \delta) v^\alpha v^\beta], \quad (1.68)$$

où le contraste de densité  $\delta = (\rho - \rho_0)/\rho_0$  a été introduit en fonction de la densité moyenne  $\rho_0$  et  $H = \dot{a}/a$  est le paramètre de Hubble.

Pour de faibles contrastes de densité  $\delta \ll 1$  et en supposant que les vitesses sont suffisamment faibles ( $vt/d \ll \delta$  avec  $v$  la vitesse typique du fluide,  $d$  la longueur de cohérence du système et  $t \approx (G\rho_0)^{-1/2}$  le temps d'expansion), on peut linéariser ces équations. On obtient de l'équation (1.67) :

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla \cdot \mathbf{v} = 0 \quad (1.69)$$

et de l'équation (1.68) :

$$\frac{\partial^2 \delta}{\partial t^2} + 2H \frac{\partial \delta}{\partial t} = V_s^2 \frac{\nabla^2 \delta}{a^2} + 4\pi G \rho_0 \delta, \quad (1.70)$$

avec  $V_s^2 = P/\rho$  la vitesse du son. Il devient possible de montrer que les solutions pour le champ de densité peuvent s'écrire en général sous la forme :

$$\delta(\mathbf{x}, t) = A(\mathbf{x}) D_1^+(t) + B(\mathbf{x}) D_1^-(t), \quad (1.71)$$

où  $A(x)$  et  $B(x)$  sont deux fonctions quelconques représentant le champ de densité initial. Dans cette équation,  $D_1^+(t)$  et  $D_1^-(t)$  sont appelées facteur de croissance linéaire. Pour différents modèles, on obtient (voir Bernardeau *et al.* [4]) :

- ( $\Omega_m = 1, \Omega_\Lambda = 0$ )

$$D_1^+(t) = a \quad (1.72)$$

$$D_1^-(t) = a^{-3/2} \quad (1.73)$$

- ( $\Omega_m < 1, \Omega_\Lambda = 0$ )

$$x = \frac{1}{\Omega_m} - 1$$

$$D_1^+(t) = 1 + \frac{3}{x} + \sqrt{\frac{1+x}{x^3}} \ln(\sqrt{1+x} + \sqrt{x}), \quad (1.74)$$

$$D_1^-(t) = \sqrt{\frac{1+x}{x^3}}, \quad (1.75)$$

$$(1.76)$$

- ( $\Omega_m, \Omega_\Lambda$ )

$$\begin{aligned} D_1^+(t) &= H(a) \frac{5\Omega_m}{2} \int_0^a \frac{da}{a^3 H(a)} \\ &\approx \frac{5}{2} \frac{a\Omega_m}{\Omega_m^{4/7} - \Omega_\Lambda + (1 + \Omega_m/2)(1 + \Omega_\Lambda/70)}, \end{aligned} \quad (1.77)$$

$$D_1^-(t) = \frac{H}{a}. \quad (1.78)$$

De plus, pour une perturbation de vecteur d'onde  $\mathbf{k}$  donné, l'équation (1.70) se réécrit en fonction de  $\epsilon$  :

$$\ddot{\epsilon} + 2H\dot{\epsilon} + \left[ \left( \frac{kV_S}{a} \right)^2 - \frac{3}{2}H^2 \right] \epsilon = 0, \quad (1.79)$$

avec  $\epsilon$  tel que  $\delta(\mathbf{x}, t) = \epsilon(t) \exp(i\mathbf{k}\mathbf{x})$ . A décalage spectral  $z \gg 1$ , la matière dominant largement, il est possible de définir une longueur d'onde de Jeans  $\lambda_J = \sqrt{6\pi}V_S t$  (car alors  $H = 2t^{-1}/3$  et  $a = t^{2/3}$ ) qui annule le membre de droite et il devient possible de différencier deux cas. Si  $\lambda \ll \lambda_J$ , on obtient :

$$\ddot{\epsilon} + \left( \frac{2\pi V_S}{\Lambda} \right)^2 \epsilon = 0 \quad (1.80)$$

qui est l'équation d'une onde sonore vouée à s'évanouir. Mais si  $\lambda \gg \lambda_J$ , alors :

$$\ddot{\epsilon} + 2H\dot{\epsilon} - \frac{3}{2}H^2\epsilon = 0 \quad (1.81)$$

et il est facile de montrer que le contraste de densité va augmenter irrémédiablement avec une solution du type  $\delta(\mathbf{x}, t) = A(\mathbf{x})D_1^+(t) + B(\mathbf{x})D_1^-(t)$  précédemment décrit.

La description des surdensités en effondrement est quand à elle bien plus complexe car elle suppose de prendre en compte les non-linéarités de l'équation de Vlasov (1.64). Citons comme exemple le modèle "top hat" d'effondrement sphérique qui, en faisant l'approximation d'une symétrie sphérique et en traitant la surdensité comme un ensemble de coquilles de densités indépendantes permet de prédire analytiquement l'évolution du contraste de densité jusqu'au croisement des dites coquilles.

### 1.2.2 Évolution non linéaire : simulations numériques

Prédire l'évolution des structures de manière analytique lorsque les non-linéarités sont importantes devient vite extrêmement complexe. Une autre approche consiste à utiliser les énormes capacités de calcul des ordinateurs. Nous présentons ici les différentes méthodes couramment utilisées en cosmologie pour effectuer des simulations N-Corps de l'évolution de la répartition de la matière dans l'univers. Nous nous limiterons principalement aux simulations de matière noire non collisionnelle que nous avons utilisées pour cette thèse et où il n'est pas nécessaire de prendre en compte la physique complexe des interactions baryoniques.

#### Méthode Particule-Particule (PP)

La méthode PP est la méthode la plus simple que l'on puisse imaginer [17]. Elle consiste à calculer la force  $\mathbf{F}_i(j)$  exercée par la particule  $i$  de position  $\mathbf{r}_i$  sur chaque particule  $j$  de position  $\mathbf{r}_j$ . Dans le cas de la force gravitationnelle Newtonienne :

$$\mathbf{F}_i(j) = m_j \mathbf{g}_i = \frac{Gm_i m_j}{\|\mathbf{r}_{ij}\|^3} \mathbf{r}_{ij}, \quad (1.82)$$



où  $m_i$  est la masse de la particule  $i$ ,  $\mathbf{g}_i$  le potentiel gravitationnel de la particule  $i$  et  $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ . Il ne reste alors qu'à intégrer les équations du mouvement  $\sum_j \mathbf{F}_j(i) = m_i \gamma_i$  où  $\gamma_i$  est l'accélération de la particule  $i$ . Le compteur de temps  $t$  est alors incrémenté d'un temps  $\Delta t$  et l'opération renouvelée.

La méthode d'intégration est le plus souvent celle de Runge-Kutta. Il y a cependant quelques points auxquels il faut bien faire attention. Tout d'abord, le mouvement des particules étant chaotique, il faut prendre soin de bien choisir les temps. Un pas de temps constant donnera forcément des résultats imprécis, soit pour des raisons numériques, soit tout simplement parce qu'il existera toujours un cas où l'approximation d'un déplacement en ligne droite entre deux pas de temps ne suffira plus. La solution habituelle est d'utiliser un pas de temps adaptatif et spécifique à chaque particule. Habituellement, plus les particules seront éloignées, plus le pas de temps sera grand et inversement. Un autre problème est la divergence de la force gravitationnelle quand la distance tend vers zéro. La précision numérique sur le calcul de la force est limitée et lorsque deux particules sont très proches, la précision d'un nombre à virgule flottante est dépassée. La solution consiste à modifier légèrement l'expression de la force en introduisant une distance de lissage  $\epsilon$  :

$$\mathbf{F}_i(j) = -\frac{Gm_i m_j}{\epsilon + \|\mathbf{r}_{ij}\|^3} \mathbf{r}_{ij}. \quad (1.83)$$

L'avantage principal de la méthode *PP* est qu'elle ne nécessite aucune approximation sur le calcul des forces. Son principal inconvénient est bien sûr le temps de calcul proportionnel au carré du nombre de particules. Pour ces raisons, la méthode *PP* est principalement utilisée dans les simulations d'orbites de planètes dans le système solaire où le nombre de corps est faible mais où la précision du calcul doit être très élevée. Les applications cosmologiques sont quant à elles très rares mais il existe des ordinateurs parallèles appelés *GRAPE* (pour "GRAvity PipE") spécifiquement adaptés à cette méthode et permettant d'augmenter significativement le nombre de particules simulées.

### Dynamique dans un Univers en expansion

L'équation (1.82) est valable seulement dans le cas d'une approximation non relativiste, ce qui revient à supposer que les vitesses des particules sont négligeables devant la vitesse de la lumière  $c$ , que le potentiel gravitationnel est faible devant  $c^2$  ou encore que l'échelle de la simulation est petite devant la longueur de Hubble  $c/H$ . Dans ces conditions, il est justifié d'utiliser l'équation (1.82). Une résolution efficace des problèmes *N*-corps nécessite quelques modifications. L'équation utilisée en pratique pour le calcul de l'évolution des particules sous un champ gravitationnel est l'équation de Poisson :

$$\begin{cases} \nabla \mathbf{g} = -4\pi G \rho(\mathbf{r}, t), \\ \nabla \times \mathbf{g} = 0 \end{cases} \quad (1.84)$$

avec  $\mathbf{g}_i = \mathbf{g}(\mathbf{r}_i, t)$  la solution pour la particule  $i$ . On peut en déduire l'expression du potentiel Newtonien  $\Phi$  tel que  $\mathbf{g} = -\partial\Phi/\partial\mathbf{r}$  :

$$\nabla^2\Phi = 4\pi G\rho \quad (1.85)$$

ou, pour être plus général, en tenant compte des équations de la relativité générale :

$$\nabla^2\Phi = 4\pi G\left(\rho + \frac{3P}{c^2}\right) - \Lambda \quad (1.86)$$

avec  $P$  la pression et  $\Lambda$  l'éventuelle constante cosmologique.

Afin de tenir compte de l'expansion de l'univers, la coordonnée  $\mathbf{r}$  désigne la coordonnée comobile  $\mathbf{r} = a(t)\mathbf{x}$  avec  $a(t)$  le facteur d'expansion et  $\mathbf{x}$  les coordonnées physiques. Il en découle l'équation de Hubble :

$$\frac{\partial\mathbf{r}}{\partial t} = H(t)\mathbf{r} + a(t)\frac{\partial\mathbf{x}}{\partial t} \quad (1.87)$$

avec

$$\mathcal{H}(t) = \frac{\dot{a}}{a}. \quad (1.88)$$

Les équations du mouvement d'une particule se déduisent de son Lagrangien

$$\mathcal{L} = \frac{1}{2}m\left(\frac{\partial\mathbf{r}}{\partial t}\right)^2 - m\Phi. \quad (1.89)$$

Le Lagrangien étant défini à une différentielle totale près (l'ajout de celle-ci ne changeant rien aux équations du mouvement), il est possible de remplacer  $\mathcal{L}$  par  $\mathcal{L}'$  qui lui est équivalent et s'écrit :

$$\mathcal{L}' = \mathcal{L} - \frac{1}{2}\frac{d}{dt}(ma\dot{\mathbf{x}}^2) = \frac{1}{2}ma^2\dot{\mathbf{x}}^2 - m\phi \Leftrightarrow \mathcal{L} \quad (1.90)$$

avec

$$\phi = \Phi + \frac{1}{2}a\ddot{a}\mathbf{x}^2 \quad (1.91)$$

Ce nouveau potentiel  $\phi$  obéissant à une équation de Poisson un peu différente de (1.84) :

$$\nabla_x^2\phi = 4\pi Ga^2\langle\rho\rangle\delta \quad (1.92)$$

où apparaît le contraste de densité

$$\delta(\mathbf{x}, t) = \frac{\rho - \langle\rho\rangle}{\langle\rho\rangle}, \quad (1.93)$$

et où  $\nabla_x = a\nabla$ . On peut en déduire une forme d'expression plus simple pour les équations du mouvement dans un contexte cosmologique :

$$\begin{cases} \mathbf{p} = ma^2\dot{\mathbf{x}} \\ \frac{d\mathbf{p}}{dt} = -m\nabla_x\phi \end{cases} \quad (1.94)$$

ou, en utilisant la vitesse propre des particules  $\mathbf{v} = a\dot{\mathbf{x}}$  :

$$\frac{d\mathbf{v}}{dt} + H(t)\mathbf{v} = -\frac{1}{a}\nabla_{\mathbf{x}}\phi. \quad (1.95)$$

L'équation (1.95), à quelques redéfinitions du paramètre  $t$  près, est celle utilisée dans la plupart des programmes de simulations cosmologiques.

### Méthode Particule-Grille (“Particle-Mesh” ou PM)

La méthode PM consiste à échantillonner le champ de densité  $\rho(\mathbf{r}, t)$  sur une grille cartésienne. Cela revient à attribuer en chacun des nœuds de la grille la valeur moyenne du champ de densité dans les environs. Toutes les grandeurs physiques sont alors calculées sur la grille et interpolées afin d'obtenir leur valeur pour chaque particule. Un pas de temps se déroule de la manière suivante :

1. Échantillonnage du champ de densité sur la grille.
2. Résolution de l'équation de poisson (1.95).
3. Calcul du champ de force sur la grille
4. Interpolation de la force pour chaque particule
5. Intégration de la force, identique à la méthode PP.

Lors de la première étape, un choix doit être fait dans la méthode de répartition des particules sur les nœuds de grille. La méthode NGP (pour “Nearest Grid Point”) est la plus simple et consiste à attribuer le poids de chaque particule au nœud de grille le plus proche. Une méthode plus raffinée telle que CIC (pour “Cloud In Cell”) permet cependant d'améliorer la continuité du champ et donne de meilleurs résultats (voir le chapitre 3.3.1 pour plus de détails). Le problème lié à l'utilisation d'une grille est cependant que celle-ci introduit des directions privilégiées. Il s'agit d'un problème “d'aliasing” qui peut être amélioré par filtrage, par exemple en utilisant la méthode TSC (pour “Triangular Shape Cloud”) qui confère aux particules une extension spatiale.

La deuxième étape est le plus souvent effectuée dans l'espace de Fourier à l'aide d'une transformée de Fourier rapide (ou FFT pour “Fast Fourier Transform”). Elle consiste simplement à multiplier le champ de densité par la fonction de Green appropriée afin de résoudre l'équation de Poisson. Un des premiers articles faisant mention de cette méthode est Miller and Prendergast [30] mais de nombreuses versions ont vu le jour dans les années quatre-vingt, en deux dimensions tout d'abord comme dans Doroshkevich *et al.* [9] puis en trois dimensions dans Klypin and Shandarin [22], Miller [29] ou encore Bouchet and Kandrup [6].

Finalement, lors de la quatrième étape, une interpolation de même ordre est utilisée pour des raisons de cohérence avec la première étape. Le principal avantage du code PM est bien sûr sa vitesse, proportionnelle à  $N + N_g \log(N_g)$ , où  $N$  est le nombre de particules et  $N_g$  le nombre de nœuds de grille. Cela ne va cependant pas

sans inconvénients, le principal étant que, contrairement aux codes PP, la force est calculée de manière approximative. En effet, la taille de grille est fortement limitée par des contraintes de mémoire disponible et la résolution à une échelle très inférieure à la taille d'une cellule est très imprécise. On peut donc considérer que le code PM est une bonne approximation du code PP au delà d'une échelle de la taille de 1 à 2 cellules.

### Particule-Grille, Grille-Grille (P3M)

La méthode P3M est faite pour pallier aux défauts de la méthode PM. Pour cela, elle mélange l'utilisation d'une grille pour les interactions à grande échelle à une méthode PP pour les interactions à courte distance. La rapidité du code PM est donc mise à profit lorsque l'approximation est suffisamment précise et la précision du code PP à courte distance permet cependant de garder une résolution élevée. D'abord largement utilisés pour la physique des plasmas, ils furent adaptés aux simulations cosmologiques au milieu des années 80 [11].

En pratique, une grille large est utilisée pour le calcul des forces à grande échelle et une grille plus précise sert à identifier les particules proches les unes des autres. Le schéma d'un pas de temps est identique à celui de la méthode PM, mis à part qu'après la résolution de l'équation de Poisson sur la grille large, les forces exercées par les particules appartenant aux cellules proches de la grille fine sont calculées à la manière PP. Une version adaptée des équations du mouvement est alors intégrée pour déduire le déplacement des particules à partir de ces deux forces.

Les performances atteintes sont très bonnes, le seul problème provenant de l'apparition de fortes non-linéarités à faible redshift ralentissant considérablement le calcul. Ainsi, à l'échelle d'un halo dont la taille est de l'ordre de celle de la grille, le calcul est aussi lent qu'un code PP classique.

### Codes à raffinement de grille

De nombreuses méthodes ont été imaginées dans le but d'améliorer les codes P3M, la plupart utilisant des sous-grilles incluses dans la grille initiale. La méthode HPM (pour "Hierarchical Particle-Mesh") développée par Villumsen [48] tente de pallier au problème de ralentissement du code P3M pour des contrastes de densité élevés. Pour cela, des sous-grilles sont rajoutées aux endroits surdenses et les particules rentrant dans ces zones sont séparées en sous-particules de masses plus faibles. Ce code fait la supposition qu'une structure à une échelle  $L$  donnée n'influence pas une autre structure de la même taille  $L$  à une distance supérieure à  $L$ , de plus, l'évolution à une échelle donnée n'est pas influencée par celle à une échelle significativement plus petite. Le code HPM inspirera plus tard les codes en arbres présentés dans la section suivante.

Une autre tentative moins fructueuse mais néanmoins intéressante est celle de Gnedin [15] ou encore Pen [37] qui utilisent un système de coordonnées dynamiques

suivant les déformations du champ de densité afin d’obtenir une grille Lagrangienne qui suit le mouvement des particules et se déforme en conséquence. Ce type de code offre une résolution très élevée pour un temps de calcul faible, malheureusement, une forte anisotropie est introduite dans le calcul des forces.

La méthode ART (pour “Adaptative Refinement Tree”) développée par Kravtsov *et al.* [24], propose quant à elle d’utiliser une méthode PM sur une grille mère raffinée hiérarchiquement et de manière adaptative en fonction d’un critère dépendant du champ de densité. Grâce à cela, la précision du calcul augmente dans les zones surdense (là où c’est nécessaire) et autorise la simulation de larges portions d’univers tout en conservant une bonne précision à petite échelle. Les codes AMR (pour “Adaptative Mesh refinement”) de Teyssier [47] ou Norman and Bryan [33] sont fortement inspirés de cette méthode et sont, avec les codes en arbre, les méthodes les plus utilisées actuellement pour les simulations cosmologiques.

### Codes en arbre (Tree-codes)

Les codes en arbres sont basés sur un découpage hiérarchique de l’espace permettant de profiter du fait qu’il est possible d’ignorer les détails précis de la distribution interne d’un groupe de particules suffisamment distant. Il devient alors possible de remplacer une grande partie des interactions entre particules de type PP par des interactions particule-groupe tout en contrôlant la précision des calculs. Le premier code utilisant un arbre hiérarchique est sans doute celui de Barnes and Hut [2] qui fut parallélisé de manière à pouvoir profiter des architectures de calcul vectoriel. Les premières adaptations au cas cosmologique (univers en expansion, conditions aux limites périodiques, ... ) datent du début des années 90 par Sugihara *et al.* [45].

Le principe est assez simple. Une cellule mère contient l’ensemble des particules. Cette cellule est alors découpée en  $2^3$  cellules filles contenant chacune une partie de ces particules et l’opération est renouvelée jusqu’à ce que chacune des cellules filles ne contienne plus qu’une seule particule. A chaque division de l’espace, le centre de masse des particules contenues ainsi que les éventuels premiers ordres du développement multipolaire (habituellement jusqu’au quadrupole) du potentiel de la distribution sont calculés et attribués à chaque cellule fille. Le réseau des cellules filles constitue alors une sorte d’arbre d’où le nom donné à l’algorithme. Lors du calcul des interactions, pour chaque particule considérée, l’arbre est parcouru à partir de la cellule mère jusqu’à satisfaction d’un critère dépendant de la taille de la cellule et de sa distance à la particule considérée. L’interaction est alors approximée en utilisant les différents multipoles précédemment calculés.

Le critère d’ouverture des cellules est habituellement du type  $l/D < \theta$  où  $D$  est la distance au centre de masse de la cellule,  $l$  la taille de la cellule et  $\theta$  un paramètre arbitraire fixant la précision des calculs. Pour  $\theta = 0$ , le code en arbre redevient un simple code PP mais dans le cas contraire, le calcul des forces peut être effectué en un temps  $t \propto N \log(N)$  dans le meilleur des cas (pour une distribution homogène), où

$N$  est le nombre de particules considérées. Dans le cas d'une distribution non homogène, les performances sont cependant moins bonnes ce qui explique l'augmentation significative du temps de calcul avec l'évolution des simulations (la distribution de matière étant de moins en moins homogène). Les particules ayant tendance à former des groupes, le critère d'ouverture des cellules est en effet plus souvent vérifié.

L'utilisation de conditions aux limites périodiques étant nécessaire dans un contexte cosmologique (le système n'étant pas lié, contrairement à une galaxie par exemple), la méthode habituellement utilisée est basée sur la sommation d'Ewald (voir Hernquist *et al.* [20]). Le principe est d'exprimer le potentiel gravitationnel et la distribution des particules sous forme de séries de Fourier (ce qui est possible étant donné la périodicité de la distribution). Il devient alors possible de calculer le potentiel total du système infini sous forme d'une double somme sur l'ensemble des particules qui peut être remplacée par une suite rapidement convergente en utilisant la sommation d'Ewald.

Dans cette thèse, l'ensemble des simulations (sauf mention spéciale) ont été réalisées en utilisant le code en arbre GADGETII de Springel [44]. Il s'agit probablement actuellement du code libre le plus rapide pour les simulations de matière noire seule (par opposition aux simulations hydrodynamiques tenant compte de la matière baryonique).

## Conditions initiales

Afin de simuler l'évolution de la répartition de la matière, des conditions initiales réalistes doivent être générées. Comme nous avons vu au chapitre 1.2.1, les modèles d'inflation prédisent, en accord avec les observations, une distribution des structures présentant un spectre de puissance en loi de puissance du type  $P(k) \propto k^{(n-1)}$ . On désire déterminer les positions d'un grand nombre de particules pour que leur répartition spatiale échantillonne correctement le champ de densité. La méthode numérique la plus couramment utilisée pour générer ces champs *discrets* est décrite dans Bertschinger [5]. De nombreux programmes basés sur cette méthode sont d'ailleurs disponibles et un lien vers celui implémentant directement la méthode (appelé COSMICS) est donné dans Bertschinger [5].

Cette méthode consiste principalement à générer un champ gaussien aléatoire de spectre de puissance correct puis répartir l'ensemble des particules de manière à obtenir une réalisation de ce champ. Pour cela, on utilise l'approximation de Zel'dovich [51], celle-ci consistant à répartir les particules sur une grille et à les déplacer de manière à ce que leurs positions génèrent le spectre de puissance désiré. Dans le régime linéaire, la position Eulérienne  $\mathbf{x}$  des particules en fonction de leur position initiale Lagrangienne  $\mathbf{q}$  est donnée par :

$$\mathbf{x}(t) = \mathbf{q} + b(t)\Psi(\mathbf{q}). \quad (1.96)$$

Il est possible de montrer que, pour correspondre à la théorie linéaire, il faut tout

d’abord que la fonction  $b(t)$  décrit l’évolution du mode croissant des structures :

$$\frac{\delta\rho}{\rho}(\mathbf{x}, t) = b(t)\delta_i(\mathbf{x}) \quad (1.97)$$

où  $\delta_i$  est le contraste de densité initial des structures. De plus, la fonction  $\Psi(\mathbf{q})$ , appelée champ de déplacement, s’écrit :

$$\Psi(\mathbf{q}) = \sum_{\mathbf{k}} \frac{i\mathbf{k}}{k^2} A_k \exp(i\mathbf{k}\cdot\mathbf{q}) \quad (1.98)$$

et peut être exprimée en fonction du potentiel gravitationnel de la théorie linéaire,  $\Psi(\mathbf{q})$  pouvant alors être interprété comme le champ de vitesse particulière, d’où le nom “champ de déplacement” puisqu’il suffit de déplacer les particules selon ce champ pour qu’elles suivent linéairement le potentiel gravitationnel (voir Padmanabhan [34] p.294 pour plus de détails).

En pratique, les particules étant réparties sur une grille uniforme (le champ est asymptotiquement homogène), la fonction  $\Psi(\mathbf{q})$  est calculée pour un champ gaussien aléatoire de spectre de puissance correct et les particules sont déplacées de manière à suivre les prédictions de la théorie linéaire, les vitesses étant :

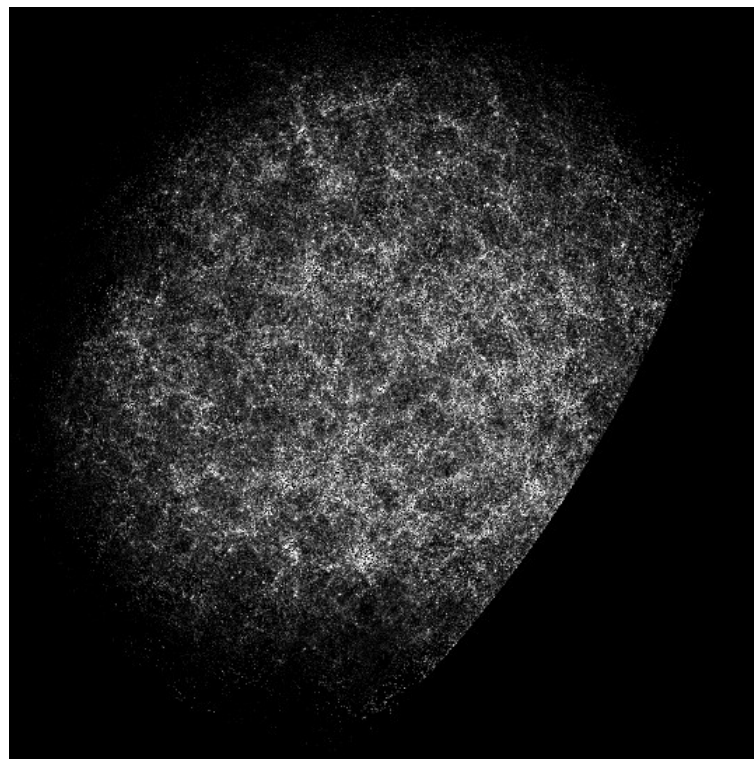
$$\dot{\mathbf{x}} = -\dot{b}\Psi(\mathbf{q}). \quad (1.99)$$

Il suffira de bien prendre garde à générer les conditions initiales à un décalage spectral suffisamment élevé pour que l’approximation linéaire soit valide. Un inconvénient certain de la méthode provient de l’utilisation d’une grille qui introduit des symétries. En pratique, celles-ci disparaissent après un temps suffisamment long d’évolution (lorsque les particules ont parcouru plusieurs mailles de grilles).

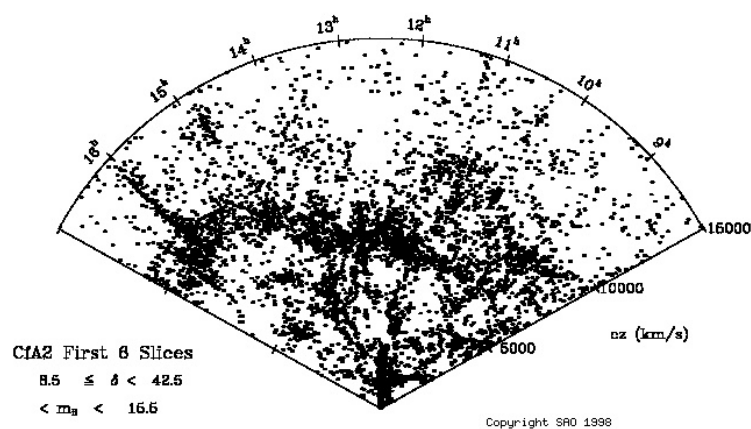
## 1.3 Les catalogues de galaxies

Depuis les années 80 les astronomes font des campagnes de mesure des vitesses radiales et des magnitudes des galaxies dans différentes régions du ciel : les catalogues de galaxies. Le premier et le plus remarquable a été le catalogue LICK qui effectue un recensement angulaire uniquement des galaxies, c’est à dire que l’on ne possède pour ces galaxies que leurs coordonnées sur le ciel mais pas leurs distances (voir la figure 1.5).

Le premier catalogue tri dimensionnel (avec mesures des décalages spectraux) fut CfA1. Ce catalogue permet pour la première fois de mettre en évidence la structuration de la répartition des galaxies sur de très grandes distances, notamment avec la découverte du “Grand Mur” en 1983 (voir la figure 1.6). Ces vingt dernières années ont été marquées par une explosion du nombre de catalogues, chacun rivalisant soit pas sa couverture angulaire, son nombre de galaxies mesurées, sa profondeur en distance, sa profondeur en magnitude ou son domaine de longueur d’onde. Dans les



**Fig. 1.5:** Image du catalogue *LICK*, un des tout premiers catalogues de galaxies ne recensant que les coordonnées angulaires de celles-ci. Source de l'image :<http://groth2005.princeton.edu/~groth/>.



**Fig. 1.6:** Image du catalogue *CfA*, le premier à mesurer les distances des galaxies grâce à leurs décalages spectraux. La structure entre 8 heure et 17 heure d'angle et entre 5,000 and 10,000 km/s est la première grande structure de galaxies identifiée, le "grand mur". Source de l'image :<http://cfa-www.harvard.edu/~huchra/zcat/>.



années à venir, nous devrions assister à une croissance exponentielle du nombre de galaxies mesurées avec des catalogues comme PANSTARS par exemple. Ces catalogues sont a priori faits pour l'équipe d'astronomes les réalisant, mais après plusieurs mois ou années, ils deviennent publics et tout le monde peut alors utiliser les données telles qu'elles sont distribuées. Les chercheurs n'ayant pas participé à la réalisation du catalogue, n'ayant jamais accès aux données non réduites (c'est à dire telles que mesurées), il faut donc compter avec les erreurs induites par les traitements de réduction de données et par les méthodes d'observation : ce sont des erreurs dites "systématiques". Ces erreurs ne sont pas contrôlables, et il est important d'en tenir compte avant d'émettre des conclusions.

Pour l'étude présentée dans cette thèse, qui concerne la façon dont se structurent spatialement les galaxies, nous avons sélectionné quelques-uns de ces catalogues, ceux offrant les meilleures mesures disponibles aujourd'hui. Voici rapidement quelques points forts et points faibles des catalogues que nous avons utilisés, ou choisis de ne pas étudier. Il faut noter que si certaines caractéristiques sont des avantages pour nous, elles peuvent s'avérer des inconvénients dans d'autres types d'études.

**SDSS** : <http://www.sdss.org/dr4/index.html>

Référence : Adelman-McCarthy [1] (2005)

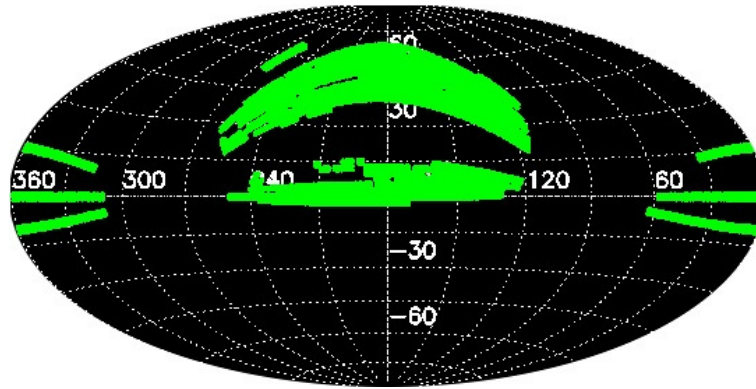
**Avantages** : grande couverture angulaire : environ 15% du ciel, grand nombre de galaxies mesurées en redshift dans les données publiques : environ un demi-million, erreurs systématiques bien contrôlées sur la mesure des magnitudes permettant de construire des sous-échantillons en volume limité (voir cette section : le squelette de SDSS).

**Inconvénients** : la stratégie d'observation est compliquée et le catalogue n'est réalisé que par petits morceaux sur le ciel. Les masques à appliquer sont extrêmement difficiles à construire (voir la figure 1.7) et des incertitudes demeurent sur l'homogénéité de la qualité photométrique des données selon la position sur le ciel.

**XSCZ** : <http://spider.ipac.caltech.edu/staff/jarrett/index-3.html>

Référence : Jarrett, T.H. 2004, PASA, 21, 396.

**Avantages** : grande couverture angulaire : tout le ciel sauf le plan galactique. Les galaxies sont sélectionnées selon leur flux en proche infra-rouge, cela nous donne une deuxième vision par rapport aux galaxies du SDSS qui sont sélectionnées dans les bandes du visible. Le nombre d'objets est très grand : environ 1 demi-million de galaxies avec un redshift photométrique. Le fait que seulement 15% du catalogue soit des redshifts spectroscopiques et 85% des redshifts photométriques, n'est pas forcément un



**Fig. 1.7:** *Couverture angulaire du ciel par le catalogue SDSS DR4. En raison de la stratégie d'observation complexe, la géométrie du catalogue est difficile à modéliser. Source de l'image : <http://www.sdss.org/dr4/>.*

mauvais point étant donné le grand nombre de galaxies.

**Inconvénients :** la profondeur du catalogue est faible :  $z < 0.06$  et combiné à la mauvaise précision inhérente des décalages spectraux photométriques, cela nous conduit à faire des études dans des coquilles.

**6DFRS :** <http://www.aao.gov.au/local/www/6df/>

Référence : Near-Infrared and Optical Luminosity Functions from the 6dF Galaxy Survey\* D. H. Jones, B. A. Peterson, M. Colless, W. Saunders, 2006, MNRAS, in press

**Avantages :** grande couverture angulaire (un demi-ciel) et redshifts spectroscopiques, les galaxies sont sélectionnées dans 2MASS (comme XSCZ) donc en proche infra-rouge, cela nous permet de comparer avec XSCZ les différences avec le clustering des galaxies optiques. Un assez grand nombre d'objets : 267, 636 galaxies dans la version publique de août 2006. Un autre bon point est que ces galaxies sont uniquement dans l'hémisphère Sud et donc ne sont pas les mêmes que celles de SDSS. On peut considérer cet échantillon d'un point de vue statistique comme une deuxième réalisation par rapport à l'univers testé avec le SDSS. Environ la même profondeur en redshift que SDSS. Autre différence : ce sont des galaxies infra-rouges, certaines théories proposent que les propriétés de ces galaxies soient différentes (au niveau du clustering) de celles des galaxies optiques.

**Inconvénients :** Nombre un peu faible de galaxies comparé au volume du catalogue, l'échantillon n'est pas continu et homogène spatialement, ce qui devrait s'améliorer dans l'avenir.

**Tully Database, LEDA Database :** <http://leda.univ-lyon1.fr>

**Avantages :** Tout le ciel est couvert, de plus, ces catalogues contiennent le plus grand nombre de galaxies dans un volume très proche ( $z < 0.01$ ), soit environ 2 millions de galaxies avec des coordonnées angulaires et des magnitudes et approximativement un million de mesures de décalages spectraux. Les structures sont très connues et identifiées dans ce volume, ce qui permet une étude précise de l'Univers local. Il peut donc être intéressant d'utiliser ce catalogue dans le but de tester l'efficacité d'algorithmes d'identification des structures comme le squelette (en particulier pour les connections à travers le plan galactique).

**Inconvénients :** Ce ne sont pas de vrais recensements de galaxies, mais des catalogues réalisés à partir de plusieurs catalogues préexistants, difficile d'avoir un contrôle sur les erreurs systématiques, mais vu le nombre de galaxies un traitement statistique est possible : on peut sélectionner plusieurs sous-échantillons en volumes limités par exemple et comparer leurs propriétés de clustering.

**MGC (Millenium Galaxy Catalog) :** <http://www.eso.org/~jliske/mgc>

Reference : The Millennium Galaxy Catalogue : the space density and surface-brightness distribution(s) of galaxies, S.P. Driver, J. Liske, N.J.G. Cross, R. De Propris and P.D. Allen, 2005, MNRAS, 360, 81.

**Avantages :** Catalogue très complet en décalages spectraux et assez profond, très bonne qualité dans l'identification des sources comme des galaxies (contamination bien plus faible par les étoiles que dans d'autres grands catalogues). Le champ mesure  $37.5 \text{ deg}^2$  dans la bande B. La région observée est entièrement comprise dans les régions à la fois du SDSS et du 2DFGRS.

**Inconvénients :** la couverture angulaire est si faible qu'il s'agit en fait pratiquement d'un catalogue 2D.

---

# Bibliographie

- [1] J. K. & al. ADELMAN-MCCARTHY. The Fourth Data Release of the Sloan Digital Sky Survey. *ApJS*, 162 :38–48, January 2006.
- [2] J. BARNES and P. HUT. A Hierarchical  $O(N \log N)$  Force-Calculation Algorithm. *Nature*, 324 :446–449, December 1986.
- [3] A. BENOÎT, P. ADE, A. AMBLARD, R. ANSARI, É. AUBOURG, S. BARGOT, J. G. BARTLETT, J.-P. BERNARD, R. S. BHATIA, A. BLANCHARD, J. J. BOCK, A. BOSCALERI, F. R. BOUCHET, A. BOURRACHOT, P. CAMUS, F. COUCHOT, P. DE BERNARDIS, J. DELABROUILLE, F.-X. DÉSSERT, O. DORÉ, M. DOUSPIS, L. DUMOULIN, X. DUPAC, P. FILLIATRE, P. FOSALBA, K. GANGA, F. GANNAWAY, B. GAUTIER, M. GIARD, Y. GIRAUD-HÉRAUD, R. GISPERT, L. GUGLIELMI, J.-C. HAMILTON, S. HANANY, S. HENROT-VERSILLÉ, J. KAPLAN, G. LAGACHE, J.-M. LAMARRE, A. E. LANGE, J. F. MACÍAS-PÉREZ, K. MADET, B. MAFFEI, C. MAGNEVILLE, D. P. MARRONE, S. MASI, F. MAYET, A. MURPHY, F. NARAGHI, F. NATI, G. PATANCHON, G. PERRIN, M. PIAT, N. PONTHEIU, S. PRUNET, J.-L. PUGET, C. RENAULT, C. ROSSET, D. SANTOS, A. STAROBINSKY, I. STRUKOV, R. V. SUDIWALA, R. TEYSSIER, M. TRISTRAM, C. TUCKER, J.-C. VANEL, D. VIBERT, E. WAKUI, and D. YVON. Cosmological constraints from Archeops. *A&A*, 399 :L25–L30, March 2003.
- [4] F. BERNARDEAU, S. COLOMBI, E. GAZTAÑAGA, and R. SCOCCIMARRO. Large-scale structure of the Universe and cosmological perturbation theory. *Phys. Rep.*, 367 :1–3, September 2002.
- [5] E. BERTSCHINGER. Cosmics : Cosmological initial conditions and microwave anisotropy codes. *ApJ*, June 1995.
- [6] F. R. BOUCHET and H. E. KANDRUP. Particle-mesh simulations of clustering in cosmology. *ApJ*, 299 :1–4, December 1985.
- [7] S. M. CARROLL. The Cosmological Constant. *Living Reviews in Relativity*, 4 :1–+, February 2001.
- [8] M. COLLESS, G. DALTON, S. MADDOX, W. SUTHERLAND, P. NORBERG, S. COLE, J. BLAND-HAWTHORN, T. BRIDGES, R. CANNON, C. COLLINS, W. COUCH, N. CROSS, K. DEELEY, R. DE PROPRIIS, S. P. DRIVER, G. EFSTATHIOU, R. S. ELLIS, C. S. FRENK, K. GLAZEBROOK, C. JACKSON, O. LAHAV, I. LEWIS, S. LUMSDEN, D. MADGWICK, J. A. PEACOCK, B. A. PETERSON,

- I. PRICE, M. SEABORNE, and K. TAYLOR. The 2dF Galaxy Redshift Survey : spectra and redshifts. *MNRAS*, 328 :1039–1063, December 2001.
- [9] A. G. DOROSHKEVICH, E. V. KOTOK, A. N. POLIUDOV, S. F. SHANDARIN, I. S. SIGOV, and I. D. NOVIKOV. Two-dimensional simulation of the gravitational system dynamics and formation of the large-scale structure of the universe. *MNRAS*, 192 :321–337, August 1980.
- [10] R. DURRER, A. GABRIELLI, M. JOYCE, and F. SYLOS LABINI. Bias and the Power Spectrum beyond the Turnover. *ApJ*, 585 :L1–L4, March 2003.
- [11] G. EFSTATHIOU, M. DAVIS, S. D. M. WHITE, and C. S. FRENK. Numerical techniques for large cosmological N-body simulations. *ApJS*, 57 :241–260, February 1985.
- [12] D. J. FIXSEN, E. S. CHENG, J. M. GALES, J. C. MATHER, R. A. SHAFER, and E. L. WRIGHT. The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Set. *ApJ*, 473 :576–+, December 1996.
- [13] W. L. FREEDMAN, B. F. MADORE, B. K. GIBSON, L. FERRARESE, D. D. KELSON, S. SAKAI, J. R. MOULD, R. C. KENNICUTT, Jr., H. C. FORD, J. A. GRAHAM, J. P. HUCHRA, S. M. G. HUGHES, G. D. ILLINGWORTH, L. M. MACRI, and P. B. STETSON. Final Results from the Hubble Space Telescope Key Project to Measure the Hubble Constant. *ApJ*, 553 :47–72, May 2001.
- [14] M. J. GELLER and J. P. HUCHRA. Mapping the universe. *Science*, 246 :897–903, November 1989.
- [15] N. Y. GNEDIN. Softened Lagrangian hydrodynamics for cosmology. *ApJS*, 97 :231–257, April 1995.
- [16] J. R. I. GOTT, M. JURIC, D. SCHLEGEL, F. HOYLE, M. VOGLEY, M. TEGMARK, N. BAHCALL, and J. BRINKMANN. A Map of the Universe. *ApJ*, 624 :463–484, May 2005.
- [17] H. GOULD and J. TOBOCHNIK. *An Introduction to Computer Simulation Methods*. Addison Wesley, third edition, 1988.
- [18] A. H. GUTH. Inflationary universe : A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23 :347–356, January 1981.
- [19] S. HANANY, P. ADE, A. BALBI, J. BOCK, J. BORRILL, A. BOSCALERI, P. DE BERNARDIS, P. G. FERREIRA, V. V. HRISTOV, A. H. JAFFE, A. E. LANGE, A. T. LEE, P. D. MAUSKOPF, C. B. NETTERFIELD, S. OH, E. PASCALE, B. RABII, P. L. RICHARDS, G. F. SMOOT, R. STOMPOR, C. D. WINANT, and J. H. P. WU. MAXIMA-1 : A Measurement of the Cosmic Microwave Background Anisotropy on Angular Scales of  $10^{-5\circ}$ . *ApJ*, 545 :L5–L9, December 2000.

- [20] L. HERNQUIST, F. R. BOUCHET, and Y. SUTO. Application of the Ewald method to cosmological N-body simulations. *ApJS*, 75 :231–240, February 1991.
- [21] N. KAISER. Clustering in real space and in redshift space. *MNRAS*, 227 :1–21, July 1987.
- [22] A. A. KLYPIN and S. F. SHANDARIN. Three-dimensional numerical model of the formation of large-scale structure in the Universe. *MNRAS*, 204 :891–907, September 1983.
- [23] R. A. KNOP, G. ALDERING, R. AMANULLAH, P. ASTIER, G. BLANC, M. S. BURNS, A. CONLEY, S. E. DEUSTUA, M. DOI, R. ELLIS, S. FABBRO, G. FOLATELLI, A. S. FRUCHTER, G. GARAVINI, S. GARMOND, K. GARTON, R. GIBBONS, G. GOLDHABER, A. GOOBAR, D. E. GROOM, D. HARDIN, I. HOOK, D. A. HOWELL, A. G. KIM, B. C. LEE, C. LIDMAN, J. MENDEZ, S. NOBILI, P. E. NUGENT, R. PAIN, N. PANAGIA, C. R. PENNYPACKER, S. PERLMUTTER, R. QUIMBY, J. RAUX, N. REGNAULT, P. RUIZ-LAPUENTE, G. SAINTON, B. SCHAEFER, K. SCHAHMANECHE, E. SMITH, A. L. SPADAFORA, V. STANISHEV, M. SULLIVAN, N. A. WALTON, L. WANG, W. M. WOOD-VASEY, and N. YASUDA. New Constraints on  $\Omega_M$ ,  $\Omega_\Lambda$ , and  $w$  from an Independent Set of 11 High-Redshift Supernovae Observed with the Hubble Space Telescope. *ApJ*, 598 :102–137, November 2003.
- [24] A. V. KRAVTSOV, A. A. KLYPIN, and A. M. KHOKHLOV. Adaptive Refinement Tree : A New High-Resolution N-Body Code for Cosmological Simulations. *ApJS*, 111 :73–+, July 1997.
- [25] A. D. LINDE. *The physics of elementary particles and inflationary cosmology*. Moscow, Izdatel'stvo Nauka, 1990, 280 p. In Russian., 1990.
- [26] D. H. D. H. LYTH and A. A. RIOTTO. Particle physics models of inflation and the cosmological density perturbation. *Phys. Rep.*, 314 :1–2, June 1999.
- [27] D. MCCAMMON. ROSAT, Rockets, and a First Look at Suzaku – Observational Constraints On Hot Gas In The Galaxy. In *American Astronomical Society Meeting Abstracts*, pages 39.02–+, June 2006.
- [28] S. S. MCGAUGH. Constraints on the radial mass distribution of dark matter halos from rotation curves. In P. NATARAJAN, editor, *The shapes of galaxies and their dark halos, Proceedings of the Yale Cosmology Workshop "The Shapes of Galaxies and Their Dark Matter Halos", New Haven, Connecticut, USA, 28-30 May 2001. Edited by Priyamvada Natarajan. Singapore : World Scientific, 2002, ISBN 9810248482, p.186*, pages 186–+, 2002.
- [29] R. H. MILLER. Numerical experiments on the clustering of galaxies. *ApJ*, 270 :390–409, July 1983.
- [30] R. H. MILLER and K. H. PRENDERGAST. Stellar Dynamics in a Discrete Phase Space. *ApJ*, 151 :699–+, February 1968.

- [31] C. W. MISNER, K. S. THORNE, and J. A. WHEELER. *Gravitation*. San Francisco : W.H. Freeman and Co., 1973, 1973.
- [32] T. MONTROY, P. A. R. ADE, A. BALBI, J. J. BOCK, J. R. BOND, J. BORRILL, A. BOSCALERI, P. CABELLA, C. R. CONTALDI, B. P. CRILL, P. DE BERNARDIS, G. DE GASPERIS, A. DE OLIVEIRA-COSTA, G. DE TROIA, G. DI STEFANO, K. GANGA, E. HIVON, V. V. HRISTOV, A. IACOANGELI, A. H. JAFFE, T. S. KISNER, W. C. JONES, A. E. LANGE, S. MASI, P. D. MAUSKOPF, C. MAC TAVISH, A. MELCHIORRI, F. NATI, P. NATOLI, C. B. NETTERFIELD, E. PASCALE, F. PIA-CENTINI, D. POGOSYAN, G. POLENTA, S. PRUNET, S. RICCIARDI, G. ROMEO, J. E. RUHL, E. TORBET, M. TEGMARK, and N. VITTORIO. Measuring CMB polarization with Boomerang. *New Astronomy Review*, 47 :1057–1065, December 2003.
- [33] M. L. NORMAN and G. L. BRYAN. Cosmological Adaptive Mesh Refinement. pages 19–+, 1999.
- [34] T. PADMANABHAN. *Books-Received - Structure Formation in the Universe*, volume 103. August 1993.
- [35] P. J. E. PEEBLES. *The large-scale structure of the universe*. Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p., 1980.
- [36] P. J. E. PEEBLES. *Principles of physical cosmology*. Princeton Series in Physics, Princeton, NJ : Princeton University Press, |c1993, 1993.
- [37] U.-L. PEN. A Linear Moving Adaptive Particle-Mesh N-Body Algorithm. *ApJS*, 100 :269–+, September 1995.
- [38] A. A. PENZIAS and R. W. WILSON. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, 142 :419–421, July 1965.
- [39] S. PERLMUTTER, G. ALDERING, G. GOLDHABER, R. A. KNOP, P. NUGENT, P. G. CASTRO, S. DEUSTUA, S. FABBRO, A. GOOBAR, D. E. GROOM, I. M. HOOK, A. G. KIM, M. Y. KIM, J. C. LEE, N. J. NUNES, R. PAIN, C. R. PENNYPACKER, R. QUIMBY, C. LIDMAN, R. S. ELLIS, M. IRWIN, R. G. MCMAHON, P. RUIZ-LAPUENTE, N. WALTON, B. SCHAEFER, B. J. BOYLE, A. V. FILIPENKO, T. MATHESON, A. S. FRUCHTER, N. PANAGIA, H. J. M. NEWBERG, W. J. COUCH, and THE SUPERNOVA COSMOLOGY PROJECT. Measurements of Omega and Lambda from 42 High-Redshift Supernovae. *ApJ*, 517 :565–586, June 1999.
- [40] A. RIOTTO. Theories of Baryogenesis. In A. MASIERO, G. SENJANOVIC, and A. SMIRNOV, editors, *High Energy Physics and Cosmology, 1998 Summer School*, pages 326–+, 1999.
- [41] H. P. ROBERTSON. Kinematics and World-Structure. *ApJ*, 82 :284–+, November 1935.

- [42] P. SCHNEIDER. Part 3 : Weak gravitational lensing. In G. MEYLAN, P. JETZER, P. NORTH, P. SCHNEIDER, C. S. KOCHANÉK, and J. WAMBSGANSS, editors, *Saas-Fee Advanced Course 33 : Gravitational Lensing : Strong, Weak and Micro*, pages 269–451, 2006.
- [43] D. N. & al SPERGEL. Wilkinson microwave anisotropy probe (wmap) three year results : Implications for cosmology. *ApJ*, 2006.
- [44] V. SPRINGEL. The cosmological simulation code GADGET-2. *MNRAS*, 364 :1105–1134, December 2005.
- [45] T. SUGINOHARA, Y. SUTO, F. R. BOUCHET, and L. HERNQUIST. Cosmological N-body simulations with a tree code - Fluctuations in the linear and nonlinear regimes. *ApJS*, 75 :631–643, March 1991.
- [46] I. TERENO, O. DORÉ, L. VAN WAERBEKE, and Y. MELLIER. Joint cosmological parameters forecast from CFHTLS-cosmic shear and CMB data. *A&A*, 429 :383–398, January 2005.
- [47] R. TEYSSIER. Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES. *A&A*, 385 :337–364, April 2002.
- [48] J. V. VILLUMSEN. A new hierarchical particle-mesh code for very large scale cosmological N-body simulations. *ApJS*, 71 :407–431, November 1989.
- [49] A. G. WALKER. Distance in an expanding universe. *MNRAS*, 94 :159–+, December 1933.
- [50] S. WEINBERG. *Gravitation and Cosmology : Principles and Applications of the General Theory of Relativity*. *Gravitation and Cosmology : Principles and Applications of the General Theory of Relativity*, by Steven Weinberg, pp. 688. ISBN 0-471-92567-5. Wiley-VCH , July 1972., July 1972.
- [51] Y. B. ZEL'DOVICH. Gravitational instability : an approximate theory for large density perturbations. *A&A*, 5 :84–89, March 1970.



---

# Chapitre 2

## Catalogues virtuels de galaxies

### 2.1 Modèles semi-analytiques pour la formation de galaxies

#### 2.1.1 GalICS

GalICS (acronyme de "GALaxies in Cosmological Simulations") est un modèle de formation et d'évolution panchromatique des galaxies développé par Steve Hatton, Julien Devriendt, Jérémy Blaizot et Stéphane Ninin sur une initiative de François Bouchet et Bruno Guiderdoni [6]. Il suit l'approche hybride qui consiste à appliquer un traitement semi-analytique aux structures détectées dans des simulations de matière noire. Le but d'une telle méthode est principalement de permettre la confrontation des théories de formation galactique dans un contexte cosmologique aux observations, mais aussi d'offrir la possibilité de créer de faux catalogues de galaxies pouvant aussi bien être utilisée pour la conception de futurs instruments de mesure que pour la calibration de mesures statistiques sur les distributions de galaxies. L'avantage principal d'une méthode hybride est sa rapidité qui autorise la simulation de larges portions d'Univers local, contrairement aux simulations numériques pures (avec des codes AMR notamment). Celles-ci utilisent les lois de la mécanique des fluides pour former de manière réaliste les galaxies, ce qui demande une puissance de calcul incomparable. Le but de cette section n'est pas d'expliquer précisément la méthode GalICS mais d'en exposer le principe général afin de garder en tête ses qualités et ses limites. Pour une description plus détaillée et complète, on pourra se référer à Hatton *et al.* [6] ainsi qu'aux thèses de S. Ninin, J. Devriendt et J. Blaizot.

Le premier postulat des modèles semi-analytiques et de GalICS en particulier est que, la masse totale de l'univers provenant majoritairement de la matière noire, les baryons ne font que suivre le potentiel gravitationnel des grandes structures formées par celle-ci. Le lieu de formation des galaxies est alors principalement influencé par la matière noire. Les simulations de matière noire obtenues grâce aux algorithmes N-Corps sont constituées d'un ensemble de cubes ayant des tailles pouvant aller de

quelques mégaparsecs à quelques centaines de mégaparsecs. Pour une simulation, le résultat obtenu est donc constitué d'un ensemble de cubes, chacun échantillonnant par quelques dizaines de millions de particules (typiquement de  $256^3$  à  $512^3$ , voire  $1024^3$ ) le champ de densité de matière noire en un même endroit de l'univers mais à des temps  $a_i$  différents.

### Arbres de fusion des halos

La première étape de GalICS consiste à identifier les halos de matière noire pour chaque  $a_i$  puis à déterminer les arbres de fusion de ces halos en identifiant descendants et progéniteurs d'un pas de temps à l'autre. L'identification des halos est réalisée par un algorithme de percolation de type Friend-of-Friend (FOF). La version utilisée pour GalICS est celle disponible sur le serveur public de l'Université de Washington (<http://www-hpcc.astro.washington.edu/tools/>), mais le lecteur pourra se référer à l'annexe ?? pour une explication de l'algorithme FOF développé pendant cette thèse pour d'autres applications. Les halos identifiés grâce à cette méthode sont constitués de groupes pouvant contenir de vingt à des centaines de milliers de particules (les halos de moins de vingt particules étant rejetés car potentiellement dûs au hasard). Une première limitation importante apparaît ici : aucun halo de moins de vingt particules ne pouvant être détecté, il ne pourra y avoir formation de galaxies que dans des halos d'une masse minimale égale à vingt fois la masse d'une particule de la simulation de matière noire. Ce n'est pas une limitation physique mais bien numérique dont il faudra tenir compte dans l'interprétation des résultats.

Une fois tous les halos identifiés, l'arbre de fusion peut être calculé en identifiant tous les progéniteurs et descendants de chaque halo à tous les temps  $a_i$ . Pour cela, un halo identifié à un temps  $a_i$  contenant une particule au moins appartenant à un autre halo identifié à un temps  $a_{i-1}$  sera identifié comme son descendant (et réciproquement, l'autre sera identifié comme son progéniteur). Le descendant principal est, parmi tous les descendants d'un halo, celui contenant le plus de particules ayant appartenu à celui-ci (et réciproquement pour le géniteur principal).

### Modélisation de la matière baryonique

La première hypothèse de GalICS étant que la matière noire gouverne la formation des halos, ceux-ci doivent être peuplés de baryons sous forme de gaz selon un modèle simple. La nucléosynthèse primordiale prédit la formation d'une fraction  $\Omega_B/\Omega_0$  de baryons, composés à 75% d'hydrogène et à 25% d'hélium en masse. C'est donc cette même fraction de matière baryonique qui devrait être présente dans les halos, les baryons étant supposés suivre le potentiel de la matière noire. En supposant que les halos sont virialisés et que leurs profils de densité sont ceux de sphères

isothermes :

$$\rho(\mathbf{r}) = \frac{M_v}{4\pi R_v} \frac{1}{r^2} \quad (2.1)$$

avec  $M_v$  la masse du viriel du halo et  $R_v$  le rayon du viriel du Halo, la masse  $M_v \Omega_B / \Omega_0$  de baryons "chauds" est donc attribuée à chaque halo. Sa température est donnée par la température du viriel  $T_v$  :

$$T_v = \frac{\mu m_p}{2k_B} \frac{GM_v}{R_v}. \quad (2.2)$$

Par la suite, ce gaz ionisé va refroidir radiativement et tomber dans le puits de potentiel du halo. Le temps caractéristique de ce refroidissement  $t_{cool}(\mathbf{r})$  pour la fraction du gaz se trouvant à une distance  $\mathbf{r}$  du centre du halo peut être exprimé en fonction de l'énergie et du taux de perte par radiation. Dans GalICS, la prescription suivie est celle de Sutherland and Dopita [11].

### Formation de galaxies

La formation de galaxies est modélisée en supposant que la masse baryonique refroidie se dépose sur un disque proto-galactique en conservant son moment angulaire spécifique jusqu'au rayon où le gaz a eu le temps de refroidir. En supposant que le profil de densité du disque est exponentiel et en définissant :

$$\lambda = \frac{E^{1/2} \|\mathbf{J}\|}{GM^{5/3}} \quad (2.3)$$

où  $\mathbf{J}$  est le moment angulaire du halo,  $E$  son énergie et  $M$  sa masse, le rayon du disque est donné, d'après Fall and Efstathiou [5], par :

$$r_d = \frac{\lambda}{\sqrt{2}} R_v. \quad (2.4)$$

Dans GalICS, toutes les galaxies sont donc formées de cette manière, et restent des disques purs tant qu'il n'y a pas de fusion ou que leur masse ne dépasse pas un seuil les rendant instables (ce qui conduit à la formation d'un bulbe par un mécanisme séculaire de type "barre").

Une galaxie est composée d'étoiles et de gaz. La formation d'étoiles suit la loi de Kennicutt [7]. Un paramètre d'efficacité  $\beta^{-1}$  est donc défini. Il caractérise la formation d'étoiles - la physique de la formation d'étoiles restant à ce jour très mal connue - et le taux de formation est donné par :

$$\frac{dM_\star}{dt} = \frac{M_g(t)}{\beta t_{\text{dyn}}} \quad (2.5)$$

avec  $M_g(t)$  la masse de gaz dans le disque au temps  $t$ . La répartition en masse de ces étoiles est quant à elle modélisée par une loi universelle appelée "fonction de masse initiale". Le cas le plus typique est celui de la fonction de masse de Salpeter [9] :

$\phi(m) \propto m^{-2.35}$ , normée sur le domaine des masses possibles d'étoiles (de 1 à  $120 M_{\odot}$ ). Il est à noter que la rétroaction des supernovae est aussi prise en compte dans la formation d'étoiles et dans les échanges de gaz avec le milieu interamas, d'après le modèle de rétroaction Silk [10].

La morphologie des galaxies est représentée par trois composantes : le bulbe, le disque et le sursaut. Le sursaut constitue une sorte de composante transitoire lorsque le disque est instable (d'après le critère de van den Bosch [12] sur la valeur du rapport des vitesses circulaires du disque et du halo) ou lors de fusions. Une partie du gaz et des étoiles est alors transférée dans cette composante où elle entraîne une formation élevée d'étoiles qui seront transférées dans une composante bulbe. La proportion d'étoiles et de gaz dans chaque composante est déterminée par l'historique de fusion des halos à l'intérieur desquels les galaxies fusionnent à leur tour par collision directe ou plus fréquemment en tombant au centre du puits de potentiel de matière sombre sous l'effet de la friction dynamique.

### Spectres synthétiques et magnitudes

Connaissant pour chaque galaxie l'historique de fusions ainsi que l'évolution en contenu gazeux et métallique, GalICS permet de créer des spectres synthétiques pour chaque galaxie. Ces spectres synthétiques,  $F_{\lambda}^*(t)$ , sont calculés en utilisant la formule suivante :

$$F_{\lambda}^*(t) = \int_0^t \int_1^{120} \frac{dM_{\star}(t_{\tau})}{dt} \phi(m) f_{\lambda}(m, \tau) dm d\tau, \quad (2.6)$$

qui lie le flux  $F_{\lambda}^*(t)$  de la galaxie à une longueur d'onde  $\lambda$  donnée à la fonction de masse initiale  $\phi(m)$  ainsi qu'au taux de formation d'étoiles  $dM_{\star}/dt$  et au flux  $f_{\lambda}(m, \tau)$  d'une étoile de masse initiale  $m$  et d'âge  $\tau$ . Pour être calculée, cette intégrale est discrétisée dans le temps et les spectres stellaires sont tirés d'une librairie de spectres théoriques. Ces spectres sont principalement tirés de Kurucz [8] et le lecteur peut se référer à Devriendt *et al.* [4] pour une explication détaillée de la procédure.

Les spectres étant calculés à toutes les époques et pour toutes les galaxies, il est alors facile de calculer les luminosités bolométriques des galaxies mais aussi, par simple convolution avec les filtres adéquats, de récupérer les magnitudes apparentes dans une grande quantité de bandes observées (celles utilisées par SDSS notamment).

### Résultats obtenus

GalICS permet d'obtenir, à partir d'une simulation de résolution raisonnable, une distribution de galaxies réaliste reproduisant les principales caractéristiques des distributions observées, notamment les fonctions de luminosité (Hatton *et al.* [6]) et les fonctions de corrélation (voir Cattaneo *et al.* [3] et Blaizot *et al.* [2]). La description de ces galaxies est de plus assez complète puisqu'elle contient aussi bien

les informations morphologiques que les caractéristiques spectrales. Certaines limitations doivent cependant être prises en compte :

- La masse maximale d’un halo étant de vingt fois la masse d’une particule  $M_p$  de la simulation de matière noire, des galaxies ne peuvent se former que dans des halos d’une masse supérieure à  $20 M_p$ .
- La limite inférieure sur la masse des galaxies est donc la masse de gaz contenue dans ces halos soit  $20M_p\Omega_B/\Omega_0$
- Cette limite induit une limite sur la luminosité minimale des galaxies plus délicate à calculer précisément.
- La formation de la première galaxie est déterminée par la détection du premier halo, par conséquent, la résolution de la simulation utilisée détermine le temps à partir duquel la distribution de galaxies est réaliste.
- L’historique de formation des galaxies ne commence que lorsqu’un halo de  $20 M_p$  est formé. L’accrétion de matière en dessous de cette masse n’est pas prise en compte et seules les galaxies dans les halos ayant subi plusieurs fusions ont donc des propriétés physiques réalistes (voir Blaizot *et al.* [1]).

### 2.1.2 Méthodes de biais

## 2.2 MoLUSC

Le but de MoLUSC (acronyme de “MOck Local Universe Survey Constructor”) est d’autoriser la construction de catalogues virtuels de galaxies à grande échelle en utilisant une approche hybride entre les méthodes semi-analytiques et les méthodes de biais. Comme il a été expliqué dans la section 2.1.1, la qualité des résultats obtenus grâce à GalICS dépend fortement de la résolution des simulations de matière noire. Cette résolution est donnée par le rapport du volume de la boîte de simulation au nombre de particules utilisées. La puissance de calcul étant limitée (et surtout l’espace mémoire), il n’est malheureusement pas possible actuellement de simuler des volumes de l’ordre du gigaparsec cube tout en conservant une résolution suffisante pour obtenir des galaxies réalistes avec GalICS. Par exemple, pour une simulation LCDM de  $1 Gpc^3$  contenant  $512^3$  particules, la masse d’un halo de 20 particules est de l’ordre de  $\approx 2 \times 10^{13} M_\odot$  soit une masse minimale pour les galaxies de  $\approx 2.7 \times 10^{12} M_\odot$ . Des problèmes apparaissent alors si l’on désire simuler de grands catalogues de galaxies tels que le SDSS : soit la taille des simulations est faible, auquel cas le catalogue simulé ne peut fournir aucune information sur une échelle dépassant la taille de la boîte, soit la taille de la simulation est suffisante mais on ne peut étudier que les galaxies les plus massives (il manque les galaxies les moins lumineuses).

Les méthodes de biais ne connaissent pas ce genre de contraintes mais produisent des catalogues bien moins réalistes que GalICS et les spectres des galaxies ne peuvent pas être calculés. MoLUSC propose une approche hybride entre les deux dont le principe est d’utiliser GalICS sur de petites simulations à haute résolution et de re-

produire les caractéristiques de la distribution obtenue sur des simulations de grande taille en utilisant une approche statistique.

### 2.2.1 Calcul des distributions de galaxies

Tout comme GalICS, MoLUSC fait l’hypothèse que la distribution spatiale des galaxies est principalement influencée par la distribution de la matière noire sous-jacente et que tous les autres phénomènes physiques l’influencant peuvent être considérés comme stochastiques et donc négligeables. Partant de ce principe, MoLUSC permet la création de catalogues virtuels de galaxies de grande taille (de l’ordre du gigaparsec) à partir d’une simulation de matière noire de grande taille ainsi qu’une autre simulation plus petite mais de meilleure résolution à laquelle GalICS a été appliqué.

Appelons  $S_g$  la grande simulation de matière noire,  $S_p$  la petite simulation de résolution élevée et  $G_p$  la distribution de galaxies obtenues en appliquant GalICS à  $S_p$ . Le processus de création de la distribution de galaxies  $G_g^*$  à partir de  $S_g$  en utilisant MoLUSC se déroule en deux étapes :

1. Le calcul du biais entre galaxies et matière noire à partir de  $S_p$  et  $G_p$  :
  - (a) Échantillonnage du champ de densité de  $S_p$  et  $G_p$  sur une grille (noté  $\rho_{S_p}(\mathbf{r}_i)$  et  $\rho_{G_p}(\mathbf{r}_i)$ ) .
  - (b) Calcul de la probabilité  $P(\rho_{G_p}(\mathbf{r}_i) | \rho_{S_p}(\mathbf{r}_i))$  qu’en un nœud  $i$  de grille, une densité  $\rho_{S_p}(\mathbf{r}_i)$  donnée de matière noire corresponde à une densité  $\rho_{G_p}(\mathbf{r}_i)$  de galaxies.
  - (c) Calcul, pour une valeur donnée de  $\rho_{S_p}(\mathbf{r})$  et de  $\rho_{G_p}(\mathbf{r})$ , de la probabilité qu’une galaxie située en  $\mathbf{r}$  ait un spectre donné parmi tous ceux de  $G_p$ .
2. La création à partir de  $S_g$  d’une distribution de galaxies  $G_g^*$  respectant les probabilités calculées à la première étape et suivant la distribution du champ de matière noire :
  - (a) Échantillonnage du champ de densité de  $S_g$  sur une grille (noté  $\rho_{S_g}(\mathbf{r}_i)$ ).
  - (b) Construction d’un champ de densité  $\rho_{G_g^*}(\mathbf{r}_i)$  à partir de  $\rho_{S_g}(\mathbf{r}_i)$  et respectant la distribution  $P(\rho_{G_p}(\mathbf{r}_i) | \rho_{S_p}(\mathbf{r}_i))$ .
  - (c) Création d’une distribution de galaxies avec des spectres associés à partir de  $\rho_{G_g^*}(\mathbf{r}_i)$ .

#### Calcul du biais à partir de GalICS

La première étape consiste à calculer le biais simulé par GalICS entre la distribution de matière noire et de galaxies. Typiquement, afin d’obtenir des résultats suffisamment précis, la simulation  $S_p$  de petite taille utilisée doit avoir une masse maximale pour une particule de l’ordre de  $10^8 M_\odot$ . Une fois GalICS appliqué à  $S_p$  pour obtenir  $G_p$ , chacun des deux champs de densité est échantillonné sur une grille

dont chaque cellule est un cube mesurant  $\sigma$  Mpc. Il existe de nombreuses méthodes pour échantillonner un champ représenté par une distribution discrète de particules. La méthode CIC et les techniques de lissage sont décrites dans le chapitre 3.3.1. Le cas de MoLUSC est cependant un peu particulier. On désire en effet ici garder une trace de l'identité des particules ayant contribué à chaque nœud de grille afin de conserver l'information spectrale. C'est impossible si l'on se contente d'utiliser un lissage par FFT (voir le chapitre 3.3.1) certes rapide mais inadapté. La solution choisie est de considérer chaque particule  $i$  comme une nuage de densité  $W(\mathbf{r} - \mathbf{r}_i)$  centré sur la position de la  $i^{\text{ième}}$  particule. La densité en nombre attribuée au nœud  $k$  est donc donnée par :

$$n(\mathbf{r}_k) = \sum_{i=0}^N W(\mathbf{r}_k - \mathbf{r}_i), \quad (2.7)$$

où  $N$  est le nombre de particules total et  $\mathbf{r}_k$  les coordonnées du  $k^{\text{ième}}$  nœud de grille. De manière similaire la densité en masse au nœud  $k$  est décrite par :

$$\rho(\mathbf{r}_k) = \sum_{i=0}^N m_i W(\mathbf{r}_k - \mathbf{r}_i), \quad (2.8)$$

avec  $m_i$  la masse de la particule  $i$ . De nombreux choix sont possibles pour le noyau  $W(\mathbf{r})$  utilisé, par exemple les splines cubiques couramment utilisés par les méthodes de simulation SPH. En pratique, il ne s'agit pas pour nous de calculer des interactions, et la fonction gaussienne tronquée convient parfaitement :

$$W(\mathbf{r}) = \frac{1}{(4\pi L^2)^{3/2}} \exp\left(-\frac{\|\mathbf{r}\|^2}{2L^2}\right) \Pi(\Delta\sigma - \|\mathbf{r}\|), \quad (2.9)$$

la valeur de  $L$  fixant la taille du lissage effectué lors de l'échantillonnage et  $\Pi(x)$  représentant la fonction de Heavyside, nulle si  $x < 0$  et valant 1 dans les autres cas. En pratique, la fonction gaussienne ayant une extension infinie il n'est en effet pas possible d'utiliser directement le noyau gaussien et c'est donc une version tronquée qui est appliquée. Dans les équations (2.8) et (2.7), les sommes seront alors restreintes aux particules se situant à une distance  $d < \Delta\sigma$  du nœud  $k$ . L'expérience montre qu'une valeur  $L = \sigma$  donne un lissage suffisant du champ pour qu'il soit continu. En prenant alors  $\Delta = 5$ , l'erreur commise sur la valeur du champ de densité pour un champ homogène est inférieure à  $6 \times 10^{-5} \%$  en tout point, ce qui est tout à fait acceptable pour notre application.

Une fois les champs de densité  $\rho_{S_p}(\mathbf{r}_i)$  et  $\rho_{G_p}(\mathbf{r}_i)$  échantillonnés, la deuxième étape consiste à mesurer comment ils sont reliés l'un à l'autre. Pour cela, la probabilité  $P(n_G | \rho_S)$  que la densité en nombre de galaxie soit  $n_G(\mathbf{r})$  en un point quelconque  $\mathbf{r}$  dans la distribution de galaxies  $G$  est mesurée, sachant que la densité en masse dans la simulation de matière noire correspondante est  $\rho_S(\mathbf{r})$ . Cette mesure est effectuée

directement sur la grille d'échantillonnage en appliquant l'équation suivante :

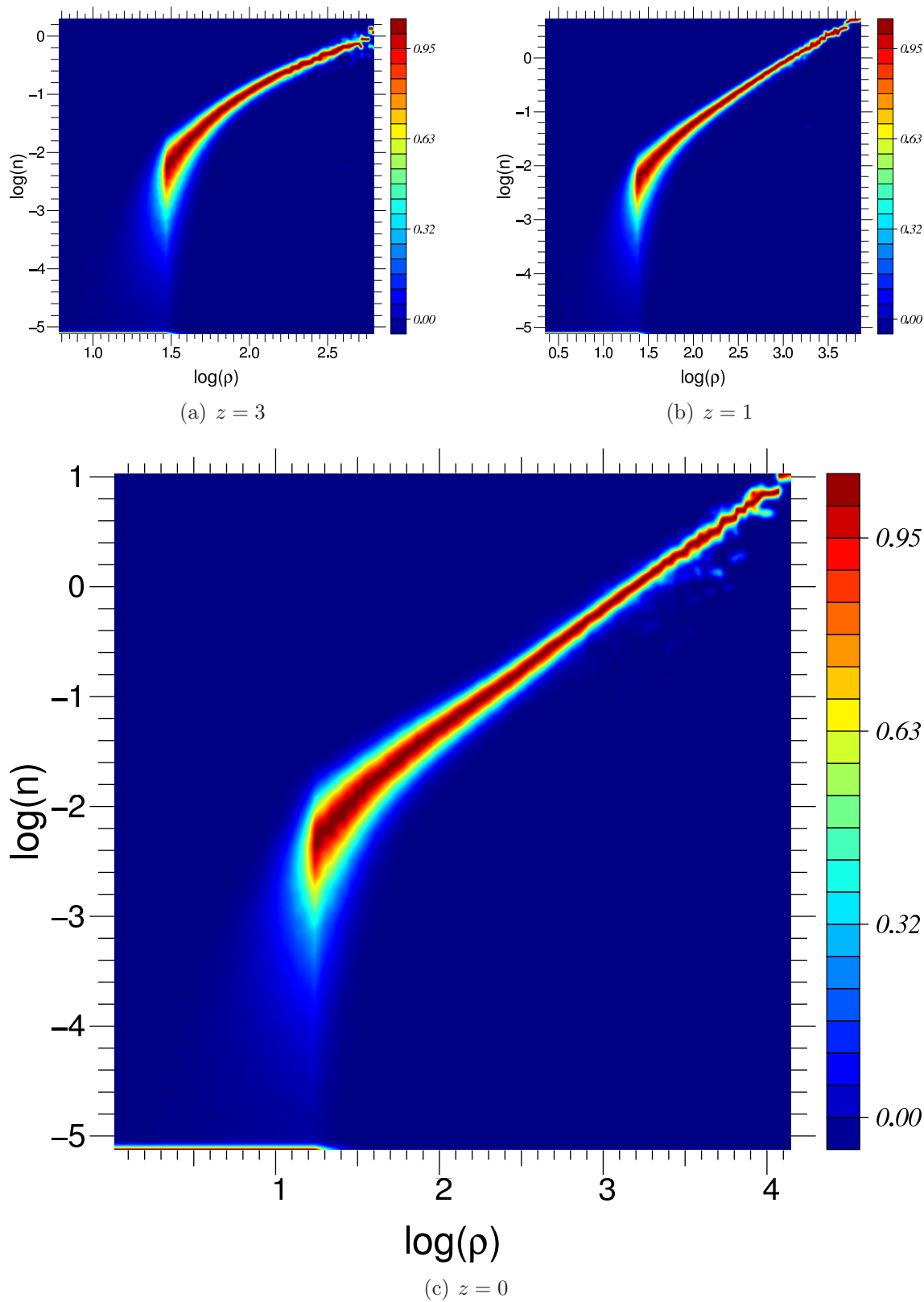
$$\begin{cases} P(n_G | \rho_S) \propto \sum_{k=1}^{N_n} \delta(\rho_S(\mathbf{r}_k) - \rho_S) \delta(n_G(\mathbf{r}_k) - n_G) \\ \int_0^\infty P(n_G | \rho_S) dn_G = 1 \end{cases} \quad (2.10)$$

où la somme est effectuée sur les  $N_n$  nœuds de la grille et où  $\delta(\mathbf{r})$  désigne la fonction de Dirac.

La figure 2.1 présente les fonctions  $P(n_G | \rho_S)$  calculées à différents décalages spectraux à partir d'une simulation de taille  $100h^{-1}$  Mpc comportant  $512^3$  particules et du champ de galaxies obtenu après utilisation de GalICS. Comme on pouvait s'y attendre, le biais apparaît constitué de deux régimes distincts. Pour une faible densité de matière, la formation de galaxie ne peut avoir lieu, et la densité de galaxies est donc nulle. Au contraire, lorsque la densité est suffisamment élevée, la densité en nombre de galaxies lui est directement proportionnelle. Entre ces deux régimes, une sorte de transition apparaît, où des densités de galaxies très variées peuvent correspondre à une même densité de matière, selon l'historique de la formation de galaxies dans ces régions. Un bon ajustement du rapport entre densité galactique  $n$  et de matière  $\rho$  dans le régime linéaire est donné par  $n = b\rho$  où  $b$  est le paramètre de biais et  $\delta = \rho / \langle \rho \rangle - 1$ . L'examen des figures 2.1(a) et 2.1(b) montre cependant un changement de comportement à grand décalage spectral. Ce changement peut être expliqué par les limitations en résolution de GalICS (masse minimale des halos de vingt particules). Lorsque  $z = 3$  par exemple, l'effondrement gravitationnel des halos massifs de matière noire n'a pas encore eu lieu. La fonction de masse des halos est donc dominée par les petits objets or ceux-ci sont les moins bien résolus par la simulation. Les plus gros halos résultent alors de la fusion de petits halos, et contiennent moins de galaxies que ce qu'ils devraient.

Finalement, afin de pouvoir attribuer de manière cohérente des spectres aux galaxies qui seront recrées à partir de  $S_g$ , la probabilité pour une galaxie donnée d'avoir un spectre donné doit être calculée. En partant du postulat que la distribution des galaxies est principalement dictée par la distribution de la matière noire, il est cohérent de faire l'approximation que la proportion de chaque type de galaxies à un endroit donné ne dépendra que de la densité en matière noire et en galaxies à ce même endroit. Sous cette hypothèse, la distribution des spectres de galaxies est donnée par la probabilité  $P(F_i(\lambda) | n_G, \rho_S)$  qu'une galaxie se trouvant en un point où la densité de matière noire est  $\rho_S$  et la densité en nombre de galaxies  $n_G$  se voit attribuer le spectre  $F_i(\lambda)$ . Pour chacune des  $N$  galaxies, GalICS crée un spectre synthétique  $F_i(\lambda)$  en fonction de son histoire. Une bonne approximation de





**Fig. 2.1:** Représentation en échelle logarithmique de la probabilité  $P(n_G | \rho_S)$  (non normalisée) qu'à une densité en masse  $\rho_S$  dans la simulation de matière noire corresponde une densité en nombre  $n_G$  de galaxies dans la simulation obtenue après utilisation de GalICS. Les différentes figures montrent l'évolution du biais avec le décalage spectral.

$P(F_i(\lambda) | n_G, \rho_S)$  peut alors être donnée par la formule suivante :

$$\left\{ \begin{array}{l} P(F_i(\lambda) | n_G, \rho_S) \propto \sum_{j=1}^{N_n} \delta(\rho_S(\mathbf{r}_j) - \rho_S) \delta(n_G(\mathbf{r}_j) - n_G) W(\mathbf{r}_i - \mathbf{r}_j) \\ \sum_{i=1}^N P(F_i(\lambda) | n_G, \rho_S) = 1 \end{array} \right. \quad (2.11)$$

où  $\mathbf{r}_i$  désigne la position de la galaxie  $i$ ,  $\mathbf{r}_j$  la position du nœud de grille  $j$  et où  $F_i(\lambda)$  est le spectre associé à la galaxie  $i$ . L'équation (2.11) exprime en quelques sortes le fait que la probabilité qu'un type de galaxies donné corresponde à des valeurs de  $n_G$  et  $\rho_S$  est proportionnelle au nombre de galaxies de ce type observées, dans la simulation GalICS, aux alentours de nœuds ayant des densités de galaxies et matière noire  $n_G$  et  $\rho_S$ . De plus, cette probabilité dépend de l'éloignement entre les galaxies et les nœuds concernés, dans la proportion de la contribution de ces galaxies à la valeur de la densité  $n_G$  mesurée en ces nœuds de grille (d'où le facteur  $W(\mathbf{r}_i - \mathbf{r}_j)$ ). En pratique, pour chacune des figures de 2.1, une liste des index des spectres de galaxies ayant contribué au calcul de couples  $(n_G, \rho_S)$  est attribuée à chaque pixel, avec les facteurs  $P(F_i(\lambda) | n_G, \rho_S)$  associés.

### Création de la distribution de galaxies

Une fois extraites les informations des simulations  $S_p$  et  $G_p$ , il reste encore à les utiliser afin de créer  $G_g^*$ , la distribution de galaxies correspondante à la simulation de matière noire à grande échelle  $S_g$ . La première étape consiste à calculer le champ de densité en nombre  $n_{G_g^*}$  des galaxies pour  $S_g$ . Pour cela, le champ de densité de matière noire de  $S_g$  est échantillonné en utilisant l'équation (2.8) et le noyau (2.9), tout en conservant les mêmes paramètres que précédemment (même valeurs de  $\sigma$ ,  $L$  et  $\Delta$ ). A partir de la valeur de  $P(n_G | \rho_S)$  calculée précédemment, il est alors facile de générer  $n_{G_g^*}$  en partant de  $\rho_{S_g}$ . Pour chacun des nœuds de la grille de densité  $\rho_i$ , la densité de galaxies correspondante  $n_i$  est tirée au hasard en suivant la densité de probabilité  $P(n | \rho_i)$ , une valeur  $n$  ayant la probabilité  $P(n | \rho) dn$  d'être choisie. D'un point de vue numérique, les ordinateurs ne sont capables de générer que des séries aléatoires de nombres décrivant une loi uniforme. Un nombre  $a$  tel que  $0 \leq a \leq 1$  est donc donné par l'ordinateur, et la valeur de  $n_i$  est alors choisie telle que :

$$\int_0^{n_i} P(n | \rho_i) dn = a. \quad (2.12)$$

De cette façon, chaque valeur de  $n_i$  est bien tirée avec une probabilité  $P(n_i | \rho_i) dn$  (voir l'annexe A).

La dernière étape consiste à créer la distribution *discrète* de galaxies respectant le champ de densité  $n_{G_g^*}$  calculé de la manière décrite dans le paragraphe précédent. Pour cela, il faut commencer par évaluer le nombre  $N_{G_g^*}$  de galaxies à générer. Lors du calcul du champ de densité, un noyau gaussien tronqué a été utilisé afin d'alléger le calcul numérique. L'inconvénient est que ce noyau n'est pas normé, de

plus, selon la position des particules, la répartition du poids des particules n'est pas forcément équivalente. Si, par exemple, la particule se situe sur un nœud de grille, elle contribuera à ce nœud pour  $W(0)$ , sur les nœuds directement voisins pour  $W(\sigma)$  ... Si elles se situent au milieu d'une cellule, elle contribuera pour  $W(\sqrt{3}\sigma)$  aux cellules voisines ... Du fait de la grille, la couverture du champ de densité n'est pas parfaitement homogène (une distribution de particules homogène ne donnerait pas exactement un champ de densité homogène mais périodique à l'échelle des cellules). Il se trouve qu'en choisissant  $L \leq \sigma$  comme paramètre du noyau, ce problème n'en n'est pas un. En effet, on peut montrer (de manière empirique) que dans ce cas, si l'on note :

$$\|W(\mathbf{r})\| = \sum_{i=1}^N W(\mathbf{r} - \mathbf{r}_i) \quad (2.13)$$

où  $i$  parcourt les  $N$  nœuds de grille, alors

$$\frac{\max_{\mathbf{r}}(\|W(\mathbf{r})\|) - \min_{\mathbf{r}}(\|W(\mathbf{r})\|)}{\min_{\mathbf{r}}(\|W(\mathbf{r})\|)} \leq 1\%, \quad (2.14)$$

où  $\min_{\mathbf{r}}(f(\mathbf{r}))$  (resp.  $\max_{\mathbf{r}}(f(\mathbf{r}))$ ) représente la valeur minimale (resp. maximale) prise par  $f(\mathbf{r})$ . Cela signifie en pratique que l'on peut considérer que  $\|W(\mathbf{r})\|$  est une constante et le nombre de galaxies est donc simplement donné par :

$$N_{G_g^*} = \frac{1}{\|W(\mathbf{r})\|} \sum_{i=1}^{N_g} n_{G_g^*}(\mathbf{r}_i) \quad (2.15)$$

avec  $N_g$  le nombre de nœuds dans la grille.

La distribution de galaxies étant, selon notre modèle, principalement influencée par la distribution de la matière noire, il semble judicieux de s'aider de la distribution des particules de matière noire pour générer celle des galaxies. La méthode utilisée pour placer les  $N_{G_g^*}$  galaxies consiste à parcourir l'ensemble des particules de la simulation de matière noire  $S_g$  et à trouver un critère pour, soit les transformer en galaxies et leur attribuer un spectre ainsi que la vitesse particulière de la particule de matière noire, soit les rejeter. De cette manière, la distribution des galaxies est assurée de suivre la distribution de matière noire. Étant donné que l'on connaît déjà le champ de densité des galaxies, la probabilité  $Q_i$  qu'une particule de matière noire située en  $\mathbf{r}_i$  devienne une galaxie est donc :

$$Q_i \propto \frac{n_G(\mathbf{r}_i)}{\rho_S(\mathbf{r}_i)} \quad (2.16)$$

qui peut facilement être normalisée, connaissant  $N_{G_g^*}$  :

$$Q_i = N_{G_g^*} \left( \frac{n_G(\mathbf{r}_i)}{\rho_S(\mathbf{r}_i)} \right) \left( \sum_{j=1}^{N_S} N_S \frac{n_G(\mathbf{r}_j)}{\rho_S(\mathbf{r}_j)} \right)^{-1} \quad (2.17)$$

où  $i$  parcourt l'ensemble des  $N_s$  particules de matière noire. Connaissant les densités en chaque nœud de grille, la densité en chaque particule est simplement interpolée linéairement. Il suffit en pratique de parcourir l'ensemble des particules de matière noire et de tirer, pour chacune d'elle, un nombre  $x \in [0, 1]$  au hasard, si  $x < Q_i$ , elle est transformée en galaxie, dans le cas contraire elle est effacée. L'attribution des spectres se passe de la même façon que la fabrication du champ de densité des galaxies : si une galaxie est créée en un point de densités  $\rho_S$  et  $n_G$ , un spectre est tiré au hasard selon la densité de probabilité  $P(F_i(\lambda) | n_G, \rho_S)$ . Il arrive parfois que  $Q_i$  soit supérieur à un lorsque la densité en galaxies attendue est supérieure à celle en particules de matière noire. C'est un cas très rare, se produisant uniquement lorsque la différence de résolution entre la petite simulation  $S_p$  et la grande  $S_g$  est très importante. Si toutefois il se produit, un nombre de galaxies égal à la partie entière de  $Q_i$  sont préalablement créées en tirant au hasard leurs coordonnées sur une sphère de rayon  $d$  centrée sur  $\mathbf{r}_i$ . Le rayon  $d$  suit une distribution Gaussienne de largeur  $L$  pour des raisons de cohérence avec l'échelle de lissage du champ de densité.

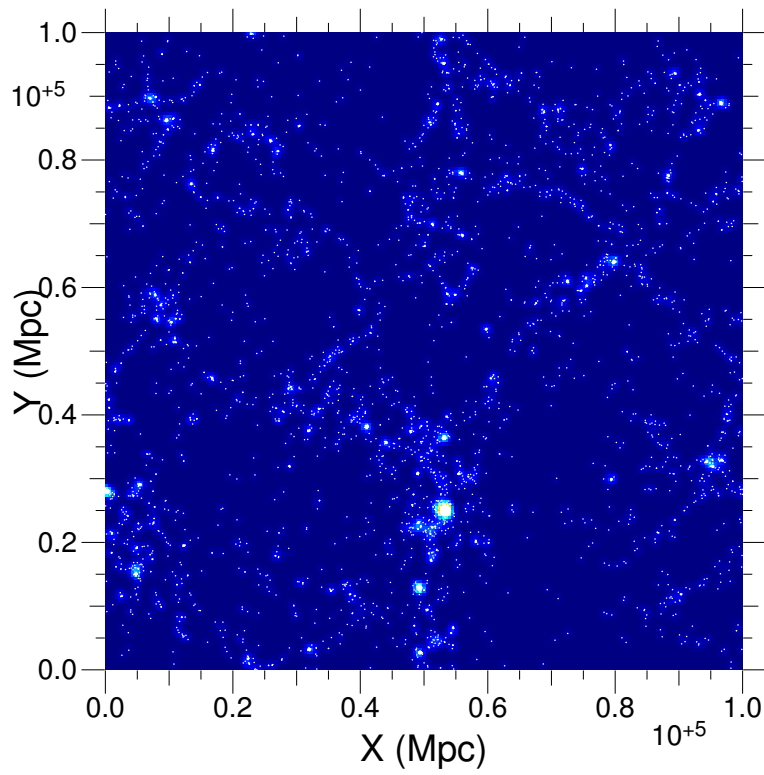
### Comparaison des résultats de GalICS et MoLUSC

Les figures 2.2(a) et 2.2(b) représentent les projections d'une tranche de  $40h^{-1}$  Mpc des galaxies générées avec GalICS et MoLUSC respectivement, à partir de la même simulation de matière noire. Cette simulation de type  $^3\text{LCDM}_{100}^{256}$  est celle disponible sur le site web de GalICS<sup>1</sup>, et la figure 2.2(b) est obtenue en utilisant MoLUSC calibré sur la distribution de galaxies de la figure 2.2(a). Le nombre de galaxies trouvées dans les deux cas est à peu près identique (30,765 avec GalICS et 30,941 avec MoLUSC) mais surtout la similitude des distributions est frappante : il est facile de retrouver les mêmes structures (halos, filaments et vides) exactement au même endroit. Les amas de galaxies créés par MoLUSC sont cependant plus étalés, ce problème est dû au lissage Gaussien des champs de densité nécessaire à l'application de l'algorithme. Une autre différence apparente réside dans la forme de ces amas mais constitue cette fois un atout de MoLUSC. GalICS faisant l'approximation de sphéricité des halos identifiés par FOF, les amas de galaxies ont forcément une géométrie sphérique (ce qui n'est pas très réaliste) ; contrairement aux amas créés par MoLUSC qui suivent parfaitement le champ de matière noire sous-jacent (de part la méthode utilisée).

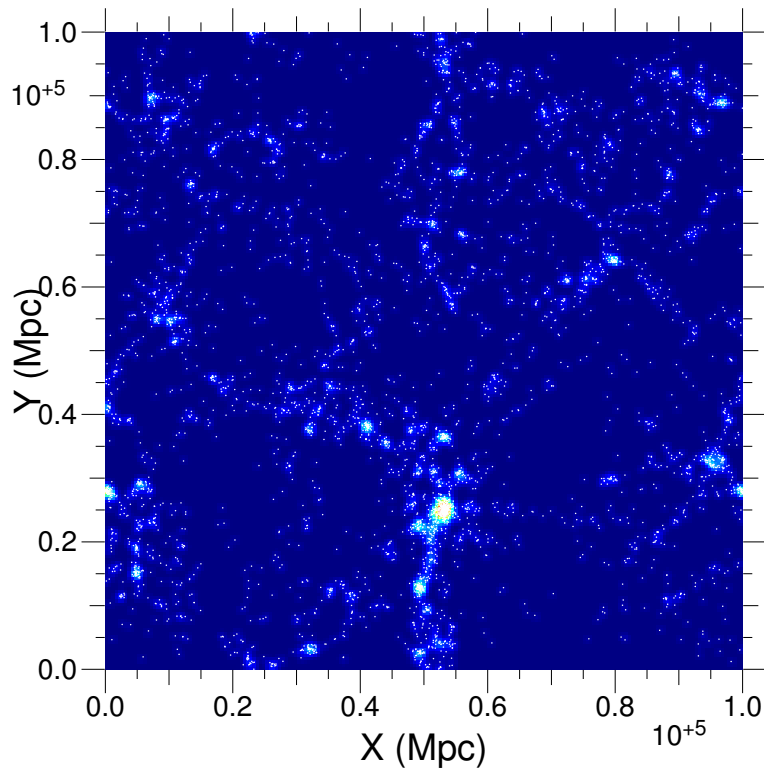
L'examen des fonctions de corrélation à deux points de la figure 2.3 confirme les observations. Le but de MoLUSC serait à priori de reproduire la fonction de corrélation des galaxies générées par GalICS, ce qui est parfaitement le cas pour des échelles suffisamment grandes : typiquement au-delà de 2 Mpc. Les différences apparaissant à plus petite échelle sont dues à l'utilisation du lissage (sur une distance de 1 Mpc ici) nécessaire pour MoLUSC qui, par construction, colle à la distribution de matière sombre. Le lissage entraîne donc un manque de corrélations pour MoLUSC à des échelles inférieure à celle du lissage mais le fait que la distribution de galaxies

---

<sup>1</sup><http://www.galics.iap.fr>



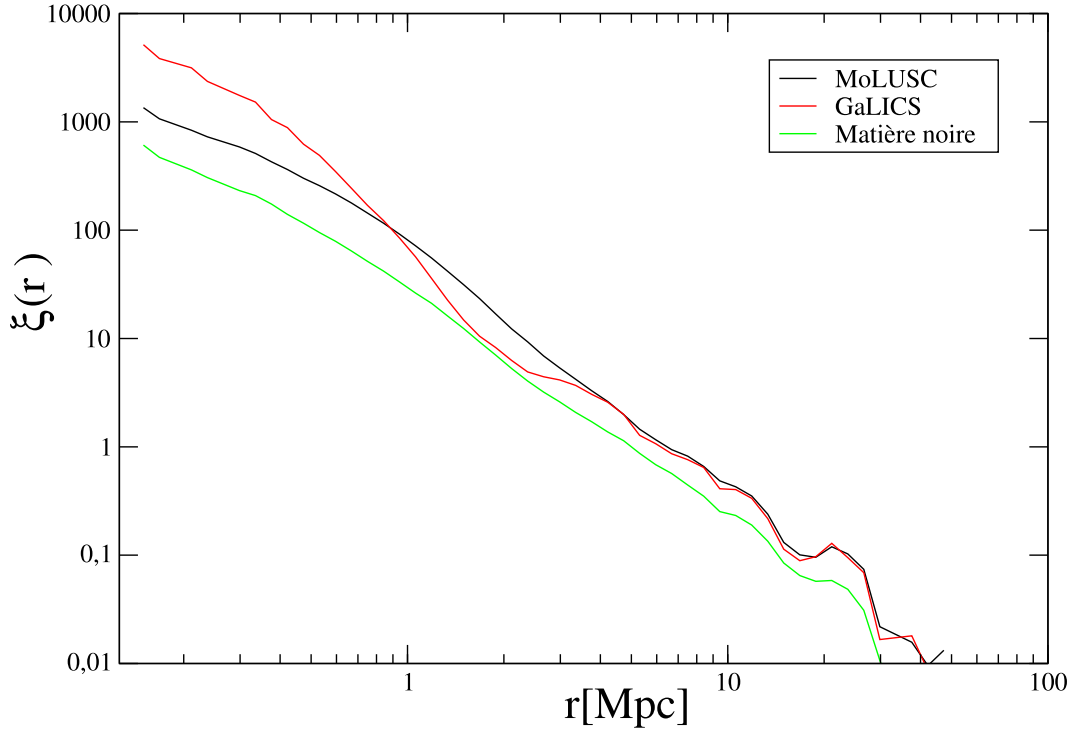
(a) GalICS (30,765 galaxies)



(b) MoLUSC (30,941 galaxies)

**Fig. 2.2:** Comparaison des distributions de galaxies obtenues avec GalICS et MoLUSC à partir de la même distribution de matière noire. L'étalonnage de MoLUSC est fait à partir de cette même simulation de matière noire et de la distribution de galaxies GalICS de la figure 2.2(a).

créée par MoLUSC respecte la distribution des sous-halos de matière noire permet la conservation des corrélations à une échelle de l'ordre du mégaparsec, ce que n'autorise pas GalICS en raison de l'utilisation de l'approximation sphérique pour la distribution galactique intra-halo. Ces résultats sont donc très probants sachant que l'objectif principal de MoLUSC est la production de catalogues de galaxies à très grande échelle.



**Fig. 2.3:** *Comparaison des fonctions de corrélation à deux points de la distribution de matière noire (courbe verte) et des distributions générées avec GalICS (courbe rouge) et MoLUSC (courbe noire). Une partie des particules de la simulation de matière noire a été retirée aléatoirement afin que le nombre de particules restantes soit le même dans les trois cas.*

### 2.2.2 Construction de catalogues fictifs

L'une des difficultés majeures dans l'étude des grandes structures réside dans la difficulté de la comparaison des résultats théoriques aux données observationnelles. Il est actuellement impossible d'observer directement la matière noire, d'où l'intérêt de modèles semi-analytiques comme GalICS permettant de générer rapidement des distributions de galaxies réalistes. MoLUSC est plus spécifiquement orienté vers l'étude des grandes structures en autorisant la création de volumes cohérents sur de plus grandes distances. Ces galaxies étant issues de simulations N-corps de matière noire, les distributions sont générées sous la forme de cubes de données représentant des sortes de photographies de portions d'univers à des moments donnés (que nous appellerons "pas de temps"). Un catalogue tel que le SDSS par exemple représente, quant à lui, une distribution de galaxies vues par un observateur, avec tout ce

que cela implique en terme de biais observationnels. Construire de faux catalogues consiste, à partir des cubes synthétiques de galaxies, à reproduire des catalogues tels qu'ils auraient été observés. Il se trouve qu'avant de nous parvenir, la lumière des galaxies parcourt un long chemin à travers l'Univers et est, par conséquent, fortement influencée par les propriétés de ce dernier. L'expansion permanente notamment influence énormément les propriétés observées des galaxies mais c'est aussi le cas de la constante cosmologique ou de la quantité de matière totale. Pour reproduire ces effets dans nos catalogues virtuels (c'est à dire passer du référentiel propre d'une galaxie à celui de l'observateur), la modélisation utilisée est celle présentée dans le chapitre d'introduction 1.1.3.

### Méthode du pavage aléatoire

Le but est ici de construire de faux catalogues observés à partir de boîtes contenant une image de la distribution des galaxies à différents pas de temps (nous appellerons ces boîtes "catalogues initiaux" dans la suite). Chacun des catalogues initiaux contient les informations suivantes :

- La valeur du décalage spectral correspondant a chaque pas de temps.
- Le nombre ainsi que la position et la vitesse de chacune des galaxies.
- Pour chaque galaxie, un spectre associé.

Il suffit alors de placer un observateur virtuel en un point quelconque  $\mathbf{r}$  dans le catalogue initial pris à décalage spectral nul, définir une ligne de visée et calculer les propriétés observées de chacune des galaxies se trouvant dans le volume du catalogue à reproduire (en prenant soin de les choisir dans le catalogue initial à décalage spectral correspondant le mieux possible à la distance entre la galaxie et l'observateur). Le problème est que la plupart du temps le catalogue virtuel à produire est plus volumineux que la distribution de galaxies générée. La seule solution est donc de le reproduire plusieurs fois et ce dernier ayant des conditions aux limites périodiques, il est alors possible de créer une distribution continue de taille arbitraire. Malheureusement, cette méthode entraîne des effets de réplication provenant de la répétition régulière des structures contenues dans les catalogues initiaux. Ce n'est pas un problème dans le cas où la géométrie est celle d'un pinceau étroit (c'est à dire dont la section est petite devant la taille du catalogue initial). Il suffit dans ce cas de choisir un axe de visée non parallèle aux bords de la boîte pour que les galaxies contenues dans le cône ne soient pas plusieurs fois les mêmes. Mais le but est ici de reproduire des catalogue couvrant une grande portion du ciel, la méthode adoptée est donc celle du pavage aléatoire.

Elle consiste en pratique à effectuer une série de rotations, translations et symétries de paramètres arbitraires à chaque réplication du catalogue initial. Appelons  $z_i < z_{i+1}$  les différents pas de temps disponibles pour les catalogues initiaux. L'observateur est alors placé dans le catalogue  $z_0 = 0$  et une série de paramètres des différentes transformations est tirée au hasard :

- L'angle, le centre et le vecteur donnant l'axe de la rotation.
- Le vecteur donnant la translation à effectuer.

- Trois nombres  $n_i \in \{0, 1\}$  déterminant si les coordonnées sont inversées selon chacun des axes.

Ce catalogue initial subit alors cette série de transformations et les propriétés des galaxies dont le décalage spectral  $z$  par rapport à l'observateur est plus proche de  $z_0$  que de  $z_1$  sont gardées. Si une galaxie se trouve à un décalage spectral plus proche de  $z_1$  que de  $z_0$ , alors le catalogue initial n°1 subit les mêmes transformations et les propriétés des galaxies sont prises dans le catalogue 1 et ainsi de suite. Une fois que le décalage spectral des galaxies observées dépasse son équivalent en taille de boîte, une nouvelle série de paramètres de transformation est tirée et l'opération est renouvelée jusqu'à atteindre la distance à l'observateur souhaitée.

Cette opération pose cependant un nouveau problème de réplication. En effet, entre deux catalogues initiaux à des pas de temps différents, les amas de galaxies se sont déplacés. Par conséquent, il est possible qu'un même amas se retrouve deux fois côte à côte dans le catalogue virtuel. Une option permet donc de désactiver l'utilisation de plusieurs pas de temps tant que les galaxies sont à une distance de l'observateur inférieure à la taille de la boîte, ce qui peut être utile pour la conservation des grandes structures dans des catalogues peu profonds. L'effet de réplication peut alors apparaître mais de manière beaucoup moins fréquente (uniquement sur les bords d'un cube de même taille que celle du catalogue initial).

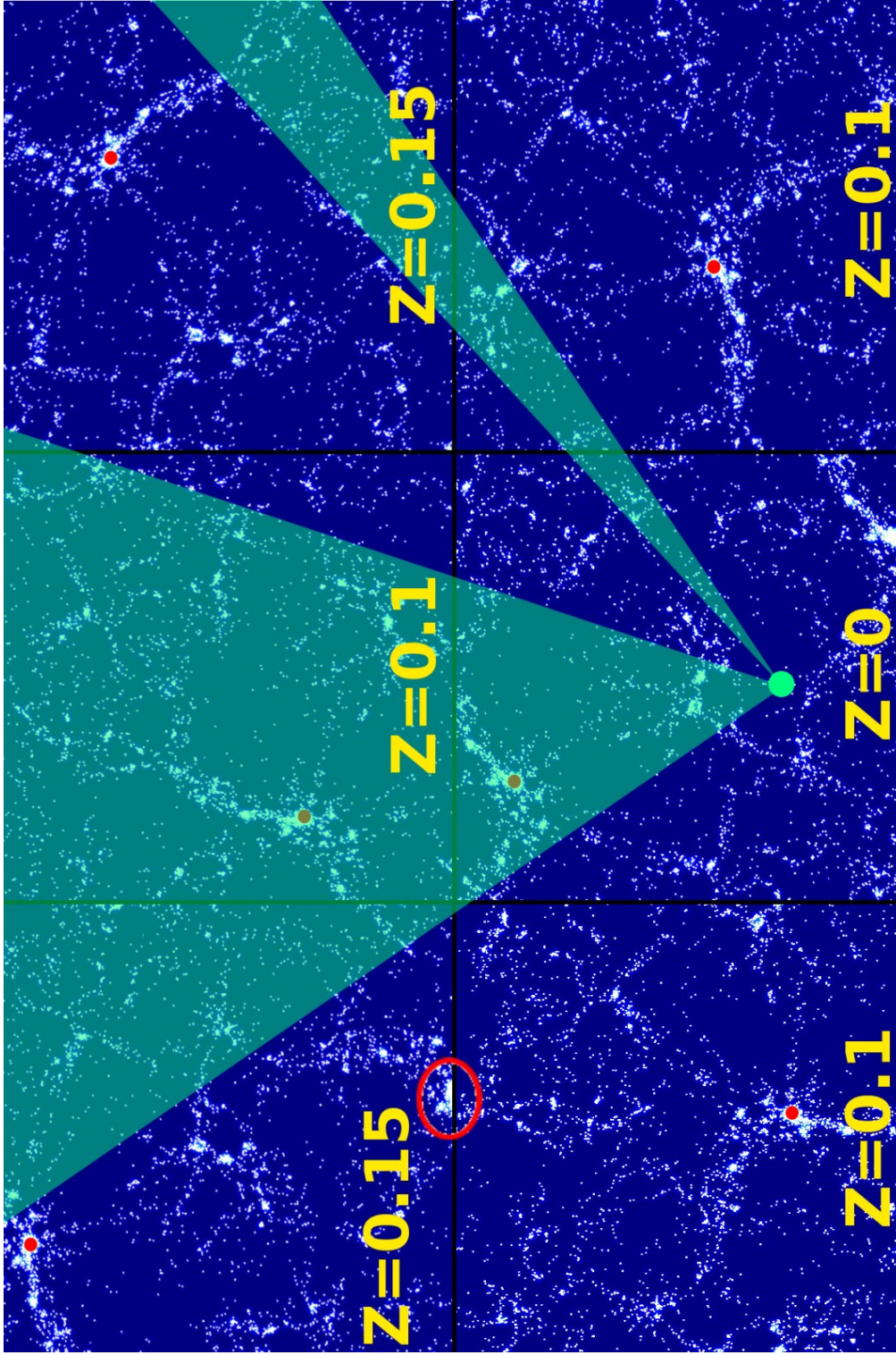
La figure 2.4 illustre la technique du pavage avec utilisation d'un pas de temps unique sur le volume d'un catalogue initial. Sur cette figure, le cercle rouge met en évidence un autre défaut de la méthode : si le changement de catalogue initial s'effectue brutalement, il est tout à fait possible qu'un amas de galaxies se trouve coupé en deux faisant apparaître une structure tout à fait irréaliste dans la distribution de galaxies. Pour pallier à cela, les amas sont identifiés par la méthode FOF (voir l'annexe ??) et les transformations sont effectuées sur les coordonnées des centres des amas au lieu des galaxies seulement. Si le centre d'un amas est identifié comme appartenant au catalogue virtuel, c'est alors la totalité de cet amas qui est recopiée.

## Découpage des cônes

Grâce à la technique du pavage aléatoire, il est possible de disposer d'une distribution de galaxies dans un volume aussi grand que souhaité. Les catalogues de galaxies ont cependant souvent une géométrie assez complexe et il est important de bien la reproduire afin de pouvoir tenir compte des effets de bords. La figure 1.7 donne une idée de cette géométrie pour le catalogue SDSS qui à l'évidence n'est pas facilement modélisable.

La méthode que nous utilisons est numérique afin de rester la plus généraliste possible. Partant d'un catalogue existant quelconque, elle consiste à échantillonner sur une grille les coordonnées angulaires  $(x, y)$  de chacune des galaxies dans un système de coordonnées bien choisi. Le système de coordonnées utilisé ainsi que la taille angulaire des pixels sont deux paramètres très importants. En effet, il est impossible,





**Fig. 2.4:** Illustration de la technique du pavage aléatoire et du découpage des faux catalogues à partir des catalogues initiaux. L'espace est constitué d'un pavage des catalogues initiaux à différents décalages spectraux  $z$  et ayant subi des transformations aléatoires. Les points rouges désignent le même amas de galaxies et le cercle rouge illustre le problème du découpage des amas sur les bords des boîtes. Pour créer le catalogue virtuel, seules les galaxies appartenant à la géométrie du catalogue à imiter sont sélectionnées (ici en vert).

lors de la projection, de conserver à la fois la surface angulaire des pixels et les distances angulaires entre les galaxies. De plus, la taille des pixels doit être telle qu'il ne soit pas possible de confondre une zone non observée avec une zone du ciel observée mais dépourvue de galaxies.

Afin que toutes les régions du ciel puissent être identifiées comme appartenant ou non au catalogue avec la même précision, une projection de Flamsteed des coordonnées des galaxies est utilisée (voir la figure 2.5 pour une projection du catalogue SDSS). Elle transforme les coordonnées angulaires  $(\alpha, \delta)$  des galaxies en coordonnées  $(x, y)$  telles que :

$$\begin{aligned} x &= \alpha \cos(\delta) \\ y &= \delta. \end{aligned} \tag{2.18}$$

L'avantage est que dans le système de coordonnées  $(x, y)$ , la surface d'un pixel centré en  $(x_i, y_i)$  et de largeur  $(dx, dy)$  ne dépend pas de la valeur de  $x$  et de  $y$ . Par conséquent, si l'on découpe l'espace en une grille cartésienne et que l'on donne la valeur 1 à chaque pixel contenant au moins une galaxie après projection, et 0 à tout autre pixel, on peut obtenir une carte de précision constante de la géométrie du catalogue. Il reste cependant à fixer les valeurs de  $dx$  et  $dy$ . En pratique ces valeurs dépendront de chaque catalogue et seront choisies pour que le produit de  $dx$  et  $dy$  soit inférieur à la surface de la plus petite région non observée mais supérieure à une dizaine de fois le carré de la distance angulaire moyenne entre les galaxies.

Après échantillonnage, il devient facile de découper un faux catalogue ayant la bonne géométrie dans la distribution de galaxies créée par pavage aléatoire : les coordonnées des galaxies par rapport à l'observateur sont calculées et seules celles tombant dans un pixel non nul sont gardées.

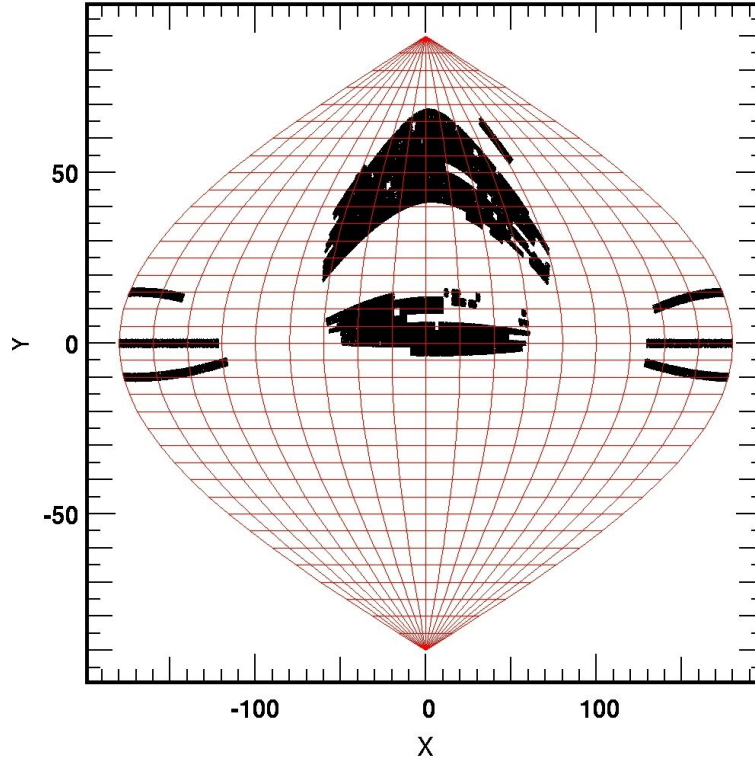
### Calcul des propriétés des galaxies

L'ultime étape de la création de catalogues simulés est le calcul des propriétés des galaxies telles que mesurées par un observateur. La distance observée tout d'abord est déduite du décalage spectral lui-même lié aux vitesses relatives de l'observateur et de la galaxie observée. Si l'on note  $\mathbf{x}_i$  la position de la galaxie  $i$  dans le repère de l'observateur, alors sa distance comobile exacte est  $D_i = \|\mathbf{x}_i\|$ . Soit  $\mathbf{v}_i$  la vitesse particulière de cette même galaxie, alors le décalage spectral observé se décompose en deux termes : le premier dû à l'expansion de l'Univers noté  $z_{ie}$  et le second à la vitesse particulière de la galaxie  $z_{ip}$ , ce dernier étant un terme parasite faussant la mesure. Le terme  $z_{ie}$  est facilement calculé en résolvant numériquement l'équation :

$$D_i = D(z_{ie}) \tag{2.19}$$

où l'expression de  $D(z_{ie})$  est donnée par l'équation (1.40). Le terme parasite vaut quant à lui :

$$z_{ip} = \sqrt{\frac{1 + \mathbf{v}_p}{1 - \mathbf{v}_p}} - 1 \tag{2.20}$$



**Fig. 2.5:** *Projection de Flamsteed des coordonnées angulaires des galaxies de SDSS. Le quadrillage rouge représente l'image d'une grille cartésienne sur les coordonnées angulaires non transformées. La projection de Flamsteed conservant les aires, en échantillonnant sur une grille cartésienne (ici d'axes X et Y) les coordonnées des galaxies après projection, la surface angulaire couverte par chaque pixel ne dépend plus de la déclinaison et la précision d'échantillonnage est donc constante.*

avec

$$\mathbf{v}_p = \frac{\mathbf{v}_i \mathbf{x}_i}{cD_i}. \quad (2.21)$$

L'erreur commise sur la mesure de la vitesse de la galaxie dans le flot de Hubble est égale à la projection de sa vitesse particulière sur l'axe de visée. Le décalage spectral observé de la galaxie  $i$  est alors  $z_i = z_{ie} + z_{ip}$ . Le terme parasite  $z_{ip}$  est à l'origine de fortes distorsions de la distribution de galaxies observée par rapport à la distribution réelle et il est possible de le supprimer dans le but d'étudier son impact.

L'autre caractéristique principale des galaxies est donnée par leurs spectres. En pratique, chaque catalogue de galaxies utilise une série de filtres spécifiques pour lesquels sont mesurées les magnitudes apparentes. Afin de calculer de manière réaliste les magnitudes de chaque galaxie, il faut à nouveau tenir compte du décalage spectral vers le rouge qui provoque un décalage entre les longueurs d'ondes  $\lambda$  observées et celles émises. Ainsi, une galaxie observée à un décalage spectral  $z_i$  aura un spectre

observé  $P_i^{\text{obs}}(\lambda)$  tel que :

$$P_i^{\text{obs}}(\lambda) = \frac{P_i((1+z_i)\lambda)}{1+z_i}, \quad (2.22)$$

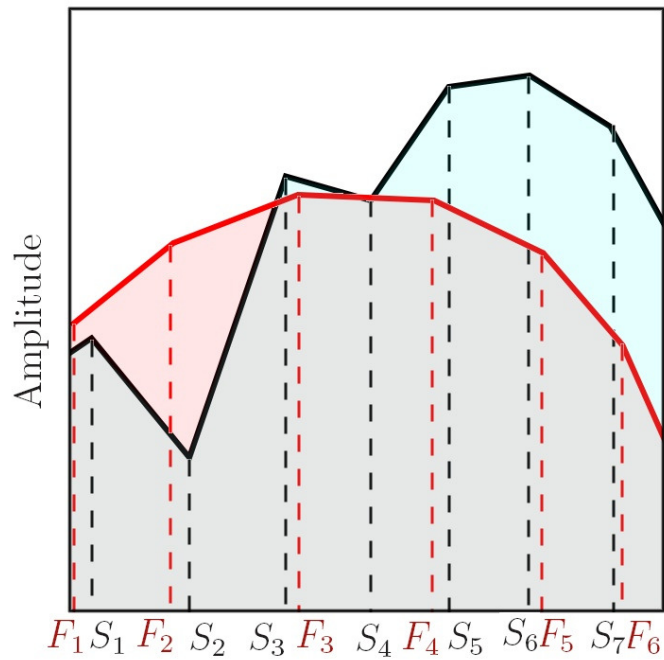
avec  $(1+z_i)\lambda$  la longueur d'onde observée et le facteur  $(1+z_i)^{-1}$  assurant la conservation de l'énergie totale rayonnée. La magnitude absolue  $M_i$  de la galaxie vu dans un filtre  $F(\lambda)$  est calculée par convolution du spectre et du filtre :

$$M_i = \int_0^\infty P_i(\lambda)F(\lambda) d\lambda \quad (2.23)$$

et la magnitude apparente vaut donc :

$$m_i = \int_0^\infty P_i^{\text{obs}}(\lambda)F((1+z_i)\lambda) d\lambda. \quad (2.24)$$

La base de donnée de filtres utilisés est celle de GaLCIS qui contient un grand nombre de filtres utilisés dans les plus grands catalogues de galaxies. Ces filtres ainsi que les spectres obtenus sont échantillonnés à des échelles différentes, les caractéristiques des filtres étant parfaitement connues mais celles des spectres étant plus où moins précises selon la longueur d'onde. Le calcul numérique de l'équation (2.24) se doit donc d'être à la fois précis et rapide (les magnitudes étant calculées plusieurs fois pour chacune des galaxies observées). La méthode utilisée est illustrée par la figure 2.6 et consiste à effectuer simplement la convolution dans l'espace réel par la méthode des trapèzes mais avec un pas variable afin de garder une précision maximale.



**Fig. 2.6:** Illustration de la méthode de calcul des intégrales (2.23) et (2.24) lorsque le spectre (noir) et le filtre (rouge) ont des échelles d'échantillonnage différentes. A partir de la longueur d'onde échantillonnée la plus petite  $\lambda = \min(S_1, F_1)$ , la méthode des trapèzes est utilisée entre les bornes  $\lambda$  et  $\lambda' = \min(S_2, F_2)$ , la valeur de l'amplitude de la composante non échantillonnée pour  $\lambda'$  étant interpolée linéairement. L'opération est renouvelée jusqu'à atteindre la valeur maximale de longueur d'onde du spectre ou du filtre.

---

# Bibliographie

- [1] J. BLAIZOT, B. GUIDERDONI, J. E. G. DEVRIENDT, F. R. BOUCHET, and S. HATTON. Predicting multi-wavelength properties of Lyman break galaxies with GALICS. *Ap&SS*, 284 :373–376, 2003.
- [2] J. BLAIZOT, B. GUIDERDONI, J. E. G. DEVRIENDT, F. R. BOUCHET, S. J. HATTON, and F. STOEHR. GALICS- III. Properties of Lyman-break galaxies at a redshift of 3. *MNRAS*, 352 :571–588, August 2004.
- [3] A. CATTANEO, A. DEKEL, J. DEVRIENDT, B. GUIDERDONI, and J. BLAIZOT. Modelling the galaxy bimodality : shutdown above a critical halo mass. *MNRAS*, 370 :1651–1665, August 2006.
- [4] J. E. G. DEVRIENDT, B. GUIDERDONI, and R. SADAT. Galaxy modelling. I. Spectral energy distributions from far-UV to sub-mm wavelengths. *A&A*, 350 :381–398, October 1999.
- [5] S. M. FALL and G. EFSTATHIOU. Formation and rotation of disc galaxies with haloes. *MNRAS*, 193 :189–206, October 1980.
- [6] S. HATTON, J. E. G. DEVRIENDT, S. NININ, F. R. BOUCHET, B. GUIDERDONI, and D. VIBERT. GALICS- I. A hybrid N-body/semi-analytic model of hierarchical galaxy formation. *MNRAS*, 343 :75–106, July 2003.
- [7] R. C. KENNICUTT, Jr. Star Formation in Galaxies Along the Hubble Sequence. *ARA&A*, 36 :189–232, 1998.
- [8] R. L. KURUCZ. Model Atmospheres for Population Synthesis. pages 225–+, 1992.
- [9] E. E. SALPETER. The Luminosity Function and Stellar Evolution. *ApJ*, 121 :161–+, January 1955.
- [10] J. SILK. A new prescription for protogalactic feedback and outflows : where have all the baryons gone? *MNRAS*, 343 :249–254, July 2003.
- [11] R. S. SUTHERLAND and M. A. DOPITA. Cooling functions for low-density astrophysical plasmas. *ApJS*, 88 :253–327, September 1993.
- [12] F. C. VAN DEN BOSCH. The Formation of Disk-Bulge-Halo Systems and the Origin of the Hubble Sequence. *ApJ*, 507 :601–614, November 1998.

---

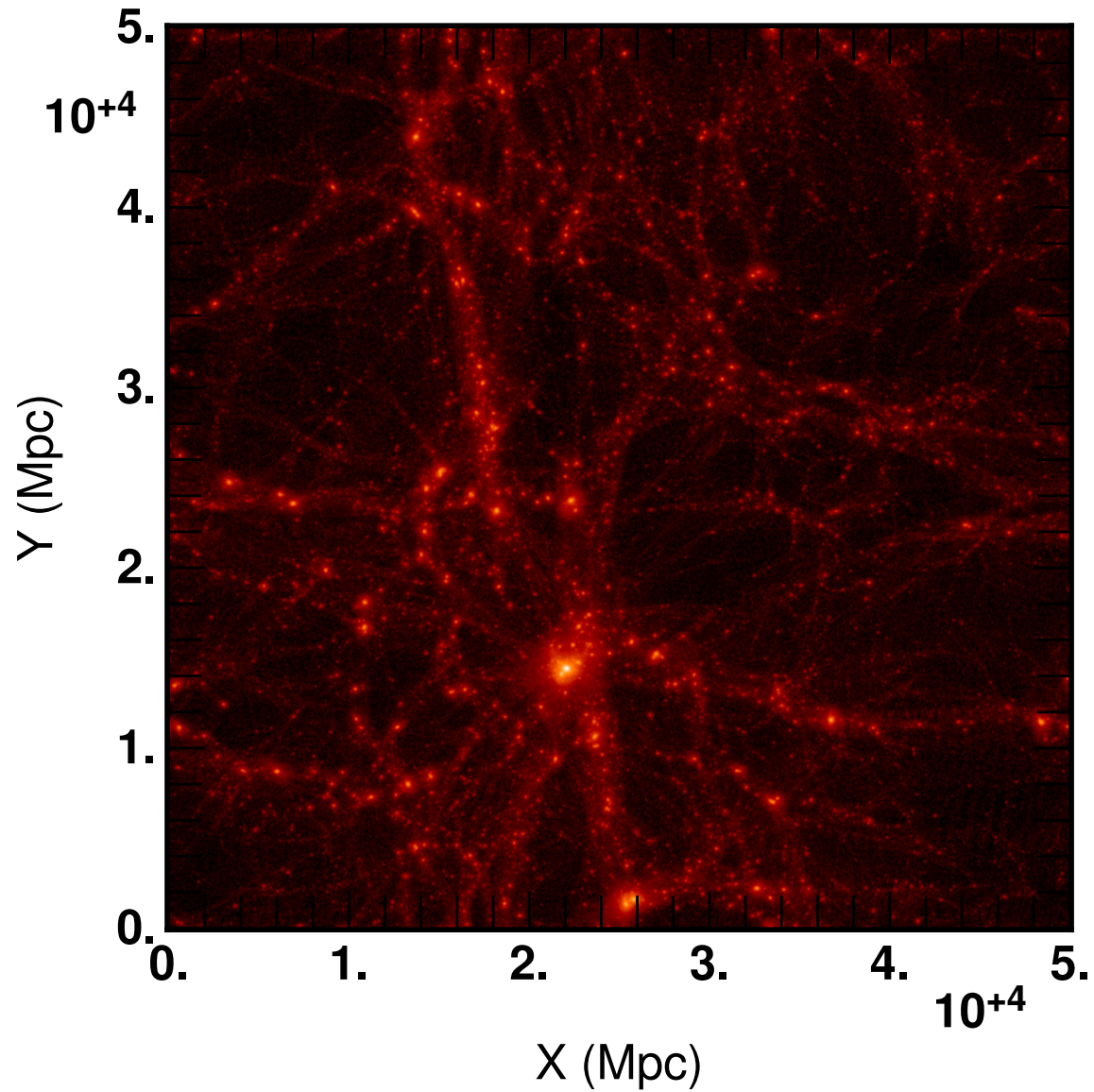
## Chapitre 3

# Topologie des grandes structures

En examinant le champ de densité dans une simulation de matière noire (figure 3.1), on est frappé par la structure filamentaire de celle-ci. La matière semble être répartie en trois zones distinctes :

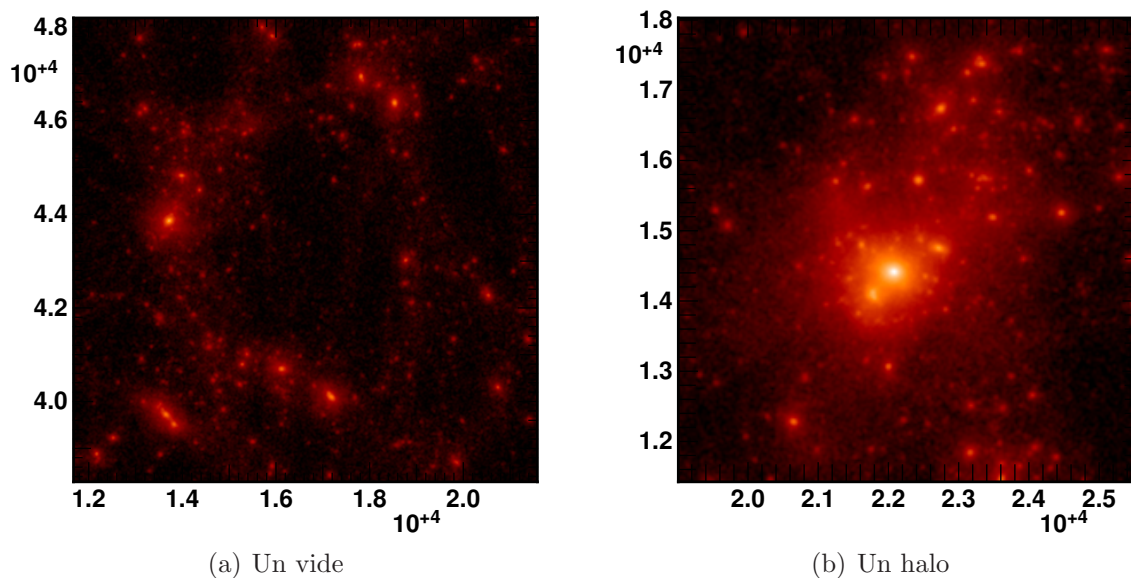
- **Les vides** : des zones contenant peu de matière et formant des sortes de bulles sous-denses.
- **Les filaments** : tout un réseau sur-dense de structures filamentaires entourant les vides.
- **Les halos** : des zones surdenses à la confluence des filaments, contenant elles même d'autres zones surdenses de plus petite taille (les sous-halos). Souvent, les halos et sous-halos ont une géométrie plus ou moins ellipsoïdale.

La physique des halos est actuellement la mieux connue. En effet, l'étude des halos présente un grand intérêt car ce sont les zones les plus denses et donc celles où le taux de formation des galaxies est le plus élevé. De plus ces régions se prêtent bien aux études théoriques (voir la section 1.2.1 pour un rapide rappel des bases), ainsi qu'à la modélisation en raison de leur géométrie pouvant être supposée sphérique voire ellipsoïdale au premier ordre. Enfin, leur identification dans des catalogues ou des simulations est relativement aisée grâce à des algorithmes de type friend-of-friend (voir l'annexe ??) ou plus évolués de type HOP (voir Aubert *et al.* [2] ou Eisenstein and Hut [14]), permettant l'identification de sous-halos. La distribution de matière dans les vides ainsi que leur répartition a fait l'objet de plus de recherches ces dernières années mais reste bien moins développée. Par définition, les vides sont des zones de faible densité où le taux de formation de galaxies est quasiment nul, rendant toute observation détaillée relativement délicate. Caractériser la nature des vides reste cependant très intéressant car ils constituent une source d'information sur l'évolution et la formation de notre Univers. Ainsi, leur taille, leur forme ou encore leur distribution constituent autant d'empreintes permettant de contraindre les modèles. Enfin, les filaments en eux-mêmes n'ont fait l'objet que de peu d'études, principalement car leur identification et leur caractérisation sont bien



**Fig. 3.1:** *Coupe de la projection du champ de densité de matière noire d'une simulation  ${}^3\text{LCDM}_{50}^{256}$ . L'image représente le logarithme de la densité projetée selon l'axe  $Z$  restreint entre 0 et 20 Mpc. L'aspect filamenteux de la distribution est évident sur cette image.*





**Fig. 3.2:** Agrandissements de deux zones de la figure 3.1 comportant un vide (fig. 3.2(a)) et un halo ainsi que ses sous-halos (fig. 3.2(b)).

plus délicate. Les principales méthodes utilisées sont le “minimal spanning tree” et dans une moindre mesure les excursions du champ de densité, introduites par la suite.

Dans ce chapitre, nous nous attacherons à présenter l’ensemble des outils les plus couramment utilisés dans l’analyse du champ de densité en nous efforçant de mettre en avant leurs avantages et leur inconvénients ainsi que leur champ principal d’application. Nous présenterons par la suite le concept du squelette, un nouvel outil dédié à l’étude des filaments visant à permettre leur extraction et leur étude. Finalement, l’implémentation pratique de cette méthode sera présentée en détails.

### 3.1 Méthodes actuelles pour étudier la topologie

Dans cette section nous présenterons les principales méthodes utilisées en cosmologie pour les études topologiques de la distribution de matière dans l’Univers. Cette présentation ne se veut cependant pas exhaustive mais a plutôt pour but de donner une idée du type d’approches imaginées. Parmi celles non décrites ici, on pourra notamment trouver les études de probabilités de vides relativement anciennes mais toujours d’actualité qui connurent un certain succès avec l’avènement de catalogues de galaxies de grande taille tels le SDSS, 2dFGRS ou 6dF. Certains articles récents abordent le sujet tel Benson *et al.* [4] et le lecteur pourra s’y référer pour de plus amples informations ainsi qu’aux nombreuses références citées dans l’introduction de cet article pour plus de détails (dont par exemple Peebles [30]). D’autres approches intéressantes existent encore, basées sur les diagrammes de voronï par exemple (Icke and van de Weygaert [18]), ou plus récemment les analyses multi-fractales (Querre

*et al.* [32] ou Martínez *et al.* [23]).

### 3.1.1 Fonctions de corrélation

#### Principe

La fonction de corrélation à deux points  $\xi(r)$  a certainement été le premier outil utilisé dans le but de quantifier les propriétés des structures à grande échelle dans l'Univers (voir Totsuji and Kihara [38] pour le premier article abordant le sujet dans un contexte cosmologique et Peebles [29] pour une description détaillée). En partant d'une réalisation discrète d'un processus aléatoire, la fonction  $\xi(r)$  peut être définie comme caractérisant l'excès (ou le défaut) de probabilité de trouver, dans une distribution discrète, un point à une distance  $r$  d'un autre point par rapport à une distribution homogène et uniforme de type Poissonienne. Ainsi, si  $\bar{n}$  est la densité moyenne de points, la probabilité  $\delta P$  de trouver un objet dans un volume  $\delta V$  situé à une distance  $r$  d'un autre objet s'écrit :

$$\delta P = \bar{n}^2(1 + \xi(r))\delta V. \quad (3.1)$$

La valeur de  $\xi(r)$  est donc une estimation de la tendance qu'ont les objets de la distribution à se regrouper (resp. se repousser) à une distance  $r$  donnée, suivant l'intensité de la fonction  $\xi(r)$  lorsqu'elle est positive (resp. est négative). Une distribution où les points ont tendance à se regrouper en tas verra sa fonction de corrélation fortement augmenter en dessous de l'échelle correspondant à la taille de ces tas. Au contraire, une distribution pour laquelle tous les points sont placés de manière à maximiser le volume occupé<sup>1</sup> aura une fonction de corrélation très négative en dessous de la distance interparticulaire moyenne.

Si l'on définit le contraste de densité  $\delta$  :

$$\delta = \frac{\rho(\mathbf{r}) - \langle \rho(\mathbf{r}) \rangle_{\mathbf{r}}}{\langle \rho(\mathbf{r}) \rangle_{\mathbf{r}}}, \quad (3.2)$$

alors on peut montrer que

$$\xi(r) = \langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle, \quad (3.3)$$

où  $\xi(r)$  ne dépend effectivement pas de  $\mathbf{x}$  pour une distribution homogène et isotrope. L'Univers étant supposé statistiquement invariant sous translation, il est pratique d'utiliser la transformée de Fourier du contraste de densité :

$$\delta(x) = \int \frac{d^3k}{(2\pi)^3} \delta_k \exp ikx, \quad (3.4)$$

les différents modes  $\delta_k$  étant non corrélés. On peut alors montrer que

$$\langle \delta_{k_1} \delta_{k_2} \rangle = \delta(k_1 + k_2)(2\pi)^3 P(k) \quad (3.5)$$

---

<sup>1</sup>de type "verre" par exemple, qui peut être obtenue en simulant les interactions gravitationnelles à l'aide d'un code N-corps dans lequel la flèche du temps aura été inversée.

où le spectre de puissance  $P(k)$  est la transformée de Fourier de la fonction de corrélation à deux points :

$$\xi(r) = \int d^3\mathbf{k} P(k) \exp(i\mathbf{k}\mathbf{r}). \quad (3.6)$$

Dans le cas de la distribution des galaxies, on mesure en général que la fonction de corrélation est une loi de puissance  $\xi(r) \propto r^{-\gamma}$  mais que son intensité chute énormément à petite échelle pour devenir négative. Cela s'explique par le fait que localement, la présence d'une galaxie empêche la formation d'une autre galaxie très proche (si c'est le cas, il y a de grandes chances qu'elle fusionnent). La loi de puissance traduit, quant à elle, le fait que la force en jeu, la gravitation, est purement attractive. Par conséquent les galaxies ont tendance à se regrouper là où il y a déjà d'autres galaxies, dans les amas de matière noire. Le spectre de puissance est une quantité très intéressante puisqu'il est caractéristique de la distribution initiale de matière, dans les modèles d'inflation par exemple (voir 1.2.1) qui suggèrent  $P(k) \propto k^n$ .

La fonction de corrélation peut être étendue à  $N$  points. D'après le théorème de Wick (voir par exemple Bernardeau *et al.* [5]), on peut montrer que

$$\delta P_{123} = \bar{n}^3 [1 + \xi(x_{12}) + \xi(x_{23}) + \xi(x_{13}) + \xi^{(3)}(w_{12}, x_{13}, x_{23})] \delta V_1 \delta V_2 \delta V_3. \quad (3.7)$$

où  $\delta P_{123}$  désigne la probabilité que trois points soient dans une configuration où la distance entre 1 et 2 soit  $x_{12}$ , celle entre 2 et 3 soit  $x_{23}$  et celle entre 1 et 3 soit  $x_{13}$ . Cette probabilité tient donc compte des fonctions de corrélation à deux points pour chacune des trois paires possibles (les termes  $\xi(x_{ij})$ ) mais aussi de ce que l'on définit comme la fonction de corrélation à trois points  $\xi^{(3)}(w_{12}, x_{13}, x_{23})$ . Ce genre de relation est valable jusqu'à n'importe quel ordre et l'on peut donc définir de cette manière la fonction de corrélation à  $N$  points  $\xi^{(N)}(r_i, \dots, r_N) = \langle \delta_1, \dots, \delta_N \rangle_c$  où  $\langle \rangle_c$  désigne la valeur moyenne sur l'ensemble des diagrammes connectés (c'est à dire où les  $N$  points sont *tous* à la bonne distance  $r_i = x_{ij}$  avec  $j \neq i$  les uns des autres en même temps, voir Bernardeau *et al.* [5]).

## Implémentations

En pratique, il existe plusieurs méthodes pour calculer les fonctions de corrélations d'une distribution de Galaxies, mais aussi et surtout plusieurs estimateurs. Celui qui est le plus souvent utilisé et semble être le plus approprié dans les contextes des catalogues de galaxies semble être celui de Landy et Szalay (voir Kerscher *et al.* [21]) :

$$\xi_{\text{LS}}(r) = \frac{DD - 2DR + RR}{RR} \quad (3.8)$$

où  $DD$  désigne la fonction de corrélation des données,  $DR$  celle des données à une distribution de Poisson et  $RR$  celle d'une distribution de Poisson. Le temps de calcul est aussi un problème, celui-ci étant déjà, si l'on utilise l'algorithme le plus simple consistant à mesurer toutes les paires, proportionnel au nombre de points dans la

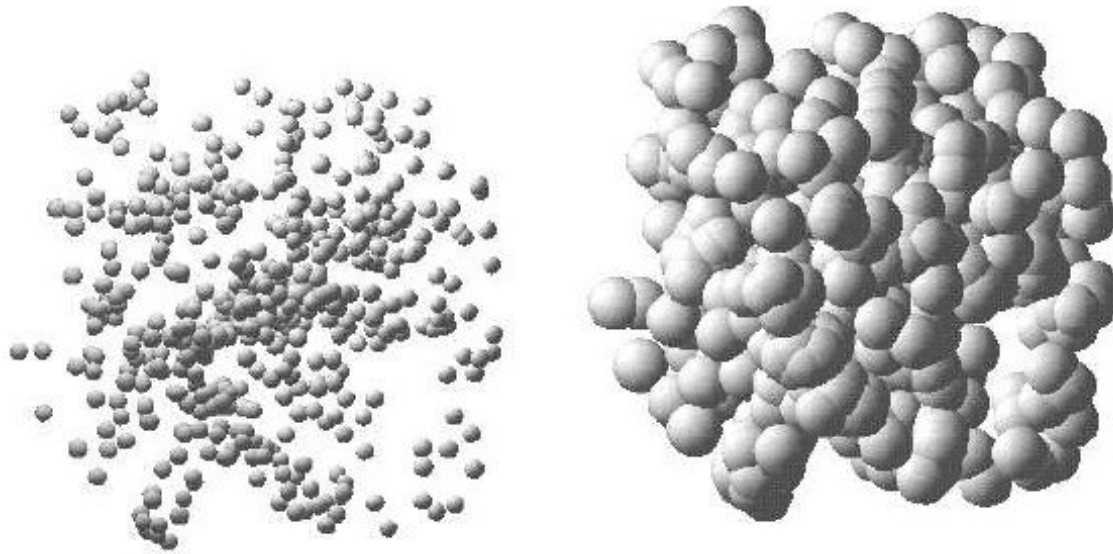
distribution à la puissance  $N$  pour la fonction de corrélation à  $N$  points. Ceci devient très rapidement ingérable (le catalogue SDSS par exemple contient de l'ordre de 400,000 galaxies) et plusieurs méthodes rapides ont été mises au point (voir Szapudi [37]).

Citons par exemple l'utilisation de codes en arbres (ou *kd-tree*) décrits dans Moore *et al.* [26]. Ces codes sont du même type que ceux utilisés pour les simulations N-Corps (décrits au chapitre 1.2.2) et permettent de ne pas prendre en compte chaque point individuellement mais plutôt en groupes compris dans un intervalle de distance donné. Une autre méthode courante consiste à utiliser les transformées de Fourier rapides (FFT) et c'est celle-ci que nous avons implémentée pour les calculs de corrélations effectués pendant la thèse. La fonction de corrélation pouvant être exprimée comme un produit d'autoconvolution du contraste de densité  $\delta$  (équation (3.3)), cela se traduit par un simple produit dans l'espace de Fourier. Les algorithmes des FFT étant très performants, un simple échantillonnage sur une grille suffisamment fine suivi d'une FFT, du calcul du produit des coefficients puis d'une FFT inverse permet d'obtenir un résultat en un temps  $t \propto N \log N$  où  $N$  est la taille de la grille. La méthode demande cependant une certaine quantité de mémoire, surtout si l'on veut corriger les effets de bords dus à la non périodicité de la distribution. Il existe aussi des versions hybrides entre FFT et codes en arbres qui semblent d'ailleurs être actuellement les plus efficaces pour le calcul de fonctions de corrélations d'ordre élevé. Cette méthode est utilisée dans Jenkins *et al.* [19]). Beaucoup d'autres méthodes existent encore comme les comptages en cellules (expliqués en détails dans Szapudi [37]) qui s'avèrent très adaptées aux mesures sur de petites échelles et quand le nombre de dimensions de l'espace est faible. La méthode de décomposition en base principale par transformée de Karhunen-Loève est aussi très utilisée et permet un calcul très précis du spectre de puissance lorsque la géométrie est complexe (voir Pope *et al.* [31]).

### Résultats obtenus et limites

La mesure des fonctions de corrélation et du spectre de puissance de la distribution des galaxies a sûrement été l'une des sources principales d'information de ces vingt dernières années. Les mesures du spectre de puissance tout d'abord ont permis la confirmation des modèles d'inflation. Récemment, les mesures de paramètres cosmologiques sur le CMB (chapitre 1.1.3) et dans les catalogues de galaxies grâce à la découverte de nouveaux algorithmes [31] ont aussi permis de contraindre fortement les modèles et même de confirmer des théories, avec notamment la mesure du pic acoustique des baryons (empreinte laissée dans le spectre de puissance par le couplage radiation/matière avant la recombinaison) dans le catalogue SDSS dans Eisenstein *et al.* [15].

Mais les problèmes principaux posés par le calcul des fonctions de corrélation sont certainement leur temps de calcul élevé et leur forte sensibilité aux effets de biais très complexes à corriger (déformation dans l'espace des décalages spectraux



**Fig. 3.3:** Illustration de l'ensemble  $K$  des sphères pour deux rayons donnés, à partir d'un catalogue discret de galaxies. Image extraite de Schmalzing *et al.* [35].

notamment). De plus, les fonctions de corrélations à  $N$  points ne contiennent qu'une partie de l'information sur la distribution (lorsque celle-ci n'est pas purement Gaussienne) et il est toujours possible de trouver des distributions très différentes ayant les mêmes fonctions de corrélation pour les ordres faibles (voir Querre *et al.* [33], figure 15 par exemple). Le plus souvent, il n'est malheureusement pas possible de calculer correctement les ordres élevés en raison de la complexité de la géométrie ou du nombre d'échantillons trop faible.

### 3.1.2 Fonctionnelles de Minkowsky

Afin d'étudier les structures de la matière, on souhaiterait pouvoir exprimer la grande quantité d'information contenue dans le champ de densité en fonction d'une petite quantité de paramètres significatifs. C'est bien sûr une tâche très délicate mais H. Minkowsky a montré en 1903 qu'il suffisait de trois grandeurs scalaires en 2D (quatre en 3D) pour extraire la totalité de l'information topologique du champ. De bonnes présentations du sujet peuvent être trouvées dans Schmalzing *et al.* [35] ou Mecke *et al.* [25].

#### Principe

Pour résumer, il s'agit, à partir d'un catalogue de galaxies (ou d'une simulation de matière noire), de remplacer chaque point par une sphère de rayon  $r$  donné (voir figure 3.3) et d'étudier l'ensemble  $K$  formé par l'union de ces sphères. Les fonctionnelles de Minkowsky  $M$  peuvent alors être définies comme étant les seules fonctionnelles de l'ensemble  $K$  des sphères telles que :

- Elles sont invariantes par translation ou rotation de  $K$ .
- Elles sont additives, c'est à dire que les fonctionnelles de Minkowsky de l'union de deux ensembles  $K$  et  $K'$  est la somme des fonctionnelles de Minkowsky de  $K$  et  $K'$  moins la fonctionnelle de Minkowsky de leur intersection.
- Elles sont continues dans le sens où la fonctionnelle de Minkowsky d'une approximation  $K_i$  de  $K$  tend vers la fonctionnelle de Minkowsky de  $K$  quand  $K_i$  tend vers  $K$  ( $K_i \rightarrow K \Rightarrow M(K_i) \rightarrow M(K)$ ).

On montre alors qu'il est possible de définir  $d + 1$  fonctionnelles de Minkowsky  $M_\mu$  dans un espace à  $d$  dimensions. Les  $M_\mu$  ne sont pas pratiques à utiliser mais en utilisant le volume des sphères  $\omega_\nu$  dans un espace de dimension  $\nu$ , il est possible de définir des quantités utilisées plus couramment dans la littérature :

$$V_\mu = \frac{\omega_{d-\mu}}{\omega_d} M_\mu \quad (3.9)$$

$$W_\mu = \frac{\omega_\mu \omega_d}{\omega_{d-\mu}} M_\mu \quad (3.10)$$

$$(3.11)$$

avec

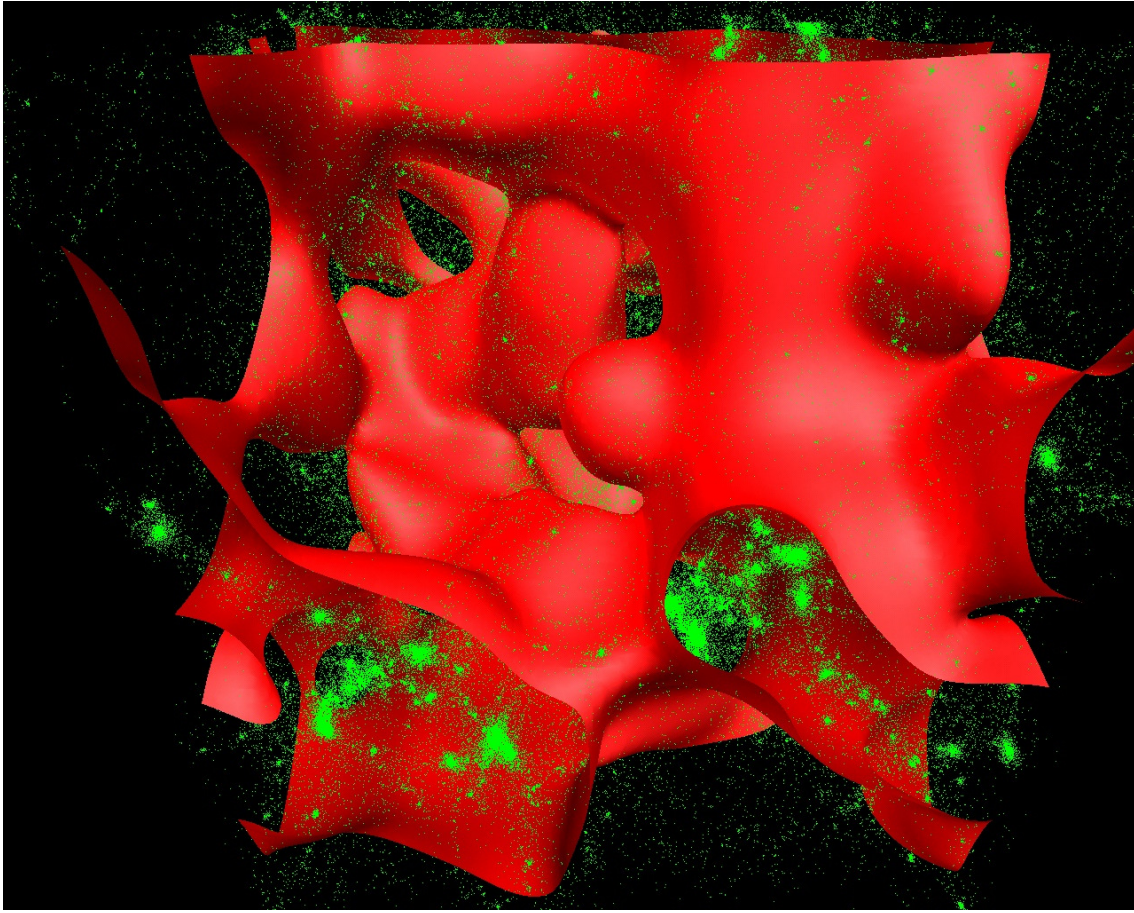
$$\omega_\nu = \frac{\pi^\nu / 2}{\Gamma(1 + d/2)}. \quad (3.12)$$

Dans un espace 3D, celles-ci peuvent s'exprimer en fonction de quantités géométriques facilement calculables : le volume  $V$ , l'aire  $A$ , la courbure moyenne  $H$  et la caractéristique d'Euler  $\chi$ . Par exemple,  $V_\mu = (V, A/6, H/3\pi, \chi)$ . La caractéristique d'Euler est une grandeur moins couramment utilisée que les autres mais peut être simplement définie comme étant le nombre de composantes (c'est à dire le nombre de régions séparées) moins le nombre de tunnels plus le nombre de cavités :

$$\chi = C_c - C_f + C_p - C_v. \quad (3.13)$$

Dans cette expression,  $C_h$  (resp.  $C_f$ ,  $C_m$  et  $C_v$ ) désigne le nombre de points critiques de type halos (resp. de type filament, mur et vide). Chacun de ces types peut simplement être identifié par le nombre de valeurs propres négatives du Hessien (la matrice des dérivées secondes, voir aussi le chapitre 3.2.2).

De nos jours, les moyens de calculs étant plus performants, il n'est plus nécessaire d'utiliser des sphères mais l'on peut simplement définir les fonctionnelles de Minkowsky à partir des isosurfaces du champ de densité Sheth *et al.* [36], il suffit alors de mesurer leur volume, surface, courbure et caractéristique d'Euler (voir figure 3.4). Le calcul du champ de densité introduit cependant une fréquence d'échantillonnage (c'est-à-dire une distance minimale en dessous de laquelle on considère le champ comme constant) ainsi qu'une distance de lissage afin que ce champ soit continu et non contaminé par le bruit dû à la distribution discrète des particules. Ces deux sujets seront abordés par la suite dans le chapitre 3.3.1.



**Fig. 3.4:** Une isosurface du champ de densité d'une simulation de matière noire  ${}^3\text{LCDM}_{50}^{256}$ . En vert est représenté un sous-échantillon des particules de matière noire et en rouge une surface d'isodensité servant à calculer la valeur des fonctionnelles de Minkowsky à une densité donnée. Ces dernières dépendent de l'aire, du volume, de la courbure et de la caractéristique d'Euler de la surface rouge.

### Ensembles d'excursions et seuil de percolation

Une excursion dans un champ  $\rho(\mathbf{r})$  est définie comme l'ensemble des points tels que  $\rho(\mathbf{r}) = \rho_e$ . La figure 3.4 représente une excursion dans le champ de densité d'une simulation de matière noire. L'étude d'ensembles d'excursions permet de fournir des informations importantes sur la nature du champ étudié, au même titre que les fonctionnelles de Minkowsky. Il est ainsi possible d'utiliser des indicateurs topologiques pour décrire l'évolution des conditions initiales gaussiennes sous l'influence des instabilités gravitationnelles [10]. Si l'on définit le contraste de densité de la même manière qu'en (3.2) :

$$\delta(\mathbf{r}) = \frac{\rho(\mathbf{r}) - \langle \rho(\mathbf{r}) \rangle_{\mathbf{r}}}{\langle \rho(\mathbf{r}) \rangle_{\mathbf{r}}}, \quad (3.14)$$

alors l'étude des ensembles d'excursion  $E_{\delta_{\text{TH}}}^+ = \{\mathbf{r} \mid \delta(\mathbf{r}) \geq \delta_{\text{TH}}\}$  et  $E_{\delta_{\text{TH}}}^- = \{\mathbf{r} \mid \delta(\mathbf{r}) \leq \delta_{\text{TH}}\}$  permet de caractériser les structures à grande échelle de la distribution de ma-

tière dans l’Univers. Une valeur particulière de  $\delta_{TH}$  appelée seuil de percolation  $\delta_c$  apparaît d’ailleurs particulièrement intéressante. Par définition, pour cette valeur, les ensembles d’excursions forment en effet une structure connectée infinie.

Tout comme les fonctionnelles de Minkowsky, le seuil de percolation était tout d’abord calculé en remplaçant les galaxies de la distribution par des sphères de rayons variables et en trouvant la chaîne la plus longue formée par des sphères connectées (voir Bhavsar and Barrow [6] par exemple) mais d’autres auteurs ont aussi utilisé l’algorithme friend-of-friend (voir annexe ??) en faisant varier la longueur de lien pour obtenir une courbe de percolation (i.e la longueur de la plus grande structure en fonction de  $\delta_{TH}$ , voir par exemple Dekel and West [11]). Tous s’accordent cependant à dire que la mesure du seuil de percolation seul n’est pas très discriminante pour le type de structures formées par la matière et de nos jours le seuil de percolation est plutôt utilisé en tant que mesure complémentaire aux fonctionnelles de Minkowsky plus générales (la caractéristique d’Euler constituant une mesure équivalente à la courbe de percolation).

### Résultats obtenus

La principale application des fonctionnelles de Minkowsky est sans aucun doute la mise en évidence des déviations à la gaussianité de la distribution de matière. La forme des fonctionnelles de Minkowsky en fonction du niveau de densité choisi est en effet connue de manière analytique et peu même être prédite pour de légères déviations non-gaussiennes (voir Matsubara [24]). Un autre avantage sur les fonctions de corrélation à deux points est qu’à une distribution donnée peut être associé un jeu unique de fonctionnelles de Minkowsky qui contient donc toute l’information. La caractéristique la plus utilisée est le génus et sa mesure sur le catalogue CfA par exemple par Vogeley *et al.* [39] lui permet de mesurer les déviations à la gaussianité sur des échelles de moins de  $10h^{-1}$  Mpc mais aussi de montrer qu’un modèle de type CDM standard ( $\Omega_m = 1$ ,  $\Omega_\Lambda = 0$ ) est incompatible avec les résultats obtenus (rappelons que ce modèle était considéré comme le plus probable à l’époque). De la même manière, Colley [9] et Canavezes *et al.* [8] mettent en évidence la compatibilité des distributions mesurées dans les catalogues LCRS et PSCz respectivement avec une distribution initiale gaussienne de la matière ainsi que les déviations aux échelles inférieures à  $10h^{-1}$  Mpc, dûes aux effets gravitationnels non linéaires. Ces effets non linéaires dont la forme est prédite analytiquement par Matsubara [24] ont d’ailleurs aussi permis à Park *et al.* [28] de contraindre différents modèles de biais entre la distribution galactique et de matière noire en utilisant le catalogue SDSS.

Les fonctionnelles de Minkowsky constituent donc d’excellents indicateurs de gaussianité et permettent aussi de caractériser la nature d’une distribution (filamenteuse, de type éponge ou “meatball” ...) même lorsque la géométrie de la distribution est complexe (comme celle du SDSS par exemple). Il s’agit cependant de caractéristiques générales de la distribution et il est impossible en utilisant les fonctionnelles de Minkowsky d’identifier des régions caractéristiques de la distribution comme les



filaments de matière par exemple.

### 3.1.3 Arbre à recouvrement minimal

La méthode de l'arbre à recouvrement minimal (ou MST pour Minimal spanning Tree) est inspirée de la théorie des graphes et son utilisation en cosmologie a pour la première fois été suggérée dans Barrow *et al.* [3] qui l'a appliqué à l'identification des motifs dans la distribution des galaxies (filaments, murs ...).

#### Principe

Pour une distribution donnée de points, trouver le MST revient à construire l'arbre reliant l'ensemble de points et ayant les propriétés d'être de longueur totale minimale et de ne jamais faire de boucle (i.e. en le parcourant linéairement, on ne retombe jamais sur le même point). Il existe principalement deux méthodes de calcul pour extraire l'arbre d'une distribution : l'algorithme de Kruskal (1956) et celui de Prim (1957), ce dernier étant le plus couramment utilisé. La méthode est la suivante :

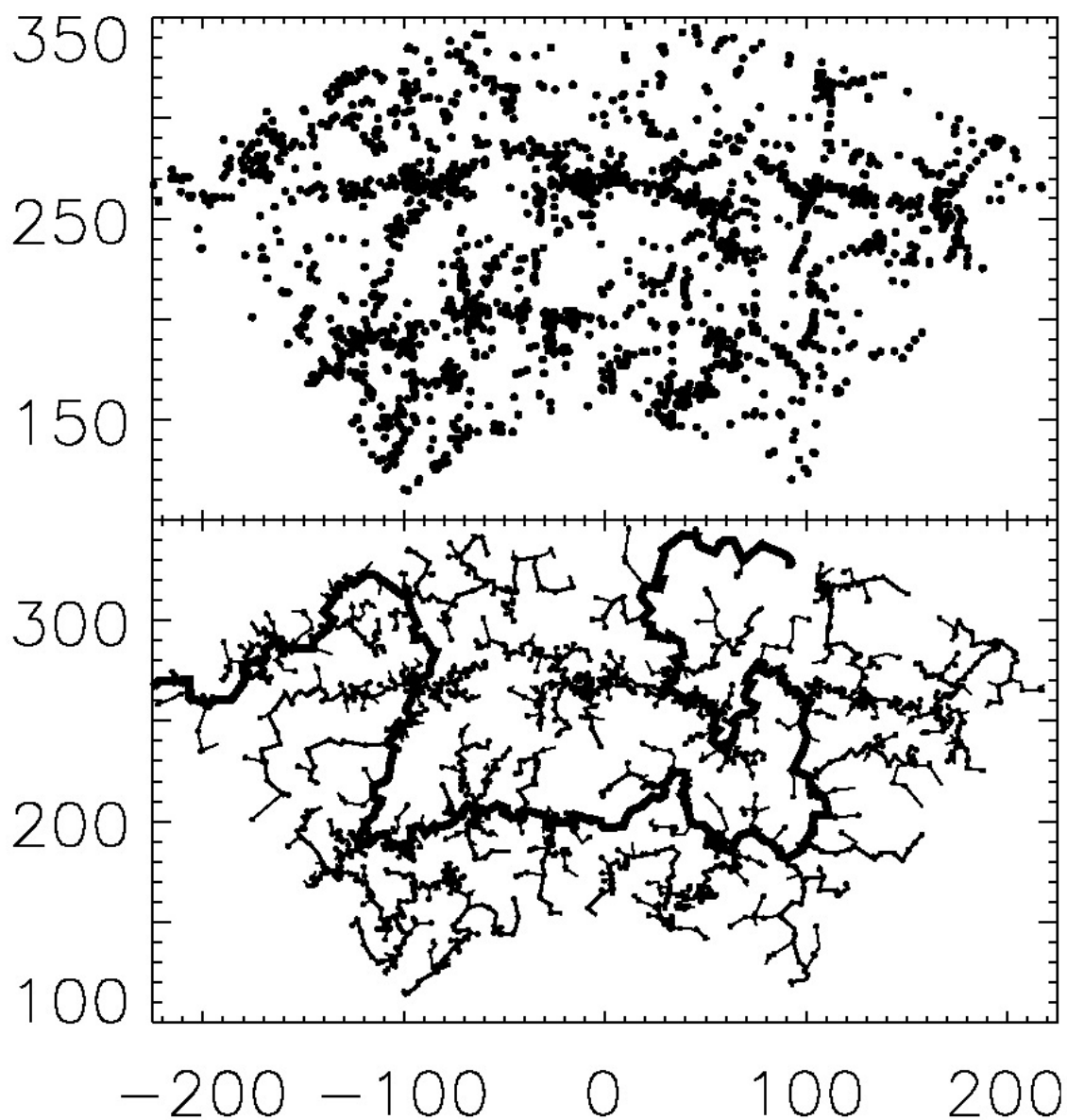
1. On choisit un point de départ
2. Il est relié par un segment au point le plus proche formant un arbre partiel
3. De ces deux points, celui qui a le plus proche voisin hors de l'arbre lui est relié par un nouveau segment
4. L'opération est répétée jusqu'à ce que tous les points soient dans l'arbre
5. Un arbre est calculé en prenant chacun des points de la distribution comme point de départ, et l'arbre le plus court est gardé.

Appliqué directement, le temps de calcul d'un tel algorithme est  $t \propto N^3$  où  $N$  est le nombre de points mais peut être réduit à  $N \log(N)$  en utilisant judicieusement un système de pile afin de ne jamais rechercher plusieurs fois les plus proches voisins d'un même point. L'implémentation d'un tel algorithme est cependant relativement complexe.

La figure 3.5 présente le résultat obtenu pour la distribution des galaxies du catalogue "Las Campanas" par Doroshkevich *et al.* [13]. Le MST semble effectivement tracer la structure filamenteuse de la distribution mais puisque, par définition, l'arbre contient *toutes* les galaxies, l'identification des halos et des filaments individuellement est difficile. Certaines méthodes existent pour réaliser cet objectif et consistent par exemple à couper tous les segments de l'arbre plus grand qu'une certaine longueur. Cela permet en pratique de séparer les morceaux d'arbre liés gravitationnellement. Si il existe un segment de longueur inférieur à  $l_{\text{TH}}$ , cela signifie en effet que la densité locale est :

$$n_{\text{TH}} \geq \left( \frac{4\pi}{3} l_{\text{TH}}^3 \right)^{-1}. \quad (3.15)$$

Cette formule est couramment utilisée par les algorithmes de recherche de groupes de type friend-of-friend (voir Sahni and Coles [34] et l'annexe ??). Il s'agit d'ailleurs



**Fig. 3.5:** La distribution des galaxies du catalogue “Las Campanas” (en haut) et l’arbre de recouvrement minimal correspondant calculé par Doroshkevich et al. [13]. Il est clair visuellement que l’arbre trace les filaments observés et le filament principal en gras représente le plus long chemin (en nombre de segments) qu’il est possible de parcourir dans l’arbre (ce chemin est appelé tronc)..

d'un des avantages du MST : il contient l'ensemble des informations de percolation d'une distribution (i.e. comment les structures de densité supérieure à un seuil donné se comportent en fonction de la valeur de ce seuil, voir la section 3.1.2 ). Une fois calculé, il est possible d'isoler en temps réel (par simple parcours de l'arbre) les amas et filaments dont la densité dépasse un seuil donné. L'information contenue dans le MST est, par conséquent, au moins équivalente à toute l'information contenue dans une courbe de percolation (voir Fairall *et al.* [16], Bhavsar and Splinter [7]).

### Résultats obtenus

A ce jour, la technique du MST a permis d'obtenir des résultats intéressants en quantifiant par exemple le pourcentage de galaxies présentes dans les filaments et les murs du catalogue SDSS DR1 [12]. La méthode utilisée consiste à calculer l'histogramme des longueurs de segments du MST, qui peut être ajustée par des profils spécifiques selon la dimensionnalité de l'arbre (une seule dimension dans les filaments, deux dimensions dans les murs). Il est ainsi montré que la proportion de galaxies dans les deux types de structure est à peu près égale mais que les murs sont largement majoritaires dans les régions de haute densité et les filaments dans celle de basse densité. Dans Krzewina and Saslaw [22], plusieurs méthodes de quantification des propriétés du MST sont proposées et sont utilisées pour démontrer la propension des galaxies lumineuses à se concentrer plus que les galaxies faibles (voir aussi Adami and Mazure [1]).

Le principal avantage de l'arbre de recouvrement minimal reste sa capacité à identifier et à localiser différents types de structures dans la distribution des galaxies permettant ainsi de faire des mesures de quantités dépendamment du type d'environnement (Fairall *et al.* [17] par exemple appliquent le MST à l'identification des structures dans l'univers local). La méthode souffre cependant de plusieurs défauts. Tout d'abord, le MST étant calculé sur un ensemble discret de galaxies, sa structure est forcément irrégulière car elle suit la distribution des galaxies. Ensuite, sa définition elle-même fait qu'il est assez compliqué d'établir un lien avec des propriétés physiques, la définition d'un filament par exemple restant ad-hoc. Enfin, il est non local, c'est à dire que les propriétés du MST en un point dépendent de l'ensemble de la distribution, en enlevant une seule galaxie on peut changer l'ensemble de l'arbre. L'étude du MST ne peut, par conséquent, que rester statistique. Il semble de plus quasiment impossible de modéliser mathématiquement les caractéristiques du MST en raison de la nature même de sa définition qui ne s'y prête pas du tout.

## 3.2 Le squelette

Le concept de squelette a pour la première fois été introduit en 2003 par D. Novikov, S. Colombi et O. Doré [27]. Dans cet article intitulé "Skeleton as a probe of the cosmic web : the 2D case", les auteurs présentent le formalisme, l'implémentation ainsi que quelques applications dans le cas bidimensionnel. Dans cette thèse,

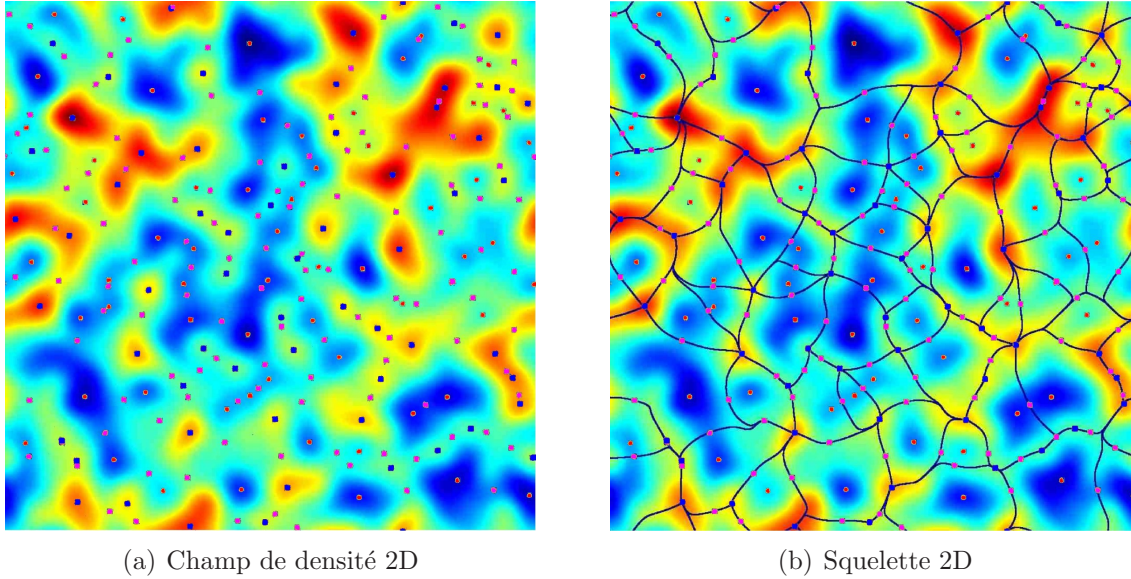
le travail principal a été de transposer les idées au cas tri-dimensionnel et d'implémenter un programme les mettant en pratique. L'idée directrice est de trouver une méthode, si possible ne nécessitant l'introduction d'aucun paramètre arbitraire, permettant l'extraction de la structure filamenteuse et de ses propriétés. C'est ce que fait déjà le minimal spanning tree (MST), cependant les inconvénients de cette méthode ont fortement limité son utilisation (voir le chapitre 3.1.3). Le squelette propose une définition mathématique simple des filaments à partir du champ de densité, basée sur la théorie de Morse (voir Colombi *et al.* [10] ou Jost [20]). Cette définition lui permet de pallier à la plupart des inconvénients du MST, et son principal défaut notamment (i.e son irrégularité intrinsèque due à la nature discrète de la distribution à partir de laquelle il est calculé).

Grâce à cet atout, le calcul de grandeurs analytiques est rendu possible (ce qui était quasiment infaisable avec le MST). À partir de cette définition, de nombreuses quantités caractéristiques de la distribution de matière (et donc de la manière dont l'Univers a évolué) peuvent être calculées. Parmi elles, la longueur totale (ou moyenne) des filaments, l'évolution de la longueur des filaments en fonction du seuil de densité, leur courbure ou leur torsion ... Une autre application prometteuse est le calcul des propriétés de la matière (baryonique ou non) en fonction de sa situation géographique. En effet les filaments sont en quelque sorte le lieu où se passent la plus grande partie des interactions gravitationnelles et pouvoir les modéliser en tant qu'entités à part entière est très intéressant.

### 3.2.1 Vrai squelette

La figure 3.6(a) représente un champ de densité gaussien  $\rho(\mathbf{r})$  dans le cas bi-dimensionnel (i.e.  $\mathbf{r} \in \mathbb{R}^2$ ). L'échelle de couleur varie de bleu (peu dense) à rouge (très dense) et les extrema locaux du champ sont représentés par des points. Soit  $\lambda_i$  les valeurs propres du Hessien  $\mathcal{H} = \partial^2 \rho / \partial r_i \partial r_j$ , telles que  $\lambda_i \geq \lambda_j$  si  $i < j$ . Alors, les maxima (bleu), les minima (rouge) et les points selles (rose) sont les points du champ où  $\nabla \rho = \frac{\partial \rho}{\partial \mathbf{r}} = 0$  et où  $0 > \lambda_1 \geq \lambda_2$ ,  $\lambda_1 \geq 0 \geq \lambda_2$  et  $\lambda_1 \geq \lambda_2 \geq 0$  respectivement. Si l'on demandait à une personne quelconque de tracer les filaments du champ de densité de la figure 3.6(a), celle-ci tracerait sûrement le chemin de la figure 3.6(b) : l'ensemble des lignes reliant les maxima du champ de densité entre eux et entourant les zones sous-denses. Intuitivement, le lieu géométrique des filaments (que nous appellerons squelette) est un ensemble de courbes partant des maxima du champ et les reliant entre eux en passant par les points selles. Deux maxima sont de plus toujours reliés en passant par un seul et unique point selle et entre un maximum et un point selle, le squelette doit obéir à une équation dépendant de  $\rho$  et de ses dérivées le forçant à "rester sur les crêtes".

Une autre manière de voir les choses serait de définir les "void patch" et les "peak patch". Les "void patch" (resp. les "peak patch") sont par définition les zones du champ de densité qui convergent vers un même minimum (resp. maximum) lorsque

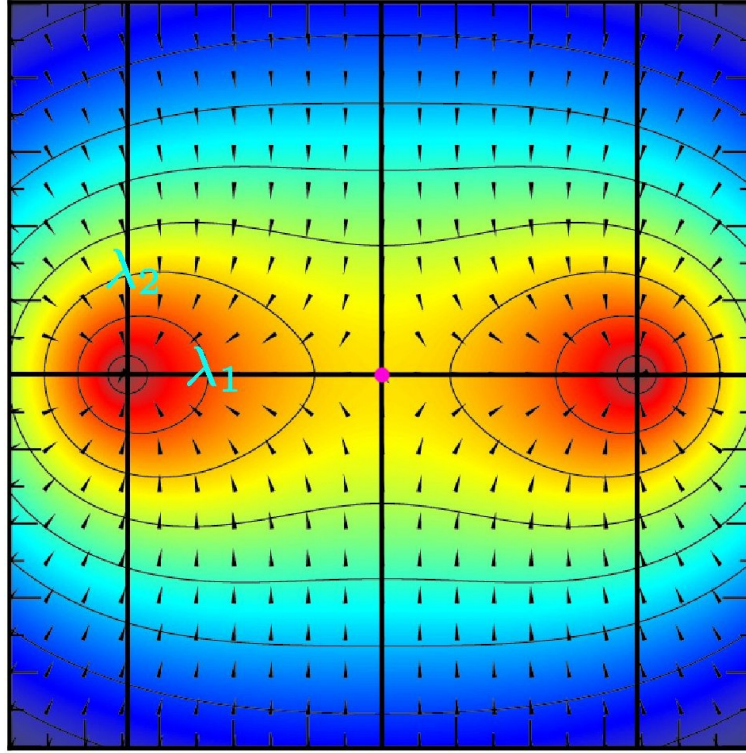


**Fig. 3.6:** Illustration d'un champ de densité gaussien 3.6(a) et de son squelette 3.6(b). L'échelle de couleur varie de bleu (faible densité) à rouge (densité élevée). Sur la figure 3.6(a) sont représentés les extrema du champ, à savoir les maxima (bleu), les minima (rouge) et les points selles (rose). Sur la figure 3.6(b), le vrai squelette de ce champ est tracé, ce dernier reliant les maxima du champ en passant uniquement par les points selles. L'accord avec ce que l'on devine intuitivement comme étant les filaments est remarquable. L'image est extraite de Novikov et al. [27].

l'on suit les lignes de champ dans la direction opposée au gradient (resp. dans la direction du gradient). Pour mieux se rendre compte, il suffit de se représenter le champ de densité 2D comme un paysage de montagnes et de vallées. Si on lâchait une balle à un endroit donné, elle roulerait dans une vallée pour finalement s'arrêter dans un minimum local. Les "void patch" seraient alors l'ensemble des points tels que, si on lâche une balle, elle se retrouvera finalement sur un même minimum local. Sur la figure 3.6(b), les "void patch" sont les zones délimitées par les courbes en noir. Si l'on définit le squelette comme étant les limites externes des "void patch", cela nous permet de compléter notre définition. Ces limites relient bien les maxima entre eux en passant par les points selles. De plus, si l'on considère un point sur cette limite et que l'on suit le gradient, on ne va jamais tomber vers un minimum ou l'autre (par définition du "void patch") et donc converger vers un maximum du champ. La méthode pour tracer le squelette consiste à partir de l'ensemble des points selles du champ et à suivre le gradient jusqu'à converger vers un maximum local, c'est à dire résoudre l'équation :

$$\mathbf{v} \equiv \frac{\partial \mathbf{r}}{\partial t} = \nabla \rho. \quad (3.16)$$

Il est intéressant à ce stade d'essayer de prédire quelques propriétés du squelette.



**Fig. 3.7:** *Champ de densité 2D résultant de la somme de deux gaussiennes. Les vecteurs représentent la direction du gradient de ce champ, le point violet est un point selle et en rouge apparaissent deux extrema locaux. Les axes noirs représentent les directions des vecteurs propres du Hessien au niveau des extrema,  $\lambda_i$  étant les valeurs propres correspondantes sachant que  $\lambda_1 > \lambda_2$ . Résoudre l'équation (3.16) revient à suivre le champ de vecteur.*

La figure 3.7 sert d'illustration et représente un champ de densité de la forme :

$$\rho(\mathbf{r}) = \exp(-\|\mathbf{r} - \mathbf{r}_0\|) + \exp(-\|\mathbf{r} - \mathbf{r}_1\|). \quad (3.17)$$

Sur la figure, le champ de vecteur normalisé  $\nabla\rho$  est représenté par des flèches noires et résoudre l'équation (3.16) revient à tracer l'ensemble des trajectoires suivant la direction de ces flèches. Ce champ de densité comporte trois extrema : un point selle au centre et deux maxima, sur la gauche et la droite. Les axes noirs représentent les directions des vecteurs propres du Hessien  $\mathcal{H}$ , l'axe noté  $\lambda_1$  représentant l'axe de valeur propre maximale. Au voisinage des extrema,  $\nabla\rho = 0$  et il est possible en toute généralité de faire un développement du champ du type (voir aussi le chapitre 3.2.2) :

$$\rho(\mathbf{r} + d\mathbf{r}) = \rho(\mathbf{e}_0) + d\mathbf{r}^T \mathcal{H} d\mathbf{r} \quad (3.18)$$

ce qui signifie que les propriétés déduites dans le cas du champ (3.17) autour des extrema sont vraies dans le cas général. Ainsi on remarquera les points suivants :

- Les extrema du champ constituent les nœuds du squelette où peuvent se connecter plusieurs branches.

- Les connexions du squelette aux extrema se font principalement le long du vecteur propre du Hessien de valeur propre maximale. Toutes les trajectoires (sauf celles dans l’axe du deuxième vecteur propre) arrivent au maximum parallèlement à cette direction.
- Seule une branche de squelette passe par un point selle, toute branche provenant d’un autre point selle “évitera” un autre point selle.

On peut imaginer des cas particuliers où toutes ces conditions ne sont pas respectées mais ce sont toujours des cas pathologiques où un changement infinitésimal du champ  $\rho$  résoudre le problème [20]. Deux hypothèses sont cependant nécessaires quant à la nature du champ  $\rho$ . Premièrement,  $\rho$  est continuellement dérivable une fois. Deuxièmement,  $\nabla\rho$  doit s’annuler en un ensemble *discret* de points. Cela signifie que si  $\rho$  est défini sur  $E$ , alors il n’existe pas de sous ensemble  $S \subset E$  tels que  $\rho(\mathbf{r})$  soit constant pour  $\mathbf{r} \in S$ .

Nous venons d’obtenir une nouvelle définition du lieu géométrique des filaments (i.e. le squelette) qui a l’avantage d’être simple et parfaitement définie d’un point de vue mathématique. Elle présente cependant aussi quelques inconvénients auxquels il serait bon de pallier :

- Cette définition est *non locale* : pour reconstituer le squelette, il faut partir des points selles et rejoindre les maxima en suivant le gradient du champ (équation (3.16)). Ainsi si l’on définit un champ  $\rho$  sur un espace  $E$ , le squelette calculé sur un sous-espace  $S \subset E$  ne correspondra pas à l’intersection du squelette calculé sur  $E$  et de  $S$ . C’est un problème car on souhaiterait appliquer la méthode à des catalogues de galaxies dont la géométrie est complexe et l’on aimerait que le squelette obtenu ne dépende pas de la géométrie du catalogue mais uniquement de son contenu.
- L’équation (3.16) est simple mais sa résolution numérique est très complexe. En effet, elle suppose d’être capable de détecter l’ensemble des extrema d’une distribution sans commettre d’erreur. Ne pas détecter un maximum ou un point selle signifie ne pas détecter un ensemble de filaments or la détection efficace et robuste des extrema sur une grille n’est pas facile. De plus, la convergence des lignes du squelette n’est pas assurée car par définition le gradient du champ est faible à proximité des extrema, rendant la résolution numérique de l’équation (3.16) de plus en plus lente et hasardeuse. Un algorithme dans le cas 2D a été développé par S. Colombi [27], mais celui-ci connaît encore quelques problèmes de convergence et sa généralisation à 3D n’a pas encore été faite.

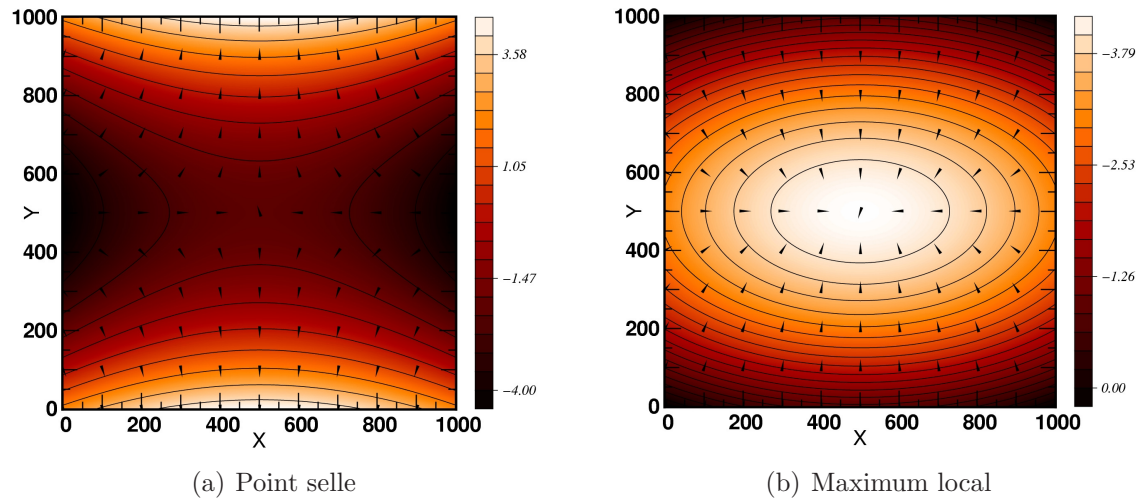
L’idée pour résoudre ces problèmes est de calculer une approximation *locale* du squelette qui est présentée dans la prochaine section. A partir d’ici, nous appellerons le squelette défini précédemment *vrai* squelette et celui défini par la suite squelette

*local* ou, par abus de langage, squelette.

### 3.2.2 Approximation locale

#### Cas d'un espace 2D

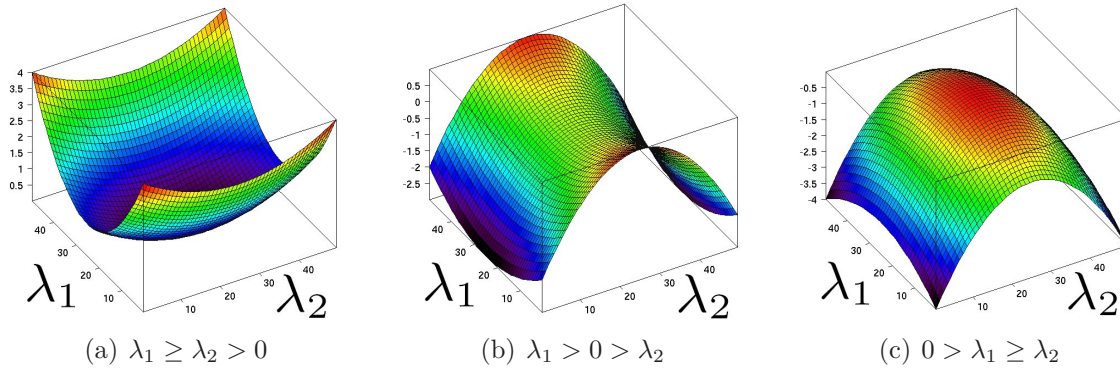
Afin de rendre les choses plus concrètes, nous allons dans un premier temps définir le squelette local dans le cas d'un espace 2D. La figure 3.8 présente une approximation de Taylor à l'ordre 2 (voir l'équation (3.18)) du champ de densité autour d'un maximum (3.8(b)) et autour d'un point selle (3.8(a)).



**Fig. 3.8:** *Approximation locale du champ autour d'un point selle (3.8(a)) et d'un maximum (3.8(b)). Les courbes noires représentent des isocontours du champ, c'est à dire des points pour lesquels  $\rho = C$ , avec  $C \in \mathbb{R}$  une constante. Par définition, le gradient du champ est orthogonal aux isocontours (représenté par les flèches noires). Ici, le champ est représenté dans le repère orthonormé des vecteurs propres du Hessien  $\mathcal{H}$ .*

Sur ces figures, les contours d'isodensité sont représentés par les courbes noires. Par définition, le gradient du champ de densité  $\nabla\rho$  (flèches noire sur les figures) est en tout point orthogonal aux contours d'isodensité (i.e. la valeur du champ ne varie pas le long d'un isocontour). De plus, le champ étant représenté par un polynôme d'ordre 2, son Hessien  $\mathcal{H}$  est une constante et ses vecteurs propres, qui représentent les axes principaux de courbure, sont donc identiques dans tout l'espace. La direction de ces vecteurs propres est donnée par les axes des coniques que définissent les isocontours du champ (ellipses pour la figure 3.8(b) et hyperboles pour la figure 3.8(a)). Ainsi, pour la figure 3.8 qui est représentée dans le repère des vecteurs propres de  $\mathcal{H}$ , ce sont les axes  $X$  et  $Y$ . Il s'agit alors de trouver une définition du squelette local qui soit équivalente à la définition (3.16) dans le cadre de notre approximation au second ordre.





**Fig. 3.9:** Représentation d'un champ 2D en fonction du signe des valeurs propres du Hessien  $\mathcal{H}$ . Les valeurs propres de  $\mathcal{H}$ , qui sont proportionnelles à l'inverse du carré du rayon de courbure, déterminent la présence ou non de "crête" suivant leur signe. Ainsi, alors que le champ présente une crête dans la direction du vecteur propre principal (associé à  $\lambda_1$ ) lorsque  $\lambda_2 < 0$ , ce n'est plus le cas si  $\lambda_2 > 0$  (il s'agit d'une vallée dans ce cas). Afin de respecter la définition intuitive d'un filament, il convient donc de sélectionner uniquement les régions où  $\lambda_2 > 0$  et où le gradient est vecteur propre de  $\mathcal{H}$  avec pour valeur propre  $\lambda_1$ .

Si le maximum local représente un halo de matière noire, on aimerait alors que les filaments soient dans la direction où la densité reste la plus élevée possible (décroisse le moins rapidement). Pour la figure 3.8(b), le filament serait l'axe horizontal qui passe par le maximum du champ et pour la figure 3.8(a), ce serait l'axe vertical qui passe par le centre. Une manière d'imposer cette condition serait de dire que le gradient (dans l'espace 2D) du champ de densité doit être de norme *minimale* le long d'un isocontour. On souhaite effectivement choisir la direction vers laquelle la densité reste la plus élevée possible donc, pour l'ensemble des points de même densité, on sélectionnera celui autour duquel la densité sera la moins susceptible de diminuer rapidement. Une autre manière de voir cela consiste à se représenter le champ comme un paysage de montagne où le maximum local serait un sommet. Suivre un filament reviendrait alors à suivre une crête, c'est à dire la direction où la pente (i.e. le gradient) est aligné avec l'axe *principal* de courbure (i.e. le vecteur propre de valeur propre maximale). Il ne reste alors qu'à imposer une condition pour se restreindre aux crêtes (et éliminer les "vallées"), à savoir que la plus petite des valeurs propres doit être négative. En termes plus mathématiques, si  $\lambda_1$  et  $\lambda_2$  sont valeurs propres de  $\mathcal{H}$  et que  $\lambda_1 > \lambda_2$ , on chercherait alors l'ensemble des points tels que  $\nabla\rho$  soit vecteur propre de  $\mathcal{H}$  de valeur propre  $\lambda_1$ , tout en imposant  $\lambda_2 < 0$  (voir la figure 3.9).

Définissons le squelette local *total* comme l'ensemble des points où le gradient du champ  $\nabla\rho(x, y)$  est extrémal le long d'un isocontour  $\mathbf{r}(u) = (r_1(u), r_2(u))$ . Par définition, le long de cet isocontour,

$$\frac{\partial\rho}{\partial u} = \frac{\partial\rho}{\partial r_1} \frac{\partial r_1}{\partial u} + \frac{\partial\rho}{\partial r_2} \frac{\partial r_2}{\partial u} = 0 \quad (3.19)$$

et le fait que le gradient soit extrémal sur  $\mathbf{r}(u)$  s'écrit :

$$\frac{\partial}{\partial u} (\|\nabla\rho(\mathbf{r}(u))\|^2) = 0. \quad (3.20)$$

L'ensemble des points du squelette *total* est donc donné par :

$$\mathcal{S} \equiv \frac{\partial}{\partial u} \left( \frac{\partial\rho}{\partial x}(\mathbf{r}(u))^2 + \frac{\partial\rho}{\partial y}(\mathbf{r}(u))^2 \right) = 0 \quad (3.21)$$

$$\Leftrightarrow \frac{\partial\rho}{\partial x} \frac{\partial}{\partial u} \frac{\partial\rho}{\partial x} + \frac{\partial\rho}{\partial y} \frac{\partial}{\partial u} \frac{\partial\rho}{\partial y} = 0. \quad (3.22)$$

En utilisant (3.19), il est alors possible de montrer que :

$$\begin{aligned} \mathcal{S} &\equiv \frac{\partial\rho}{\partial x} \frac{\partial\rho}{\partial y} \left( \frac{\partial^2\rho}{\partial x^2} - \frac{\partial^2\rho}{\partial y^2} \right) \\ &+ \frac{\partial^2\rho}{\partial x\partial y} \left( \left[ \frac{\partial\rho}{\partial y} \right]^2 - \left[ \frac{\partial\rho}{\partial x} \right]^2 \right) = 0. \end{aligned} \quad (3.23)$$

Nous obtenons une équation locale ne dépendant que des dérivées première et seconde du champ. Cependant, cette équation est celle du squelette *total* et sa solution est l'ensemble des points pour lesquels le gradient est *extrémal* sur son isocontour. Imposer la condition que le gradient soit minimal n'est pas facile. Une simple réécriture de  $\mathcal{S}$  peut cependant nous aider à y parvenir, ainsi :

$$\begin{aligned} \mathcal{S} &\equiv \frac{\partial\rho}{\partial y} \left( \frac{\partial\rho}{\partial x} \frac{\partial^2\rho}{\partial x^2} + \frac{\partial\rho}{\partial y} \frac{\partial^2\rho}{\partial xy} \right) \\ &- \frac{\partial\rho}{\partial x} \left( \frac{\partial\rho}{\partial y} \frac{\partial^2\rho}{\partial y^2} + \frac{\partial\rho}{\partial x} \frac{\partial^2\rho}{\partial xy} \right) = 0. \end{aligned} \quad (3.24)$$

ce qui peut être mis sous la forme plus compacte :

$$\mathcal{S} \equiv \det(\mathcal{H}\nabla\rho, \nabla\rho) = 0. \quad (3.25)$$

Nous retrouvons la deuxième définition (i.e. que le gradient doit être vecteur propre du Hessien) qui est parfaitement équivalente à la première. Celle-ci présente cependant à ce stade un avantage certain : la condition à imposer pour obtenir le squelette se ramène à une simple comparaison des valeurs propres du Hessien. Il suffit donc de rajouter les conditions suivantes à l'équation (3.23) :

$$\begin{cases} \lambda_2 < 0, \\ \mathcal{H}\nabla\rho = \lambda_1 \nabla\rho \end{cases} \quad (3.26)$$

### Cas d'un espace 3D

Le but du squelette est d'extraire la structure filamentaire de la distribution de matière. Bien qu'une méthode pour le faire en 2D soit d'un intérêt certain, c'est en 3D que les applications sont les plus nombreuses. En partant de la même définition que précédemment, nous allons dériver les équations définissant le squelette. Nous voulons trouver l'ensemble des points tels que le gradient 3D soit extrémal sur une *surface* d'isodensité. Par la suite nous adopterons la convention de notation  $\mathbf{r} = (r_1, r_2, r_3)$  et le vecteur  $\mathbf{r}_i$  sera le  $i^{\text{ème}}$  de la base orthonormée, c'est à dire tel que  $(r_i)_j = 0$  pour  $i \neq j$  et  $(r_i)_i = 1$ .

Soit  $(u, v)$  un système de coordonnées paramétrant une isosurface  $I$  du champ  $\rho(r)$  telle que :

$$I(u, v) = (r_1(u, v), r_2(u, v), r_3(u, v)). \quad (3.27)$$

Par définition d'une isosurface nous avons :

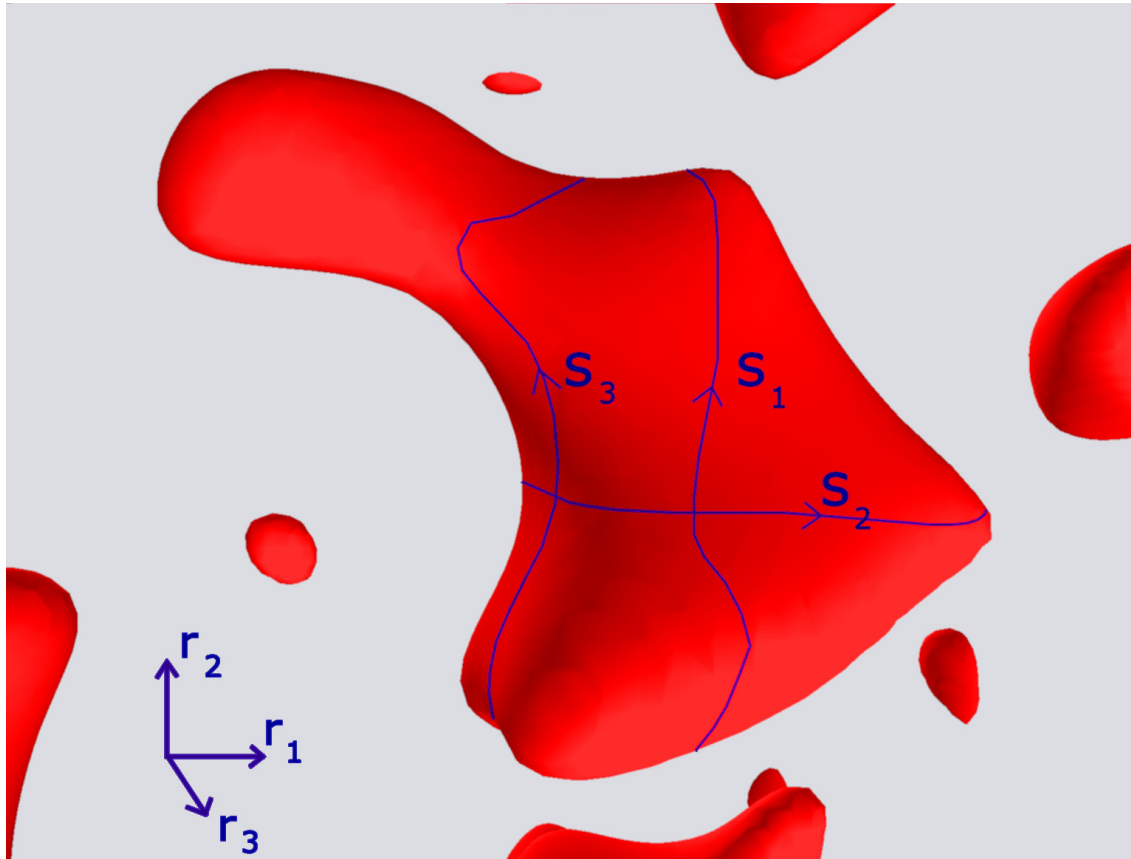
$$\begin{cases} \frac{\partial \rho}{\partial u} = \frac{\partial \rho}{\partial r_1} \frac{dr_1}{du} + \frac{\partial \rho}{\partial r_2} \frac{dr_2}{du} + \frac{\partial \rho}{\partial r_3} \frac{dr_3}{du} = 0 \\ \frac{\partial \rho}{\partial v} = \frac{\partial \rho}{\partial r_1} \frac{dr_1}{dv} + \frac{\partial \rho}{\partial r_2} \frac{dr_2}{dv} + \frac{\partial \rho}{\partial r_3} \frac{dr_3}{dv} = 0. \end{cases} \quad (3.28)$$

De plus, nous voulons que le gradient du champ soit extrémal sur cette isosurface :

$$\begin{cases} \frac{d}{du}(\|\nabla \rho\|^2) = 0 \\ \frac{d}{dv}(\|\nabla \rho\|^2) = 0. \end{cases} \quad (3.29)$$

Il ne reste plus qu'à utiliser les équations (3.28) et (3.29) afin de dériver l'équation du squelette, similairement à ce qui a été fait précédemment. Une différence importante existe cependant entre les deux cas : alors qu'il est facile de paramétrer une courbe (isocontour), il est impossible de paramétrer une surface (isosurface) de manière générale sans que le système de coordonnées ne soit singulier en certains points. C'est gênant car en ces points, le squelette ne serait pas défini et l'implémentation d'un algorithme serait problématique.

Pour remédier à cela, posons  $s_i \in \mathbb{R}, i \in \{1, 2, 3\}$  trois paramètres définissant des systèmes de coordonnées unidimensionnels sur une isosurface. Par définition,  $s_i$  paramétrera la position sur l'isocontour intersection de la surface d'isodensité et du plan  $(\mathbf{r}_j, \mathbf{r}_k)$  tel que  $i \neq j \neq k$  et  $i, j, k \in \{1, 2, 3\}$  (voir figure 3.10). Alors chacun des paramètres  $s_i$  est singulier partout où la normale à l'isosurface (c'est à dire  $\nabla \rho$ ) est colinéaire à  $\mathbf{r}_i$  ( $\mathbf{r}_i \propto \nabla \rho$ ). Par conséquent, jamais plus d'un des paramètres ne sera singulier en un point donné. En tout point, on disposera alors de deux paramètres non singuliers pour décrire la position sur la surface d'isodensité. Il nous reste donc à résoudre les équations (3.28) et (3.29) en posant  $u \equiv s_i$  et  $v \equiv s_j$ , sachant que



**Fig. 3.10:** *Choix d'un système de coordonnées sur une surface d'isodensité  $S$  (en rouge). Une isosurface possède deux dimensions, deux coordonnées sont donc nécessaires pour déterminer de manière unique la position d'un point sur celle-ci. Un problème est qu'il est impossible de définir un système de coordonnées non dégénéré en tout point de la surface (car il n'existe pas d'application bijective entre le plan  $\mathbb{R}^2$  et  $S$ ). On définit  $s_i$  avec  $i \in \{1, 2, 3\}$  comme la coordonnée paramétrisant la position dans l'intersection de l'isosurface et des plans de normale  $\mathbf{r}_i$ . La coordonnée  $s_i$  est alors dégénérée lorsque la normale à la surface est  $\mathbf{r}_i$  (n'importe quelle valeur de  $s_i$  désignera alors le même point). En tout point, on peut toujours choisir deux axes  $s_i$  et  $s_j$  avec  $i \neq j$  tels que ces deux systèmes de coordonnées ne soit pas dégénérés.*

$i, j \in \{1, 2, 3\}$  et  $i \neq j$ . Pour une valeur de  $i$  donnée, le système d'équations s'écrit :

$$\begin{cases} \frac{d}{ds_i}(|\nabla\rho|^2) = 0 \\ \frac{\partial\rho}{\partial s_i} = \frac{\partial\rho}{\partial r_1} \frac{dr_1}{ds_i} + \frac{\partial\rho}{\partial r_2} \frac{dr_2}{ds_i} + \frac{\partial\rho}{\partial r_3} \frac{dr_3}{ds_i} = 0. \end{cases} \quad (3.30)$$

Soit, en procédant de la même manière que pour l'équation (3.23) et en prenant  $i \neq j \neq k \in \{1, 2, 3\}$  :

$$\begin{aligned} \mathcal{S}_i \equiv G_i &= \frac{\partial^2\rho}{\partial r_j \partial r_k} \left( \frac{\partial\rho}{\partial r_j}^2 - \frac{\partial\rho}{\partial r_k}^2 \right) \\ &+ \frac{\partial\rho}{\partial r_j} \frac{\partial\rho}{\partial r_k} \left( \frac{\partial^2\rho}{\partial r_k^2} - \frac{\partial^2\rho}{\partial r_j^2} \right) \\ &- \frac{\partial\rho}{\partial r_i} \left( \frac{\partial\rho}{\partial r_k} \frac{\partial^2\rho}{\partial r_i \partial r_j} - \frac{\partial\rho}{\partial r_j} \frac{\partial^2\rho}{\partial r_i \partial r_k} \right) = 0 \end{aligned} \quad (3.31)$$

On obtient l'équation  $\mathcal{S}_i$  dont la solution est l'ensemble des points pour lesquels le gradient est extrême sur l'intersection de la surface d'isodensité et du plan  $(\mathbf{r}_j, \mathbf{r}_k)$ . Une rapide vérification consiste à choisir  $\rho$  constant selon l'axe  $\mathbf{r}_i$ , ce qui revient à imposer  $\frac{\partial\rho}{\partial r_i} = 0$ . On retrouve alors  $\mathcal{S}_i \Leftrightarrow \mathcal{S}$  (équation (3.23)).

L'ensemble des points du squelette est donc défini par le système d'équations :

$$\mathbf{S} \equiv \left( \begin{array}{l} \mathcal{S}_i \equiv G_i = 0 \\ \mathcal{S}_j \equiv G_j = 0 \end{array} \right), \quad i \neq j \in \{1, 2, 3\}. \quad (3.32)$$

L'équation (3.32) a pour solution le squelette *total*, mais à nouveau il semble difficile d'imposer la condition de minimalité du gradient. On peut cependant vérifier que, tout comme dans le cas 2D, le squelette peut aussi être défini comme l'ensemble des points où le gradient est vecteur propre du Hessien et il suffira d'imposer des conditions sur les valeurs propres  $\lambda_i$  de  $\mathcal{H}$ . Si l'on part de l'équation :

$$\begin{pmatrix} T_1 \\ T_2 \\ T_3 \end{pmatrix} \equiv \nabla\rho \times \mathcal{H}\nabla\rho = 0 \quad (3.33)$$

qui impose  $\nabla\rho$  comme vecteur propre de  $\mathcal{H}$ , alors :

$$\begin{aligned} T_i \Leftrightarrow & \frac{\partial\rho}{\partial r_j} \left( \frac{\partial\rho}{\partial r_i} \frac{\partial^2\rho}{\partial r_i \partial r_j \partial r_k} + \frac{\partial\rho}{\partial r_j} \frac{\partial^2\rho}{\partial r_j \partial r_k} + \frac{\partial\rho}{\partial r_k} \frac{\partial^2\rho}{\partial r_k^2} \right) \\ & - \frac{\partial\rho}{\partial r_k} \left( \frac{\partial\rho}{\partial r_i} \frac{\partial^2\rho}{\partial r_i \partial r_j} + \frac{\partial\rho}{\partial r_j} \frac{\partial^2\rho}{\partial r_j^2} + \frac{\partial\rho}{\partial r_k} \frac{\partial^2\rho}{\partial r_j \partial r_k} \right) = 0, \end{aligned} \quad (3.34)$$

où  $i \neq j \neq k \in \{1, 2, 3\}$ . On peut alors réécrire  $T_i$  sous la forme :

$$\begin{aligned}
T_i &\Leftrightarrow \frac{\partial^2 \rho}{\partial r_j \partial r_k} \left( \frac{\partial \rho}{\partial r_j} - \frac{\partial \rho}{\partial r_k} \right) \\
&+ \frac{\partial \rho}{\partial r_j} \frac{\partial \rho}{\partial r_k} \left( \frac{\partial^2 \rho}{\partial r_k^2} - \frac{\partial^2 \rho}{\partial r_j^2} \right) \\
&- \frac{\partial \rho}{\partial r_i} \left( \frac{\partial \rho}{\partial r_k} \frac{\partial^2 \rho}{\partial r_i \partial r_j} - \frac{\partial \rho}{\partial r_j} \frac{\partial^2 \rho}{\partial r_i \partial r_k} \right) = 0 \\
&\Leftrightarrow \mathcal{S}_i.
\end{aligned} \tag{3.35}$$

On retrouve bien que les deux définitions sont équivalentes. Il ne reste donc plus qu'à définir les conditions supplémentaires pour obtenir le squelette (voir la figure 3.11) :

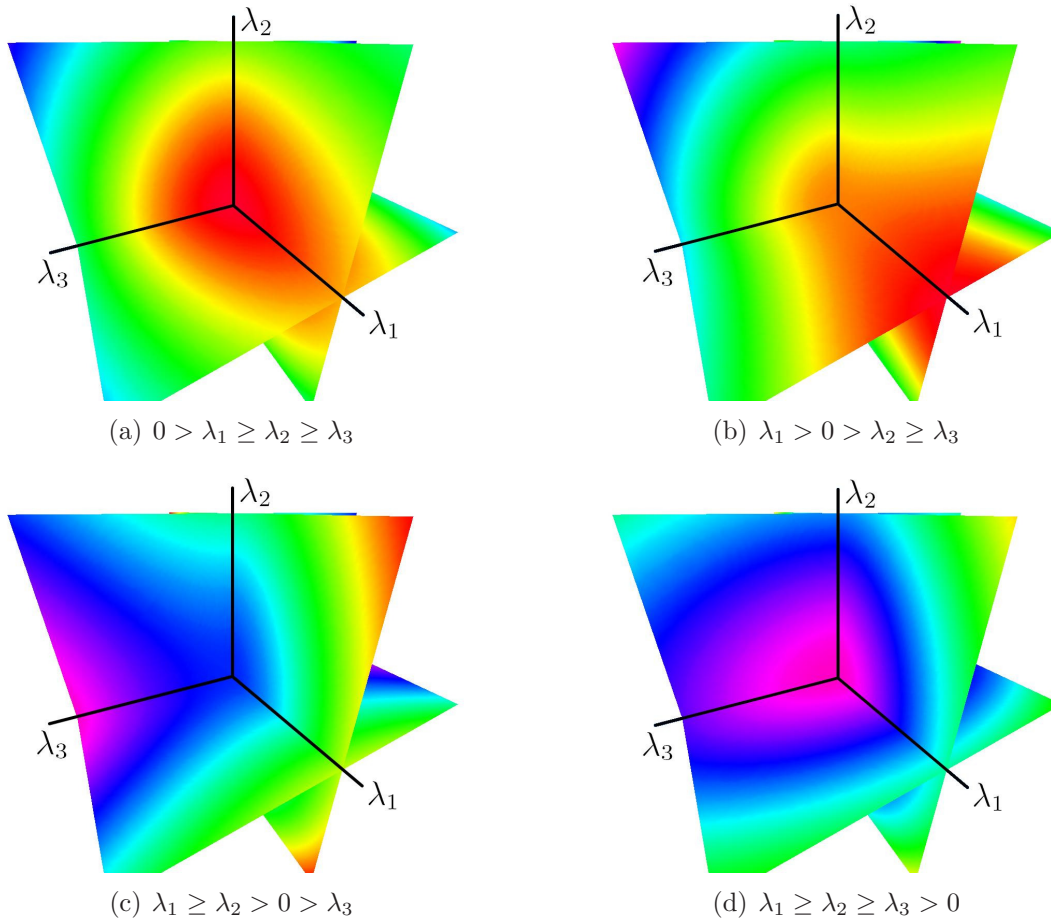
$$\begin{cases} \lambda_3 < \lambda_2 < 0, \\ \mathcal{H} \nabla \rho = \lambda_1 \nabla \rho \end{cases} . \tag{3.36}$$

### 3.3 Implémentation

Nous disposons maintenant d'une définition claire et précise du squelette. Afin de l'exploiter, il est nécessaire de pouvoir l'appliquer en pratique aussi bien aux prédictions numériques qu'aux observations. Nous utiliserons principalement deux sources d'informations : les simulations numériques N-corps de portions d'Univers et les catalogues de galaxies observées dans notre Univers local. Dans un cas comme dans l'autre, le champ de densité est représenté de manière discrète, par des particules traçant ce champ dans le premier cas et par des galaxies dans le deuxième cas. Premièrement, il convient de définir une méthode afin d'échantillonner le champ de densité  $\rho$  sur une grille. Une fois cela fait, il reste à trouver une méthode numérique permettant la résolution des équations (3.32). Celle-ci doit répondre à des contraintes techniques. Le temps de calcul et surtout la capacité mémoire des ordinateurs étant limités, il est nécessaire de trouver un algorithme à la fois efficace et peu gourmand en terme de mémoire. Nous nous sommes fixés comme objectif de pouvoir calculer le squelette sur un ordinateur de bureau standard possédant une capacité mémoire de l'ordre du Go. Les simulations numériques courantes comportent en général de  $256^3 = 16,777,216$  à  $512^3 = 134,217,744$  particules, ce qui représente une quantité de données de l'ordre de 1 Go ne serait-ce que pour stocker les trois coordonnées donnant la position de chaque particule.

#### 3.3.1 Calcul du champ de densité

Mis à part le cas des champs gaussiens qui sont générés directement, toutes nos sources d'informations sont de nature discrète. Dans tous les cas, le champ  $\rho$  n'est pas analytique et les calculs doivent donc être effectués après un échantillonnage discret. Par souci d'efficacité et puisque les contraintes ne sont pas les mêmes, la méthode utilisée est différente de celle de MoLUSC (chapitre 2.2). Afin de respecter



**Fig. 3.11:** Illustration d'un champ 3D en fonction du signe des vecteurs propres de  $\mathcal{H}$ . Le champ est représenté par trois plans orthogonaux aux vecteurs propres de  $H$ . La couleur varie de rouge (haute densité) à violet (faible densité). La figure 3.11(a) est un maximum local, le filament étant situé dans l'axe  $\lambda_1$ , la figure 3.11(b) est un point selle pour lequel une seule valeur propre est positive, à nouveau le filament est dirigé selon  $\lambda_1$ . Les figures 3.11(c) et 3.11(d) représentent un point selle avec deux valeurs propres positives et un minimum du champ. Dans ces deux cas, il n'y a pas de filaments mais plutôt des sortes d'antifilaments, c'est-à-dire des filaments de vide dans la direction de  $\lambda_3$ . Clairement, si l'on souhaite calculer le squelette, il convient de garder les deux premiers cas, c'est-à-dire d'imposer la condition  $\lambda_2 < 0$  et de sélectionner uniquement les régions où  $\nabla\rho$  est vecteur propre de  $\mathcal{H}$  de valeur propre  $\lambda_1$ .

la condition de continue dérivabilité, le champ devra être lissé à une échelle adéquate pour être de classe  $C^2$ . De plus, la condition de discrétisation des extréma du champ implique des restrictions sur le choix de la résolution de la grille d'échantillonnage. En effet, si celle-ci est trop précise, il est possible que le champ de densité reste constant sur plusieurs nœuds contigus (la précision de mesure n'étant pas arbitraire). Il existe de nombreuses manières d'obtenir un champ lisse à partir d'une réalisation discrète de ce dernier. Chacune d'elles présente des avantages et des inconvénients mais introduit forcément un arbitraire, non seulement dans la manière dont le champ est interpolé dans les zones inter-particulaires, mais aussi dans la précision avec laquelle le champ est échantillonné.

## Introduction

La méthode SPH par exemple propose de calculer la densité pour chaque particule en fonction de la position de ses  $N$  plus proches voisines ou des  $N(h)$  particules englobées dans la sphère de rayon  $2h$  centrée sur la particule selon le cas. On définit pour cela une fonction noyau ("kernel")  $W(\mathbf{r}, h)$  qui peut être n'importe quelle fonction normée. Une quantité  $A$  quelconque peut alors être calculée pour la particule  $i$  comme :

$$A_i(\mathbf{r}) = \sum_j m_j \frac{A_j}{\rho_j} W(\|\mathbf{r}_j - \mathbf{r}_i\|, h), \quad (3.37)$$

où  $m_j$  est la masse de la  $j^{\text{ème}}$  particule et  $\rho_j$  représente la densité moyenne à l'endroit où se trouve cette particule. La densité est alors facilement calculée pour la particule  $i$  comme :

$$\rho_i(\mathbf{r}) = \sum_j m_j \frac{\rho_j}{\rho_j} W(\|\mathbf{r}_j - \mathbf{r}_i\|, h) = \sum_j m_j W(\|\mathbf{r}_j - \mathbf{r}_i\|, h). \quad (3.38)$$

Habituellement,  $W(h)$  est un spline cubique qui présente l'avantage d'être de support borné, limitant la somme sur  $j$  à un petit sous-groupe de l'ensemble des particules. Les dérivées des quantités  $A_j$  se calculent aisément grâce à cette méthode, notamment lorsque  $A_j = \rho$ . En utilisant la règle d'intégration par partie, on obtient :

$$\nabla A_i(\mathbf{r}) = \sum_j m_j \frac{A_j}{\rho_j} \nabla W(\mathbf{r}_j - \mathbf{r}_i, h), \quad (3.39)$$

à condition bien sûr que  $\lim_{\|\mathbf{r}\| \rightarrow \infty} W(\|\mathbf{r}\|, h) = 0$ . Si, de plus,  $W(\|\mathbf{r}_j - \mathbf{r}_i\|, h)$  est choisi de classe  $C^2$ , alors le champ interpolé sera aussi de classe  $C^2$ . L'avantage majeur de cette méthode est qu'elle est adaptative : plus une région est dense et plus son champ de densité est échantillonné précisément. Il s'agit cependant dans notre cas d'un défaut rédhibitoire. L'échantillonnage n'est, par définition, pas effectué sur une grille régulière, par conséquent la conception d'un algorithme efficace pour la résolution des équations du squelette en devient quasiment impossible. Nous reparlerons de cette méthode dans le cadre du calcul des propriétés de la matière dans la



région des filaments, dans le chapitre 4.2.1.

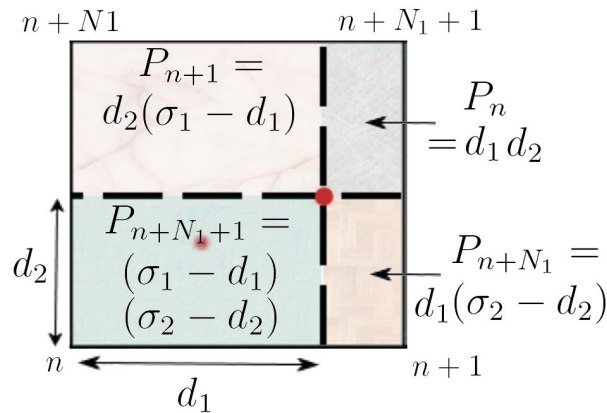
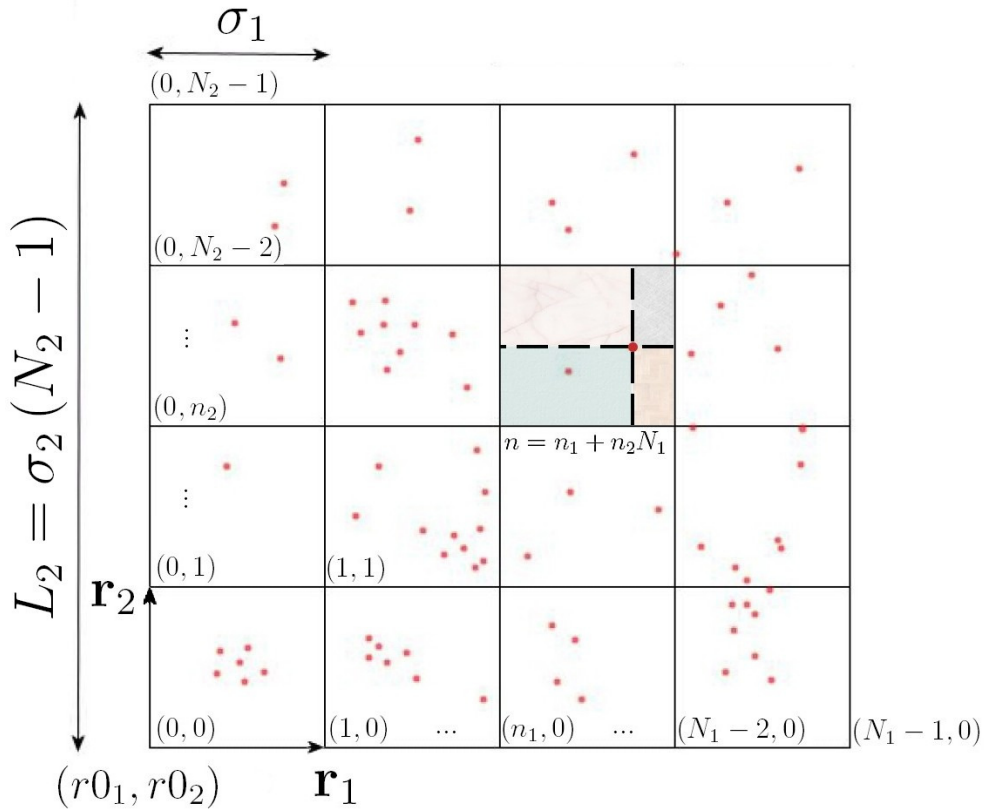
Puisque c'est le caractère Lagrangien de la méthode SPH qui pose problème, il convient d'examiner la version cartésienne du lissage SPH : le lissage à grille adaptative issu des techniques de simulation AMR. Il s'agit ici de définir une grille mère de faible résolution puis, après avoir défini un critère quelconque (par exemple le nombre de particules dans chaque cellule), de diviser les cellules en fonction de ce critère. Les cellules sont toujours cubiques ou parallélépipédiques mais ont simplement une taille plus ou moins grande selon la précision de l'échantillonnage. Le calcul des dérivées du champ se fait en utilisant les différences finies ou par transformée de Fourier (nous préciserons ces méthodes par la suite). Le problème posé par cette approche provient de l'interface entre deux zones de raffinement différent. La résolution des équations du squelette peut être faite dans chacune des deux zones mais la précision de l'échantillonnage étant différente, les résultats ne seront pas continus à l'interface. Nous avons donc décidé de nous limiter à un échantillonnage régulier, l'adaptation de l'algorithme à un échantillonnage adaptatif pouvant toujours être faite par la suite (à partir de la résolution indépendante des équations sur chacune des zones où les dimensions de la grille sont constantes). De plus, comme nous le verrons, cette méthode supporte des résolutions suffisamment élevées pour que l'ensemble des distributions de densité soient correctement calculées. Cette technique a de plus l'avantage de rendre plus simple l'interprétation des résultats : la précision étant identique dans l'ensemble de l'espace, le calcul du squelette est effectué à une même échelle quelle que soit la zone étudiée.

## Notations

Nous allons définir, dans un premier temps, un ensemble de notations qui seront utilisées afin de faciliter les explications. Le calcul du champ de densité sera effectué sur une grille cartésienne régulière  $\mathcal{G}$ , le paramètre de grille ou résolution de la grille sera noté  $\sigma_{\mathcal{G}}$  ou  $\sigma$  lorsqu'il n'y aura pas d'ambiguïtés. Dans le repère 3D  $(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ ,  $\mathcal{G}$  est donc défini par un ensemble de  $N = N_1 N_2 N_3$  nœuds de grille, les  $N_i$  étant le nombre de plans de nœuds de grille selon l'axe  $\mathbf{r}_i$ . Chaque nœud aura pour coordonnées entières  $n_i$  dans la grille,  $i \in \{1, 2, 3\}$ . Étant donné qu'il est possible d'établir une bijection entre  $\mathbb{N}$  et  $\mathbb{N}^3$ , un index unique  $n$  peut être attribué à chaque nœud de la grille. Ainsi, le nœud  $n = n_1 + n_2 N_1 + n_3 N_1 N_2$  aura pour coordonnées dans la grille  $(n_1, n_2, n_3)$ . Le vecteur position du nœud  $n$  de la grille dans le repère  $(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$  est alors donné par :

$$\mathbf{r}_n = \left( r0_1 + \frac{n_1}{N_1} L_1 \right) \mathbf{r}_1 + \left( r0_2 + \frac{n_2}{N_2} L_2 \right) \mathbf{r}_2 + \left( r0_3 + \frac{n_3}{N_3} L_3 \right) \mathbf{r}_3, \quad (3.40)$$

avec  $(r0_1, r0_2, r0_3)$  les coordonnées de l'origine de  $\mathcal{G}$  (la coordonnée du nœud  $n = 0$ ) et  $L_i = (N_i - 1)\sigma_i$  la taille de  $\mathcal{G}$  selon l'axe  $\mathbf{r}_i$ . La figure 3.12 résume l'ensemble de ces notations.



**Fig. 3.12:** Illustration des conventions de notation pour la grille d'échantillonnage (figure 3.12(a)) et de la méthode de lissage CIC (figure 3.12(b)). Les points rouges représentent la réalisation discrète du champ de densité. La figure 3.12(a) résume les principales conventions utilisées pour décrire la grille sur laquelle le champ de densité est échantillonné (une valeur de champ est calculée pour chaque nœud de la grille). La figure 3.12(b) présente la manière dont le poids de chaque particule est réparti sur les nœuds environnants. Une fraction  $P_n/V$  de la masse de chaque particule est attribuée au nœud  $n$  si le nœud  $n$  est un des sommets de la cellule contenant la particule, le volume de la cellule étant  $V = \sigma_1\sigma_2$  (dans le cas 2D). Ce poids  $P_n$  est égal au volume de la sous-partie parallélépipédique de la cellule de sommets la particule et le nœud opposé au nœud  $n$ .

## Contraintes

Au moment de définir la grille d'échantillonnage, la taille de la zone à étudier étant définie, il est obligatoire de fixer le paramètre définissant la taille des cellules,  $\sigma_i$ . La valeur des  $\sigma_i$  est cruciale pour la suite car elle donne l'échelle de résolution maximale du champ de densité. Au-dessous de cette échelle, la valeur du champ ne peut être qu'une interpolation et les propriétés du champ devront donc y être supposées continues. La valeur de  $\sigma_i$  doit, de plus, répondre à des contraintes techniques, principalement la place occupée en mémoire. Pour une grille 3D, le nombre de nœuds est en effet  $N = N_1 N_2 N_3$  et est proportionnel à  $(\sigma_1 \sigma_2 \sigma_3)^{-1}$ . Pour une grille où  $N_1 = N_2 = N_3 = 1000$ , le nombre de nœuds est ainsi de  $10^9$  soit un espace mémoire occupé d'approximativement 4 Go en simple précision (8 Go en double). Si l'on souhaite respecter les contraintes fixées sur la quantité de mémoire utilisée, on voit déjà qu'une résolution d'un millièmme de la taille de la zone étudiée sera une limite supérieure à ce que l'on pourra s'autoriser. Une autre contrainte provient de la réalisation discrète du champ de densité que l'on souhaite échantillonner : en dessous d'une certaine distance, ce dernier ne contient plus d'information et le signal est dominé par le bruit de Poisson. Typiquement, la taille de la cellule ne devra jamais descendre en dessous de la distance interparticulaire moyenne ( $d_{IM}$ ) donnée par :

$$d_{IM} = \left( \frac{V}{N} \right)^{1/3} \quad (3.41)$$

où  $V$  est le volume de la zone étudiée et  $N$  le nombre de particules dans cette zone. Idéalement, il convient même d'avoir au moins une dizaine de particules par cellule afin que le champ soit continu et une valeur de  $\sigma_i \approx 2d_{IM}$  conviendra donc parfaitement.

## Affectation des particules

Afin d'échantillonner le champ de densité, il ne reste plus qu'à calculer sa valeur en tout point de la grille. Il existe plusieurs manières de réaliser cela, la plus simple étant d'attribuer un huitième du poids de chaque particule à chaque nœud de la cellule contenant la particule. L'inconvénient de cette méthode est que l'on "gaspille" de l'information. En effet, la position exacte de la particule dans sa cellule n'est absolument pas prise en compte. Une autre approche consisterait à attribuer l'intégralité du poids de chaque particule au nœud de la grille qui lui est le plus proche. Le champ mesuré devient alors précis à une demi-taille de grille près mais cela nuit à la continuité du champ. L'idéal est donc de répartir le poids de chaque particule sur l'ensemble des huit nœuds de la cellule qui l'englobe en fonction de sa position. Cette manière de faire est appelée affectation CIC (pour "Cloud In Cell") et permet à la fois de conserver la continuité et la précision du champ à l'échelle d'une cellule. La figure 3.12(b) présente un schéma expliquant cette méthode. Considérons une particule de position  $\mathbf{q}_i$  et la cellule l'englobant et supposons que le nœud  $n$  est le nœud ayant les coordonnées  $n_1$ ,  $n_2$  et  $n_3$  individuellement les plus petites de tous

les nœuds de la cellule. Alors, la distance  $d_i$  entre ce nœud et la particule selon l'axe  $\mathbf{r}_i$  s'écrit :

$$d_i = q_i - \left( r0_i + \frac{n_i}{N_i} L_i \right). \quad (3.42)$$

Il ne reste plus qu'à attribuer une fraction de la particule à chacun des huit nœuds en fonction des  $d_i$  et des dimensions  $\sigma_i$  de la cellule. Si l'on repère chaque sommet de la cellule par ses coordonnées  $\alpha_i \in \{0, 1\}$  telles que leur index soit  $n(\alpha_i) = n + \alpha_1 + \alpha_2 N_x + \alpha_3 N_x N_y$ , alors la fraction de la particule affectée à chaque nœud est

$$P_{n(\alpha_i)} = \frac{\prod_i (\sigma_i - d_i + \alpha_i (2d_i - \sigma_i))}{\prod_i \sigma_i}. \quad (3.43)$$

Pour le nœud  $n(\alpha_i)$ , ce poids représente la fraction du volume de la cellule occupée par le parallélépipède rectangle dont les sommets opposés sont la particule et le nœud  $n(1 - \alpha_i)$ .

La figure 3.1 est une représentation de la projection du champ de densité  $\rho$  d'une simulation de matière noire calculée de cette manière là. Il s'agit d'une simulation de  $256^3$  particules échantillonnée sur une grille de paramètre  $\sigma = 20/512\text{Mpc}$ . Il s'agit plutôt du champ  $\rho' = \log(1 + \rho)$ , les contrastes de densité étant très élevés, il est nécessaire de représenter le logarithme du champ de densité afin de faire ressortir la structure filamentaire. De plus, on est obligé de rajouter une constante au champ de densité car il existe des zones où  $\rho = 0$ . Ceci constitue un problème pour la suite, en effet, pour que le squelette soit défini, il faut que  $\nabla\rho$  s'annule en un ensemble *discret* de points, ce qui n'est pas le cas lorsque le champ échantillonné est uniformément nul sur un ensemble de nœuds voisins. Enfin, il est important que le champ soit de classe  $C^2$ , c'est à dire continu sur *plusieurs* tailles de grille. La méthode pour régler ces deux problèmes est de lisser le champ.

### Lissage

Il existe bien sûr de nombreuses manières de lisser un champ échantillonné, nous désirons cependant utiliser une méthode simple afin de bien en maîtriser les conséquences. Il faut que le nombre de paramètres introduits soit le plus petit possible, typiquement une distance de lissage simplement. La méthode retenue est d'appliquer un filtre gaussien au champ  $\rho$ . Convoluer le champ à une fonction gaussienne  $G(\mathbf{r}, s)$  revient à "effacer" les détails du champ à une échelle inférieure à  $s$ , c'est à dire rendre le champ continu à cette échelle. De plus, la fonction gaussienne n'étant pas de support borné, le champ lissé ne peut être nul en aucun point. Soit :

$$G(\mathbf{r}, s) = \frac{1}{(2\pi \|s\|^2)^{3/2}} \exp\left(-\frac{\|\mathbf{r}\|^2}{2s}\right). \quad (3.44)$$

alors si l'on note  $\rho_s$  le champ lissé à une échelle  $s$ ,

$$\rho_s(\mathbf{r}) = \rho(\mathbf{r}) \otimes G(\mathbf{r}, s) = \int_{\mathbf{t}} \rho(\mathbf{t}) G(\mathbf{r} - \mathbf{t}, s) d\mathbf{t} \quad (3.45)$$

où  $\otimes$  dénote le produit de convolution. En pratique, on désire effectuer ce calcul sur un champ échantillonné. La méthode la plus efficace pour obtenir le résultat de ce produit de convolution est de le faire dans l'espace de Fourier. Alors, si l'on note :

$$G(\mathbf{r}, s) = \frac{1}{(2\pi \|\mathbf{s}\|^2)^{3/2}} \exp\left(-\frac{\|\mathbf{r}\|^2}{2s}\right). \quad (3.46)$$

sa transformée de Fourier est :

$$\hat{G}(\mathbf{k}, s) = \text{TF}[G(\mathbf{r}, s)] = \frac{1}{\sqrt{2\pi s^2}} \exp(-2\pi \|\mathbf{k}\|^2 s^2) \quad (3.47)$$

et l'équation (3.45) devient :

$$\rho_s(\mathbf{r}) = \text{TF}^{-1} \left[ \frac{1}{\sqrt{2\pi s^2}} \exp(-2\pi \|\mathbf{k}\|^2 s^2) \text{TF}[\rho(\mathbf{r})] \right] \quad (3.48)$$

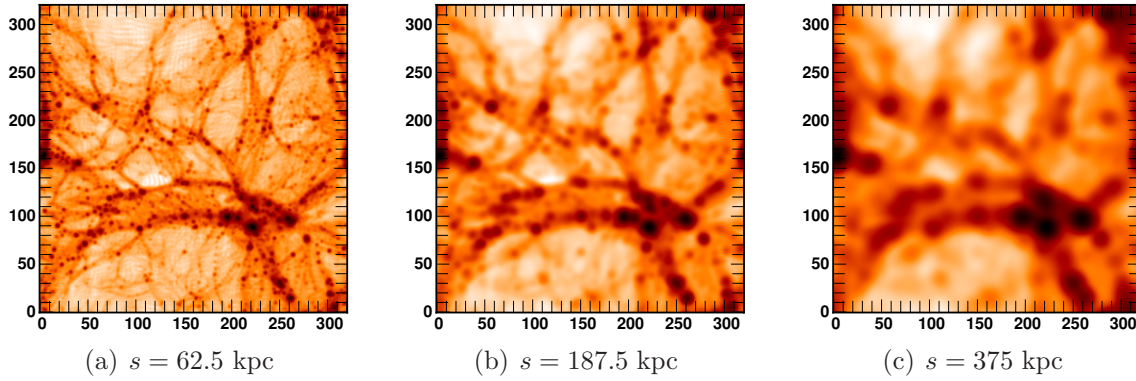
On notera  $\rho_s^\sigma$  le champ échantillonné avec un paramètre de grille  $\sigma$  et lissé à une échelle  $s$ . Alors, en utilisant l'équation (3.48),  $\rho_s^\sigma$  est facilement calculé à partir du champ non lissé  $\rho^\sigma$  par transformée de Fourier rapide (FFT). Cet algorithme très répandu ne nécessite en effet aucune mémoire supplémentaire pour effectuer la transformée de Fourier et son temps de calcul est seulement proportionnel à  $N \log(N)$  où  $N$  est le nombre de nœuds dans la grille. L'implémentation utilise la librairie FFTW3<sup>2</sup>, une librairie reconnue pour ses performances.

La figure 3.13 illustre l'évolution de la quantité de détails en fonction de l'échelle du lissage pour une simulation  ${}^3\text{LCDM}_{20}^{256}$ . L'image 3.13(a) présente un problème bien connu des simulations de petites tailles : lorsque les conditions initiales n'ont pas été calculées à un redshift suffisamment élevé, les particules se trouvant dans les plus gros vides n'ont subi jusqu'à  $z = 0$  que l'influence d'un potentiel gravitationnel très faible. Il arrive qu'elles se soient déplacées d'une distance plus faible que la taille de la grille initiale et la simulation garde donc l'empreinte de cette grille. C'est ce que l'on peut voir sur le bord haut gauche de l'image 3.13(a). Il s'agit bien sûr d'un problème pour le calcul du squelette qui détecterait cette structure régulière. Un lissage plus fort résout le problème, rendant le champ lisse et continu dans cette zone (image 3.13(b)). Il est aussi très clair que le lissage aura un rôle très important dans la sélection de l'échelle que l'on désire étudier. Ainsi, alors qu'un ensemble très complexe de filaments est visible sur la figure 3.13(a), seules les branches principales de ce réseau apparaissent sur l'image 3.13(c).

### 3.3.2 Résolution des équations

On désire maintenant parvenir à résoudre le système d'équations (3.32) de manière numérique à partir d'un champ de densité échantillonné  $\rho^\sigma$  que nous noterons simplement  $\rho$ . Il faut résoudre un système de deux équations différentielles choisies

<sup>2</sup>La librairie peut être téléchargée à l'adresse <http://www.fftw.org/>



**Fig. 3.13:** *Projection d'une tranche épaisse de 6 Mpc extraite d'une simulation  $^3\text{LCDM}_{20}^{256}$  lissée à différentes échelles sur une grille d'échantillonnage de 320 pixels de côté. Les différentes échelles de lissage font disparaître les détails dont l'ordre de grandeur est celui de la fenêtre gaussienne : 1 pixel pour l'image 3.13(a), 3 pixels pour 3.13(b) et 6 pixels pour 3.13(c).*

parmi trois selon un critère qui reste à définir exactement (le système de coordonnées correspondant à chacune des deux équations ne doit pas être dégénéré, voir la figure 3.10 pour un rappel de la définition). Ces deux équations différentielles dépendent d'un champ  $\rho$  quelconque, il n'est par conséquent pas question de procéder de manière analytique. On peut cependant remarquer que seules les dérivées première et seconde interviennent, il serait alors judicieux de décider de la meilleure méthode pour les calculer.

### Dérivées du champ

Deux méthodes sont couramment utilisées afin d'obtenir les dérivées d'une fonction quelconque de manière numérique. La première s'appelle méthode des différences finies et repose sur un développement de Taylor du champ  $\rho$ . On peut en effet écrire :

$$\begin{aligned} \rho(r_i + dr_i) &= \rho(r_i) + dr_j \frac{\partial \rho}{\partial r_j}(r_i) \\ &+ \frac{1}{2!} \frac{\partial^2 \rho}{\partial r_j \partial r_k}(r_i) dr_j dr_k \\ &+ \frac{1}{3!} \frac{\partial^3 \rho}{\partial r_j \partial r_k \partial r_l}(r_i) dr_j dr_k dr_l + O(dr^4) \end{aligned} \quad (3.49)$$

et

$$\begin{aligned} \rho(r_i - dr_i) &= \rho(r_i) - dr_j \frac{\partial \rho}{\partial r_j}(r_i) \\ &+ \frac{1}{2!} \frac{\partial^2 \rho}{\partial r_j \partial r_k}(r_i) dr_j dr_k \\ &- \frac{1}{3!} \frac{\partial^3 \rho}{\partial r_j \partial r_k \partial r_l}(r_i) dr_j dr_k dr_l + O(dr^4) \end{aligned} \quad (3.50)$$

où la convention de sommation implicite sur les indices répétés est utilisée et où  $r_i$  désigne le vecteur de position en notation tensorielle. Pour obtenir une approximation de  $\frac{\partial \rho}{\partial r_j}(r_i)$ , il suffit de faire la différence des équations (3.50) et (3.51) et d'imposer  $dr_i = h_n \delta_{in}$  avec  $n$  fixé. On obtient alors :

$$\frac{\partial \rho}{\partial r_n}(r_i) = \frac{\rho(r_i + h_n \delta_{in}) - \rho(r_i - h_n \delta_{in})}{2h_n} + O(h^3). \quad (3.51)$$

De la même manière, en faisant la différence de (3.50) et (3.51) et en posant  $dr_i = \frac{1}{2}(h_m \delta_{im} + h_n \delta_{in})$  ( $m$  et  $n$  étant fixés), on obtient une approximation de la dérivée seconde (le Hessien) :

$$\begin{aligned} \frac{\partial^2 \rho}{\partial r_m \partial r_n}(r_i) &= \frac{\rho(r_i + h_m \delta_{im} + h_n \delta_{in}) + \rho(r_i - h_m \delta_{im} - h_n \delta_{in}) + 2\rho(r_i)}{2h^2} \\ &- \frac{\rho(r_i + h_m \delta_{im}) + \rho(r_i + h_n \delta_{in}) + \rho(r_i - h_m \delta_{im}) + \rho(r_i - h_n \delta_{in})}{2h^2} \\ &+ O(h^4). \end{aligned} \quad (3.52)$$

Si l'on identifie  $h_m = \sigma_m$  où  $\sigma_m$  est la taille des cellules de la grille d'échantillonnage selon l'axe  $\mathbf{r}_m$ , il devient aisé de calculer la valeur du gradient et du Hessien du champ au nœud  $n$  en fonction de la valeur du champ aux nœuds voisins. Pour cela, il suffit de supposer que, sur une distance  $\sigma_m$ , les variations du champ sont très faibles. En d'autres termes, il faut que  $\rho$  soit continu à l'échelle de la grille. Il est intéressant de noter qu'en utilisant cette méthode, on obtient une précision à l'ordre trois pour le gradient et quatre pour le Hessien, supérieure à ce que l'on aurait obtenu avec une approche plus simple. Il existe d'ailleurs des méthodes d'ordre supérieur utilisant les nœuds de grille plus éloignés mais nous utiliserons celle-ci pour plus de simplicité.

La deuxième méthode consiste à utiliser les transformées de Fourier. Dans l'espace de Fourier les opérations de dérivation (où d'intégration) se réduisent à de simples multiplications de fonctions :

$$\frac{\partial \rho}{\partial r_n}(r_i) = \text{TF}^{-1} [-ik_n \text{TF} [\rho(r_i)]] \quad (3.53)$$

et

$$\frac{\partial^2 \rho}{\partial r_m \partial r_n}(r_i) = \text{TF}^{-1} [(-ik_m)(-ik_n) \text{TF} [\rho(r_i)]] . \quad (3.54)$$

Il est alors facile d'utiliser la FFT pour appliquer cette méthode à  $\rho^\sigma$ .

Le principal point fort de la méthode par FFT est qu'elle assure une précision maximale des valeurs obtenues pour les dérivées à une échelle d'échantillonnage donnée. De plus, cette méthode a aussi l'avantage de n'introduire aucun arbitraire, contrairement à la méthode des différences finies pour laquelle il existe plusieurs expressions possibles. En terme de temps de calcul, la méthode des différences finies est plus rapide vu que l'algorithme s'exécute en un temps proportionnel à  $N$  alors qu'une FFT s'exécute en un temps  $N \log N$ . Ce n'est cependant pas un argument

très important, le temps de calcul restant plutôt faible dans les deux cas. Il existe malgré tout un cas qui pose problème pour l'utilisation de la FFT : lorsque la distribution n'est pas périodique. En effet, l'utilisation de la FFT suppose la périodicité de la distribution et le seul moyen de pallier à cela (si l'on souhaite effectuer des calculs sur SDSS par exemple) est d'inclure la grille d'échantillonnage au centre d'une grille deux fois plus grande et valant uniformément 0. Malheureusement, cela signifie que la consommation mémoire est multipliée par 8, ce qui peut être rédhibitoire. En pratique, le champ étant lissé, l'expérience montre que les deux méthodes donnent des résultats identiques. C'est pourquoi, bien que les deux solutions soient implémentées, nous utiliserons par défaut la méthode des différences finies.

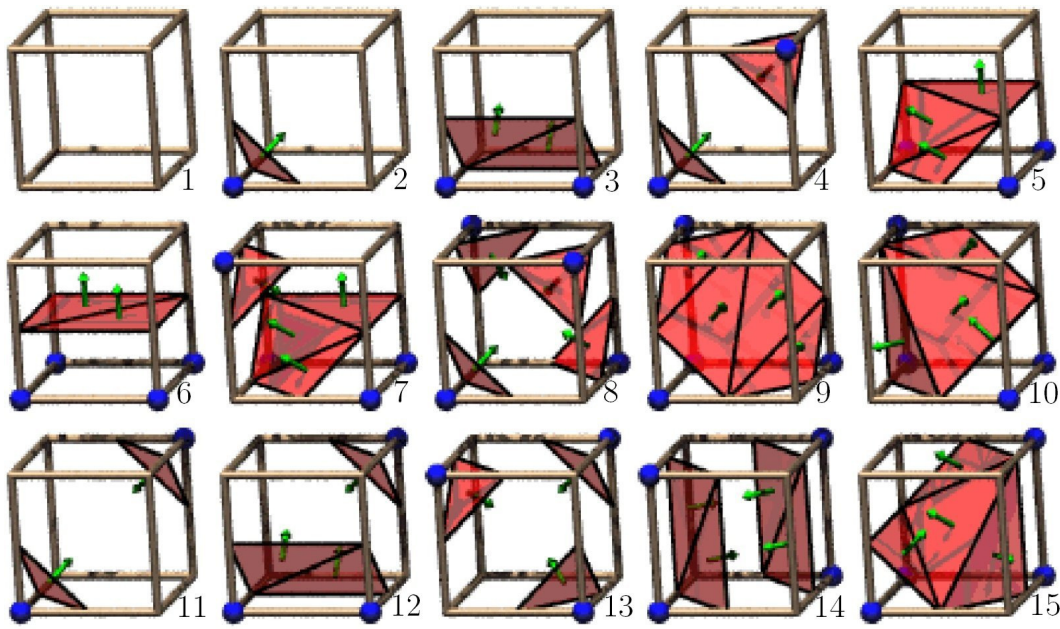
### Tracé d'isosurfaces, l'algorithme "marching cubes"

Il nous est maintenant possible de calculer n'importe quelle fonction de  $\rho$  et de ses dérivées, et notamment les champs  $G_i$  tels que  $\mathcal{S}_i \equiv G_i\left(\rho, \frac{\partial \rho}{\partial r_i}, \frac{\partial^2 \rho}{\partial r_i r_j}\right) = 0$  (voir l'équation (3.31)). Nous savons que, par définition, le squelette est un ensemble de courbes 1D plongées dans un espace 3D. Etant défini par le système de deux équations (3.32)) le squelette est donc constitué par l'intersection des deux surfaces définies par ces équations. Une autre approche consiste à remarquer que la solution de chacune de ces équations peut être interprétée comme étant la surface d'isodensité de valeur 0 du champ  $G_i$ .

Il existe plusieurs algorithmes permettant l'extraction d'une surface d'isodensité, le plus connu est appelé "marching cubes" et fut développé dans les années 70 comme outil de visualisation appliqué à la médecine. Cet algorithme est applicable à des champs 2D ou 3D mais devient trop complexe lorsque la dimensionnalité est plus élevée. Le principe est assez simple, il suffit pour l'appliquer à un champ échantillonné  $G^\sigma$  quelconque de vérifier pour chaque nœud de la grille d'échantillonnage si celui-ci a une valeur supérieure ou inférieure au seuil désiré  $G_0$ . En considérant individuellement chaque cellule, on peut remarquer alors qu'il n'existe que  $2^8 = 256$  configurations différentes, chaque sommet pouvant être dans deux états : en dessous ou au dessus du seuil, tout en sachant qu'il y a huit sommets par cellule. Un examen plus poussé réduit considérablement ce nombre. Si l'on enlève toutes les cellules où la configuration n'a subi qu'une simple rotation, il n'en reste plus que quinze qui sont décrites sur la figure 3.14. Il suffit donc de répertorier ces quinze cas, puis pour chaque cellule de la grille de trouver à quelle situation la configuration correspond.

La manière de décrire une surface géométrique dans la mémoire d'un ordinateur n'est pas unique, la plus courante est cependant de le faire sous forme de triangles (c'est la manière la plus simple et elle permet de faciliter le rendu graphique et donc la visualisation du résultat). Une surface sera donc constituée d'un ensemble de triangles (ou faces). Pour chacune de ces faces, la coordonnée des trois sommets, la normale (qui pointe vers l'extérieur), ainsi que l'identité de la cellule dans laquelle elle à été calculée sont connues. La figure 3.14 permet pour chaque configuration possible de savoir combien de triangles sont nécessaires pour chaque cellule ainsi





**Fig. 3.14:** Description des quinze configurations différentes de l’algorithme “marching cubes”. Le point bleu désigne les sommets des cellules au dessus du seuil et les surfaces rouges constituent l’intersection de la surface d’isodensité avec la cellule. Les flèches vertes montrent l’orientation des normales (l’extérieur de la surface). Cette image provient du site web <http://www.exaflop.org/docs/marchcubes/ind.html>.

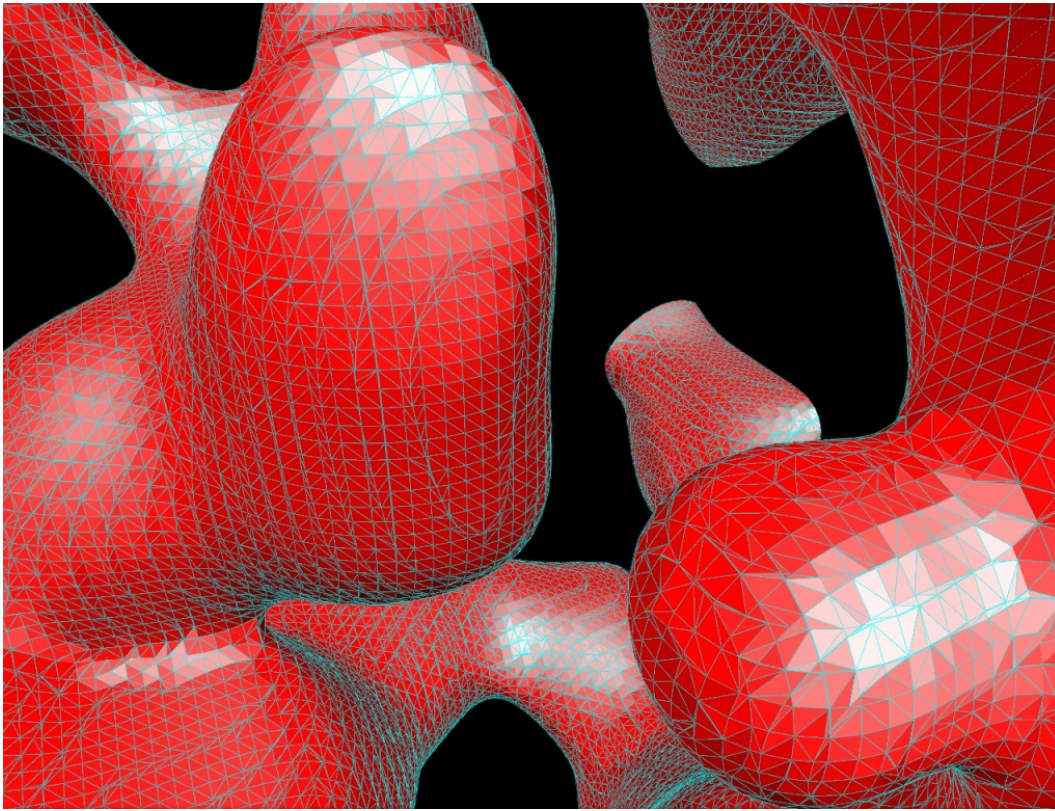
que de connaître les arêtes auxquelles appartiennent les sommets des triangles. Afin d’avoir un résultat plus précis dans le programme de calcul du squelette, les positions exactes des sommets sur l’arête sont calculées par interpolation linéaire. Ainsi, si un sommet  $\mathbf{P}_i$  du triangle se situe entre le nœud  $n$  de position  $\mathbf{r}_n$  et le nœud  $n + \Delta n$  de position  $\mathbf{r}_{n+\Delta n}$ , sa position exacte sera donnée par :

$$\mathbf{P}_i = \mathbf{r}_n + \alpha \Delta \mathbf{r} \quad (3.55)$$

avec  $\Delta \mathbf{r} = (\mathbf{r}_{n+\Delta n} - \mathbf{r}_n)$  et  $\alpha$  tel que :

$$\alpha = \frac{G_0 - G(\mathbf{r}_n)}{G(\mathbf{r}_{n+\Delta n}) - G(\mathbf{r}_n)}. \quad (3.56)$$

La figure 3.15 montre le résultat obtenu par l’application de cette méthode au calcul d’une surface d’isodensité d’un champ de densité provenant d’une simulation de matière noire et échantillonné sur une grille de  $N = 128^3$  nœuds. Bien que l’ensemble de la surface soit défini par des triangles et malgré une résolution de grille faible, la définition de la surface est excellente et ne présente pas de défauts de continuité apparents. De plus le temps de calcul est totalement négligeable puisqu’il est nettement inférieur au dixième de seconde pour cette résolution.

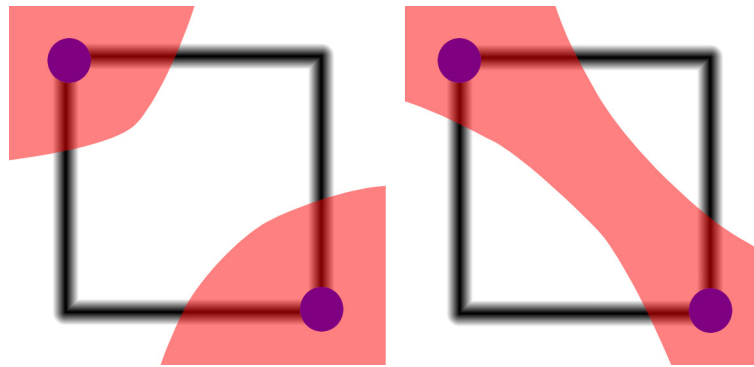


**Fig. 3.15:** Résultat obtenu pour le calcul d'une surface d'isodensité sur une simulation LCDM. Le rendu fait ressortir la nature des surfaces qui sont en fait constituées d'un ensemble de facettes. La figure 3.4 a été obtenue avec le même algorithme mais en utilisant les normales calculées en chaque sommet, on peut ainsi obtenir un rendu lissé montrant la qualité de l'approximation.

L'algorithme présente cependant quelques inconvénients dont il faut être conscient. Le principal provient du fait que certaines configurations ne sont pas bien définies. La configuration 14 par exemple est indéfinie et une rotation des facettes de  $90^\circ$  autour de l'axe vertical pourrait aussi bien convenir (voir figure 3.16). L'origine de ce problème est l'existence de plusieurs distributions de densité possibles à l'échelle d'une cellule conduisant à la même configuration des sommets de la cellule. Dans tous ces cas un choix doit être fait et l'on a conservé les solutions les plus vraisemblables.

### Calcul du squelette *total*

Il ne reste plus qu'à appliquer l'algorithme "Marching Cubes" au système d'équations (3.32). Pour cela, on résout individuellement chacune des trois équations en calculant l'isosurface de  $G_i\left(\rho, \frac{\partial \rho}{\partial r_i}, \frac{\partial^2 \rho}{\partial r_i \partial r_j}\right)$  de valeur 0 (voir la figure 3.17(a)). La solution du système est alors la courbe intersection de ces trois surfaces deux à deux. On obtiendra en réalité trois solutions qui devraient a priori être identiques en dehors des régions où la surface d'isodensité de  $\rho$  est normale à un des axes du système de coordonnées  $\mathbf{r}_i$ . Afin de voir les résultats obtenus en pratique, des grilles échan-



**Fig. 3.16:** *Cas problématique dans la définition des configurations de cellules. Il s’agit ici de la configuration 14 (figure 3.15) vue de dessus. La zone rouge représente le volume où le champ est au dessus du seuil. Il est impossible d’après la valeur du champ au sommet de la cellule de décider si la surface d’isodensité a la forme de gauche ou celle de droite.*

tillonnées à partir de champs gaussiens aléatoires ont été générées. Leur résolution est de  $128^3$  pixels et elles ont été lissées sur des échelles de 20 à 30 pixels selon les cas afin de rendre possible une visualisation claire des résultats et des différents cas exposés. Sauf mention du contraire, les images dans la suite de cette sous-section seront générées à partir de ces champs.

Le calcul de l’intersection des isosurfaces constitue un des points clés du calcul du squelette. En effet, de sa précision et de son optimisation dépendent la vitesse d’exécution et surtout la qualité du squelette obtenu. Pour ce qui est de la vitesse d’exécution, la méthode utilisée est optimale puisque, grâce à l’algorithme “marching cubes”, le temps de calcul est simplement proportionnel aux nombre de faces  $N_f$  constitutives des isosurfaces. Appelons  $I_1$  et  $I_2$  ces isosurfaces,  $N_{f_1}$  et  $N_{f_2}$  leurs nombres de faces respectifs. Une approche naïve consisterait à calculer l’ensemble des intersections des  $N_{f_1}$  faces de  $I_1$  avec les  $N_{f_2}$  faces de  $I_2$ , soit un temps de calcul proportionnel à  $N_{f_1}N_{f_2}$ . Ce n’est pas acceptable sachant que  $N_f \approx 100,000$  est une valeur raisonnable pour une grille de  $512^3$  nœuds. Une optimisation consisterait à utiliser un arbre afin de ne comparer que les faces susceptibles de s’intersecter. Le temps de calcul serait alors proportionnel à  $N_{f_1} \log(N_{f_2})$ . Il existe une solution bien meilleure : utiliser les index des cellules de grilles dans lesquelles chaque face a été calculée. Si ces index sont stockés en mémoire pour chaque face, il devient possible de ne calculer les intersections que pour des faces appartenant à la même cellule, ce nombre étant au maximum de cinq. Les faces étant de plus calculées par index de cellule croissant, aucun tri n’est nécessaire et le temps de calcul est alors proportionnel à  $N_{f_1} + N_{f_2}$ , la seule contrainte étant que  $I_1$  et  $I_2$  doivent avoir été calculées sur la même grille. Pour ce qui est du calcul d’intersection des faces, la méthode utilisée est décrite à l’adresse <http://www.inria.fr/rrrt/rr-4488.html>. Il s’agit d’une méthode élégante et optimisée permettant un temps de calcul très court tout en assurant des résultats exacts même dans les cas limites où les faces ne se touchent que par un point ou sont tangentes. Une conséquence importante de la technique

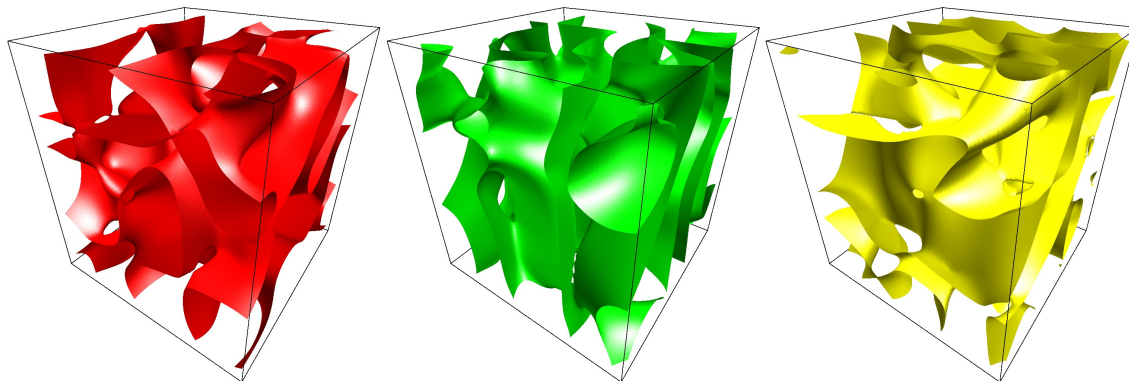
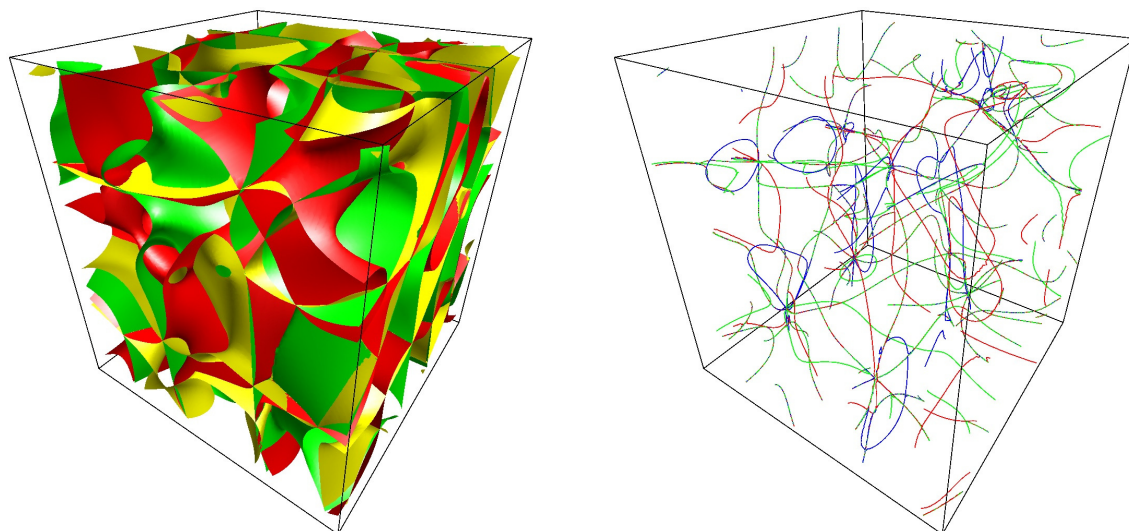
utilisée est que le squelette sera constitué d'un ensemble de segments rectilignes de taille inférieure à celle des faces, soit à peu près le tiers de celle des cellules.

La figure 3.17 présente les isosurfaces ainsi que les squelettes totaux (résultat des trois intersections possibles entre les trois isosurfaces) calculés à partir d'un champ gaussien fortement lissé. Malgré la faible résolution de la grille ( $128^3$  nœuds) et le fort lissage (30 pixels) les trois isosurfaces de la figure 3.17(a) apparaissent relativement tourmentées mais bien définies et continues. Leur superposition (figure 3.17(b)) et le résultat obtenu par le calcul de leur intersection (figure 3.17(c)) en utilisant la méthode décrite précédemment montrent que celle-ci est parfaitement définie dans la plupart des cas. Il existe cependant des zones où certains problèmes non prévus se manifestent. Les trois couleurs différentes représentent les intersections des trois paires possible d'isosurfaces et on peut remarquer que certaines parties de celles-ci ne sont pas superposées. A priori cela ne devrait pas être le cas, sauf lorsqu'un des systèmes de coordonnées est dégénéré mais ce n'est pas le cas ici. Un autre problème se manifeste sous la forme d'une oscillation que présente le squelette à certains endroits, comme si il n'y était pas bien défini. Enfin, à proximité des extréma, le squelette n'est pas connecté comme on pourrait s'y attendre.

### Artefacts numériques

Ces problèmes apparaissent pour des raisons précises et il est nécessaire de les comprendre pour connaître leur influence sur le résultat final. L'existence de morceaux de squelette solutions de seulement deux des trois équations (3.32) possibles n'était pas prévue. L'image 3.18(a) illustre ce phénomène et en donne l'explication. Alors que dans la partie basse du squelette, les trois systèmes d'équations donnent des solutions identiques, il existe une branche bleue unique dans la partie haute formant une sorte de boucle. L'examen des deux isosurfaces (rouge et verte) dont l'intersection est le squelette bleu montre qu'il s'agit d'une zone où celles-ci sont quasiment tangentes. En théorie, elles devraient être simplement tangentes en une courbe 1D (le squelette) mais les erreurs numériques de calcul ne permettent pas une telle précision. Par conséquent les deux surfaces se traversent donnant naissance à un morceau de squelette artificiel. Heureusement, la partie réellement solution est conservée (le morceau dans le bas de l'image est bien solution dans tous les cas) et cela n'empêche aucunement le calcul efficace du squelette. L'expérience montre que ce cas de figure n'est pas majoritaire et le squelette est le plus souvent une intersection d'isosurfaces qui se traversent plutôt que d'être simplement tangentes. Il est tout de même nécessaire d'enlever ces morceaux "parasites" et donc d'en tenir compte lors de la sélection des morceaux significatifs du squelette décrite dans la sous-section suivante.

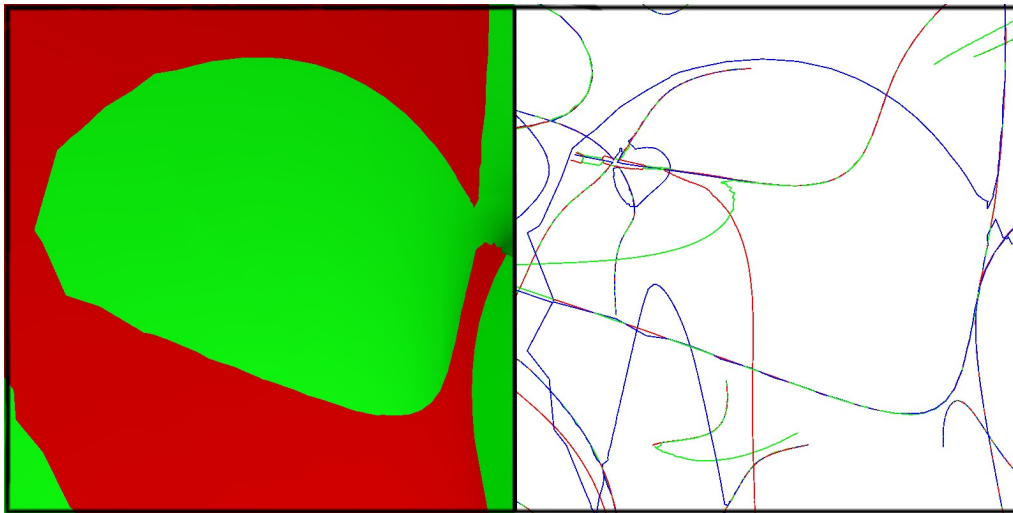
L'observation du squelette obtenu fait apparaître un autre type de défaut : en certains rares endroits, celui-ci paraît être mal défini et osciller de manière aléatoire autour de sa position réelle. Deux manifestations sont illustrées par les figures 3.18(b) et 3.18(c). Dans le premier cas, assez courant, les isosurfaces sont mal définies

(a) Solutions de  $G_i = 0$ 

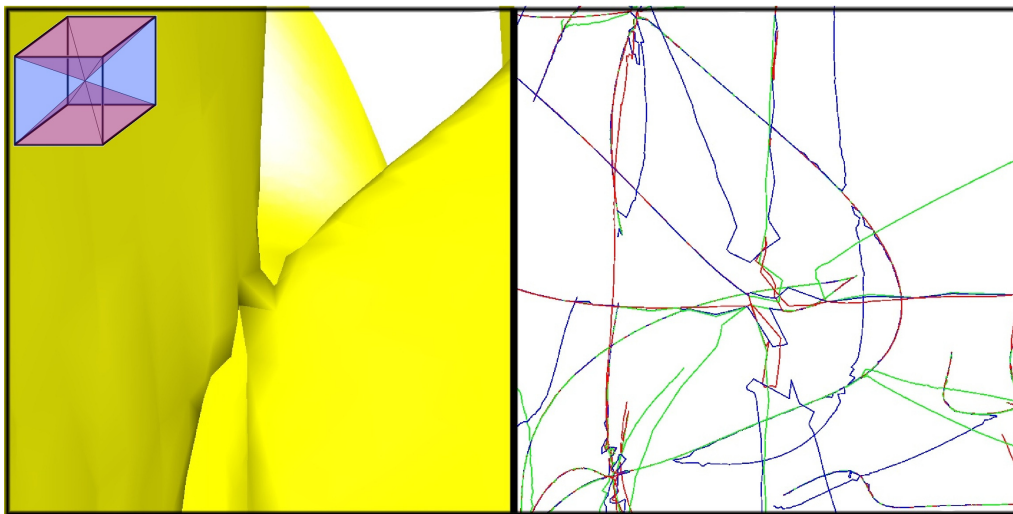
(b) Les isosurfaces dont l'intersection constitue le squelette total

(c) Résultat de l'intersection d'isosurfaces

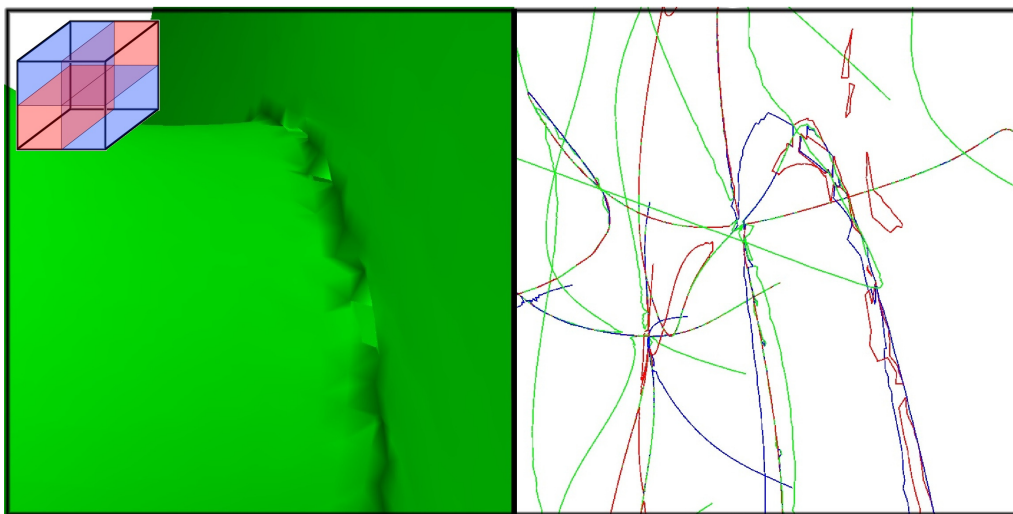
**Fig. 3.17:** Les trois isosurfaces de la figure 3.17(a) sont les solutions des équations (3.32) pour un champ gaussien. Le champ est lissé sur une grande échelle, ces trois isosurfaces, bien que parfaitement définies, sont donc plutôt complexes. La figure 3.17(b) montre la superposition de ces trois isosurfaces, le résultat de leurs intersections constitue le squelette total représenté sur la figure 3.17(c). Sur cette figure, chaque couleur représente l'intersection de deux isosurfaces différentes. Bien qu'en théorie les trois squelettes devraient être identiques, certaines parties n'apparaissent que dans une seule couleur. Il s'agit là d'un problème numérique qui sera expliqué par la suite, le résultat restant cependant remarquable de continuité.



(a) Deux isosurfaces tangentes



(b) Alentours d'un extréma du champ



(c) Problème de l'algorithme "marching cube"

**Fig. 3.18:** Les différents problèmes numérique apparaissant lors du calcul du squelette. Les cubes colorés représentent les changements de signe de  $G_i$ . Voir les explications dans le texte.

autour des extréma et provoquent des erreurs évidentes sur le calcul du squelette. Le deuxième cas est beaucoup moins répandu et se manifeste le long de certains morceaux de squelette. Il s'agit à nouveau ici d'un problème de définition approximative des isosurfaces. Dans le cas de l'extréma, le problème vient toujours du fait que les isosurfaces de  $G_i$  sont singulières. En ces points, elles sont tangentes à elles-mêmes or ce point de tangence constitue un détail dont la résolution est inférieure à celle des cellules de la grille. Il est facile de comprendre qu'un extrémum étant par définition un point où la dérivée s'annule, celle-ci change de signe dans son voisinage. Il arrive par conséquent que les  $G_i$  (voir l'équation (3.31)) changent plusieurs fois de signe autour d'un extrémum donnant cette forme particulière aux isosurfaces. Le cas de la figure 3.18(c) présente les mêmes symptômes. Ici, l'isosurface s'intersecte elle-même, sa coupe dans le plan orthogonal au squelette formant une sorte de croix. On se retrouve dans le cas de la figure 3.16 que l'algorithme ne peut pas gérer de manière appropriée.

Pour estimer la fréquence de ces défauts algorithmiques, plaçons nous dans le repère des vecteurs propres du Hessien. Ce n'est bien sûr, sauf coïncidence extraordinaire, jamais le cas en pratique et les observations qui suivent n'ont pas valeur de démonstration mais permettent toutefois de se faire une idée. Dans ce repère,  $\mathcal{H}$  est par définition diagonal et donc  $\frac{\partial^2 \rho}{\partial r_i \partial r_j} = 0$  si  $i \neq j$ . L'équation (3.31) se réécrit simplement :

$$G_i = \frac{\partial \rho}{\partial r_j} \frac{\partial \rho}{\partial r_k} \left( \frac{\partial^2 \rho}{\partial r_k^2} - \frac{\partial^2 \rho}{\partial r_j^2} \right). \quad (3.57)$$

Sur les extréma du champ,  $\frac{\partial \rho}{\partial r_i}$  est nul quelle que soit  $i$  et  $\frac{\partial \rho}{\partial r_i}$  change de signe dans le plan orthogonal à  $\mathbf{r}_i$  (qui est vecteur propre du Hessien). Autour des extréma,  $G_i$  change donc de signe en définissant trois zones (car  $\frac{\partial \rho}{\partial r_j}$  et  $\frac{\partial \rho}{\partial r_k}$  changent de signe) séparées par ses isosurfaces de valeur nulle : les isosurfaces forment alors une sorte de cône dont le sommet est l'extrémum, soit exactement ce qui se passe sur la figure 3.18(b). On s'attend à ce que cela se produise souvent. En effet, c'est le cas pour les trois isosurfaces dans ce repère et il n'y a pas de raison que ce soit différent dans un repère quelconque. Il faudra prévoir que les morceaux de squelette proches des extréma ne soient pas détectés et éventuellement en tenir compte lorsque ce sera nécessaire. Pour ce qui est du problème de la figure 3.18(c), la fréquence d'occurrence devrait être a priori bien plus faible (conformément aux observations). Par définition du squelette, le gradient est aligné avec le vecteur propre principal de  $\mathcal{H}$  soit  $r_1$  dans notre cas, celui-ci est donc nul dans le plan  $(\mathbf{r}_2, \mathbf{r}_3)$ . Il en découle que, dans l'équation (3.57), seul  $\frac{\partial \rho}{\partial r_3}$  s'annule (et change de signe) pour  $G_2$  et seul  $\frac{\partial \rho}{\partial r_2}$  s'annule pour  $G_3$ . L'isosurface nulle de  $G_2$  et  $G_3$  est constituée d'une simple surface parfaitement définie et non singulière, contrairement à celle de  $G_1$  dont la coupe dans le plan de normale  $\mathbf{r}_1$  forme une croix, les deux dérivées  $\frac{\partial \rho}{\partial r_2}$  et  $\frac{\partial \rho}{\partial r_3}$  changeant de signe et forment quatre régions où  $G_1$  a un signe différent dans le voisinage du squelette (voir le petit schéma en haut à gauche de la figure 3.18(c)). Dans tous les cas, il existe toujours au moins un des trois systèmes d'équations dont la solution numérique est parfaitement correcte et l'expérience montre que la plupart du temps, dans un repère quelconque, c'est le

cas des trois. Il conviendra néanmoins de bien sélectionner le système approprié afin de minimiser la probabilité d'apparition de ce type de problèmes.

### Sélection du squelette

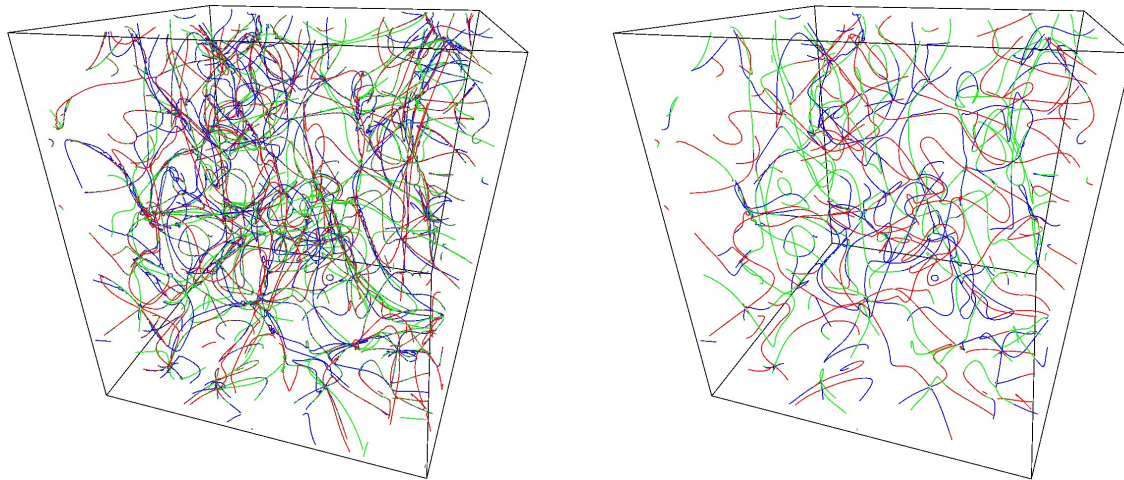
A ce stade, il nous est donc possible de résoudre numériquement les équations du squelette *total* et d'en obtenir trois versions de plus où moins bonne qualité selon les régions concernées. Pour en déduire le squelette, la procédure est de trouver un critère de sélection dépendant du champ et éventuellement de ses dérivées afin de décider lequel des squelettes est de “meilleure qualité”. Enfin, il ne restera plus qu'à appliquer les critères de sélection sur les valeurs propres de l'équation (3.36).

Afin de trouver la manière de choisir quel système d'équations est le plus adapté à une résolution numérique efficace, il faut d'abord définir les critères d'une “résolution numérique efficace”. On désire en premier lieu éliminer les équations dont les isosurfaces présentent les défauts décrits dans le paragraphe précédent (figure 3.18). Il faudra éviter de sélectionner un bout de squelette isolé (c'est à dire solution d'un système d'équations mais pas des deux autres). Pour cela, seuls les segments de squelette appartenant à des cellules contenant les trois types différents de squelettes (solutions des trois systèmes d'équations) pourront être gardés. De plus, il conviendrait d'éviter autant que possible les zones du squelette où l'algorithme “marching cubes” n'est pas adapté à la géométrie des isosurfaces (figures 3.18(b) et 3.18(c)). Pour ce qui est des extréma, il n'y a malheureusement pas grand chose à faire, la géométrie des isosurfaces étant plus ou moins la même dans tous les cas. Il faudra se résigner à une mauvaise définition du squelette au voisinage direct des extréma. Ce n'est cependant pas grave : le voisinage des extréma représentant seulement une faible fraction du squelette total, il sera toujours possible de reconstruire artificiellement ces zones à posteriori (voir chapitre 3.3.3).

En conséquence, les deux critères retenus pour la sélection seront de pouvoir éviter autant que possible l'apparition du problème de la figure 3.18(c) et de faire en sorte de choisir les systèmes de coordonnées les mieux définis. Il a été montré au chapitre 3.2.2 que le système de coordonnées  $s_i$  défini sur les surfaces d'isodensité et aboutissant à l'équation  $G_i = 0$  est singulier lorsque le gradient est parallèle à l'axe  $\mathbf{r}_i$ . Il semble judicieux de choisir les deux équations  $G_i = 0$  et  $G_j = 0$  avec  $i \neq j$  telles que  $\mathbf{r}_i$  et  $\mathbf{r}_j$  soient le plus orthogonal possible au gradient. Par bonheur, il se trouve que ce choix permet aussi d'améliorer les problèmes de mauvaise définition des isosurfaces le long du squelette. Comme il a été montré précédemment, dans le repère des vecteurs propres de  $\mathcal{H}$ , si  $\mathbf{r}_i$  est le vecteur propre principal, l'isosurface est singulière lorsque l'on choisit le système d'équation  $G_i = 0$ . Or il se trouve que par définition du squelette, le gradient et le vecteur propre principal sont parallèles. Choisir les deux équations  $G_i = 0$  et  $G_j = 0$  avec  $i \neq j$  telles que  $\mathbf{r}_i$  et  $\mathbf{r}_j$  soient le plus orthogonal possible au gradient revient donc à choisir les deux équations les plus éloignées possible du cas problématique. En pratique, il suffira alors de calculer :

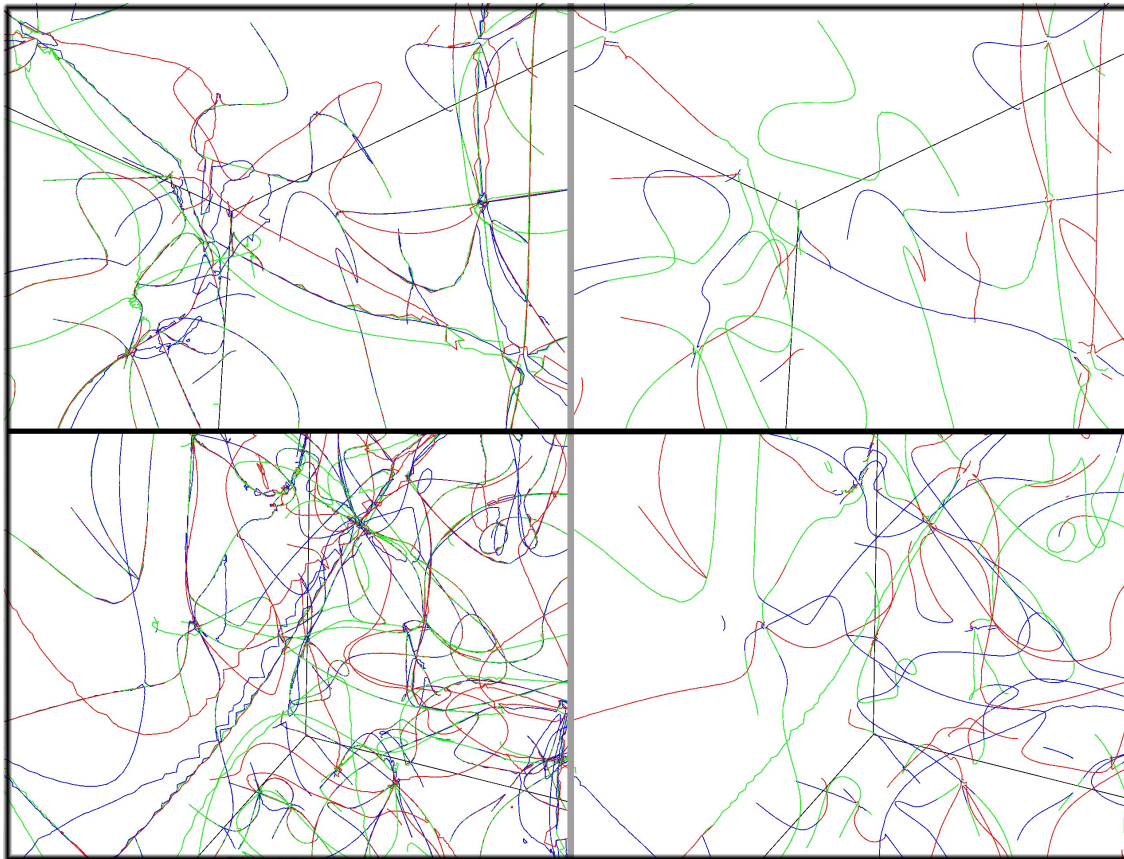
$$C_i = \nabla \rho \cdot \mathbf{r}_i. \quad (3.58)$$





(a) Squelette total sans sélection

(b) Squelette total avec sélection



(c) Gros plan sur le squelette total avant et après sélection

**Fig. 3.19:** Influence du critère de sélection (équations (3.58) et (3.59)) sur le squelette total. Le squelette apparaît beaucoup plus lisse et continu après sélection et la plupart des problèmes numériques ont disparu. Cela se voit par l'élimination des parties bruitées et des zones où le squelette est artificiel à cause de la tangence des isosurfaces.

En prenant  $l \neq m \neq n \in \{1, 2, 3\}$  tels que  $C_l < C_m < C_n$ . On choisira donc de résoudre le système d'équations (voir (3.32))

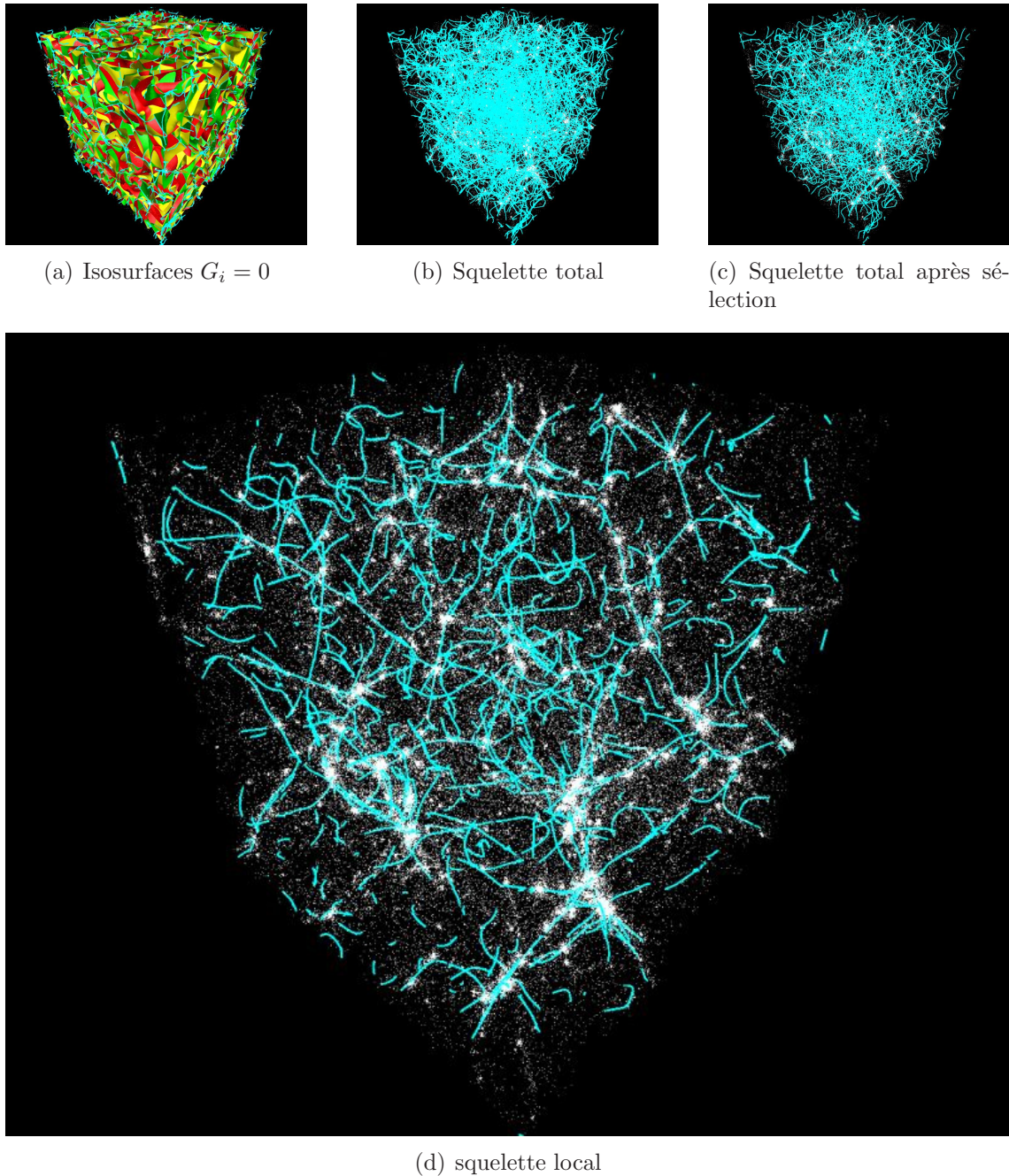
$$\mathbf{S} \equiv \begin{pmatrix} \mathcal{S}_m \equiv G_m = 0 \\ \mathcal{S}_n \equiv G_n = 0 \end{pmatrix}. \quad (3.59)$$

Les figures 3.20(d) et 3.19(a) montrent l'influence de ce critère sur l'allure du squelette total. La clarification du squelette est nette et le résultat est en accord parfait avec les prédictions théoriques. Ainsi, les parties sélectionnées sont toujours les mieux définies. La figure 3.19(c) montre l'élimination de tous les problèmes numériques dûs aussi bien à la tangence des isosurfaces qu'à leur incompatibilité avec l'algorithme "marching cubes".

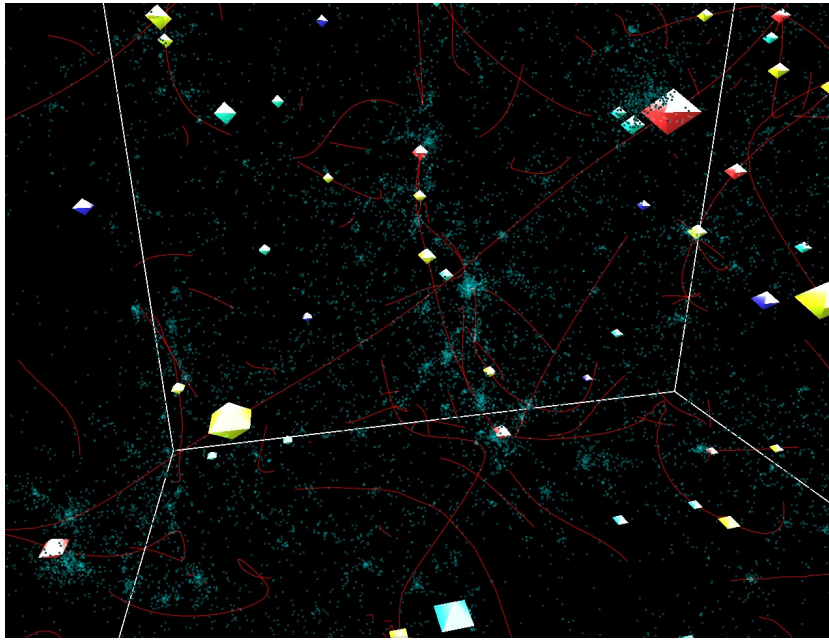
Une fois le squelette *total* obtenu, il ne reste plus qu'à imposer deux conditions pour obtenir le squelette local (i.e. les filaments). La valeur propre principale de  $\mathcal{H}$  doit être  $\lambda_1$  et de plus,  $\lambda_2$  ainsi que  $\lambda_3$  doivent être négatifs. La méthode numérique utilisée consiste simplement à diagonaliser le Hessien calculé en chaque nœud de grille puis à interpoler linéairement la valeur des vecteurs et valeurs propres au centre de chaque segment constitutif du squelette total. Pour que l'algorithme reste robuste, le produit scalaire normé de  $\nabla\rho$  et de chaque vecteur propre est calculé et seuls les segments pour lesquels sa valeur est maximale avec le vecteur propre principal sont gardés. Il ne reste plus qu'à vérifier le signe de  $\lambda_2$  et  $\lambda_3$  pour obtenir le squelette. La vérification de la pertinence de cette méthode pour la détection des filaments étant difficilement faisable sur des champs Gaussiens, la figure 3.20 montre les résultats de son application à une simulation de galaxies. Une fois les critères sur les valeurs propres appliqués, la qualité de la détection est frappante (figure 3.20). Malgré le faible nombre de galaxies (30,000 seulement), tous les filaments sont correctement identifiés. Il subsiste cependant deux problèmes. Le premier provient des défauts algorithmiques décrits précédemment qui rendent les filaments discontinus par endroits (sur de petites distances), notamment lors de leur connexion aux amas de galaxies. Le deuxième a pour cause la trop grande efficacité de la méthode. De nombreux petits bouts de filaments apparaissent en effet un peu partout sans avoir de raison apparente d'exister. Bien que réelle d'un point de vue topologique, leur existence peut poser problème pour certaines applications (par exemple l'étude des propriétés de la matière au voisinage des filaments). Afin de régler ces deux problèmes, un post-traitement du squelette a été mis au point.

### 3.3.3 Post-traitement

En théorie, il existe quatre types d'extréma pour un champ 3D : les minima, les maxima, et deux types de points selles selon le signe des valeurs propres de  $\mathcal{H}$ . Un calcul parfait du squelette local total devrait conduire à ce que l'ensemble de ces extréma soient reliés, chacun d'eux dans trois directions différentes : celle des vecteurs propres de  $\mathcal{H}$ . Le squelette local ne devrait relier que les maxima du champ et les points selles ayant deux valeurs propres négatives (aussi appelés points selle de type filament), la connexion unique se faisant dans la direction du vecteur propre



**Fig. 3.20:** La méthode du squelette appliquée à une distribution de galaxies simulées par *GalICS*. Les points blancs représentent les 30,000 galaxies contenues dans cette boîte de  $100 h^{-1}$  Mpc. L'échantillonnage est effectué sur une grille de  $128^3$  nœuds ( $\sigma \approx 781 h^{-1}$  kpc) lissée sur  $6\sigma$ . La figure 3.20(d) montre l'efficacité du squelette dans la détection des filaments de galaxies. Bien que tous les filaments importants soient correctement identifiés, des défauts tels que des trous (notamment autour des halos) sont présents. De plus, beaucoup de petits morceaux correspondant à des filaments de très faible densité (et probablement gravitationnellement insignifiants) existent d'où la nécessité d'un post-traitement pour certaines applications.



**Fig. 3.21:** *Les extréma du champ de densité détectés par la méthode des isosurfaces. Les diamants rouges représentent les maxima, centrés sur les amas de galaxies (points bleus). Le positionnement n'est pas parfaitement au centre à cause du lissage qui tient compte des positions des galaxies à proximité de l'amas. Les diamants jaunes et verts représentent respectivement les points selles de type filament (figure 3.11(b)) et pancake (figure 3.11(c)). Finalement, les diamants bleus sont les minima du champ. Les avantages principaux de cette méthode sont sa robustesse et sa cohérence avec le squelette.*

de  $\mathcal{H}$ . Malheureusement, les défauts de l'algorithme numérique ne permettent pas d'obtenir ces connexions et les filaments sont par endroits discontinus (lorsque le système de coordonnées sur les isosurfaces change). De plus, il existe de nombreuses zones dans le champ de densité respectant la définition du squelette mais ne définissant pas pour autant des filaments pertinents d'un point de vue gravitationnel. Le post-traitement vise à éliminer ces défauts afin d'étendre le champ d'application du squelette. Le but est tout d'abord de pouvoir détecter uniquement les filaments pertinents mais aussi d'être capable de les parcourir. Par exemple, il pourrait être utile d'avoir la possibilité de suivre un filament joignant deux maxima du champ tout en sachant où l'on se trouve sur ce filament.

### Calcul des extréma

Pour parvenir à ce but, il est nécessaire de trouver un moyen robuste et efficace de détection des extréma du champ. Plusieurs méthodes existent. La plus simple, mais aussi la moins efficace, consiste à considérer chacun des nœuds de grille ainsi que les nœuds voisins et d'en déduire, selon la configuration des valeurs du champ, s'il s'agit d'un extrémum. Le calcul est rapide mais la détection des points selles est

hasardeuse et la précision limitée à une taille de cellule. Une méthode plus évoluée utilise un principe similaire à celui de la méthode HOP d'identification de halos (Aubert *et al.* [2] ou Eisenstein and Hut [14]). En classant chaque nœud de grille par densité et en les parcourant de la plus élevée à la plus faible, on peut en effet détecter les extréma. Un nœud isolé est ainsi un maximum, un nœud faisant jointure entre deux zones de densité plus hautes est un point selle, ... Cette technique est plus fiable que la précédente mais à nouveau imprécise. Afin d'être cohérent avec le calcul du squelette, la méthode retenue consiste donc à considérer les extréma comme étant à l'intersection des trois isosurfaces  $\frac{\partial \rho}{\partial r_i} = 0$ . Cette manière de faire est très robuste et présente l'énorme avantage d'être cohérente avec les calculs de squelette. La figure 3.21 présente le résultat de la détection des extréma par cette méthode sur une distribution de galaxies. Les extréma (diamants rouges) correspondent aux amas de galaxies mais apparaissent un peu décalés de leur centre à cause du fort lissage qui prend en compte la distribution de galaxies dans l'environnement du halo. La méthode utilisée permet aux extréma et points selle de type filament (diamants jaunes) d'être précisément détectés en des points du squelette, ce qui sera particulièrement bénéfique à l'efficacité du post-traitement.

### Reconstruction du squelette

Pour obtenir un squelette continu et représentatif des filaments, nous procéderons en trois étapes :

1. Reconstruction des connexions aux extréma.
2. Réparation des problèmes de continuité le long du squelette.
3. Sélection des branches significatives et analyse de la connexion des segments constitutifs du squelette.

Pour arriver à un résultat probant, la seule manière est de définir un ensemble de critères permettant de déterminer les segments à conserver, ceux à rejeter et ceux à créer. Il n'est pas désirable de créer des morceaux entiers de squelette, seuls des segments reliant deux segments déjà existants pourront donc être rajoutés. De plus, on désire obtenir un squelette le plus continu possible et donc éviter les changements de direction trop brusques. Le champ étant lissé sur plusieurs tailles de grille, on ne s'attend pas à des changements rapides d'orientation sur des distances de l'ordre de la taille d'une cellule. Enfin, le squelette est, par définition, plutôt orienté dans le sens du gradient du champ qui est vecteur propre du Hessien.

Le squelette étant constitué d'un ensemble de segments, nous noterons  $\mathbf{P}_i$  la  $i^{\text{ème}}$  extrémité de ces segments et  $S(\mathbf{P}_i) = \mathbf{P}_j$  le point tel que  $\mathbf{P}_i$  et  $\mathbf{P}_j$  forment deux extrémités d'un même segment. Les coordonnées des extréma seront notées  $\mathbf{e}_i$  et l'on se placera dans le repère des vecteurs propres de  $\mathcal{H}$  (qui seront donc notés  $\mathbf{r}_i$ ) pour chaque extréma considéré. Afin de trouver quel segment de squelette reconnecter à quel extrémum, nous allons procéder par attribution de notes. Nous définissons pour

cela une fonction note  $N_k(\mathbf{e}_i)$  attribuée à chaque extrémité de segment telle que :

$$N_k(\mathbf{e}_i) = \sum_{n=1}^N \alpha_n f_n(\mathbf{e}_i, \mathbf{r}_j, \mathbf{P}_k, S(\mathbf{P}_k)). \quad (3.60)$$

Dans cette expression,  $f_n$  représente la note attribuée pour le  $n^{\text{ième}}$  critère de sélection parmi  $N$  et  $\alpha_n$  l'importance relative attribuée à ce critère par rapport aux autres. De plus, afin de ne pas attribuer artificiellement une importance plus ou moins prononcée à chaque critère, les fonctions  $f_n$  sont toujours choisies telles que  $f_n \in \{-\infty, 1\}$ , les valeurs négatives servant à éliminer directement les cas où la non satisfaction du critère est une cause de rejet.

Dans le cas de la reconnexion aux extréma, les critères sont au nombre de deux. Le premier permet de limiter la taille du segment créé pour relier l'extrémum. Soit :

$$\langle L \rangle = \frac{2 \langle \|\mathbf{P}_n - S(\mathbf{P}_n)\| \rangle_n}{N} \quad (3.61)$$

où  $\langle L \rangle$  est la taille moyenne des segments. Alors,

$$\begin{aligned} f_1(\mathbf{e}_i, \mathbf{r}_j, \mathbf{P}_k, S(\mathbf{P}_k)) &= \frac{1}{1 + \frac{\|\mathbf{P}_k - \mathbf{e}_i\|}{4\langle L \rangle}} \Pi(20 \langle L \rangle - \|\mathbf{P}_k - \mathbf{e}_i\|) \\ &- 10 \Pi(\|\mathbf{P}_k - \mathbf{e}_i\| - 20 \langle L \rangle) \end{aligned} \quad (3.62)$$

où  $\Pi(x)$  est la fonction de Heavyside telle que  $\Pi(x > 0) = 1$ , et  $\Pi(x \leq 0) = 0$ . Ce premier critère signifie simplement qu'un point ne peut être lié à un extrémum que s'il est distant de moins de vingt fois la taille moyenne des segments (soit cinq tailles de cellule sachant qu'en moyenne il y a quatre segments par cellule). Si c'est le cas, alors la probabilité de lier ce point à l'extrémum est proportionnelle à l'inverse de sa distance à cet extrémum (mesurée en nombre de cellules) plus un. Cette fonction a été choisie empiriquement et l'expérience montre qu'elle donne d'excellents résultats. Le deuxième critère concerne l'orientation du segment à créer qui doit être dans l'axe du vecteur propre principal (ou de chaque vecteur propre si l'on désire calculer le squelette total). Soit

$$A_2 = \frac{\mathbf{r}_j(\mathbf{P}_k - \mathbf{e}_i)}{\|\mathbf{r}_j(\mathbf{P}_k - \mathbf{e}_i)\|} \quad (3.63)$$

l'angle entre le vecteur propre de  $\mathcal{H}$  et le segment à créer. Alors,

$$\begin{aligned} f_2(\mathbf{e}_i, \mathbf{r}_j, \mathbf{P}_k, S(\mathbf{P}_k)) &= \Pi\left(A_2 - \frac{\pi}{3}\right) 10^{-4A_2/\pi} \\ &- \Pi\left(\frac{\pi}{3} - A_2\right) - 1000 \Pi\left(\frac{\pi}{2} - A_2\right). \end{aligned} \quad (3.64)$$

Ce critère favorise la création de segments dans la direction des vecteurs propres de  $\mathcal{H}$  et empêche la création de segments formant un angle de plus de  $\pi/3$  (un angle de  $\pi/2$  où plus étant quand à lui interdit). Étant donnée l'importance égale du respect de ces deux critères de sélection, les poids respectifs des fonctions  $f_1$  et  $f_2$  sont

choisis égaux et l'on a  $\alpha_1 = \alpha_2 = 0.5$ . En pratique, pour chaque extrémum, tous les  $N_k(\mathbf{e}_i)$  sont calculés et le segment  $[\mathbf{e}_i \mathbf{P}_n]$  tel que  $N_n(\mathbf{e}_i)$  soit maximal est rajouté.

La deuxième étape du post-traitement sert à reconstituer le squelette et utilise le même principe que lors de la reconnexion des extréma. Simplement, la fonction attribuant une note à chaque connexion possible entre le point  $\mathbf{P}_i$  et le point  $\mathbf{P}_k$  s'écrit maintenant :

$$N_k(\mathbf{P}_i) = \sum_{n=1}^N \alpha_n f'_n(\mathbf{P}_i, S(\mathbf{P}_i), \mathbf{P}_k, S(\mathbf{P}_k)). \quad (3.65)$$

La procédure consiste ici à partir de chacun des segments dont une extrémité est un extrémum (uniquement des maxima et points selles de type filament pour le squelette non total). Alors, les segments adéquats sont créés un à un en décidant à quel point  $\mathbf{P}_k$  relier le point courant  $\mathbf{P}_i$  et en excluant alors de la liste des candidats possibles aux prochaines connexions le point  $\mathbf{P}_i$ . L'opération est renouvelée en prenant  $\mathbf{P}_k$  comme point courant et jusqu'à convergence vers un extrémum ou jusqu'à ce que les notes attribuées  $N_k(\mathbf{P}_i)$  soient toutes négatives. Pour savoir à quel point  $\mathbf{P}_k$  doit être connecté le point  $\mathbf{P}_i$ , on définit trois fonctions  $f'_1$ ,  $f'_2$  et  $f'_3$ . La première est similaire à  $f_1$  et permet de connecter préférentiellement des points proches :

$$\begin{aligned} f'_1(\mathbf{P}_i, \mathbf{P}_k, S(\mathbf{P}_k)) &= \frac{1}{1 + \frac{\|\mathbf{P}_i - \mathbf{P}_k\|}{4\langle L \rangle}} \Pi(10 \langle L \rangle - \|\mathbf{P}_i - \mathbf{P}_k\|) \\ &- 10 \Pi(\|\mathbf{P}_i - \mathbf{P}_k\| - 20 \langle L \rangle) \\ &- 1000 \Pi(\langle L \rangle / 100 - \|\mathbf{P}_i - \mathbf{P}_k\|). \end{aligned} \quad (3.66)$$

Le rajout du dernier terme permet d'éviter de créer un segment reliant deux points situés exactement à la même position. En effet, l'algorithme de calcul du squelette crée des segments ayant pour la plupart un point commun avec un autre segment or le but final du post-traitement est d'éviter ces duplications et d'obtenir un ensemble de segments se suivant. Ainsi si deux segments  $[\mathbf{P}_a \mathbf{P}_b]$  et  $[\mathbf{P}_c \mathbf{P}_d]$  avec  $\mathbf{P}_b = \mathbf{P}_c$  sont présents, le point  $\mathbf{P}_b$  pourra directement être relié au point  $\mathbf{P}_d$  sans passer par  $\mathbf{P}_c$ . Les deuxième et troisième fonctions sont quant-à-elles similaires à  $f_2$  et permettent d'éviter des changements trop brutaux de direction du squelette. Soit

$$A'_2 = \frac{(\mathbf{P}_i - S(\mathbf{P}_i))(\mathbf{P}_i - \mathbf{P}_k)}{\|\mathbf{P}_i - S(\mathbf{P}_i)\| \|\mathbf{P}_i - \mathbf{P}_k\|} \quad (3.67)$$

l'angle entre le segment courant  $[\mathbf{P}_i S(\mathbf{P}_i)]$  et le segment à créer  $[\mathbf{P}_i \mathbf{P}_k]$  est :

$$A'_3 = \frac{(\mathbf{P}_k - S(\mathbf{P}_k))(\mathbf{P}_i - \mathbf{P}_k)}{\|\mathbf{P}_k - S(\mathbf{P}_k)\| \|\mathbf{P}_i - \mathbf{P}_k\|} \quad (3.68)$$

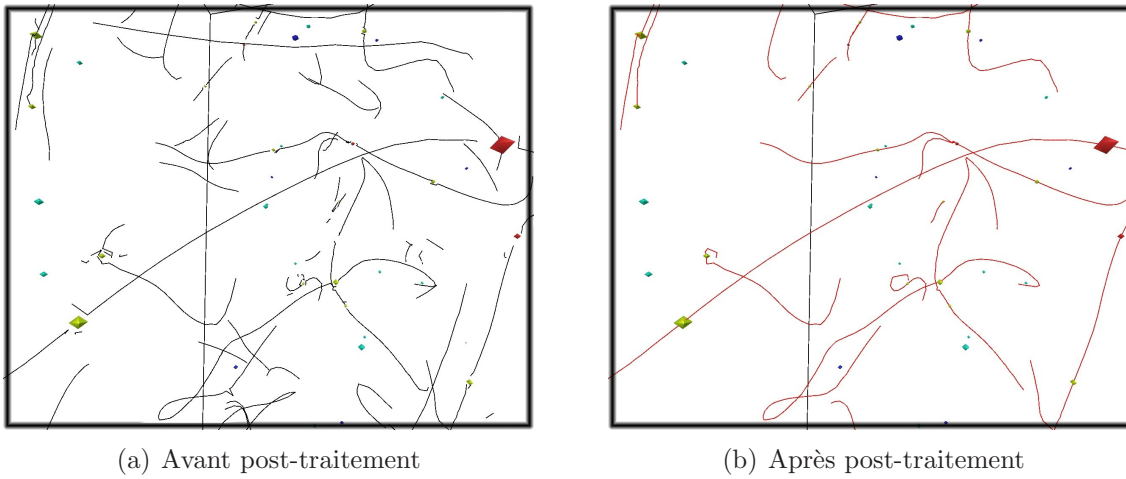
l'angle entre le segment à créer et le segment suivant  $[\mathbf{P}_k S(\mathbf{P}_k)]$ . Il est bien sûr tout à fait possible que le segment à créer soit identique au segment suivant, c'est à dire

$\mathbf{P}_i - \mathbf{P}_k = S(\mathbf{P}_k) - \mathbf{P}_k$ , et c'est même un cas qui doit être fortement avantageé. on prendra donc :

$$\begin{aligned} f'_2(\mathbf{P}_i, \mathbf{P}_k, S(\mathbf{P}_k)) &= \Pi\left(A'_2 - \frac{\pi}{3}\right)10^{-4A'_2/\pi} \\ &- \Pi\left(\frac{\pi}{3} - A'_2\right) - 1000 \Pi\left(\frac{\pi}{2} - A'_2\right) \end{aligned} \quad (3.69)$$

et

$$\begin{aligned} f'_3(\mathbf{P}_i, \mathbf{P}_k, S(\mathbf{P}_k)) &= \Pi\left(A'_3 - \frac{\pi}{3}\right)10^{-4A'_3/\pi} \\ &- \Pi\left(\frac{\pi}{3} - A'_3\right) - 1000 \Pi\left(\frac{\pi}{2} - A'_3\right). \end{aligned} \quad (3.70)$$

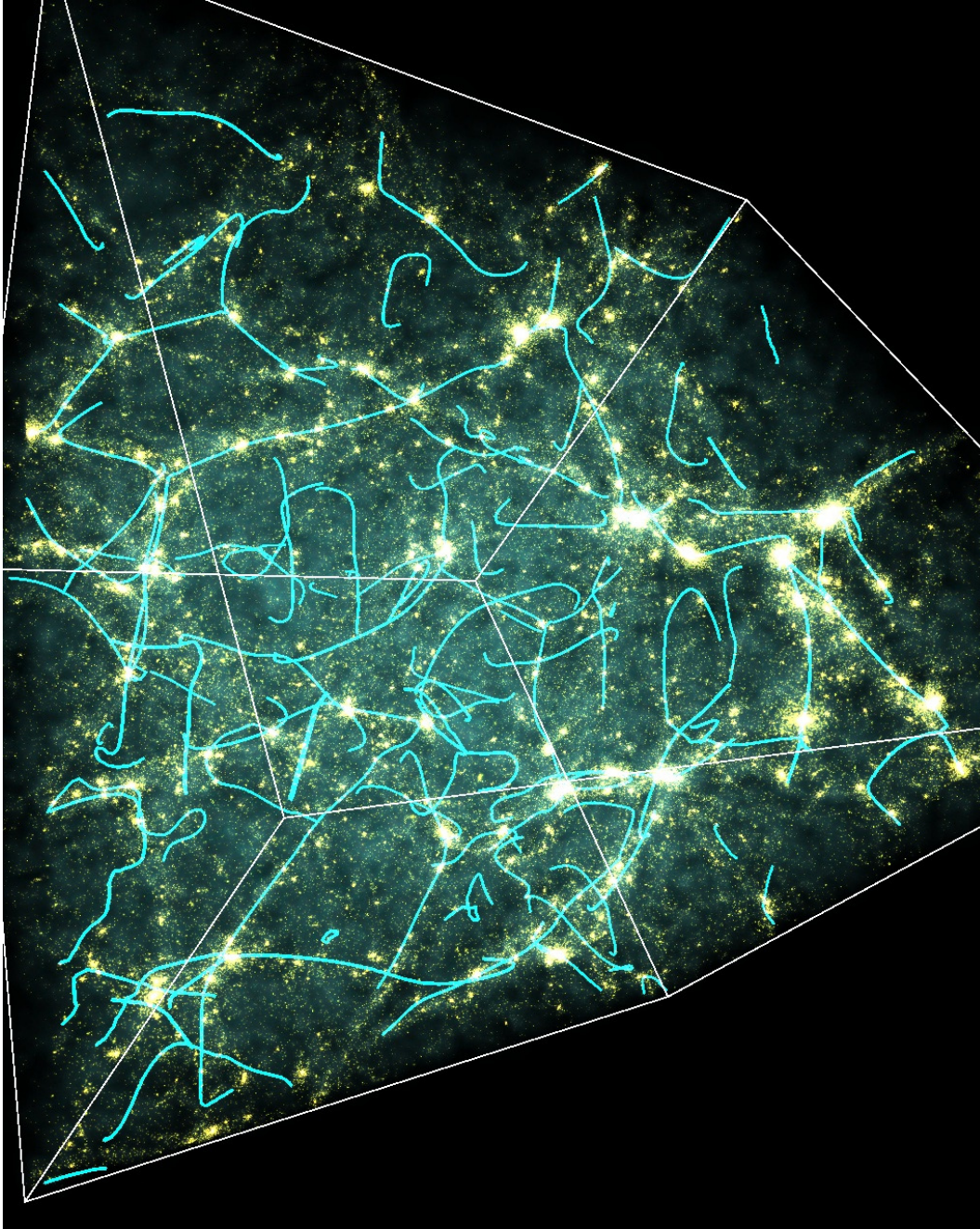


**Fig. 3.22:** Influence du post-traitement sur le squelette local. La différence entre les figures 3.22(a) et 3.22(b) illustre l'influence du post-traitement qui permet d'obtenir un squelette à la fois continu et lisse. La comparaison du squelette avant et après permet l'identification des segments ayant été artificiellement rajoutés ou enlevés lors de la reconstruction. Ces segments sont en petit nombre, ceux rajoutés étant de petite taille et toujours dans la prolongation du squelette existant. Les segments enlevés sont le plus souvent dûs au bruit de Poisson ou à des détections erronées.

Afin de favoriser les segments proches (l'algorithme sans post-traitement donnant tout de même des squelettes suffisamment continus), une importance plus grande est attribuée au critère  $f_1$  et les valeurs attribuées aux poids  $\alpha_i$  seront  $\alpha_1 = 0.5$  et  $\alpha_2 = \alpha_3 = 0.25$ . La figure 3.22 illustre l'influence de ces deux premières étapes du post-traitement sur le squelette local. La figure 3.22(a) met en évidence les lacunes de l'algorithme brut et la manière dont elles sont corrigées sur la figure 3.22(b) illustre l'efficacité du post-traitement. Ces deux images permettent aussi de mettre en évidence les segments créés lors de la deuxième étape. Ceux-ci apparaissent principalement au niveau des extréma du champ, ou plus rarement sous forme de petits segments.



Une fois les deux premières étapes achevées, il ne reste plus qu'un squelette lisse et continu ainsi qu'un ensemble de segments n'ayant pas été affectés (ceux n'ayant aucune connexion aux extréma du champ). La troisième et dernière étape du post-traitement consiste donc à éliminer ces segments et à stocker les coordonnées des points de passage du squelette  $P_i$  sous forme d'une liste liée donnant pour chaque point l'index du suivant et du précédent sur le squelette. La figure 3.23 montre l'influence du post traitement sur le squelette local par comparaison à la figure 3.20(d). La partie restante du squelette semble ainsi tracer exactement ce que l'œil identifie comme des filaments dans la distribution de galaxies. La possibilité de suivre le squelette offre l'opportunité d'effectuer de nombreuses mesures inédites de propriétés de la distribution de matière dans l'Univers.



**Fig. 3.23:** *Le squelette local post-traité de la même distribution de galaxies que celle de la figure 3.20. Le squelette est maintenant parfaitement lisse et continu et seul les morceaux correspondant à ce qui peut être identifié visuellement comme des filaments sont conservés. La reconstruction aura aussi permis de récupérer les informations de connectivité des segments constitutifs du squelette. Il sera ainsi possible de calculer des quantités le long du squelette.*

---

# Bibliographie

- [1] C. ADAMI and A. MAZURE. The use of minimal spanning tree to characterize the 2D cluster galaxy distribution. *A&AS*, 134 :393–400, January 1999.
- [2] D. AUBERT, C. PICHON, and S. COLOMBI. The origin and implications of dark matter anisotropic cosmic infall on  $L_*$  haloes. *MNRAS*, 352 :376–398, August 2004.
- [3] J. D. BARROW, S. P. BHAVSAR, and D. H. SONODA. Minimal spanning trees, filaments and galaxy clustering. *MNRAS*, 216 :17–35, September 1985.
- [4] A. J. BENSON, F. HOYLE, F. TORRES, and M. S. VOGLEY. Galaxy voids in cold dark matter universes. *MNRAS*, 340 :160–174, March 2003.
- [5] F. BERNARDEAU, S. COLOMBI, E. GAZTAÑAGA, and R. SCOCCIMARRO. Large-scale structure of the Universe and cosmological perturbation theory. *Phys. Rep.*, 367 :1–3, September 2002.
- [6] S. P. BHAVSAR and J. D. BARROW. Percolation analyses of observed and simulated galaxy clustering. *MNRAS*, 205 :61P–66P, November 1983.
- [7] S. P. BHAVSAR and R. J. SPLINTER. The superiority of the minimal spanning tree in percolation analyses of cosmological data sets. *MNRAS*, 282 :1461–1466, October 1996.
- [8] A. CANAVEZES, V. SPRINGEL, S. J. OLIVER, M. ROWAN-ROBINSON, O. KEEBLE, S. D. M. WHITE, W. SAUNDERS, G. EFSTATHIOU, C. S. FRENK, R. G. MCMAHON, S. MADDOX, W. SUTHERLAND, and H. TADROS. The topology of the IRAS Point Source Catalogue Redshift Survey. *MNRAS*, 297 :777–793, July 1998.
- [9] W. N. COLLEY. Two-dimensional Topology of Large-Scale Structure in the Las Campanas Redshift Survey. *ApJ*, 489 :471–+, November 1997.
- [10] S. COLOMBI, D. POGOSYAN, and T. SOURADEEP. Tree Structure of a Percolating Universe. *Physical Review Letters*, 85 :5515–+, December 2000.
- [11] A. DEKEL and M. J. WEST. On percolation as a cosmological test. *ApJ*, 288 :411–417, January 1985.

- [12] A. DOROSHKEVICH, D. L. TUCKER, S. ALLAM, and M. J. WAY. Large scale structure in the SDSS galaxy survey. *A&A*, 418 :7–23, April 2004.
- [13] A. G. DOROSHKEVICH, D. L. TUCKER, R. FONG, V. TURCHANINOV, and H. LIN. Large-scale galaxy distribution in the Las Campanas Redshift Survey. *MNRAS*, 322 :369–388, April 2001.
- [14] D. J. EISENSTEIN and P. HUT. HOP : A New Group-Finding Algorithm for N-Body Simulations. *ApJ*, 498 :137–+, May 1998.
- [15] D. J. EISENSTEIN, I. ZEHAVI, D. W. HOGG, R. SCOCCIMARRO, M. R. BLANTON, R. C. NICHOL, R. SCRANTON, H.-J. SEO, M. TEGMARK, Z. ZHENG, S. F. ANDERSON, J. ANNIS, N. BAHCALL, J. BRINKMANN, S. BURLES, F. J. CASTANDER, A. CONNOLLY, I. CSABAI, M. DOI, M. FUKUGITA, J. A. FRIEMAN, K. GLAZEBROOK, J. E. GUNN, J. S. HENDRY, G. HENNESSY, Z. IVEZIĆ, S. KENT, G. R. KNAPP, H. LIN, Y.-S. LOH, R. H. LUPTON, B. MARGON, T. A. MCKAY, A. MEIKSIN, J. A. MUNN, A. POPE, M. W. RICHMOND, D. SCHLEGEL, D. P. SCHNEIDER, K. SHIMASAKU, C. STOUGHTON, M. A. STRAUSS, M. SUBBARAO, A. S. SZALAY, I. SZAPUDI, D. L. TUCKER, B. YANNY, and D. G. YORK. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *ApJ*, 633 :560–574, November 2005.
- [16] A. P. FAIRALL, D. TURNER, M. L. PRETORIUS, M. WIEHAHN, V. MCBRIDE, G. DE VAUX, and P. A. WOUTD. Percolation Properties of Nearby Large-Scale Structures : Every Galaxy has a Neighbour. In A. P. FAIRALL and P. A. WOUTD, editors, *ASP Conf. Ser. 329 : Nearby Large-Scale Structures and the Zone of Avoidance*, pages 229–+, January 2005.
- [17] A. P. FAIRALL, D. TURNER, M. L. PRETORIUS, M. WIEHAHN, V. MCBRIDE, G. DE VAUX, and P. A. WOUTD. Percolation Properties of Nearby Large-Scale Structures : Every Galaxy has a Neighbour. In A. P. FAIRALL and P. A. WOUTD, editors, *ASP Conf. Ser. 329 : Nearby Large-Scale Structures and the Zone of Avoidance*, pages 229–+, January 2005.
- [18] V. ICKE and R. VAN DE WEYGAERT. The galaxy distribution as a Voronoi foam. *QJRAS*, 32 :85–112, June 1991.
- [19] A. JENKINS, C. S. FRENK, F. R. PEARCE, P. A. THOMAS, J. M. COLBERG, S. D. M. WHITE, H. M. P. COUCHMAN, J. A. PEACOCK, G. EFSTATHIOU, and A. H. NELSON. Evolution of Structure in Cold Dark Matter Universes. *ApJ*, 499 :20–+, May 1998.
- [20] j. JOST. *Riemannian Geometry and Geometric Analysis*. Springer, third edition, 2002.
- [21] M. KERSCHER, I. SZAPUDI, and A. S. SZALAY. A Comparison of Estimators for the Two-Point Correlation Function. *ApJ*, 535 :L13–L16, May 2000.

- [22] L. G. KRZEWINA and W. C. SASLAW. Minimal spanning tree statistics for the analysis of large-scale structure. *MNRAS*, 278 :869–876, February 1996.
- [23] V. J. MARTÍNEZ, J.-L. STARCK, E. SAAR, D. L. DONOHO, S. C. REYNOLDS, P. DE LA CRUZ, and S. PAREDES. Morphology of the Galaxy Distribution from Wavelet Denoising. *ApJ*, 634 :744–755, November 2005.
- [24] T. MATSUBARA. Phase Correlations in Non-Gaussian Fields. *ApJ*, 591 :L79–L82, July 2003.
- [25] K. R. MECKE, T. BUCHERT, and H. WAGNER. Robust morphological measures for large-scale structure in the Universe. *A&A*, 288 :697–704, August 1994.
- [26] A. W. MOORE, A. J. CONNOLLY, C. GENOVESE, A. GRAY, L. GRONE, N. I. KANIDORIS, R. C. NICHOL, J. SCHNEIDER, A. S. SZALAY, I. SZAPUDI, and L. WASSERMAN. Fast Algorithms and Efficient Statistics : N-Point Correlation Functions. In A. J. BANDAY, S. ZAROUBI, and M. BARTELMANN, editors, *Mining the Sky*, pages 71–+, 2001.
- [27] D. NOVIKOV, S. COLOMBI, and O. DORÉ. Skeleton as a probe of the cosmic web : the two-dimensional case. *MNRAS*, 366 :1201–1216, March 2006.
- [28] C. PARK, Y.-Y. CHOI, M. S. VOGLEY, J. R. I. GOTT, J. KIM, C. HIKAGE, T. MATSUBARA, M.-G. PARK, Y. SUTO, and D. H. WEINBERG. Topology Analysis of the Sloan Digital Sky Survey. I. Scale and Luminosity Dependence. *ApJ*, 633 :11–22, November 2005.
- [29] P. J. E. PEEBLES. *The large-scale structure of the universe*. Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p., 1980.
- [30] P. J. E. PEEBLES. The Void Phenomenon. *ApJ*, 557 :495–504, August 2001.
- [31] A. C. POPE, T. MATSUBARA, A. S. SZALAY, M. R. BLANTON, D. J. EISENSTEIN, J. GRAY, B. JAIN, N. A. BAHCALL, J. BRINKMANN, T. BUDAVARI, A. J. CONNOLLY, J. A. FRIEMAN, J. E. GUNN, D. JOHNSTON, S. M. KENT, R. H. LUPTON, A. MEIKSIN, R. C. NICHOL, D. P. SCHNEIDER, R. SCRANTON, M. A. STRAUSS, I. SZAPUDI, M. TEGMARK, M. S. VOGLEY, D. H. WEINBERG, and I. ZEHAVI. Cosmological Parameters from Eigenmode Analysis of Sloan Digital Sky Survey Galaxy Redshifts. *ApJ*, 607 :655–660, June 2004.
- [32] P. QUERRE, J.-L. STARCK, and V. J. MARTINEZ. Analysis of the Galaxy Distribution using Multiscale Methods. In J.-L. STARCK and F. D. MURTAGH, editors, *Astronomical Data Analysis II. Edited by Starck, Jean-Luc; Murtagh, Fionn D. Proceedings of the SPIE, Volume 4847, pp. 285-296 (2002).*, pages 285–296, December 2002.

- [33] P. QUERRE, J.-L. STARCK, and V. J. MARTINEZ. Analysis of the Galaxy Distribution using Multiscale Methods. In J.-L. STARCK and F. D. MURTAGH, editors, *Astronomical Data Analysis II. Edited by Starck, Jean-Luc; Murtagh, Fionn D. Proceedings of the SPIE, Volume 4847, pp. 285-296 (2002).*, pages 285–296, December 2002.
- [34] V. SAHNI and P. COLES. Approximation methods for non-linear gravitational clustering. *Phys. Rep.*, 262 :1–135, November 1995.
- [35] J. SCHMALZING, M. KERSCHER, and T. BUCHERT. Minkowski Functionals in Cosmology. In S. BONOMETTO, J. R. PRIMACK, and A. PROVENZALE, editors, *Dark Matter in the Universe, Italian Physical Society, Proceedings of the International School of Physics Course CXXXII, Varenna on Lake Como, Villa Monastero, 25 July - 4 August 1995, Edited by S. Bonometto, J.R. Primack, and A. Provenzale Oxford, GB : IOS Press, 1996, p.281*, pages 281–+, 1996.
- [36] J. V. SHETH, V. SAHNI, S. F. SHANDARIN, and B. S. SATHYAPRAKASH. Measuring the geometry and topology of large-scale structure using SURFGEN : methodology and preliminary results. *MNRAS*, 343 :22–46, July 2003.
- [37] I. SZAPUDI. Introduction to Higher Order Spatial Statistics in Cosmology. *astro-ph/0505391*, May 2005.
- [38] H. TOTSUJI and T. KIHARA. The Correlation Function for the Distribution of Galaxies. *PASJ*, 21 :221–+, 1969.
- [39] M. S. VOGLEY, C. PARK, M. J. GELLER, J. P. HUCHRA, and J. R. I. GOTT. Topological analysis of the CfA redshift survey. *ApJ*, 420 :525–544, January 1994.

---

## Chapitre 4

# Mesurer l'univers avec le squelette 3D

Le squelette constitue un outil unique dans l'analyse des caractéristiques de la distribution de matière dans l'Univers car il offre la possibilité d'analyser les filaments de la distribution en tant qu'objets à part entière. Dans ce chapitre, nous allons présenter un certain nombre d'applications que nous avons développées au cours de la thèse ainsi que certains travaux préliminaires qui donneront lieu à de futures publications. Nous présenterons tout d'abord deux applications concrètes : la contrainte du paramètre  $\Omega_m$  par la longueur des filaments par unité de volume dans le SDSS et l'étude des propriétés du flux de matière noire dans la région des filaments. Il s'agit dans les deux cas de mesures inédites rendues possibles par la méthode du squelette. Par la suite, nous présenterons d'autres applications futures qui semblent très prometteuses parmi lesquelles un test de gaussianité, le test d'Alcock-Paczynski, la mesure de la section des filaments ...

### 4.1 La longueur totale du squelette comme mesure des structures à grande échelle

La motivation originale du squelette est de donner la possibilité de tracer la structure filamenteuse de la distribution de matière à grande échelle afin d'en mesurer les caractéristiques. Comme le montre la théorie de la croissance linéaire des structures (voir la section 1.2.1), le nombre et la taille des halos de matière noire dépendent fortement des paramètres cosmologiques tels que la quantité de matière  $\Omega_m$  ou la quantité d'énergie totale  $\Omega$ . Il semble donc évident que les caractéristiques des filaments qui constituent des sortes de passerelles entre les halos doivent aussi être sensibles à la valeur de ces paramètres. Il existe de nombreuses manières de caractériser ces filaments mais il est important de garder à l'esprit que l'on désire effectuer ces caractérisations sur des données observées et par conséquent biaisées.

Dans ce chapitre nous allons montrer qu'il est possible de contraindre la valeur de la quantité de matière dans l'univers  $\Omega_m$  en mesurant simplement la longueur

totale du squelette par unité de volume dans le catalogue de galaxies SDSS DR4. La mesure de quantités physiques dans un catalogue de données observationnelles pose de nombreux problèmes aussi bien techniques (la géométrie complexe du catalogue par exemple) que théoriques (l'influence des erreurs de mesures observationnelles notamment). Afin de s'affranchir des erreurs de mesure, la démarche que nous adoptons consiste à simuler un ensemble de catalogues virtuels aussi réalistes que possible et couvrant le domaine des valeurs envisageables de  $\Omega_m$ , puis à comparer la longueur de leurs squelettes à celle mesurée dans le SDSS.

### 4.1.1 Sélection des galaxies dans le catalogue SDSS

Le SDSS redshift survey a pour projet de mesurer les décalages spectraux de toutes les galaxies situées à une distance inférieure à 600 Mpc de nous dans un volume couvrant 25% du ciel et chaque année une portion plus importante de données mesurées est mise à la disposition du public. Pour l'étude qui suit, nous avons utilisé les données contenues dans la quatrième édition publique. Une description complète de cette édition publique (DR4) peut-être trouvée dans Adelman-McCarthy [1]<sup>1</sup>.

Les données peuvent être obtenues par différents moyens, l'un d'entre eux : le système "CASJOBS", permet de faire des requêtes SQL élaborées. En effet, SDSS est un catalogue du ciel et tous les objets qui sont observés ne sont pas des galaxies. Il y a majoritairement des étoiles, mais également des nuages de gaz étendus qui peuvent être mépris pour des galaxies, des quasars, etc.... Il faut donc appliquer des critères de sélection afin de s'assurer que notre échantillon ne contienne que des galaxies. Nous utilisons par exemple le critère (*specclass* = 2 and *zconf* > 0.35) qui, d'après les caractéristiques spectrales, nous assure que cet objet peut-être classifié comme une galaxie (et non un quasar ou une étoile) et d'autre part, permet de s'assurer que le spectre obtenu est d'assez bonne qualité pour pouvoir prendre cette décision de classification, puisque l'incertitude sur la mesure du redshift est bonne (*zconf* > 0.35).

Nous demandons également via le CASJOB que l'échantillon ne soit extrait que de la table *specphoto*. Cette table est une construction liant les propriétés photométriques d'un objet à ces propriétés spectrales. Un objet étendu répertorié dans celle-ci peut encore ne pas être une galaxie, mais des critères supplémentaires sur ses magnitudes dans les différentes bandes du SDSS, sa brillance de surface, son rapport d'axe etc... permettent une bonne discrimination.

Notre sélection nous permet de construire un échantillon de 459,408 galaxies. Dans cette échantillon la complétude est atteinte pour la magnitude  $U_{\text{SDSS}} < 19$ . Nous en extrayons alors deux sous-échantillons, l'un coupé simplement en distance à 350 Mpc contient 148,012 galaxies (ci-après DR4-350), l'autre est un échantillon "volume-limité" (VL) (ci-après DR4-VL350) et contient 25,843 galaxies. Cet

---

<sup>1</sup>ou consulter <http://www.sdss.org/dr4/>



échantillon VL contient alors toutes les galaxies dont la magnitude absolue est  $M_{\text{absU}} < -17$  et la distance  $d < 350$  Mpc.

Cette sélection en volume-limité est communément utilisée en astro-statistiques afin de construire un échantillon statistiquement représentatif ou “fair sample”. L’image utilisée pour expliquer cette coupure est celle-ci : afin de décrire la distribution spatiale typique d’une foule de personnes, sachant que l’observateur verra près de lui tous les types de personnes, mais que loin il manquera les plus petites et ne verra plus que les géants. L’intérêt des catalogues en volume limité est que, si il ne contiennent pas l’ensemble des populations, celles qu’il contient sont représentées de manière exhaustive et non biaisée.

Dans notre cas, cet échantillon VL nous permettra de tester la robustesse du squelette aux biais observationnels : il sera ainsi possible de connaître l’influence d’un biais en sélection sur les mesures du squelette.

### 4.1.2 Création des catalogues virtuels

Dans le but de ne pas être influencé par les effets de biais et de passage dans l’espace des décalages spectraux, nous utilisons dans cette étude le programme MOLUSC dont le principe a été présenté au chapitre 2.2 pour générer des catalogues virtuels reproduisant ces distortions. Cet outil est particulièrement adapté à nos besoins puisqu’il permet la création de catalogues de grandes tailles tout en conservant une distribution non périodique de galaxies à grande échelle réaliste. Le catalogue SDSS s’étend sur plusieurs gigaparsecs et il est important de pouvoir comparer la taille des grandes structures de galaxies de manière réaliste.

### Simulations de matière noire

Les catalogues virtuels ont été créés à partir d’une série de dix simulations de matière noire exclusivement. Pour l’ensemble de ces simulations, une normalisation  $\sigma_8 = 0.92$  a été adoptée, la constante de Hubble est fixée à  $H_0 = 70$  km/s/Mpc et la quantité de baryons utilisée pour le calcul du spectre de puissance des conditions initiales est  $\Omega_{\text{baryons}} = 0.05$ . Dans tous les cas, la taille de la boîte est fixée à  $1000h^{-1}$  Mpc afin de couvrir une fraction significative du SDSS et la courbure est fixée à  $\Omega_k = 0$  de sorte que la quantité de constante cosmologique est toujours  $\Omega_\Lambda = 1 - \Omega_m$ . Pour neuf de ces simulations, le nombre de particules est de  $256^3$  et les valeurs de  $\Omega_m$  parcourent l’intervalle  $0.1 \leq \Omega_m \leq 0.9$  par pas de 0.1. Enfin, une dixième simulation contient  $512^3$  particules et la quantité de matière est fixée à la valeur standard de  $\Omega_m = 0.3$ . Les simulations notées  $S_i$  avec  $i \in \{1, \dots, 9\}$  seront donc de type  ${}^{i/10}\text{LCDM}_{1000}^{256}$  et la simulation notée  $S$  de type  ${}^{0.3}\text{LCDM}_{1000}^{512}$ . La distance interparticulaire moyenne des simulations  $S_i$  étant de l’ordre de 4 Mpc, il ne faudra pas s’attendre à résoudre efficacement les distributions intra-halos et on gardera à l’esprit que le seul but de ces simulations est de reproduire fidèlement les

structures à des échelles de l'ordre de 10 Mpc.

La figure 4.1 présente la projection d'une tranche de six des neuf simulations  $S_i$ . La différence aussi bien dans la distribution des halos que dans celle des filaments est évidente. Ces distributions peuvent se comprendre d'une manière intuitive si l'on se représente la constante cosmologique comme une pression négative accélérant l'expansion et faisant grandir les vides (figure 4.1(f) pour  $\Omega_\Lambda = 0.9$ ). Au contraire, une grande quantité d'énergie sous forme de matière ralentit l'expansion de l'univers et les halos ont plus rapidement tendance à se découpler de celle-ci pour s'effondrer sur eux-mêmes d'où le grand nombre de halos dans la figure 4.1(f). Un simple examen visuel permet de prédire l'influence de la valeur de  $\Omega_m$  sur le squelette. Ainsi une petite valeur de  $\Omega_m$  devrait donner naissance à un squelette plus droit et moins long alors qu'une grande valeur devrait faire apparaître un squelette très long et courbé. Pour des valeurs de  $\Omega_m \leq 0.3$  (figures 4.1(a), 4.1(b) et 4.1(c)), il est possible de voir dans les vides des zones gardant des traces des conditions initiales où la grille est encore visible. Cet effet a pour origine la faible résolution de nos simulations. En effet, lorsque la masse des particules est trop élevée, celles se trouvant dans les vides n'ont pas le temps de se déplacer pendant l'évolution et la grille reste apparente, ce qui est surtout vrai lorsque de grands vides sont présents, pour les valeurs faibles de  $\Omega_m$ . En pratique, ce n'est pas un problème si l'on lisse sur une échelle suffisamment grande, dans notre cas de l'ordre de 10 Mpc pour le pire des cas,  $\Omega_m = 0.1$ .

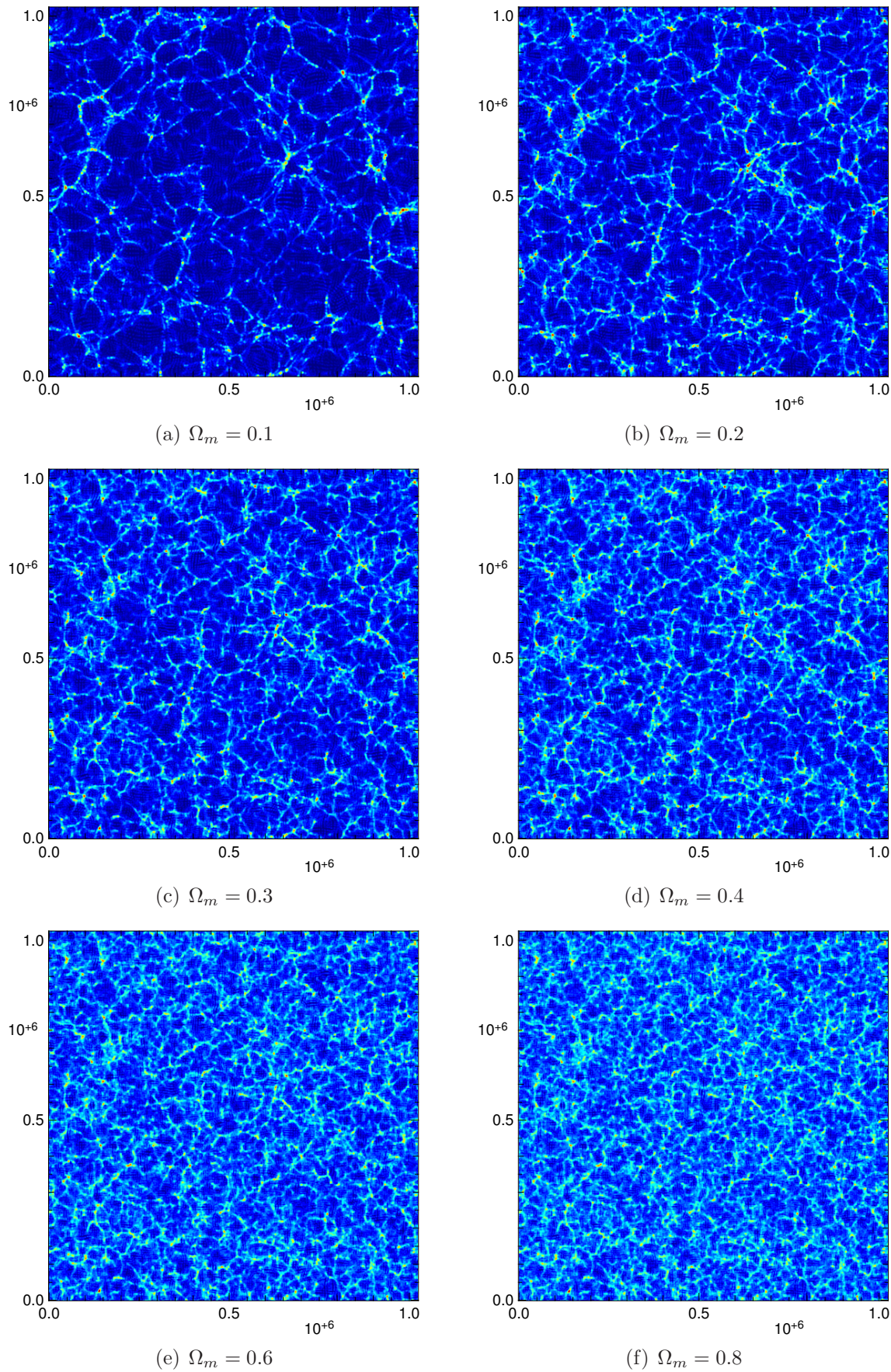
## Catalogues virtuels

La création des catalogues virtuels à partir des simulations  $S$  et  $S_i$  constitue une étape déterminante pour la qualité de l'estimation finale de  $\Omega_m$  et a donc fait l'objet d'une attention spéciale. La simulation GaLICS utilisée comme référence pour MoLUSC est la simulation galics3<sup>2</sup>, de type  $^3\text{LCDM}_{100}^{256}$ . Il a été montré dans Blaizot *et al.* [6] que sa résolution était suffisante pour décrire de manière adéquate les propriétés statistiques des galaxies du SDSS pour des magnitudes  $U_{\text{SDSS}} < 19$ . L'utilisation de MoLUSC assure, de plus, la cohérence des structures sur de grandes échelles (*cf* chapitre 2.2), évitant ainsi le problème des effets de volume fini et de réplication décrits dans Blaizot *et al.* [7].

Dans le but de faire une mesure précise de la longueur théorique des filaments, un ensemble de vingt cinq catalogues SDSS virtuels ont tout d'abord été extraits de la simulation  $S$ . Entre chacun d'eux, seule la position de l'observateur et son axe de visée change ce qui permet d'estimer la variance cosmique. Le volume occupé par le SDSS étant en effet approximativement trente fois supérieur à celui couvert par DR4-350, l'extraction de plusieurs catalogues d'une même simulation donne la possibilité d'estimer la variance cosmique de manière adéquate. A chaque fois, trois types différents de catalogues ont été simulés :

---

<sup>2</sup><http://www.galics.iap.fr>

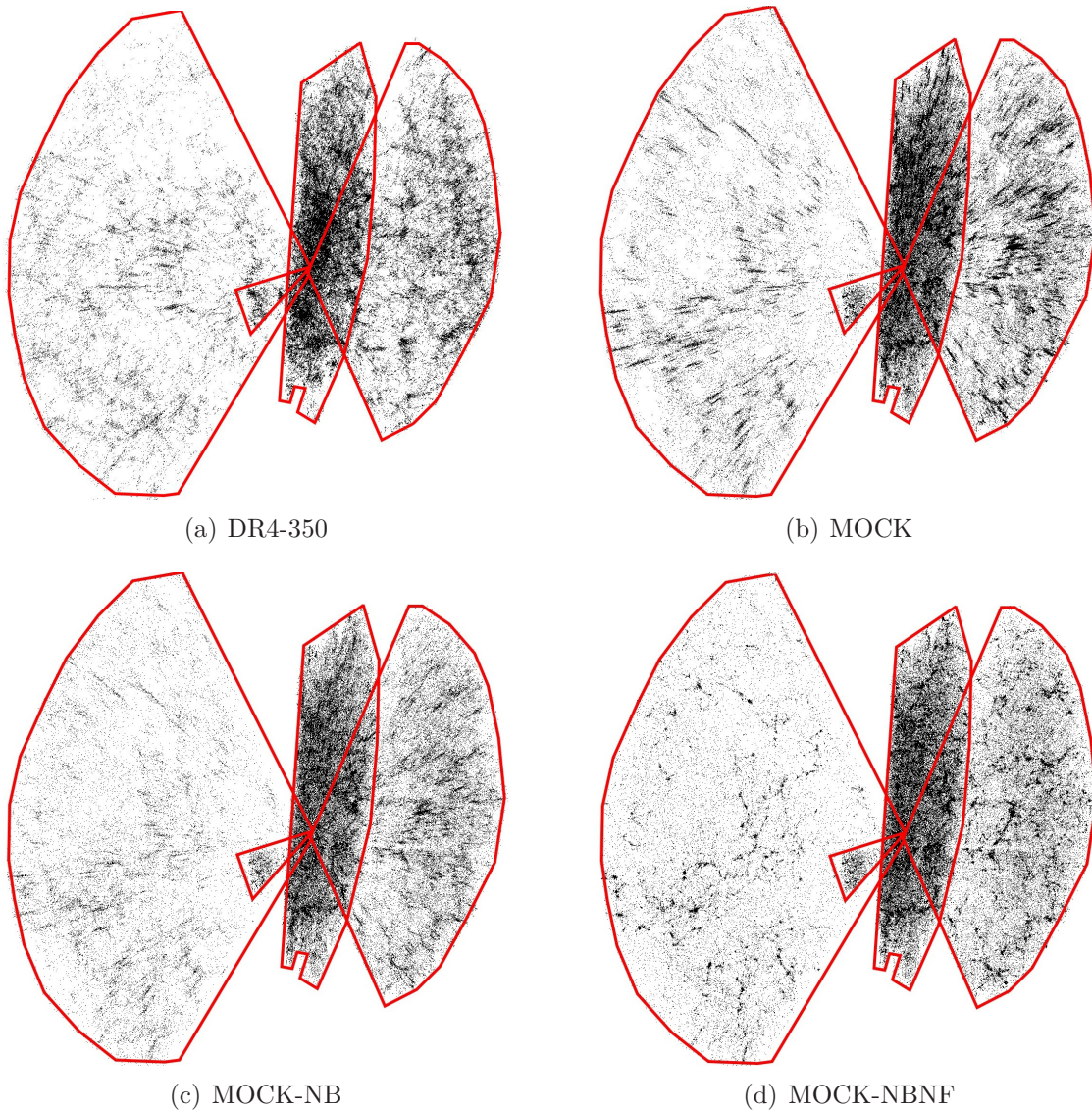


**Fig. 4.1:** *Projection d'une tranche de 20 Mpc des simulations  $S_i$  pour différentes valeurs de  $\Omega_m$ . Alors que l'Univers contient de grands vides pour les petites valeurs de  $\Omega_m$ , de nombreux filaments de petite taille se forment pour les valeurs élevées.*

- Le catalogue principal appelé MOCK, qui reproduit toutes les caractéristiques du catalogue DR4-350. Il s’agit du catalogue le plus fidèle, simulant les distortions de décalage spectral, du biais, de la géométrie et de la sélection des galaxies en volume limité.
- Le catalogue MOCK-PS identique au catalogue MOCK à ceci près que les positions des galaxies sont les positions exactes provenant de la simulation. Les déformations dues à la mesure du décalage spectral n’étant pas prises en compte, cela permet d’estimer leur influence sur nos mesures.
- Le catalogue MOCK-AS qui est une version de MOCK s’étendant sur la totalité du ciel.
- Enfin, le catalogue MOCK-NB, qui ne tient pas compte du biais. Il est réalisé en transformant une partie des particules de matière noire, sélectionnées aléatoirement sous une loi uniforme, en galaxies.

La figure 4.2 présente une image des catalogues SDSS DR4-350 ainsi que des catalogues simulés MOCK, MOCK-NB et MOCK-NBNF (pour lequel ni le biais ni les distortions ne sont prises en compte). Dans tous les cas la géométrie est identique, mais les différences sont évidentes entre les catalogues. Dans MOCK-NB tout d’abord il est clair que le contraste de densité est moins élevé que dans MOCK, rendant les vides moins vides et les amas de galaxies moins denses. La différence la plus frappante provient cependant des distortions dues au décalage spectral qui apparaissent plus importantes dans les catalogues virtuels que dans SDSS (voir la figure 4.2(d) pour une version sans ces distortions). Ce problème provient d’une part de GaLICS lui même qui a tendance à surestimer la taille des amas mais surtout du fonctionnement de MoLUSC qui utilise un champ lissé sur une échelle de 1 Mpc et attribue les vitesses des particules de matière noire aux galaxies. Le résultat est que la taille des amas est surestimée ainsi que les vitesses particulières des galaxies, entraînant un effet de “finger of god” accru et bien visible. Heureusement, comme nous le verrons par la suite, l’influence de ce problème sur la mesure de la longueur totale du squelette est minime aux échelles de lissage que nous utiliserons.

Finalement, afin de pouvoir estimer la valeur la plus probable de  $\Omega_m$ , un jeu de 25 catalogues a été produit à partir de chacune des neuf simulations  $S_i$ . Pour chacun d’eux, le maximum de biais a été pris en compte et c’est toujours la même simulation galics3, de type  ${}^3\text{LCDM}_{256}^{100}$ , qui sert de support. Il est clair que cette configuration n’est pas idéale étant donné la dépendance de la formation galactique au modèle cosmologique considéré (et donc de la valeur de  $\Omega_m$ ). Cependant, étant donné que la mesure de la longueur du squelette dans les catalogues n’est pas très sensible à la valeur du biais comme nous le verrons par la suite, l’impact de cette approximation est finalement faible.



**Fig. 4.2:** Comparaison du catalogue SDSS (figure 4.2(a)) et des trois types de catalogues virtuels créés à partir de la simulation  $S$  de type  ${}^3\text{LCDM}_{512}^{1000}$  (figures 4.2(b) à 4.2(d)). Le catalogue virtuel MOCK est celui où le plus de biais observationnels sont pris en compte, dans MOCK-NB, le biais entre galaxies et matière noire est négligé d'où une distribution en densité moins contrastée et dans MOCK-NBNF, ni le biais, ni les distortions en redshift, ne sont prises en compte, d'où la présence de halos quasi-sphériques contrairement aux halos allongés des autres figures. La grande taille des “doigts de dieu” dans les catalogues simulés provient du fait que la taille des amas de galaxies est surestimée par GaLICS et que l'effet est renforcé par l'utilisation de MoLUSC. En pratique, ce n'est cependant pas gênant pour mesurer la longueur du squelette qui est peu sensible à cet effet lorsque la taille du lissage  $s$  est grande.

### 4.1.3 Lissage et échantillonnage du champ de densité

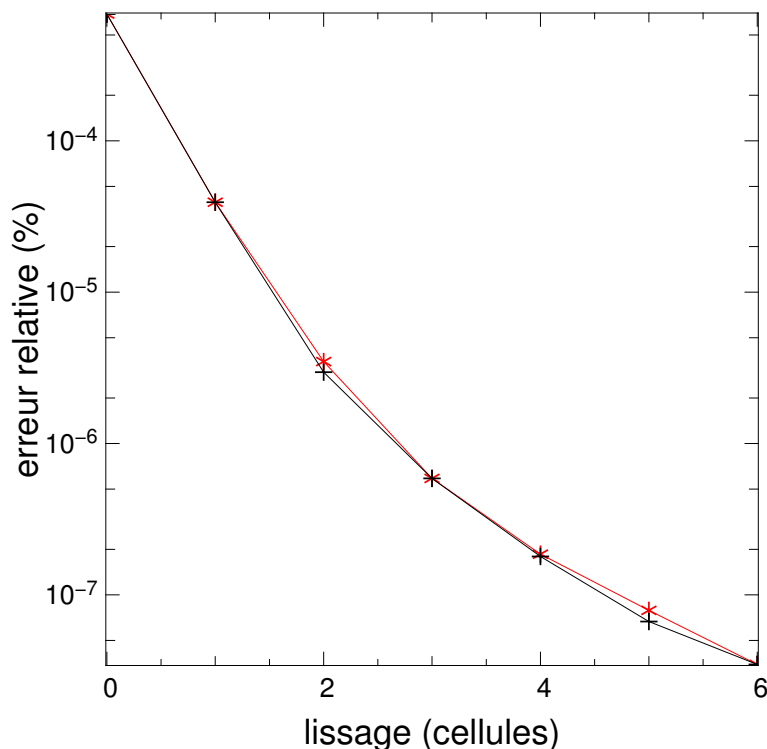
Comme il a été montré dans le chapitre 3, le calcul du squelette suppose un champ à la fois continu et dérivable deux fois. Connaître l'influence de la précision de l'échantillonnage du champ ainsi que du degré de lissage de ce dernier sur les résultats obtenus est donc de première importance. La résolution de la grille d'échantillonnage étant limitée par la capacité mémoire des ordinateurs, il est aussi important de déterminer l'intervalle des valeurs possibles pour la largeur de la fenêtre de lissage Gaussienne. C'est en effet ce paramètre qui donnera la limite de résolution inférieure que l'on pourra atteindre pour un champ de densité de taille donnée.

La continuité du champ et de ses deux premières dérivées conditionne la qualité du squelette obtenu, la méthode de calcul des dérivées est donc importante. Nous avons présenté deux méthodes dans la section 3.3.2 :

- Les différences finies, qui est une méthode rapide, simple et non contraignante mais non optimale en terme de réduction du bruit.
- La dérivation par FFT, plus lente, peu pratique dans le cas d'une distribution non périodique et demandant une plus grande quantité de mémoire, mais assurant un résultat optimal.

La figure 4.3 présente l'évolution de l'erreur maximale commise sur le calcul des dérivées premières et secondes en fonction de la taille du lissage gaussien exprimée en nombre de cellules. Le champ utilisé est un champ gaussien aléatoire d'indice spectral  $n = 1$  échantillonné sur une grille de  $128^3$  points. La valeur tracée est l'erreur maximale (en prenant le résultat par FFT comme référence) commise sur n'importe laquelle des six (respectivement trois) composantes du Hessien (respectivement du gradient) sur l'ensemble des valeurs échantillonnées sur la grille. Contrairement à ce que l'on pourrait penser, le résultat obtenu par la méthode des différences finies est donc quasiment identique à celui obtenu par FFT lorsque le champ est suffisamment continu, la différence sur chaque composante étant au maximum de l'ordre de  $10^{-5}\%$  pour un lissage sur seulement un pixel. C'est donc la méthode que nous utiliserons pour le calcul de tous les squelettes.

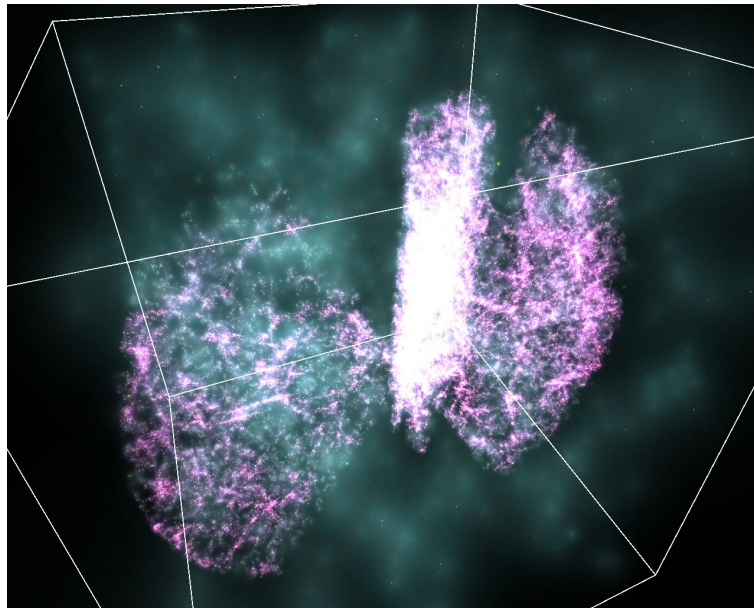
La détermination des paramètres d'échantillonnage et de lissage a déjà fait l'objet d'études dans la bibliographie. Dans Colombi *et al.* [8] notamment, les auteurs ont exploré l'ensemble des valeurs de lissage et tailles de grille possible pour le calcul du génus (équivalent de la caractéristique d'Euler, voir l'équation (3.13)). D'après leurs conclusions, les paramètres optimaux seraient une grille d'au moins  $N = N_g^3 = 256^3$  nœuds et un lissage de l'ordre de trois fois la taille de la boîte. Déterminer le génus revient en pratique à trouver l'ensemble des extréma du champ, dont la localisation ne nécessite le calcul que des dérivées premières. Dans notre cas, les dérivées secondes du champ interviennent et l'on peut s'attendre à ce qu'un lissage plus fort



**Fig. 4.3:** Comparaison des résultats obtenus pour le calcul de la dérivée première (rouge) et seconde (noir) avec la méthode des différences finies et par FFT. Les courbes donnent l'erreur maximale obtenue sur une grille de  $128^3$  cellules pour un champ gaussien aléatoire d'indice spectral  $n = 1$  en fonction de la taille du lissage gaussien exprimé en nombre de cellules.

soit nécessaire.

En procédant au calcul de plusieurs squelettes après variation des paramètres de grille et de lissage, il apparaît en pratique que le lissage doit être effectué sur un minimum de quatre pixels pour que le squelette reste lisse et continu. La résolution de la grille d'échantillonnage dépend du problème à étudier. Le critère retenu pour calculer sa taille minimale est qu'un morceau de filament devra être quasiment droit à l'échelle d'une cellule. En pratique, ce paramètre dépendra donc du problème étudié. En notant  $L$  la taille totale de la boîte, tous les squelettes calculés dans la suite le seront donc sur une grille de  $128^3$  nœuds au minimum avec un lissage  $s = L/N_g \geq 4$ . Sachant que la mémoire disponible sur les ordinateurs modernes nous limite à une grille de  $1000^3$  à  $1500^3$  nœuds, il est donc tout à fait possible en variant  $s$  et  $N_g$  de couvrir une vaste gamme d'échelles, la plus petite étant de l'ordre de trois millièmes de la taille de la boîte. Étant donné la limite de résolution des interactions gravitationnelles dans les simulations numériques, cette limite est largement suffisante. En effet, il ne sert à rien de vouloir descendre à une résolution très inférieure à la distance interparticulaire moyenne dans les simulations N-Corps (car alors le bruit de Poisson domine) et cette distance est au maximum de l'ordre du millième de la taille de la boîte.



**Fig. 4.4:** Afin de pouvoir en calculer le squelette, le catalogue de galaxies est inséré dans un cube rempli par une réalisation discrète d'un champ de densité aléatoire de valeur négligeable devant la densité moyenne à l'intérieur du catalogue. Cette image représente le champ de densité final obtenu (chaque particule étant ici remplacée par un noyau de type SPH).

#### 4.1.4 Calcul des squelettes

Jusqu'ici, nous nous sommes contentés de calculer les squelettes de distributions de densités simulées ayant une géométrie cubique bien définie. Effectuer la mesure sur le SDSS demande cependant certaines précautions. Une solution évidente consisterait simplement à inclure le catalogue dans une boîte cubique afin de pouvoir définir une grille d'échantillonnage. Le problème posé par cette solution est que, pour que le squelette existe, il est nécessaire que le gradient du champ s'annule en un nombre fini de points ce qui n'est pas le cas lorsque la densité à l'extérieur de la géométrie du catalogue est uniformément nulle. La solution que nous adopterons consiste donc à remplir l'espace hors du catalogue avec une réalisation discrète d'un champ aléatoire de densité négligeable devant celle du catalogue (voir la figure 4.4). De cette manière, la densité n'est pas uniformément nulle et la partie du squelette proche des bords du catalogue n'est absolument pas influencée par le champ extérieur.

La figure 4.5 illustre le résultat obtenu en appliquant cette méthode au calcul du squelette sur le catalogue DR4 coupé aux alentours de 500 Mpc en distance. Le catalogue a été inséré dans une boîte de 1400 Mpc de côté et une grille d'échantillonnage de  $512^3$  nœuds est utilisée. le champ de densité échantillonné est lissé sur une échelle de six tailles de cellules soit 16.4 Mpc. Un lissage de grande taille comme celui-ci permet de mettre en valeur la structure à grande échelle de la distribution des galaxies comme le montre les structures tracées par le squelette, le fameux grand mur du SDSS notamment qui est parfaitement détecté. Sur cette image, il est clair



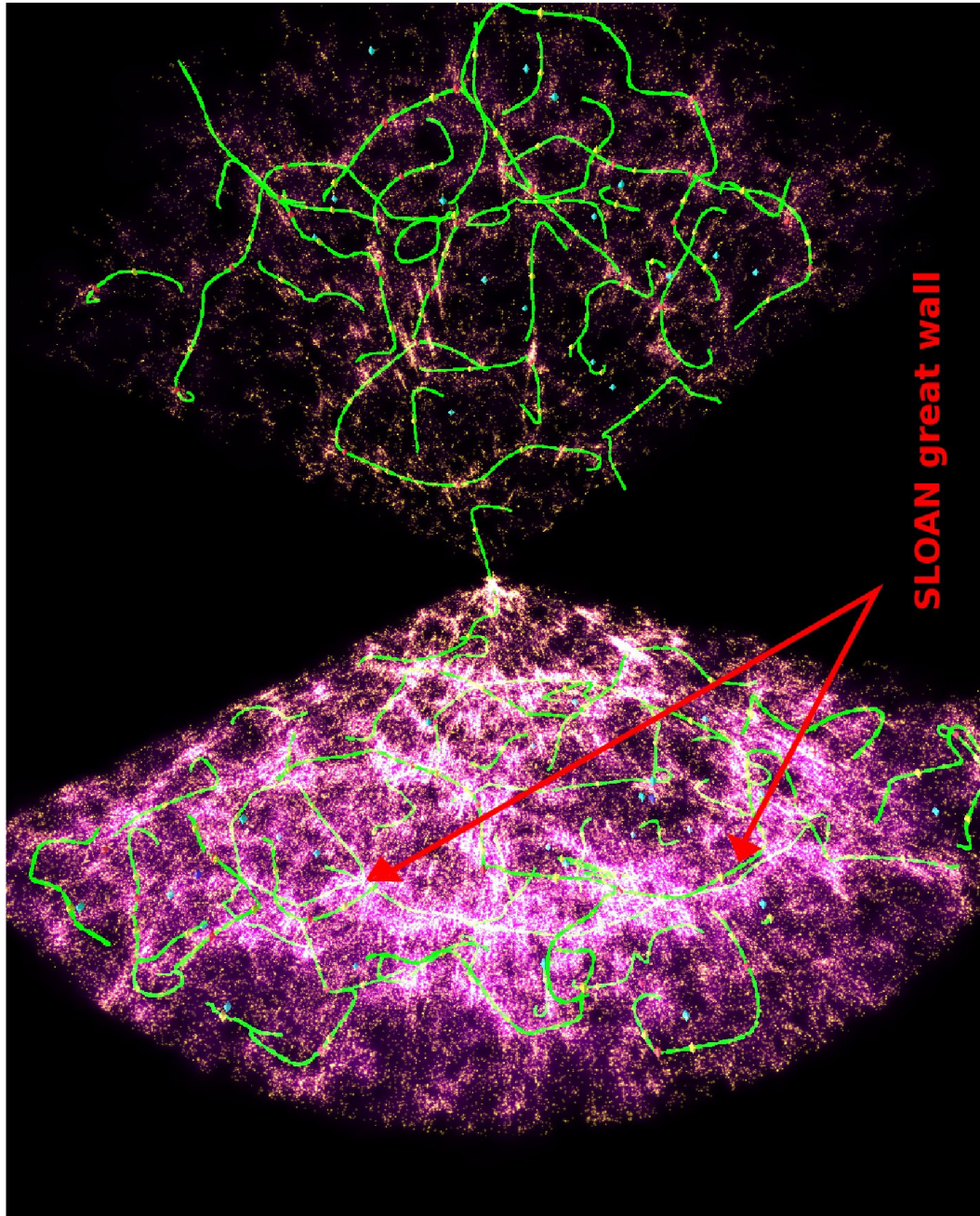
que le squelette détecte bien les structures identifiées visuellement. De plus, comme on peut le voir sur la partie droite, le chemin tracé par le squelette n'est absolument pas influencé par la présence de "doigts de dieu" puisqu'il passe littéralement au travers.

Afin de se faire une représentation plus claire de l'influence sur le squelette de la taille du lissage et de l'effet du biais de sélection des galaxies, la figure 4.6 montre les squelettes calculés sur les échantillons DR4-350 et DR4-VL350 (en haut à gauche et à droite respectivement) pour une échelle de lissage de 16.4 Mpc, ainsi que les squelettes du catalogue DR4-350 lissés à des échelles de 10.9 Mpc (en bas à gauche) et 8.2 Mpc (en bas à droite). La comparaison du squelette des 148,012 galaxies de DR4-350 à celui des 25,843 galaxies seulement du catalogue DR4-VL350 met tout d'abord en évidence le peu d'influence des effets de sélection. En sélectionnant seulement 15% des galaxies de manière à ce que celles restantes soient parfaitement représentatives d'une population, il est possible de retrouver un squelette très semblable et les filaments identiques peuvent facilement être identifiés sur les deux images. C'est un résultat auquel on pouvait s'attendre, la plupart des galaxies se trouvant de toute façon dans les filaments, une mauvaise sélection peut modifier les contrastes de densité à l'intérieur des filaments mais le lieu géométrique défini par les équations du squelette reste le même. Cela signifiera en pratique qu'il sera possible d'effectuer le calcul des squelettes sur des échantillons bruts, contenant un nombre de galaxies significativement plus élevé ce qui a pour effet de diminuer le bruitage des mesures et donc d'autoriser l'étude des filaments à de plus petites échelles (des mesures quantitatives sont faites dans la section suivante).

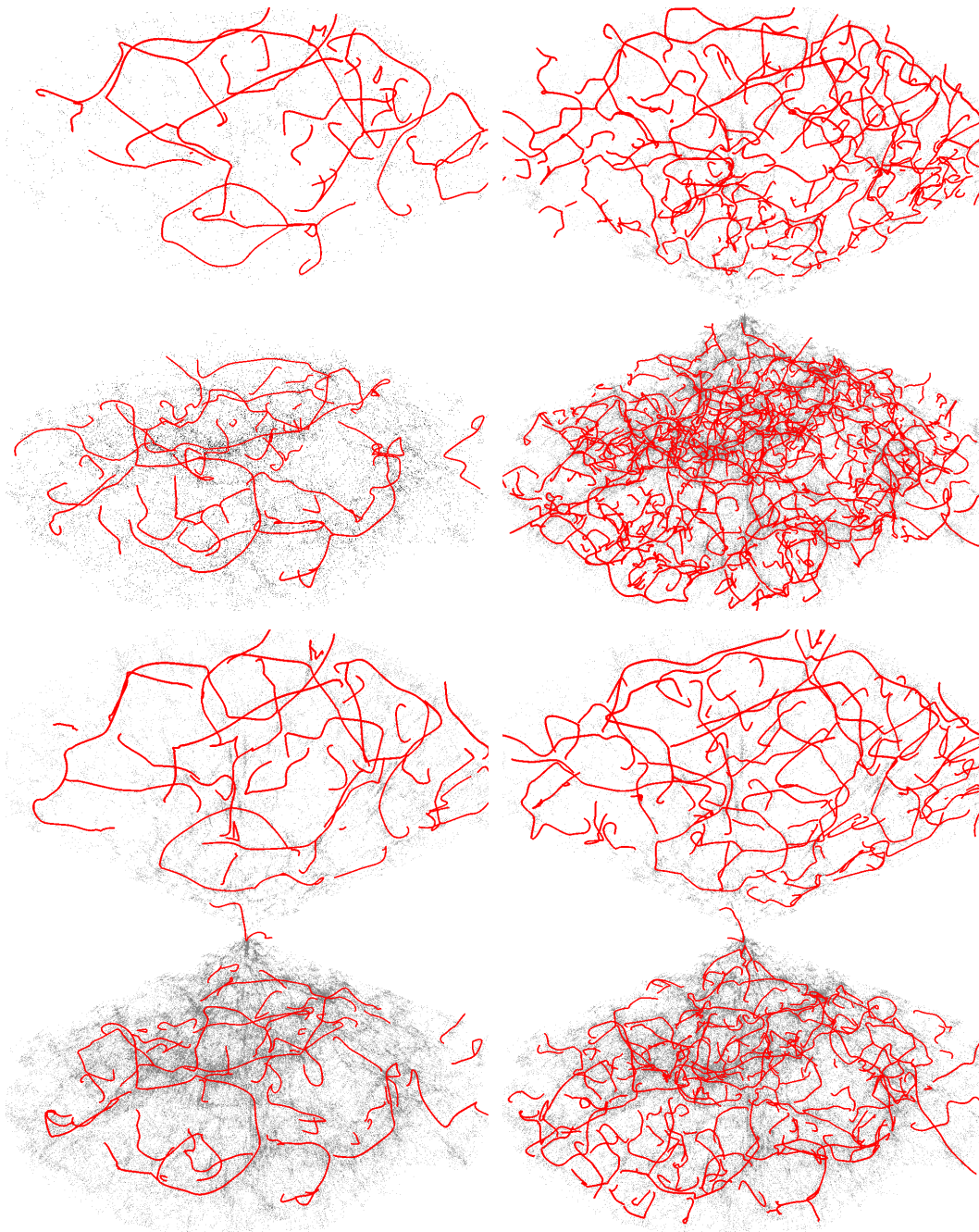
Les images obtenues pour les squelettes calculés à partir de champs lissés à de plus petites échelles sont elles aussi très encourageantes (partie inférieure de la figure 4.6). Il est clair en effet que quelle que soit l'échelle à laquelle le champ est lissé, ce sont les mêmes structures qui sont identifiées. Simplement, là où un filament plutôt continu est détecté à grande échelle (partie supérieure gauche de la figure 4.6), ce même filament est détecté avec des ramifications et plus sinueux lorsque les détails à des échelles plus petites sont pris en compte. Les conditions initiales utilisées pour les simulations N-Corps ont habituellement un spectre de puissance  $P(k) \propto k^n$  afin d'être compatibles avec les observations du CMB, ce qui les rend invariantes d'échelle. Il n'est donc pas étonnant que le squelette soit en quelque sorte auto-similaire à différentes échelles, suggérant ainsi une évolution de la longueur du squelette en puissance de l'échelle de lissage.

### 4.1.5 Résultats

Nous venons de voir qualitativement qu'il était possible de calculer le squelette de manière cohérente à partir d'un catalogue de galaxies à la géométrie complexe et ce malgré les biais observationnels. Afin de pouvoir atteindre notre objectif et utiliser le squelette comme contrainte sur la valeur de  $\Omega_m$ , une mesure plus quantitative



**Fig. 4.5:** Le squelette (en vert) calculé sur les galaxies du SDSS jusqu'à une distance approximative de 500 Mpc . Le champ de densité est lissé sur une échelle de 16.4 Mpc ce qui permet la détection des grandes structures uniquement. L'accord entre les structures détectées et celles visibles est excellent, notamment au niveau du grand mur. Le champ de densité du catalogue est ici représenté en remplaçant chaque particule par son noyau SPH calculé en prenant les 20 plus proches voisins.



**Fig. 4.6:** Ces quatre images représentent les squelettes du catalogue DR4-VL350 (haut-droit) lissés sur une échelle de 16.4 Mpc. Les deux images du bas montrent aussi le squelette du catalogue DR4-350 mais pour des échelles de lissage de 10.9 Mpc (gauche) et 8.2 Mpc (droite).

des propriétés de ce dernier est nécessaire.

### Mesures sur le SDSS et les catalogues virtuels

Nous avons choisi ici d'utiliser sa longueur par unité de volume comme caractéristique car il s'agit de loin de la grandeur la plus facilement mesurable et interprétable. Nous aurions aussi bien pu utiliser une grandeur telle que la courbure ou encore l'histogramme des longueurs de filaments mais cela aurait demandé le calcul de dérivées le long du squelette dans le premier cas, ou l'identification précise des extréma du champ dans le deuxième cas, rajoutant encore aux incertitudes de la mesure.

	$s$	DR4-350	DR4-VL350	-	-
Densité de longueur [Mpc/(100 Mpc) <sup>3</sup> ]	16.4	372	363	-	-
	10.9	795	772	-	-
	8.2	1271	1299	-	-
		MOCK	MOCK-PS	MOCK-AS	MOCK-NB
Densité de longueur [Mpc/(100 Mpc) <sup>3</sup> ]	16.4	390 ± 19	395 ± 16	362	403 ± 17
	10.9	796 ± 18	815 ± 19	740	790 ± 24
	8.2	1272 ± 25	1308 ± 27	1285	1204 ± 25

**Tab. 4.1:** *Densité de longueur du squelette mesuré dans les différents catalogues extraits du SDSS (DR4-350 et DR4-VL350) ou virtuels (MOCK, MOCK-PS, MOCK-AS et MOCK-NB). Chaque mesure est effectuée pour trois échelles de lissage “s” différentes.*

La table 4.1 présente les valeurs de la densité de longueur en Mpc/(100Mpc)<sup>3</sup> du squelette pour trois valeurs différentes du lissage et dans les différents catalogues (observés et virtuels) présentés précédemment. Chacune de ces mesures a été effectuée pour trois valeurs différentes du lissage variant de  $s = 8.2$  Mpc à  $s = 16.4$  Mpc. Cet intervalle a été choisi afin de couvrir la plus large gamme possible d'échelle tout en évitant les contaminations par les effets de grille décrits précédemment et en minimisant l'influence des fluctuations Poissoniennes. De plus, dans chacun de ces cas, seul le nombre de cellules de la grille d'échantillonnage a varié, la taille du lissage mesurée en nombre de cellules restant constante afin d'assurer une comparaison équitable des squelettes à différentes échelles. Ainsi, après avoir inclus les catalogues dans une boîte de  $700Mpc$ , les champs de densités ont été échantillonnés sur des grilles de  $256^3$ ,  $384^3$  et  $512^3$  cellules, puis lissés sur six pixels. Les mesures faites sur les catalogues virtuels sont le résultat d'une moyenne effectuée sur vingt cinq réalisations à chaque fois, afin de donner un ordre de grandeur à l'influence de la variance cosmique et aux erreurs de mesure non systématiques.

Les résultats obtenus sont très convaincants dans la mesure où l'accord entre toutes les valeurs est excellent quelle que soit l'échelle de lissage. La différence entre la densité de longueur de squelette de DR4-350 et celle de DR4-VL350 tout d'abord

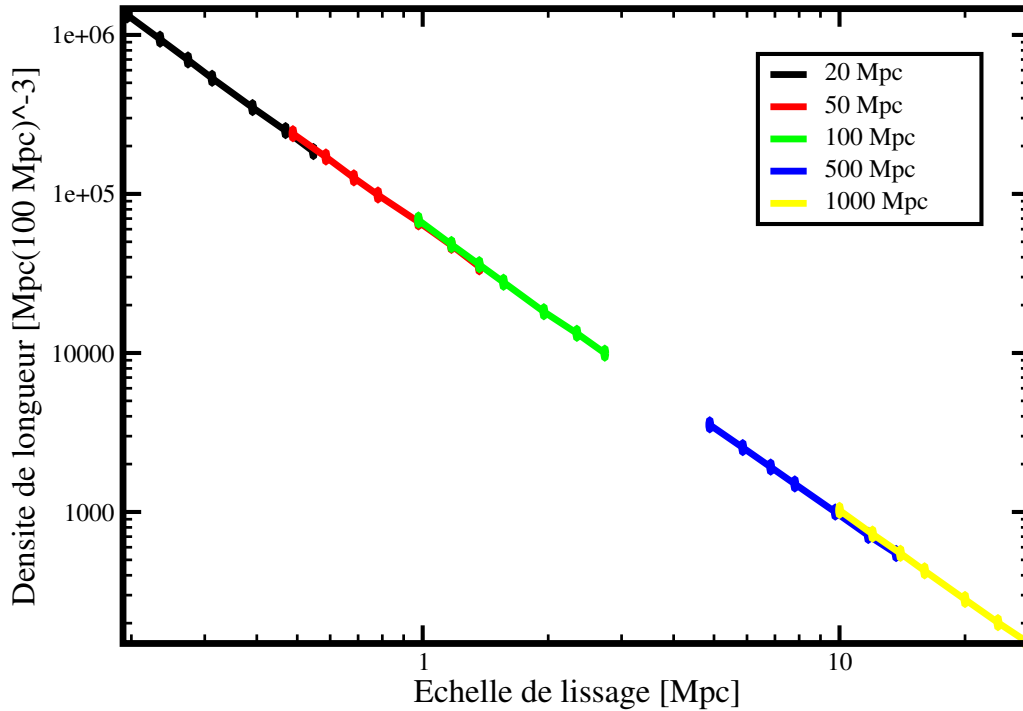
reste inférieure à 5%, démontrant le peu d'influence des effets de sélection sur la longueur du squelette. Cela signifie finalement que toutes les galaxies tracent les mêmes structures mais que leur répartition sur ces structures peuvent être différentes. Les mesures effectuées sur les catalogues virtuels mettent quant-à-elles en évidence le peu d'influence sur la mesure de la longueur du squelette des distortions de l'espace des décalages spectraux, de la géométrie du catalogue ou encore du biais entre matière noire et baryonique. Dans tous les cas, la différence entre deux catalogues est au maximum de l'ordre de 5% à 10% ce qui laisse supposer que même si certains effets n'étaient pas parfaitement bien modélisés dans nos catalogues virtuels, l'influence sur le résultat final serait minime. Finalement, les prédictions actuelles estimant la valeur de  $\Omega_m$  proche de 0.3, il est encourageant de constater que la longueur mesurée dans MOCK (pour lequel nous avons supposé  $\Omega_m = 0.3$ ) est tout à fait compatible avec la valeur mesurée dans SDSS, étant donnée la variance cosmique mesurée de l'ordre de  $20 \text{ Mpc}/(100 \text{ Mpc})^3$ .

Ce résultat ne permet cependant pas de faire une estimation quelconque des valeurs de  $\Omega_m$  acceptables. Les représentations du squelette de la figure 4.6 ainsi que les résultats obtenus pour différentes longueurs de lissage semble cependant suggérer qu'il pourrait être possible de modéliser la variation de la densité de longueur en fonction de l'échelle de lissage et par conséquent d'en extraire des grandeurs caractéristiques directement dépendantes de  $\Omega_m$ . A priori, nous nous attendons à trouver que la densité de longueur est une loi de puissance de l'échelle de lissage étant donné l'aspect invariant d'échelle du squelette. De plus, la grandeur mesurée étant un longueur par unité de volume, un exposant proche de  $-2$  est très probable et signifierait que le squelette est auto-similaire, c'est à dire qu'il possède les mêmes propriétés (tout du moins sa longueur) quelle que soit l'échelle considérée.

### Contraindre $\Omega_m$

Afin de tester cette hypothèse, nous avons effectué la mesure de la densité de longueur du squelette sur un ensemble de cinq simulations  $\Lambda$ CDM comportant toutes  $512^3$  particules mais dont la taille varie de 20 Mpc à 1,000 Mpc. Dans chacun des cas,  $\Omega_m$  est fixé à une valeur de 0.3 et le squelette est calculé sur une grille de  $512^3$  nœuds pour des échelles de lissage de cinq, six, sept, huit, dix, douze et quatorze tailles de cellules. Grâce à ce choix, les mesures effectuées sur une simulation recoupent celles effectuées sur la première simulation de taille supérieure et il est facile d'estimer la cohérence des données. La figure 4.7 présente les résultats obtenus.

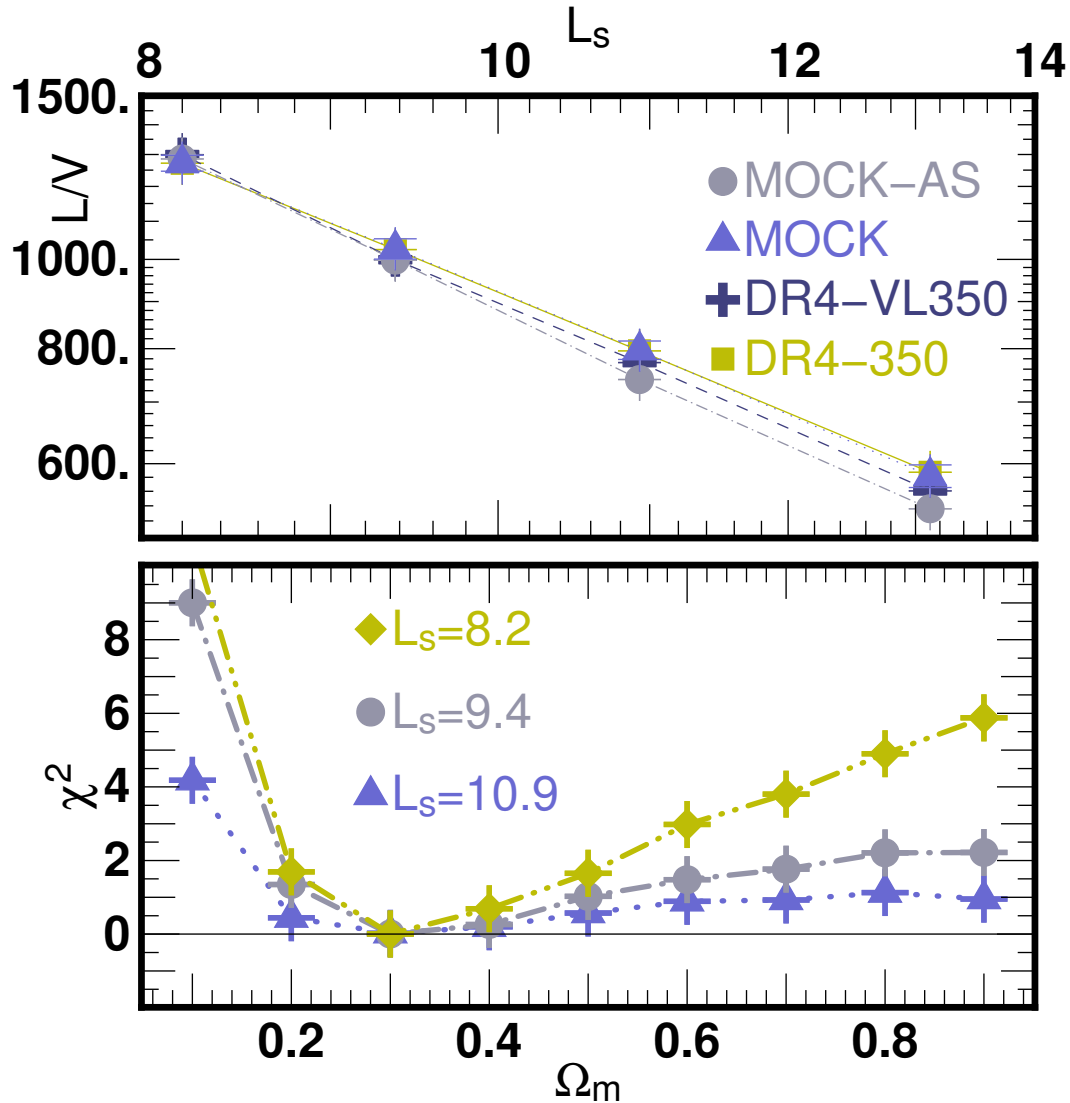
Il est assez frappant sur cette courbe de voir à quel point l'accord entre les différentes simulations est excellent : la densité de longueur varie clairement comme une puissance de l'échelle de lissage et ce, sur une vaste gamme de lissages (de 0.2 Mpc à 30 Mpc au moins, et l'on peut supposer que c'est vrai sur un domaine bien plus grand). L'ajustement des mesures permet d'extraire deux quantités caractéristiques, l'exposant de la loi de puissance et la constante de propor-



**Fig. 4.7:** *Mesure de la variation de la longueur du squelette par unités de volume en fonction de l'échelle de lissage pour un champ de matière noire simulé. Chaque couleur représente un ensemble de mesures faites pour différentes échelles de lissage sur une simulation identique. La taille des boîtes des simulations utilisées varie de 20 Mpc à 1,000 Mpc, ce qui permet d'effectuer la mesure sur une vaste gamme d'échelles.*

tionnalité. Pour les simulations décrites précédemment, l'ajustement de longueur par unité de volume  $\mathcal{L}$  en fonction de l'échelle de lissage  $L$  (en Mpc) donne  $\mathcal{L} = 63,711(L/\text{Mpc})^{-1.88} \text{Mpc}/(100 \text{Mpc})^3$ . Comme attendu, l'exposant est très proche de la valeur  $-2$ , l'écart pouvant être expliqué par le fait que le spectre de puissance effectif n'est pas exactement invariant d'échelle. Ce résultat est intéressant puisqu'il permet de prédire la longueur totale attendu des filaments à une échelle donnée dans un volume donné, ce qui constitue une grandeur jusqu'alors jamais mesurée mais qui est tout à fait caractéristique de la distribution de la matière.

Dans le but d'effectuer le même type d'ajustements sur le catalogue SDSS, les densités de longueur ont été calculées pour des échelles de lissage intermédiaires de celles répertoriées dans la table 4.1. Le résultat obtenu est que, dans le SDSS,  $\mathcal{L} = (52500 \pm 6500)(L/\text{Mpc})^{-1.75 \pm 0.06} \text{Mpc}/(100 \text{Mpc})^3$ . Il est assez remarquable que l'accord entre les simulations de matière noire et le SDSS soit aussi bon, mais cela



**Fig. 4.8:** Image du haut : Densité de longueur du squelette en  $\text{Mpc}/(100 \text{ Mpc})^3$  mesurée en fonction de l'échelle de lissage pour les deux catalogues SDSS et deux des catalogues virtuels. Les catalogues MOCK-NB et MOCK-PS ne sont pas représentés pour des raisons de clarté.

Image du bas : Statistique du  $\chi^2$  correspondant au carré de la différence entre la densité de longueur du squelette mesurée sur le SDSS et dans les catalogues virtuels réalisés à partir des simulations  $S_i$ . Cette valeur est exprimée en nombre de variance cosmique mesurée sur 25 réalisations de chaque catalogue.

s'explique par le fait que finalement, les deux tracent plus ou moins les filaments de la même distribution, à supposer que le modèle utilisé pour les simulations soit le bon. Le panneau supérieur de la figure 4.8 représente les ajustements obtenus dans le cas de certains catalogues virtuels et ceux obtenus pour DR4-350 et DR4-VL350. Les catalogues MOCK-NB et MOCK-PS ne sont pas représentés pour des raisons de clarté. Le quasi-parfait accord entre MOCK et DR4-350 suggère cependant que la valeur la plus probable de  $\Omega_m$  serait bien 0.3.

Pour évaluer le degré de contrainte sur la valeur de  $\Omega_m$ , une série de catalogues virtuels  $V_i$  a été réalisée à partir des simulations  $S_i$ . Une fonction  $\chi^2$  standard donnant le carré de l'écart sur la mesure de la densité de longueur entre le catalogue SDSS et chacun des catalogues  $V_i$  en unité de la variance cosmique mesurée sur 25 réalisations de chaque  $V_i$  est alors définie. Le panneau inférieur de la figure 4.8 représente cette fonction pour différentes échelles de lissage. Il apparaît tout d'abord que plus l'échelle de lissage est faible, plus la mesure est discriminante sur les valeurs possibles de  $\Omega_m$  et, dans le cas le plus favorable, il est donc possible d'estimer que  $0.15 < \Omega_m < 0.75$  à un niveau de confiance de deux sigmas ou encore  $0.25 < \Omega_m < 0.4$  à un niveau de confiance de un sigma.

## 4.2 Caractéristiques du flux de matière noire au voisinage du squelette

Comparée aux méthodes présentées dans la section 3.1, le squelette présente l'avantage de permettre une identification précise des filaments et de leur orientation. Dans ce paragraphe, nous allons nous servir de cette propriété afin d'essayer de comprendre qualitativement et quantitativement le rôle que jouent les filaments dans l'évolution du champ de densité de la matière noire.

### 4.2.1 Méthode de calcul

Afin de caractériser l'écoulement de la matière noire le long des filaments, il faut définir dans un premier temps une méthode efficace de calcul de ses propriétés mais aussi donner une définition à l'expression "le voisinage des filaments".

#### Moments

Nous avons décidé de nous restreindre ici aux deux premiers moments du champ de matière noire : sa vitesse moyenne et sa dispersion de vitesse. La matière noire étant non collisionnelle, la définition d'une pression n'a en effet pas de sens mais la dispersion en vitesse mesure une quantité similaire, on la notera donc abusivement  $P$  pour pression spécifique :

$$P_{ij} = \sqrt{\langle (v_i - \langle v_i \rangle)(v_j - \langle v_j \rangle) \rangle}. \quad (4.1)$$



Le squelette étant calculé à partir d'un champ de densité lissé, il paraît logique que les propriétés du flux de matière soient lissées à une même échelle. La méthode retenue consiste donc dans un premier temps à calculer la valeur moyenne des moments sur une grille dont chaque cellule est de l'ordre de grandeur de la taille de lissage utilisée pour calculer le squelette. La composante  $i$  de la vitesse moyenne des particules de matière noire dans la cellule  $n$  sera notée  $V_i^{(n)}$  et la dispersion dans cette même cellule  $P_{ij}^{(n)}$ . On a alors

$$V_i^{(n)} = \frac{1}{N^{(n)}} \sum_{q=1}^{N^{(n)}} v_i^{(q)} \quad (4.2)$$

et

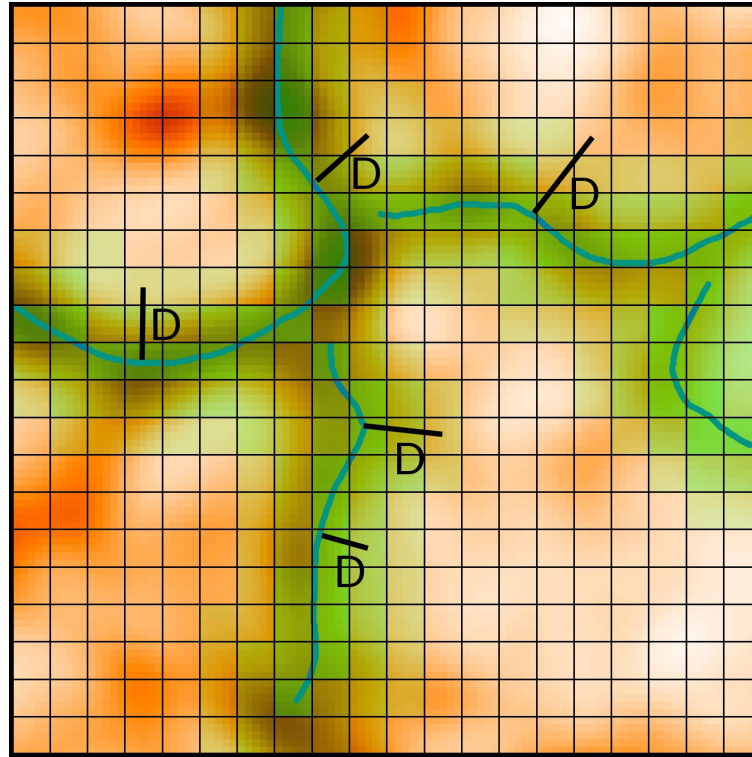
$$P_{ij}^{(n)} = \sqrt{\sum_{q=1}^{N^{(n)}} \left( v_i^{(q)} - V_i^{(n)} \right) \left( v_j^{(q)} - V_j^{(n)} \right)}, \quad (4.3)$$

où  $N^{(n)}$  désigne le nombre de particules dans la cellule  $n$  et  $v_i^{(q)}$  la composante  $i$  de la vitesse de la particule  $q$ . Il peut arriver dans les régions sous-denses que  $N^{(n)}$  soit trop petit (typiquement inférieur à 20), dans ce cas, ce ne sont plus simplement les particules de la cellule  $n$  qui sont prises en compte mais les 20 particules les plus proches du centre de cette cellule.

La question à laquelle nous désirons répondre dans ce chapitre est celle de la nature exacte des filaments et de leur rôle dans l'écoulement de la matière noire. Il est donc très important d'obtenir la mesure la moins biaisée possible de l'orientation des moments en fonction de l'orientation du squelette. Comme il a été expliqué au chapitre 3, le squelette est constitué d'un ensemble de segments dont la taille est typiquement de l'ordre du tiers de celle d'une cellule d'échantillonnage du champ de densité. Chacun de ces segments nous fournit donc localement une direction dont nous voulons étudier le rapport au comportement des particules de matière noire proches et dans une moindre mesure de celles qui sont plus éloignées. Soit un ensemble de grandeurs  $A_i^{(n)}$ , calculées pour un ensemble de cellules d'une grille  $G$  ( $n \in \{1, \dots, N_G\}$  avec  $N_G$  le nombre de cellules). On désire mesurer la fonction de distribution des probabilités (PDF)  $P(A_i(\mathbf{s}))$  des grandeurs  $A_i$  autour d'un filament orienté selon  $\mathbf{s}$ . En d'autres termes, on désire connaître la probabilité  $P(a_i)da_i$  qu'ont les différentes variables  $A_i$  de prendre les valeurs  $a_i$ . Cette PDF est estimée numériquement en considérant chaque segment du squelette puis en calculant la valeur des  $A_i^{(n)}$  pour chacune des  $N_g$  cellules de  $G$ ,  $n$  désignant l'index de la cellule. Chacune de ces valeurs contribue alors à  $P(A_i)$  selon sa distance au segment de squelette (voir la figure 4.9). Afin de tenir compte d'une pondération liée à la distance au filament, il est possible de définir un estimateur de la PDF pour les grandeurs  $A_i$  par :

$$P(A_i(\mathbf{s})) = \sum_{m=1}^{N_s} \sum_{n=1}^{N_G} A_i^{(n)}(\mathbf{s}^{(m)}) \exp(-d_{mn}/l), \quad (4.4)$$

où  $\mathbf{s}^{(m)}$  désigne l'orientation du  $m^{\text{ème}}$  segment du squelette,  $N_s$  est le nombre total de segments du squelette,  $d$  est la distance entre le segment  $m$  et la cellule  $n$  et  $l$  permet



**Fig. 4.9:** *Illustration de la méthode de calcul des moments du champ de matière noire. Le champ de densité est représenté en orange et le squelette en bleu. La valeur moyenne des quantités à mesurer est calculée pour chacune des cellules. Les cellules en transparence vertes sont celles contribuant au calcul, avec un poids dépendant de leur distance au morceau de squelette considéré.*

d'ajuster la taille de la zone autour du squelette dans laquelle la mesure est effectuée.

En pratique, si l'échelle de lissage du champ de densité est  $L$ , la valeur de  $l$  est fixée à  $L/2$  et le nombre de cellules dans la grille  $G$  est fixé de manière à ce que le nombre de particules dans chaque cellule soit en moyenne de l'ordre de 20. L'estimateur (4.4) permet alors de calculer les propriétés du flux de matière dans un rayon autour des filaments de l'ordre de grandeur de l'échelle à laquelle ces filaments ont été calculés. En effet, le champ de densité ayant été lissé sur un échelle  $L$  cela signifie que les détails du squelette d'une taille inférieure à  $L$  ont disparu.

### Moment cinétique (Spin)

Nous venons de décrire une méthode de mesure des moments du flux de matière noire. Nous savons cependant aussi que la matière noire a tendance à former des halos en rotation dans lesquels se créent la plupart des galaxies. Cette formation est influencée par le moment cinétique (spin) de ces halos et il est donc intéressant de connaître l'influence des filaments sur le spin des halos. L'identification des halos est effectuée grâce à un algorithme de type friend-of-friend (voir l'annexe ??). Une fois

celle-ci faite, le spin  $\mathcal{S}_n$  du  $n^{\text{ème}}$  halo est calculé simplement en utilisant la formule :

$$\mathbf{l}_n = \sum_{i=1}^{N_n} \mathbf{r}^{(i)} \times \mathbf{v}^{(i)}, \quad (4.5)$$

avec  $\mathbf{v}^{(i)}$  la vitesse relative de la  $i^{\text{ème}}$  particule du halo et  $\mathbf{r}^{(i)}$  sa position par rapport au centre du halo.

La méthode de calcul de la PDF consiste alors à trouver, pour chaque halo  $n$ , le segment de squelette le plus proche de son centre et à calculer son orientation moyenne  $\mathbf{s}_n$  en utilisant l'orientation des ses deux segments directement voisins. Il devient alors possible de calculer la probabilité que le spin du halo ait une orientation donnée connaissant l'orientation du filament le plus proche :

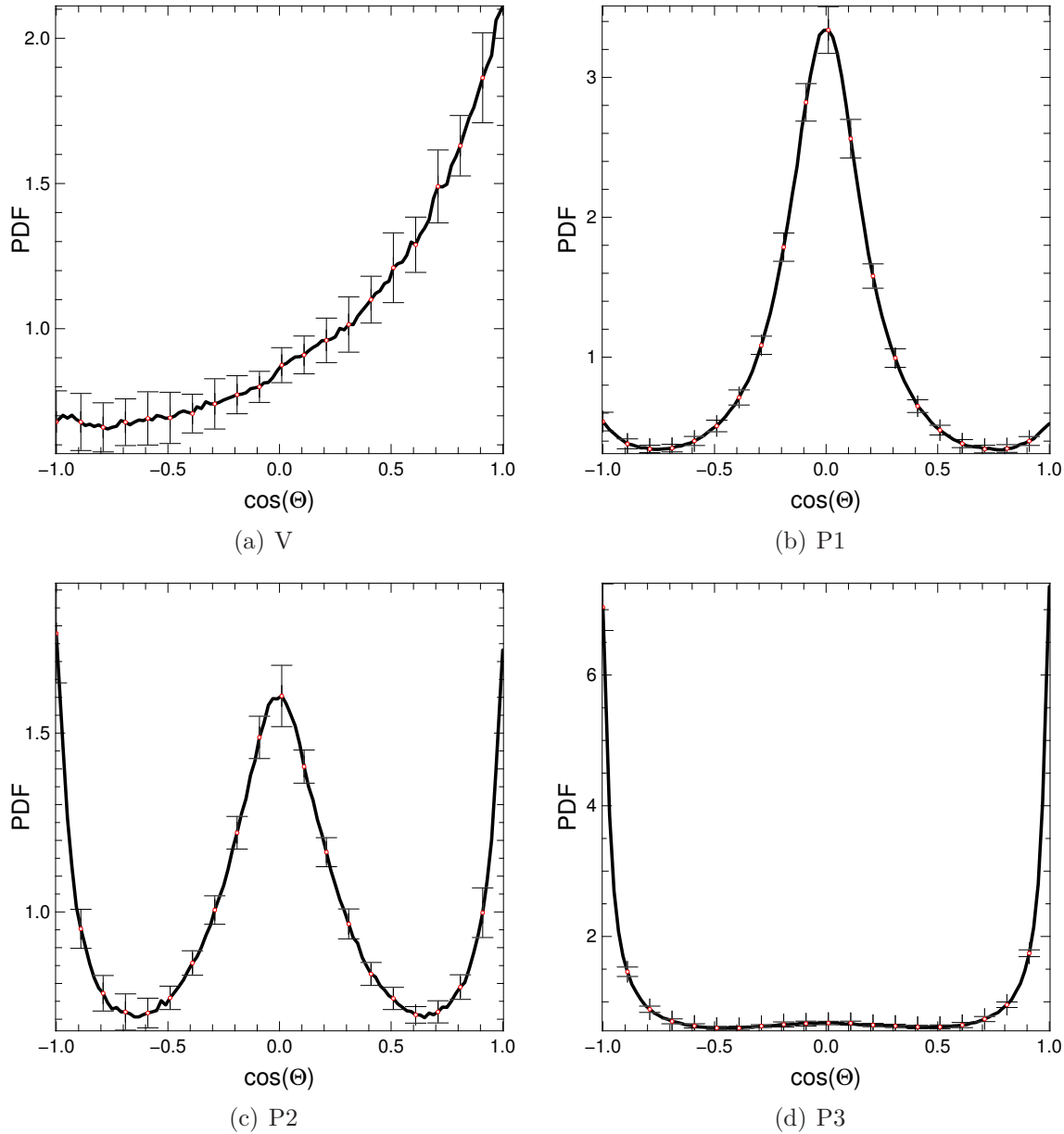
$$P(\cos(\theta)) = \sum_{n=1}^{N_H} \frac{\mathbf{l}_n \cdot \mathbf{s}_n}{\|\mathbf{l}_n\| \|\mathbf{s}_n\|}, \quad (4.6)$$

avec  $\theta$  l'angle entre le filament et le spin du halo.

## 4.2.2 Mesure des moments le long du squelettes

Afin d'étudier l'influence des filaments sur le flux de matière noire, nous avons dans un premier temps procédé à la mesure de la PDF des angles formés par le champ de vitesse et de dispersion avec la direction locale des filaments. La figure 4.10 page 137 présente la fonction de distribution de probabilité d'alignement du champ de vitesse (figure 4.10(a)) ainsi que des vecteurs propres du champ de dispersion avec le squelette local. Les figures 4.10(b) à 4.10(d) présentent les résultats obtenus pour les trois vecteurs propres du champ de dispersion,  $\mathbf{P1}$ ,  $\mathbf{P2}$  et  $\mathbf{P3}$  de valeurs propres respectives  $p1 > p2 > p3$ . Pour le calcul de ces PDF, quinze simulations de type  ${}^{0.3}\text{LCDM}_{100}^{256}$  ont été utilisées, le squelette ayant été calculé sur une grille d'échantillonnage de  $256^3$  cellules lissées sur six tailles de grille (soit  $\approx 2.3h^{-1}$  Mpc).

Dans chacun des cas, la PDF est présentée en fonction du cosinus de l'angle entre la grandeur mesurée et le filament local (noté  $\cos(\theta)$ ). Les filaments étant orientés dans le sens des densités croissantes, une valeur de  $\cos\theta = 1$  signifiera un alignement de la grandeur vectorielle avec la direction du filament, une valeur de  $\cos\theta = 0$  l'orthogonalité des deux directions et  $\cos\theta = -1$  un antialignement. La figure 4.10(a) met clairement en évidence le fait que le flot de matière noire présente une tendance marquée à s'aligner avec les filaments. Comme on pouvait s'y attendre, la matière noire a donc tendance à remonter les filaments vers les zones plus denses (i.e. les halos). La proportion de particules allant à contre-sens n'est cependant pas négligeable et nous verrons par la suite s'il est possible ou non de différencier ces deux types de populations.



**Fig. 4.10:** Fonctions de distribution des probabilités d’alignement du champ de vitesse  $\mathbf{V}$  et des vecteurs propres  $\mathbf{P1}$ ,  $\mathbf{P2}$  et  $\mathbf{P3}$  du champ de dispersion associés aux valeurs propres  $p1 > p2 > p3$ . Chacune des PDF est représentée en fonction du cosinus de l’angle (noté  $\cos(\Theta)$ ) entre le champ étudié et les filaments. Le vecteur vitesse semble préférentiellement aligné avec les filaments ce qui signifie que les particules remontent les filaments vers les halos plus denses (le sens positif est celui des densités croissante). Au contraire, l’axe de dispersion principale  $\mathbf{P1}$  est orthogonal aux filaments et l’axe de dispersion minimale  $\mathbf{P3}$  leur est parallèle : la dispersion est plus forte orthogonalement aux filaments, probablement en raison de l’accrétion de matière.

Les trois graphiques 4.10(b), 4.10(c) et 4.10(d) nous renseignent quant-à-eux sur la manière dont les filaments accrètent la matière environnante. Le vecteur principal  $\mathbf{P1}$  présente une tendance très marquée à être orthogonal aux filaments, c'est donc orthogonalement aux filaments que la dispersion de vitesse est la plus importante. Cette remarque est confirmée par le fait que l'axe de plus faible dispersion  $\mathbf{P3}$  est principalement aligné avec le filament. On peut donc penser que les filaments accrètent la matière des régions sous-denses environnantes, celle-ci acquérant en tombant une certaine vitesse orthogonalement aux filaments, d'où la forte dispersion dans cette direction. Cette dispersion serait donc une dispersion "d'arrêt" dû au fait que les particules tombent sur les filaments avec une certaine vitesse orthogonale. Le flux de matière a alors tendance à suivre la région localement sur-dense formée par le filament et l'on peut donc penser que les filaments constituent des sortes de routes empruntées par la matière se dirigeant vers les halos massifs. Il est à noter que les barres d'erreurs représentent la moyenne de l'écart à la valeur moyenne de la mesure sur l'ensemble des 15 simulations réalisées avec des conditions initiales différentes. Ces simulations sont donc toutes clairement en accord malgré le nombre relativement faible de particules ( $256^3$ ).

### 4.2.3 Influence comparée du squelette à différentes échelles

Une simple analyse angulaire ne permet pas de mettre en évidence l'existence de différents régimes dans le flux de matière noire. Les filaments existent de plus à toutes les échelles, il est donc intéressant de mesurer leur influence respective sur leur environnement à différentes échelles. Afin de répondre à ces questions, nous avons calculé les fonctions de distribution de probabilité de l'angle formé par les filaments et les différents moments mais aussi de l'intensité de ces moments. Les figures 4.11 à 4.13 (pages 140 à 142) présentent ces distributions calculées sur des simulations  $\Lambda$ CDM standards de  $512^3$  particules et de tailles  $20h^{-1}$ ,  $50h^{-1}$ ,  $100h^{-1}$ ,  $500h^{-1}$  et  $1000h^{-1}$  Mpc. Dans chacune de ces simulations, le squelette est calculé sur une grille de  $256^3$  cellules lissée sur 6 pixels soit des échelles de lissage de  $0.4h^{-1}$ ,  $1.2h^{-1}$ ,  $2.34h^{-1}$ ,  $10.7h^{-1}$  et  $23.4h^{-1}$  Mpc respectivement et les champs de vitesse et de dispersion sont échantillonnés sur une grille de  $128^3$  cellules.

La figure 4.11 met en évidence l'évolution du comportement du champ de vitesse en fonction de l'échelle, faisant apparaître une échelle de transition claire sur la simulation de  $100h^{-1}$  Mpc. En dessous de cette limite, la très grande majorité des particules ont un vecteur vitesse aligné avec les filaments alors qu'une plus faible proportion sont antialignés (ceux présentant une vitesse plus faible). Il est très probable que ces particules soient en fait influencées par les inhomogénéités à plus grandes échelles : un filament relie en effet deux halos. La composante antialignée disparaît à partir de  $100h^{-1}$  Mpc et l'on peut observer que l'alignement des vitesses des particules avec les filaments est de moins en moins bon quand l'échelle augmente, le flux est donc plus "désordonné". Ainsi, si la majorité des vitesses des particules forme un angle inférieur à 70 degrés avec les filaments dans la boîte de  $1000h^{-1}$  Mpc, presque toutes celles des particules de la simulation de  $20h^{-1}$  forment un angle inférieur à 35

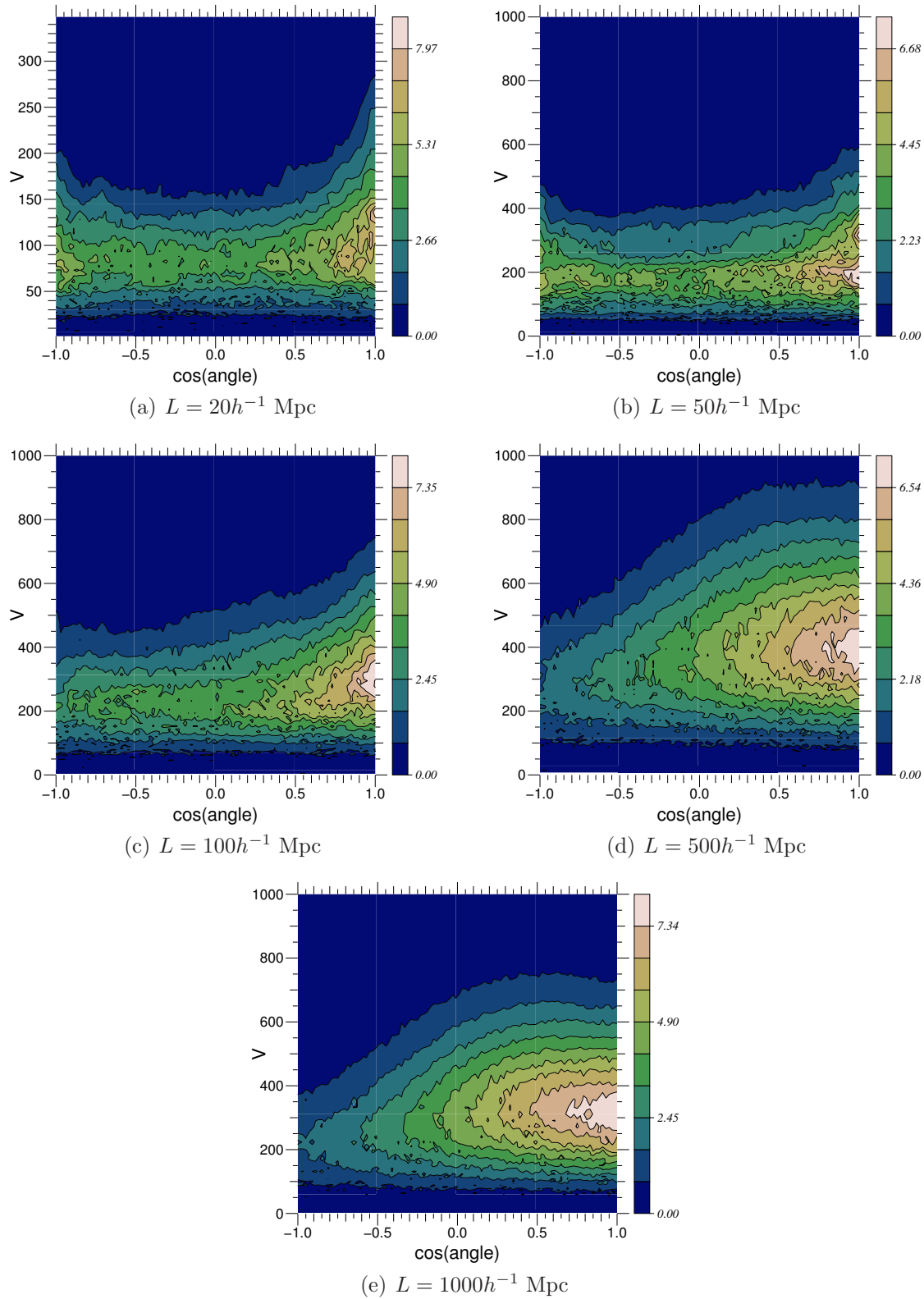
degrés avec les filaments. L'écoulement dans les filaments semble donc plus cohérent à petite échelle qu'à grande échelle.

L'examen des propriétés des vecteurs propres du champ de dispersion est tout aussi instructif et confirme la transition observée pour la simulation de  $100h^{-1}$  Mpc. Il devient clair que la PDF de l'axe de dispersion maximale (figure 4.12 page 141) présente deux comportements différents qui de plus dépendent fortement de l'échelle observée. Aux petites échelles de lissage ( $L < 2h^{-1}$  Mpc), une partie significative (et même majoritaire pour les plus petites échelles) des particules présente en effet une dispersion de vitesse **P1** alignée avec les filaments et il est clair que cette partie des particules est celle des zones où la dispersion est la plus élevée (figure 4.12(a)). A partir d'une échelle de lissage de  $\approx 2.5h^{-1}$  Mpc, la tendance s'inverse. La fraction de particules pour lesquelles le vecteur propre principal du champ de dispersion est aligné aux filaments diminue et, contrairement au cas précédent, ces zones sont celles où la dispersion est faible. On peut en conclure que, si pour les petites échelles l'alignement (ou l'orthogonalité) de **P1** avec les filaments est très marquée, pour les grandes échelles la dispersion est beaucoup plus grande. Il semble donc à priori que les filaments de petite échelle soient beaucoup plus nets que ceux à grande échelle et jouent un rôle plus important.

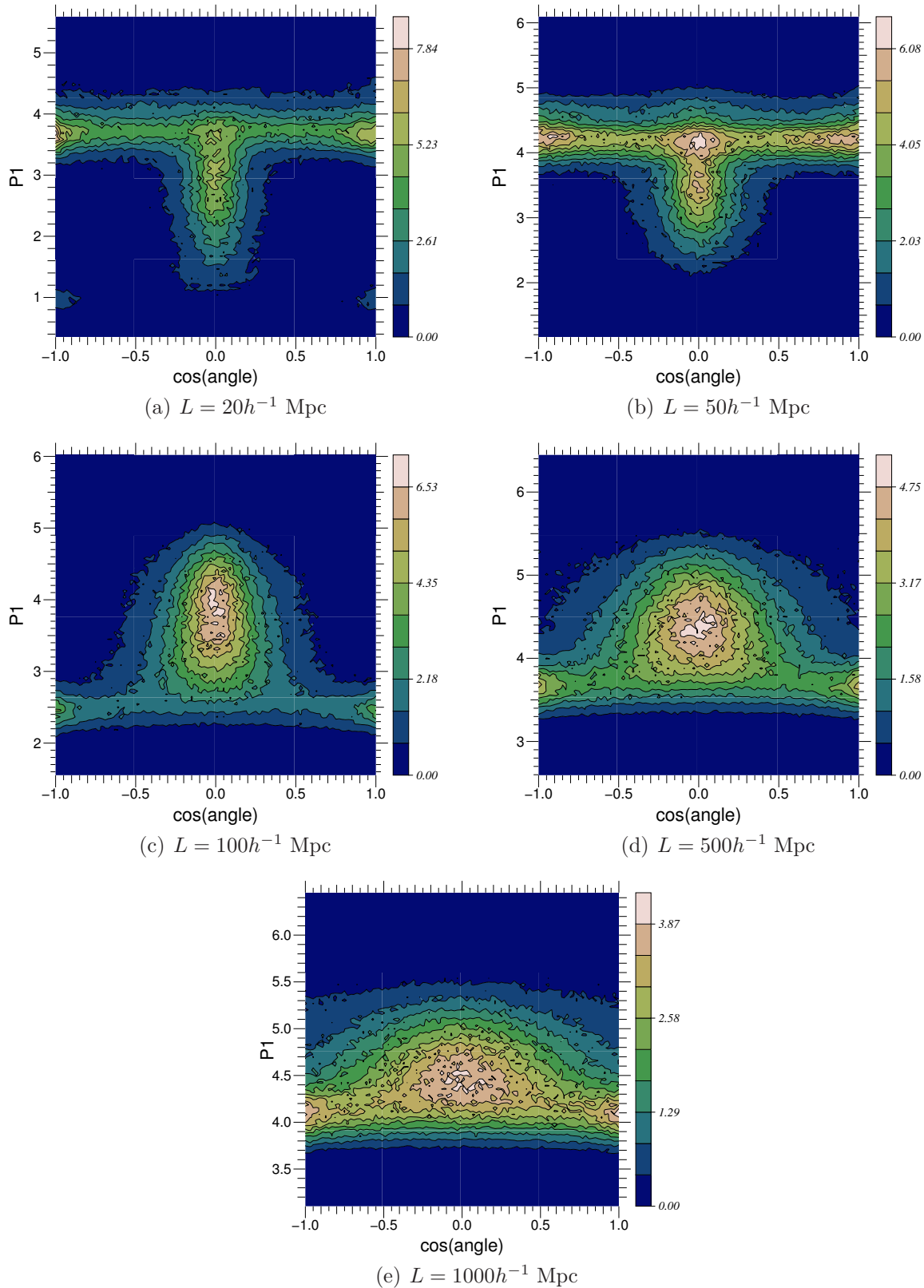
Les figures 4.13 (page 142) confirment largement ces observations pour le vecteur **P3**. Rappelons tout d'abord que  $P_{ij}$  étant une matrice symétrique, ses vecteurs propres sont forcément orthogonaux. Le vecteur **P2** est donc orthogonal à **P1** et **P3** et son comportement leur est complémentaire. La direction de **P3** est celle selon laquelle la dispersion des vitesses est minimale et il est remarquable que ce soit toujours principalement avec l'axe du filament que le vecteur **P3** est aligné. A nouveau, cette tendance est très marquée à petite échelle mais devient moins nette pour les boîtes de plus de  $100h^{-1}$  Mpc. Il reste cependant clair que, quelle que soit l'échelle considérée, l'écoulement dans les filaments est de type laminaire.

#### 4.2.4 Evolution avec le décalage spectral

La question de l'évolution avec le temps de la nature des flux et de leur relation aux filaments se pose alors. La figure 4.14 page 144 présente les PDF de la vitesse et du vecteur propre principal de la dispersion **P1**, obtenues sur la simulation de  $100h^{-1}$  Mpc pour trois décalages spectraux différents :  $z = 0$ ,  $z = 1$  et  $z = 15$ . Comme l'on pouvait s'y attendre, la tendance qu'ont les particules à se déplacer le long des filaments existe pour toutes les époques mais il est net qu'elle est beaucoup plus marquée lorsque  $z = 0$ . Le vecteur **P1** présente d'ailleurs exactement le même comportement, celui-ci restant toujours principalement orthogonal aux filaments, mais de manière de plus en plus marquée. On peut cependant remarquer que la dispersion maximale présente une plus forte tendance à s'aligner avec les filaments à grand décalage spectral. Pour résumer, on peut dire que l'écoulement de la matière observée à  $z = 15$  sur une échelle de  $100h^{-1}$  Mpc semble être similaire à celui observé à  $z = 0$  sur une échelle de l'ordre de  $1000h^{-1}$  Mpc (voir les figures 4.11(e) et 4.12(e)).

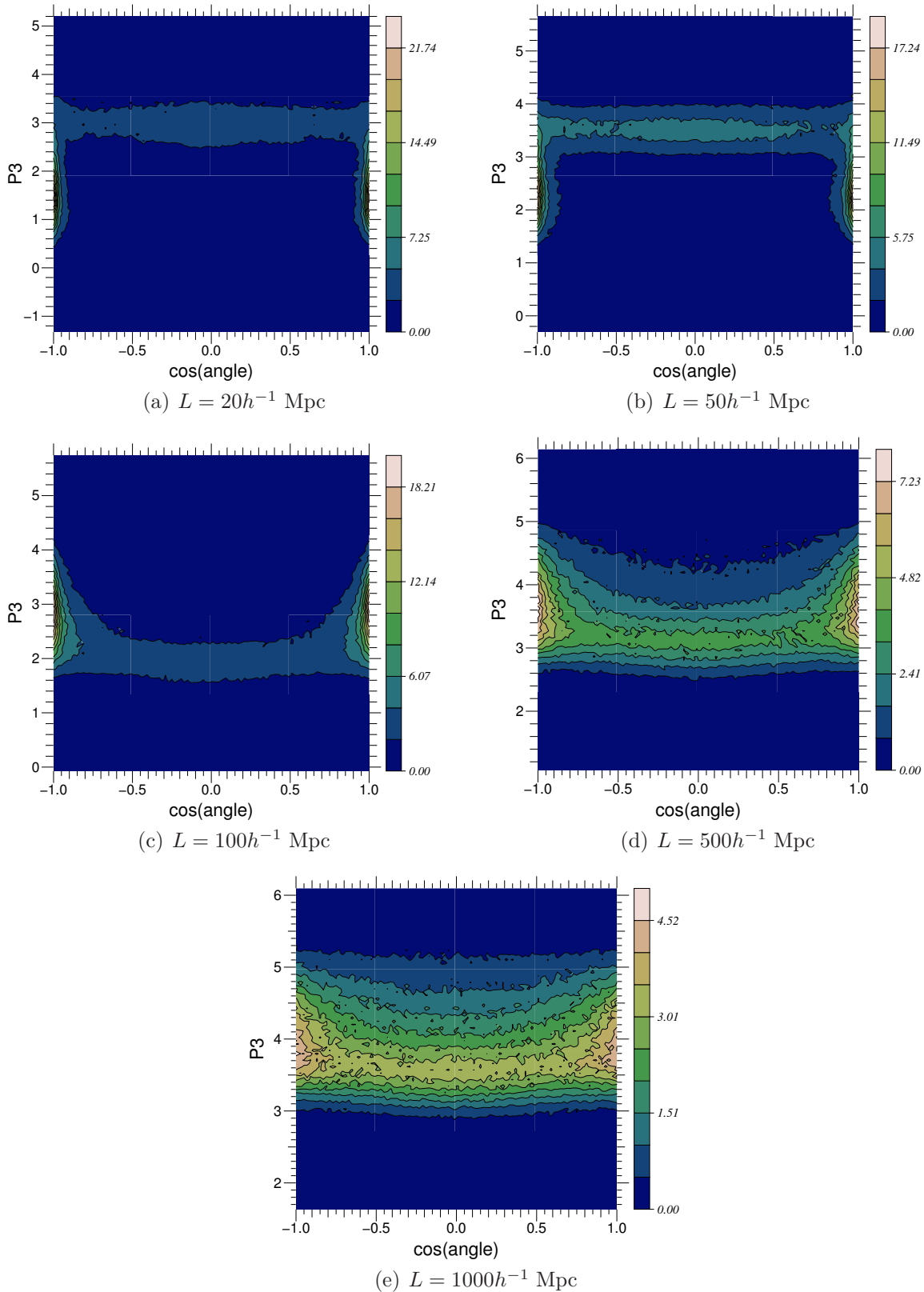


**Fig. 4.11:** Fonctions de distribution de probabilité de l'angle et de l'intensité du champ de vitesse des particules de matière noire à proximité des filaments. La PDF est donnée en fonction du cosinus de l'angle fait avec le filament, orienté dans la direction des densités croissantes. Le champ de vitesse est principalement aligné avec les filaments et cette tendance est d'autant plus marquée que l'échelle est grande. Aux petites échelles, une partie du flux de faible vitesse remonte les filaments.



**Fig. 4.12:** Fonctions de distribution de probabilité de l'angle et de l'intensité de la norme du vecteur propre principal du champ de dispersion  $\mathbf{P1}$  des particules de matière noire à proximité des filaments. La PDF est donnée en fonction du cosinus de l'angle fait avec le filament, orienté dans la direction des densités croissantes. L'axe  $\mathbf{P1}$  est principalement orthogonal aux filaments, sans doute en raison de la vitesse orthogonale de la matière noire tombant vers les halos, accrétée par les filaments avec une vitesse orthogonale élevée.





**Fig. 4.13:** Fonctions de distribution de probabilité de l'angle et de l'intensité du vecteur propres de valeur propre la plus faible du champ de dispersion  $\mathbf{P3}$  des particules de matière noire à proximité des filaments. La PDF est donnée en fonction du cosinus de l'angle fait avec le filament, orienté dans la direction des densités croissantes. L'axe de faible dispersion est clairement aligné avec les filaments suggérant un flux uniforme de la matière noire remontant les filaments.

### 4.2.5 Moment cinétique des halos

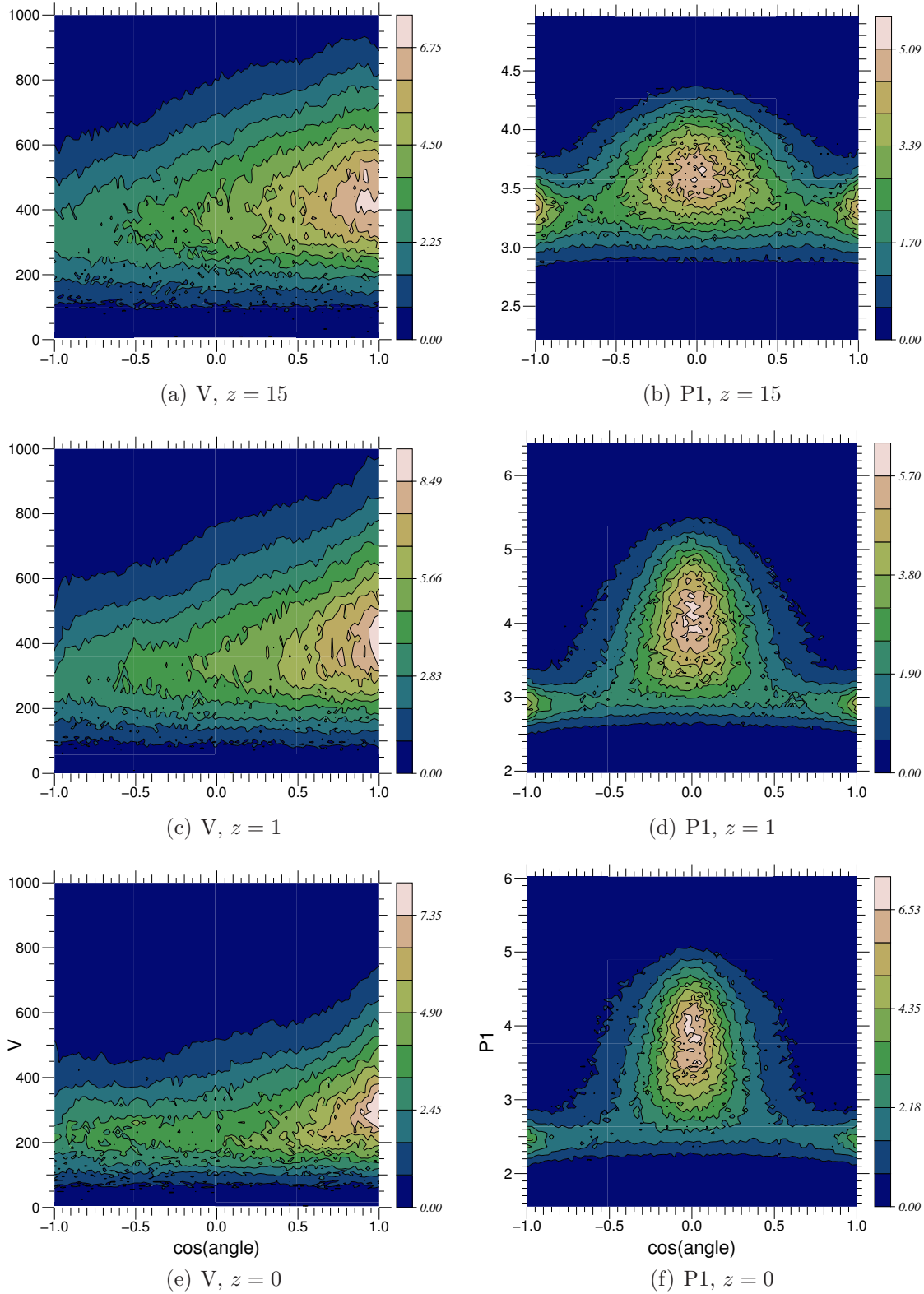
Les mesures précédentes permettent de se faire une idée du comportement du flux de matière au cours de l'évolution de l'univers. Les corrélations observées avec la position et l'orientation des filaments sont maintenant claires, mais nous n'avons pas encore mesuré leur influence sur les propriétés des halos. Une des propriétés les plus intéressantes des halos est sans doute le spin car il reflète en quelque sorte l'historique d'accrétion de la matière noire. Binney and Silk [5] ont suggéré que les interactions des amas de galaxies proches par effets de marée devaient conduire à des anisotropies et qu'il devait par conséquent être possible d'observer un alignement des axes d'élongation maximales sur des distances de l'ordre de quelques mégaparsecs. Les premières mesures numériques furent effectuées par van Haarlem *et al.* [19] qui montra que les amas ont tendance à être étirés dans la direction de la dernière fusion. Plus récemment, Faltenbacher *et al.* [10] confirmèrent ces mesures et montrèrent même qu'il existe une corrélation entre les axes d'élongation maximale des amas voisins. Ces corrélations semblent s'étendre sur des distances dépassant  $10h^{-1}$  Mpc ce qui va dans le sens d'une influence importante des filaments de matière.

Le squelette semble clairement être un outil particulièrement adapté à ce type de mesures et la figure 4.15 présente les résultats obtenus. Les figures 4.15(a) à 4.15(c) donnent la valeur de la PDF de l'angle entre le filament le plus proche et le spin en fonction de la distance à ce dernier, pour trois masses minimales de halos  $M_H$  différentes :  $M_H > 6.2 \cdot 10^{10} M_\odot$ ,  $M_H > 62.1 \cdot 10^{10} M_\odot$  et  $M_H > 621 \cdot 10^{10} M_\odot$  respectivement. L'excès de probabilité d'une orthogonalité entre le spin et le filament le plus proche est claire et paraît même augmenter avec la masse des halos. Cette corrélation diminue cependant avec la distance au filament pour disparaître au-delà de  $3h^{-1}$  Mpc.

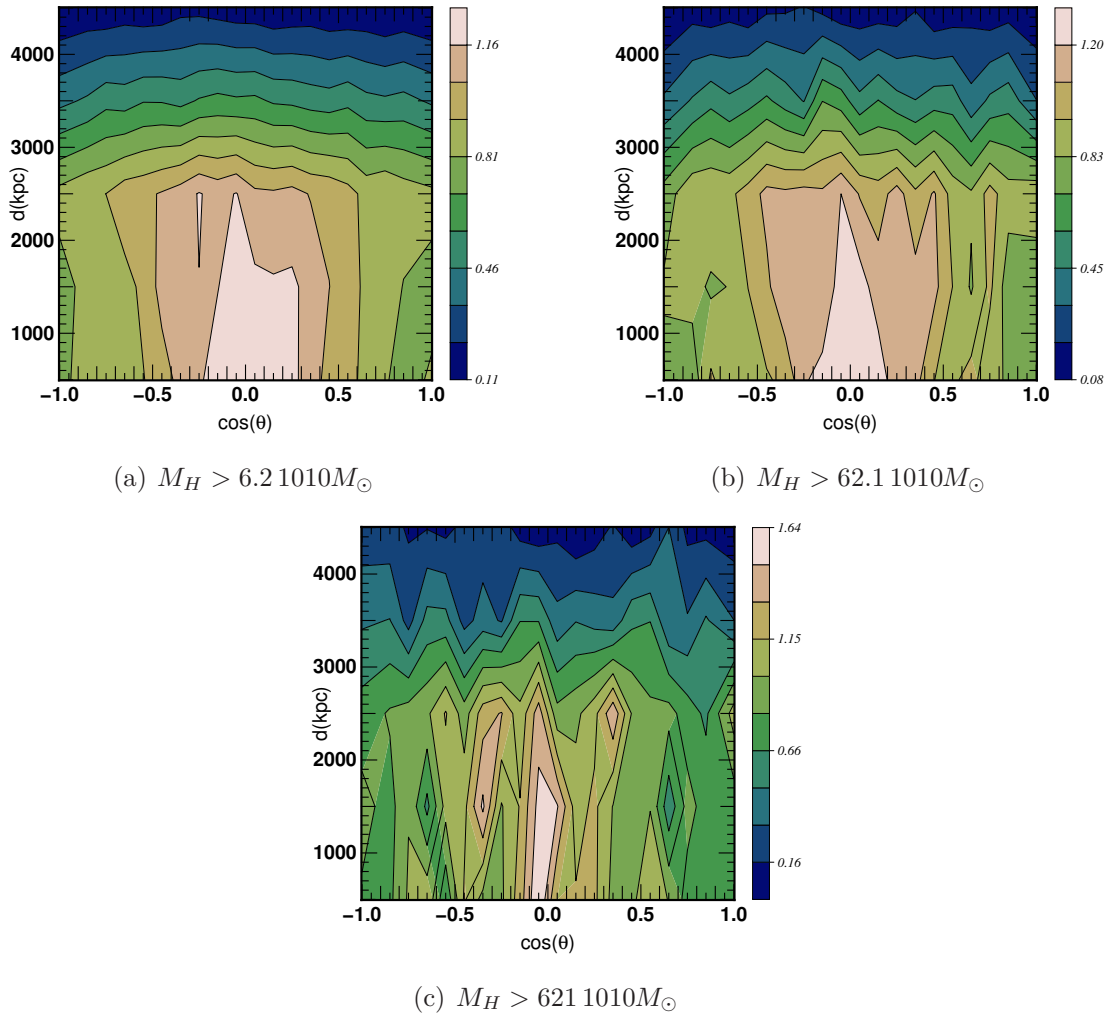
### 4.2.6 Propriétés des flux de matière dans les filaments

La connaissance de la nature et du rôle des filaments dans l'évolution de la distribution de matière est primordiale pour une modélisation efficace des processus physiques intervenant lors de la formation galactique. La méthode du squelette permet pour la première fois d'effectuer ces mesures directement sur les filaments. Les résultats obtenus, en plus de confirmer ceux obtenus précédemment sur l'orientation des spins de halos notamment (voir [3]), montrent qu'il est tout à fait possible de modéliser la nature statistique des flux de matière. Les résultats obtenus sont résumés dans le tableau 4.2.

L'intérêt de la connaissance précise de la localisation des filaments ne se limite cependant pas au calcul des propriétés du flux de matière noire. De nombreuses applications sont possibles et parmi elles, nous comptons en développer deux au moins



**Fig. 4.14:** Evolution des PDF du champ de vitesse et de l'axe principal de dispersion en fonction du cosinus de l'angle avec le filament et de l'intensité. Le filament est orienté dans le sens des densités croissantes. L'alignement (respectivement l'orthogonalité) du champ de vitesse (respectivement de dispersion principale) est d'autant plus marqué que le décalage spectral est faible. A noter que les caractéristiques de l'écoulement de la matière à  $z = 15$  sur une échelle de  $100h^{-1}$  Mpc sont assez similaires à celles mesurées sur une échelle de  $1000h^{-1}$  Mpc à  $z = 0$  (figures 4.12(e) et 4.11(e)).

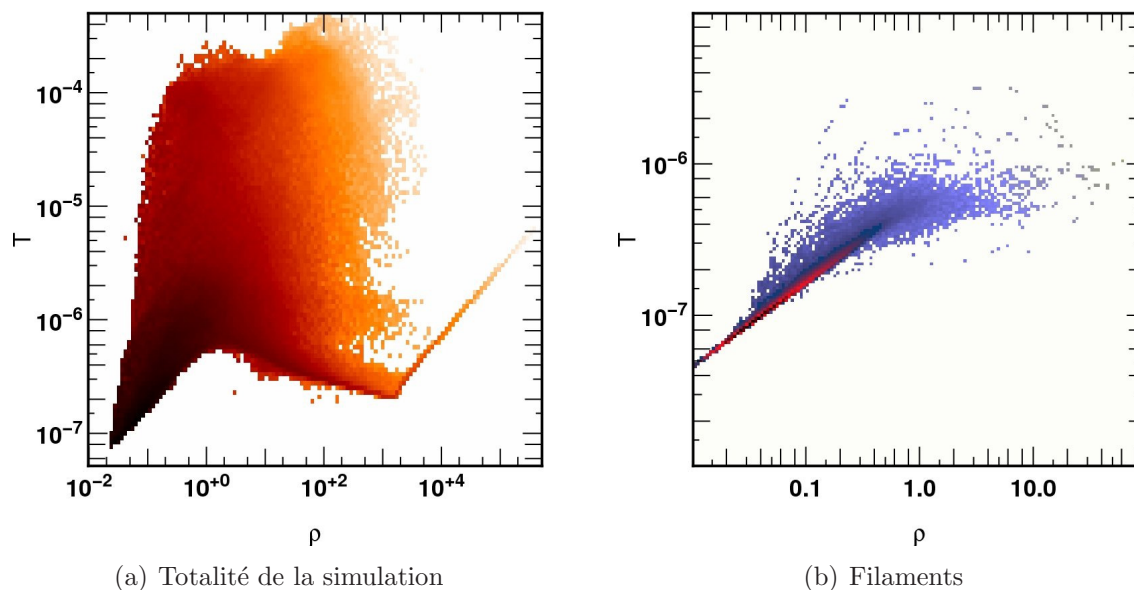


**Fig. 4.15:** *Distribution de la probabilité d'alignement du spin des halos de matière noire en fonction de la distance aux filaments pour différentes masses  $M_H$  de halos. Ces résultats ont été calculés à partir de quatre simulations de taille  $100h^{-1}$  Mpc comportant  $512^3$  particules. Dans chacun des cas, le squelette est calculé à partir d'un champ lissé à une échelle de  $2.3h^{-1}$  Mpc. Les spins des halos présentent une nette tendance à être orientés orthogonalement aux filaments.*

	<p>La matière noire est accrétée par les halos par l'intermédiaire des filaments. Avec l'évolution de l'univers, les filaments sont les zones de transit de la matière présente dans les zones sous-denses (vides), attirée par les halos massifs. La matière noire a tendance à s'écouler le long des filaments avec un flux d'autant plus régulier que les filaments sont de petits. A décalage spectral nul, on peut estimer à <math>\approx 30</math> degrés l'angle avec lequel le flux pénètre les filaments à une échelle de l'ordre de <math>1h^{-1}</math> Mpc et <math>\approx 70</math> degrés à une échelle d'une vingtaine de mégaparsecs.</p>
	<p>Le moment cinétique des halos contenus dans les filaments présente une forte tendance à être orthogonal à ceux-ci. Cette tendance est d'autant plus forte que les halos sont proches du coeur des filaments. Le flux de matière y est parallèle aux filaments, la matière est donc préférentiellement accrétée dans l'axe de ceux-ci. Plus on s'éloigne des filaments et plus le moment cinétique des halos semble aléatoire, aucune corrélation n'étant détectée au delà de l'échelle de lissage utilisée pour calculer le squelette. Les halos de forte taille, ayant accru beaucoup de matière provenant des filaments, présentent de plus une tendance plus marquée à une orthogonalité de leur spin aux filaments.</p>

Tab. 4.2: Résumé des propriétés du flux de matière noire.

dans un proche avenir. L'analyse des propriétés des galaxies en fonction de la nature de leur environnement tout d'abord. Pandey and Bharadwaj [16] par exemple ont effectué ce genre de mesures en utilisant les fonctionnelles de Minkowsky afin de mesurer le degré de filamentarité de la distribution des galaxies en fonction de leur nature (luminosité, couleur et type morphologique). Les auteurs montrent ainsi que si les galaxies elliptiques peuplent principalement les halos, les galaxies spirales se situent principalement dans le réseau de filaments. L'utilisation du squelette per-



**Fig. 4.16:** *Exemple d'application du squelette : le calcul de l'équation d'état du gaz dans la simulation AMR MareNostrum à  $z = 4$ , pour l'ensemble de la simulation et à proximité des filaments. La température apparaît ainsi clairement se comporter comme une puissance de la densité dans les filaments alors que ce n'est pas le cas dans le reste de la distribution.*

mettrait clairement une caractérisation plus précise de ce genre de propriétés, grâce notamment à la possibilité de mesurer des distances le long des filaments. Une autre application réside dans une comparaison des propriétés du flux de gaz et de matière noire à l'aide de simulations numériques hydrodynamiques. Des propriétés du gaz diffèrent en effet largement selon la région observée. La figure 4.16 présente par exemple l'équation d'état du gaz  $T = f(\rho)$  dans la simulation hydrodynamique AMR du projet Horizon<sup>3</sup> et met en évidence le comportement spécifique du gaz dans les filaments (figure 4.16(b)) par rapport au reste de la distribution (figure 4.16(a)).

## 4.3 Autres applications et perspectives

### 4.3.1 Test de Alcock-Paczynski (A-P)

#### Principe

Le principe du test d'Alcock-Paczynski (voir Alcock and Paczynski [2]) est de comparer des distances longitudinales (mesurées grâce au décalage spectral) à des distances transverses (i.e. angulaires). Pour cela, une sorte de “règle cosmique” statistiquement isotrope est nécessaire. A partir de cette règle, on peut mesurer le rapport  $\delta z / \delta \theta$  de son extension longitudinale (en décalage spectral) à son extension

<sup>3</sup><http://www.projet-horizon.fr>

transverse (angulaire). Le rapport de ces extensions exprimées en distance comobile devant être de un, il est alors possible en utilisant les équations (1.36) et (1.42) de relier la grandeur mesurée aux paramètres cosmologiques. Le test A-P a déjà connu de nombreux succès en cosmologie, principalement lors de son application aux catalogues spectroscopiques de moyen (Matsubara and Szalay [12] et Matsubara and Szalay [13]) et haut décalage spectral (da Ângela *et al.* [9]) mais aussi aux mesure spectroscopiques de forêt Lyman- $\alpha$  (McDonald and Miralda-Escudé [14] ou Rollinde *et al.* [18]). Dans chacun de ces articles, c'est l'isotropie de la fonction de corrélation qui est testée ce qui n'est pas sans poser des problèmes. En effet, aux échelles où les corrélations peuvent être considérées comme fortes, la déformation due à la distortion de l'espace des décalages spectraux est importante. Cet effet, du à l'influence des vitesses particulières de galaxies dans les régions effondrées (i.e. dans le régime non linéaire principalement), rend les mesures artificiellement anisotropes et il est donc nécessaire de le corriger. De plus, même aux échelles où la théorie linéaire est valide, il a été montré que ces déformations dépendent du biais (voir Hamilton [11]).

Le squelette trace le réseau des filaments, à priori parfaitement isotrope. Sa faible sensibilité aux effets de biais devrait donc en faire un excellent candidat pour le test A-P, à condition toutefois que le champ de densité soit lissé sur une échelle suffisamment grande pour ne pas prendre en considération les régions dans le régime non-linéaire.

### Prédictions théoriques

En pratique, le squelette est constitué d'un ensemble de filaments eux même constitués de segments s'étendant sur une fraction de la taille de chaque cellule de la grille d'échantillonnage. L'orientation des filaments doit être isotrope et pour chacun des segments, on peut mesurer le rapport de sa taille projetée sur l'axe de visée ( $\delta r$ ) à sa taille projetée sur l'axe azimuthal ( $\delta\theta$ ) où l'axe d'inclinaison ( $\delta\phi$ ). Un estimateur  $\hat{C}_{\theta/\phi}$  peut alors être défini comme l'un de ces deux rapports. Si chaque segment  $d\mathbf{l}_i$  est normalisé, on peut écrire :

$$d\mathbf{l}_i = (\sin \xi_i \cos \psi_i, \sin \xi_i \sin \psi_i, \cos \xi_i), \quad (4.7)$$

où les angles  $\xi_i$  et  $\psi_i$  sont définis relativement à l'axe de visée, et l'estimateur  $C_{\theta/\phi}$  s'exprime alors comme :

$$\hat{C}_{\theta/\phi} = \left( \sum_{i=1}^N \|\mathbf{e}_r \cdot d\mathbf{l}_i\| \right) / \left( \sum_{i=1}^N \|\mathbf{e}_{\theta/\phi} \cdot d\mathbf{l}_i\| \right). \quad (4.8)$$

Prenons un Univers de référence de paramètres cosmologiques  $(\Omega_m^{\text{ref}}, \Omega_k^{\text{ref}})$ . Si ce choix est le bon, la valeur mesurée de  $\hat{C}_{\theta/\phi}$  est de  $\hat{C}_{\theta/\phi} = 1$  et tout écart à cette mesure signifiera que le choix de valeurs de  $(\Omega_m^{\text{ref}}, \Omega_k^{\text{ref}})$  n'était pas le bon. D'après les équations (1.42) et (1.40), une petite extension radiale comobile  $\delta r$  à un décalage

spectral  $z$  s'exprime en fonction de son extension en décalage spectral  $\delta z$  comme :

$$\delta r \propto \int_z^{z+\delta z} \frac{dz}{E(z)} \propto E(z)\delta z. \quad (4.9)$$

Dans le modèle de référence, l'extension radiale d'un de nos segments en fonction de son extension mesurée est donc :

$$\delta r_i = \mathbf{e}_r \cdot d\mathbf{l}_i = \|\cos \xi_i\| \frac{E(z)}{E^{\text{ref}}(z)}. \quad (4.10)$$

avec  $\cos \xi_i$  l'extension radiale du segment  $i$ . Il est aussi possible en utilisant l'équation (1.42) d'obtenir l'extension transversale  $\delta\theta$  pour un petit angle  $\delta\alpha$  à un décalage spectral  $z$  :

$$\delta\theta \propto f_k(\chi(z))\delta\alpha. \quad (4.11)$$

Dans le modèle de référence, l'extension transversale d'un de nos segment est donc :

$$\delta\theta_i = \mathbf{e}_\theta \cdot d\mathbf{l}_i = \|\cos \psi_i \sin \xi_i\| \frac{f_k^{\text{ref}}(\chi(z))}{f_k(\chi(z))} \quad (4.12)$$

avec  $\cos \psi_i \sin \xi_i$  l'extension azimuthale du segment  $i$ .

Soit  $w(z, \theta, \phi)$  le nombre de segments de squelette par unité de volume aux coordonnées  $(z, \theta, \phi)$ , la valeur théorique attendue pour la mesure de la valeur de l'estimateur  $\hat{C}_\theta$  est donc :

$$\hat{C}_\theta = \frac{\iiint \|\cos \xi_i\| \frac{E(z)}{E^{\text{ref}}(z)} w(z, \theta, \phi) dz d \cos \theta d\phi}{\iiint \|\cos \psi_i \sin \xi_i\| \frac{f_k^{\text{ref}}(\chi(z))}{f_k(\chi(z))} w(z, \theta, \phi) dz d \cos \theta d\phi}. \quad (4.13)$$

Pour un univers isotrope,  $\langle \cos \xi_i \rangle / \langle \cos \psi_i \sin \xi_i \rangle = 1$ , par conséquent, à un décalage spectral  $z$  donné, on a :

$$C_\theta(z) \approx \frac{E(z)/E^{\text{ref}}(z)}{f_k^{\text{ref}}(\chi(z))/f_k(\chi(z))}. \quad (4.14)$$

En prenant comme paramètres de référence ( $\Omega_m^{\text{ref}} = 0.3, \Omega_k^{\text{ref}} = 0$ ) et dans l'approximation d'un faible décalage spectral, l'équation (4.14) peut être approximée par :

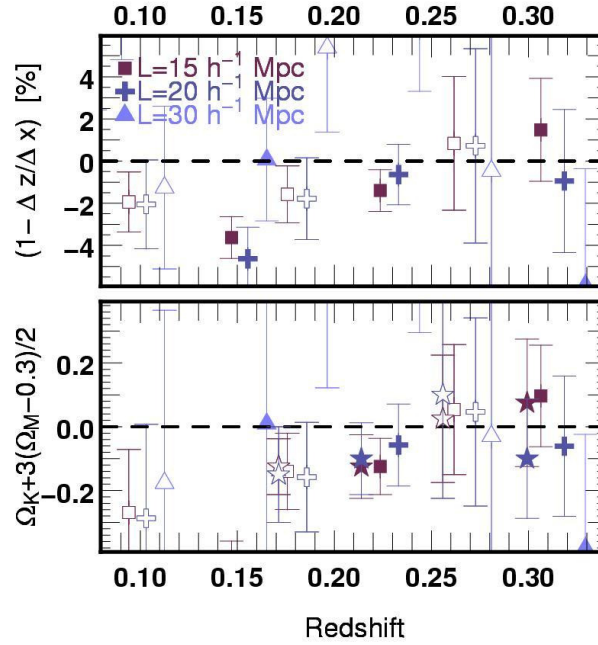
$$C_\theta(z) \approx 1 + z \left( \frac{1}{2}\Omega_k + \frac{3}{4}(\Omega_m - 0.3) \right). \quad (4.15)$$

Pour chaque intervalle de redshift considéré, il est donc possible d'obtenir une estimation des paramètres cosmologiques d'autant plus précise que  $z$  est élevé, l'erreur commise sur la mesure étant simplement :

$$\Delta(\Omega_k + 3\Omega_m/2) = 2 \frac{\Delta C_\theta(z)}{z}. \quad (4.16)$$

Ce résultat est particulièrement intéressant, en effet, la méthode présentée au chapitre 4.1 permettait de contraindre le paramètre  $\Omega_m$  seulement (indépendamment de la valeur de  $\Omega_\Lambda$ ). Etant donné que  $\Omega_k = 1 - \Omega_\Lambda - \Omega_m$ , des contraintes croisées sont donc envisageables.





**Fig. 4.17:** Résultats obtenus pour la mesure de  $C_\theta(z)$  (équation (4.14)) sur le catalogue SDSS pour trois échelles de lissage différentes. Les symboles pleins correspondent à la mesure de la médiane et de l'interquartile du rapport intervenant dans l'équation (4.8). Les symboles vides correspondent à la moyenne et l'écart type de cette même mesure. A noter que chacun des symboles a été décalé horizontalement pour des raisons de lisibilité. Le panneau supérieur présente l'écart en pourcentage de la valeur mesurée de  $C_\theta(z)$  à la valeur de référence 1. Le panneau inférieur représente quand à lui la contrainte correspondante sur la valeur de  $\Omega_k + 3(\Omega_m - 0.3)/2$ . Dans cette deuxième figure, les étoiles correspondent à la solution exacte de l'équation (4.14).

### Application au SDSS

D'après les prévisions théoriques, l'application aux données réelles du SDSS semble très prometteuse. Les résultats présentés ici ne sont cependant que des résultats préliminaires. Leur confirmation nécessite en effet de faire des mesures sur des catalogues virtuels fabriqués à partir de simulations de paramètres cosmologiques variés. Calculer le squelette d'un Univers non plat n'est pour l'instant pas faisable celui-ci étant calculé à partir d'une grille Euclidienne, la valeur de  $\omega_k$  doit donc être prise nulle par défaut. De plus, les mesures que nous avons effectuées à partir des catalogues présentés dans le chapitre 4.1 posent actuellement quelques problèmes d'interprétation non résolus à ce jour. En attendant d'en comprendre exactement les raisons, nous présenterons donc une étude préliminaire sur les galaxies du catalogue SDSS.

La version du catalogue que nous avons utilisée pour effectuer le test A-P par l'intermédiaire du squelette est identique à celle présentée au chapitre 4.1. Les sque-

lettres ont été calculés pour trois échelles de lissage différentes :  $15h^{-1}$ ,  $20h^{-1}$  et  $30h^{-1}$  Mpc, choisies grandes afin de minimiser autant que possible les effets de distorsion des échelles non-linéaires. Dans chacun des cas, quatre tranches de décalages spectraux ont été sélectionnées, centrées à  $z = [0.07, 0.14, 0.23, 0.31]$ . Il est important de noter que, dans le cas du SDSS, il est indispensable d'utiliser l'estimateur  $C_\theta(z)$  et non pas  $C_\phi(z)$  en raison de la géométrie de la distribution (voir la figure 1.7). La figure 4.17 présente les résultats de mesure obtenus pour les valeurs de  $C_\theta(z)$  (panneau supérieur) ainsi que la contrainte correspondante sur les paramètres cosmologiques (panneau du bas). D'après l'équation (4.15), une valeur mesurée de  $C_\theta(z) = 1$  indique que le modèle pris pour référence est le bon or les résultats semblent montrer que c'est le cas à quelques pourcents près et ce pour toutes les échelles de lissage considérées et avec d'autant plus de précision que le décalage spectral est élevé. Exprimés en terme de contours de vraisemblance dans l'espace  $\Omega_m - \Omega_\Lambda$  conjointement aux contraintes trouvées au chapitre 4.1, les contraintes attendues sur la valeur des paramètres cosmologiques sont présentées sur la figure 4.18.

Il semble donc réaliste d'espérer pouvoir contraindre fortement la cosmologie grâce au test A-P, rappelons cependant que ces résultats sont préliminaires et demandent des vérifications. Pour cela, une étude plus précise de l'influence des distorsions de l'espace des redshifts est nécessaire et les difficultés rencontrées lors des vérifications sur les catalogues simulés n'étant pas encore bien comprises, nous en resterons là pour l'instant.

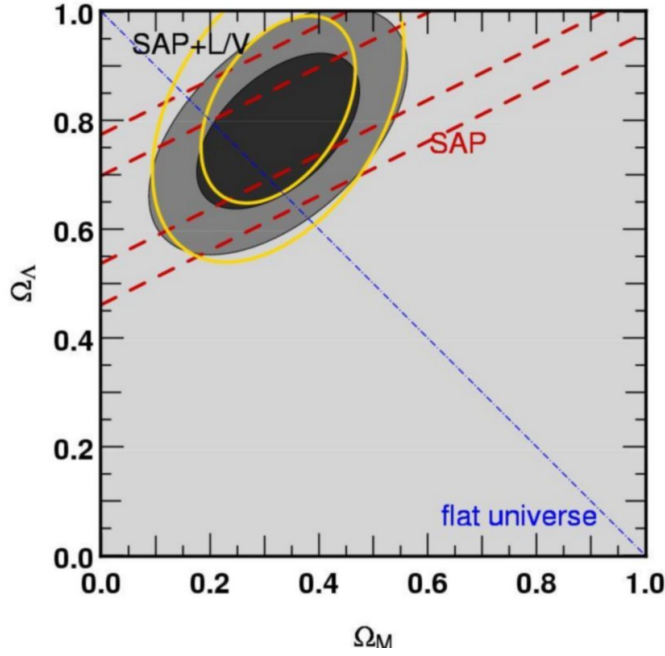
### 4.3.2 Tests de gaussianité

Le paradigme actuel veut que l'univers ait évolué à partir d'une distribution de matière pratiquement uniforme telle qu'observée dans le fond de rayonnement cosmique (voir la section 1.1.3). La grande structuration de la matière observée dans les catalogues de galaxies de l'univers local est alors expliquée par la seule influence des forces gravitationnelles. Ainsi, même pour une distribution initiale homogène, les infimes surdensités ont tendance à attirer la matière environnante et à se renforcer. Il est intuitif de comprendre que, ces zones surdenses étant indépendantes, la matière aura simplement tendance dans un premier temps à s'effondrer sur elle-même de manière linéaire. Mais avec l'augmentation du contraste en densité, plusieurs pics de densités vont commencer à interagir donnant naissance à des interactions non-linéaires beaucoup plus complexes.

La modélisation des instabilités gravitationnelles suppose habituellement que les infimes fluctuations initiales peuvent être correctement représentées par des champs gaussiens aléatoires [4]. La modélisation de ces champs passe souvent par la définition du contraste de densité :

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \rho_0}{\rho_0}, \quad (4.17)$$

où  $\rho_0 = \langle \rho(\mathbf{x}) \rangle_{\mathbf{x}}$  désigne la valeur moyenne du champ de densité  $\rho(\mathbf{x})$  ainsi que des



**Fig. 4.18:** Contours de vraisemblance dans l'espace  $\Omega_m - \Omega_\Lambda$  pour une analyse simultanée du test A-P avec le squelette et de la longueur par unité de volume du squelette dans SDSS. Les contours correspondent aux seuils de confiance à un et deux  $\sigma$  (95% et 65%). La partie remplie correspond à l'estimation non paramétrique (médiane et interquartile) alors que le contour jaune correspond à l'estimation paramétrique (moyenne et écart type). Dans chacun des cas, les valeurs sont pondérées selon leur redshift et l'échelle de lissage considérée. La région en pointillés correspond à la contrainte obtenue par le test AP seul. Dans chacun des cas, la probabilité est exprimée par la fonction  $\chi^2 = (\hat{C}_\theta - C_\theta)^2 / \sigma_{C_\theta}^2 + (\hat{\mathcal{L}} - \mathcal{L})^2 / \sigma_{\mathcal{L}}^2$  où  $\hat{C}_\theta$  et  $\hat{\mathcal{L}}$  représentent les valeurs mesurées des estimateurs  $C_\theta$  et  $\mathcal{L}$  (la longueur du squelette par unité de volume).

différents moments du champ (ici en 3D) :

$$\begin{aligned} \sigma_0^2 &= \langle \rho^2 \rangle; \quad \sigma_1^2 = 2 \langle \rho_i^2 \rangle; \quad \sigma_2^2 = 8/3 \langle \rho_{ii}^2 \rangle \\ \gamma &= \sigma_1^2 / (\sigma_0 \sigma_2). \end{aligned} \quad (4.18)$$

où  $\rho_i = d\rho/dx_i$  et  $\rho_{ij} = d^2\rho/dx_i dx_j$ . On peut montrer que la grandeur  $\gamma$  a une relation simple à l'indice spectral  $n$  du champ gaussien (voir Bardeen *et al.* [4]) :

$$\gamma^2 = \frac{n+3}{n+5}. \quad (4.19)$$

La transformée de Fourier  $\tilde{\delta}(\mathbf{k})$  de  $\delta(\mathbf{x})$  obéit à l'équation :

$$\delta(\mathbf{x}) = \sum \tilde{\delta}(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{x}) \quad (4.20)$$

et, prenant ses valeurs dans les nombre complexes, se décompose en un terme de norme  $\|\tilde{\delta}(\mathbf{k})\|$  et une phase  $\phi_{\mathbf{k}}$  :

$$\tilde{\delta}(\mathbf{k}) = \|\tilde{\delta}(\mathbf{k})\| \exp(i\phi_{\mathbf{k}}). \quad (4.21)$$

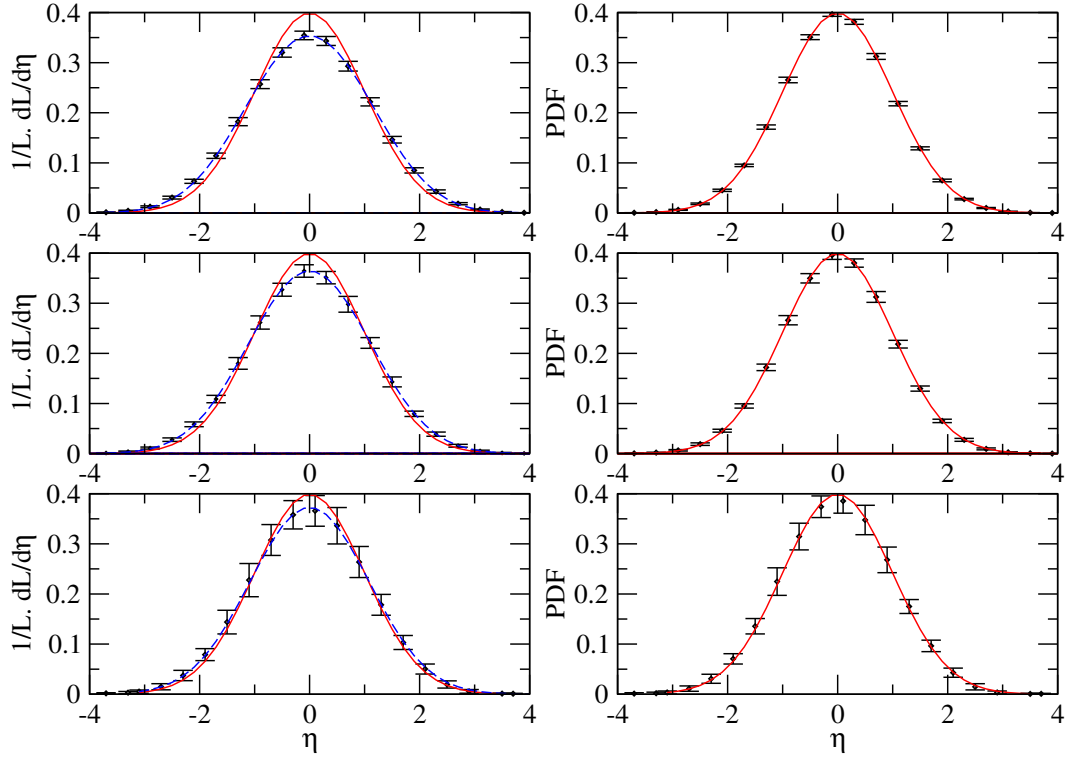
Les champs gaussiens aléatoires ont la propriété d'avoir une distribution de phases aléatoire et uniforme. En supposant que la distribution initiale de masse soit Gaussienne, il est possible de montrer [17] que tant que les fluctuations de densité restent faibles (dans le régime dit linéaire), cette distribution de phases reste aléatoire mais dès que des interactions non-linéaires apparaissent entre les pics de densités, les phases se couplent et  $\delta(\mathbf{x})$  dévie de la gaussianité.

La déviation à la Gaussianité est donc une conséquence directe de l'effondrement gravitationnel et il est extrêmement utile de pouvoir en mesurer les déviations. Le squelette semble a priori un bon outil pour atteindre ce but. Par définition, il s'agit du lieu où la majorité des interactions entre les halos (qui sont des pics du champ de densité) se déroulent. Il est alors très probable qu'en caractérisant les propriétés du squelette, on puisse en déduire les propriétés gaussiennes où non du champ de densité sous jacent. Nous nous contenterons dans cette section de présenter rapidement une mesure montrant comment le squelette pourrait être utilisé pour mesurer la Gaussianité.

Comme il a été montré dans la référence [15], la longueur différentielle du squelette  $d\mathcal{L}/d\eta$  semble être une quantité intéressante. Cette grandeur est définie par la longueur totale de squelette à un niveau de surdensité normalisée  $\eta = \delta/\sigma_0$  donné. Elle présente de plus l'avantage qu'il est possible, dans le cas bidimensionnel tout du moins, d'en faire un approximation analytique du type [15] :

$$\frac{1}{\mathcal{L}_{\text{tot}}} \frac{d\mathcal{L}}{d\eta} = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) (1 + C_2 \gamma^2 (1 - \eta^2)). \quad (4.22)$$

Les calculs étant pour le moins complexes, la démonstration de l'existence d'une formule similaire à 3D n'est pas encore achevée. Il est cependant possible de montrer numériquement que l'expression (4.22) est une bonne approximation dans ce cas là aussi. La figure 4.19 présente les résultats obtenus par la mesure de la longueur différentielle du squelette sur 25 réalisations de champs gaussiens d'indices spectraux  $n = 0$ ,  $n = -1$  et  $n = -2$ . Les ajustements obtenus grâce à l'équation (4.22) montrent qu'à 3D aussi l'approximation est toujours valable et qu'il est donc possible d'utiliser la mesure de  $d\mathcal{L}/d\eta$  comme mesure de Gaussianité (l'ajustement donnant directement la valeur de l'indice spectral par l'équation (4.19)). Une étude plus complète du problème incluant des mesures d'influence des distortions de décalage spectral ainsi qu'une étude analytique du cas tri-dimensionnel est prévue pour un futur proche.



**Fig. 4.19:** La colonne de gauche représente la PDF de la longueur différentielle du squelette mesurée en fonction de la surdensité normalisée  $\eta$  alors que la PDF du champ de densité est représenté dans la colonne de droite. De haut en bas, les mesures sont effectuées sur des champs gaussiens aléatoires d'indice spectral  $n = 0$ ,  $n = -1$  et  $n = -2$ . Les barres d'erreur sont le résultat des mesures effectuées sur 25 réalisations. Dans le panneau de gauche, la courbe bleu représente l'ajustement des données grâce à la formule 4.22 et la courbe rouge une gaussienne normalisée.

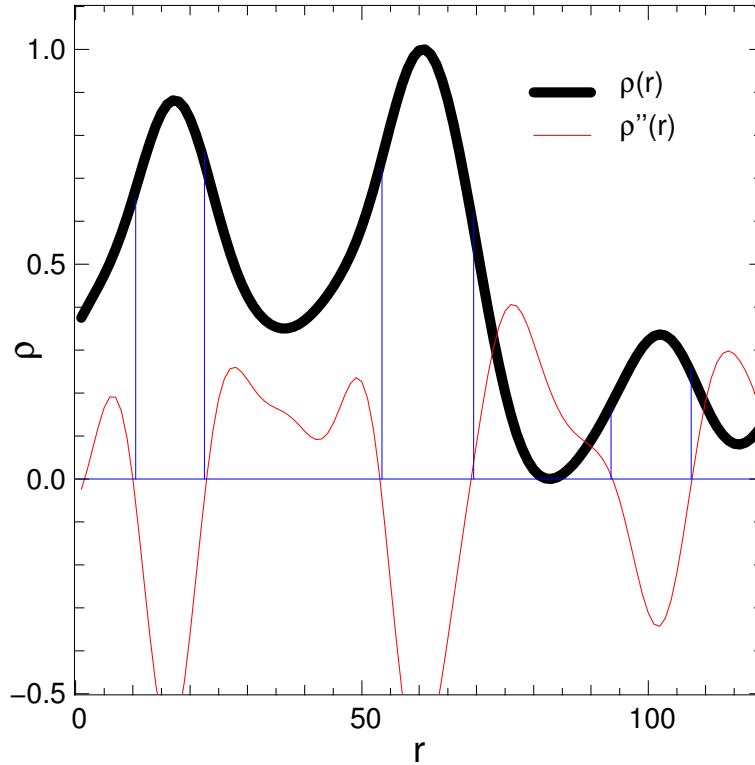
### 4.3.3 Section des filaments

Le squelette permet la localisation de la position des filaments. Afin d'en obtenir une description géographique plus complète, l'ajout de la possibilité de mesurer leur section semble très intéressante, notamment dans le but de caractériser la distribution des objets dans leur voisinage.

#### Méthode de mesure de la section des filaments

On peut imaginer plusieurs méthodes pour estimer l'extension et la géométrie des filaments. Le squelette étant calculé à partir d'une approximation au second ordre du champ de densité, il semble logique que la méthode de mesure de sa section repose sur le même type d'approximation. Nous avons donc développé deux méthodes différentes pour parvenir à ce résultat.

La première méthode est basée sur la mesure locale de la courbure orthogonalement aux filaments. Comme il a été expliqué au chapitre 3.2, le squelette est locale-



**Fig. 4.20:** *Illustration d'une des méthodes de calcul de la section des filaments. La courbe noire représente le champ de densité le long d'une droite. Si des filaments intersectent cette droite, ils le font là où  $\rho(r)$  est localement un maximum. En s'éloignant des maxima selon la droite, on voit qu'il est possible de définir la section du filament selon cette droite par la position du premier point d'inflexion rencontré soit le premier point où la dérivée seconde (courbe rouge) s'annule.*

ment aligné avec le vecteur propre principal du Hessien  $\mathcal{H}$ . Les deux autres vecteurs propres de  $\mathcal{H}$  désignent donc les deux axes de courbure principale orthogonalement aux filaments et les valeurs propres associées l'intensité de cette courbure. Soit  $\mathbf{V}_i$  le vecteur propre normalisé de  $\mathcal{H}$  associé à la valeur propre  $\lambda_i$  avec  $i \in (1, 2, 3)$ . En un point  $\mathbf{x}$  quelconque du squelette, la section du filament dans la direction  $\mathbf{V}_i$  peut alors être définie comme le rayon  $r_i$  du cercle osculateur<sup>4</sup> de la fonction  $f_i(\alpha) = \rho(\mathbf{x} + \alpha \mathbf{V}_i)$ . La courbure au point  $\mathbf{x}$  de  $f_i(\alpha)$  vaut :

$$\kappa = \frac{f_i''(0)}{(1 + f_i'(0)^2)^{3/2}} = f_i''(0) = \lambda_i. \quad (4.23)$$

Le rayon du cercle osculateur est donc  $r_i = 1/\lambda_i$ . Il est alors possible de définir la section du filament comme étant une ellipsoïde de rayons  $r_2 = 1/\lambda_2$  et  $r_3 = 1/\lambda_3$ . La seconde méthode se base aussi sur l'analyse de la dérivée seconde du champ selon les directions des vecteurs propres de  $\mathcal{H}$  mais, contrairement à la précédente, n'est

<sup>4</sup>Le cercle osculateur de la courbe paramétrique  $(x(t), y(t))$  au point  $\mathbf{p}$  est le cercle passant par  $\mathbf{p}$  et dont les dérivées première et seconde en ce point sont  $(x'(t), y'(t))$  et  $(x''(t), y''(t))$ .

pas locale. Elle consiste à définir l'épaisseur du filament en un point  $\mathbf{x}$  et dans une direction quelconque  $\mathbf{V}$  comme la distance à laquelle la dérivée seconde de la fonction  $f_{\mathbf{V}}(\alpha) = \rho(\mathbf{x} + \alpha\mathbf{V})$  change de signe. La figure 4.20 illustre cette méthode.

L'avantage principal de la première méthode réside certainement dans le fait qu'elle est locale et simple à mettre en œuvre. Malheureusement, le rayon du filament étant défini comme l'inverse de la courbure, cela peut poser des problèmes à proximité des extréma du champ. Il n'est pas rare en effet qu'à proximité des maxima, la dérivée seconde tende vers 0 ce qui a tendance à faire diverger numériquement le rayon calculé. Cette méthode est donc intéressante pour calculer rapidement le rayon moyen d'un filament mais n'est finalement pas très robuste pour obtenir une information précise sur la section des filaments à un endroit donné. La seconde méthode ne présente pas cet inconvénient et a de plus l'avantage qu'avec elle, il n'est plus nécessaire de supposer une section ellipsoïdale des filaments. La figure 4.21 illustre le résultat obtenu en l'appliquant à la distribution des 30,765 galaxies de la simulation galics3. Comme on peut le voir, la section obtenue est tout à fait crédible et englobe une grande partie des galaxies de la distribution, ne laissant apparaître que les galaxies des vides.

### Quelques applications possibles de la mesure de la section des filaments

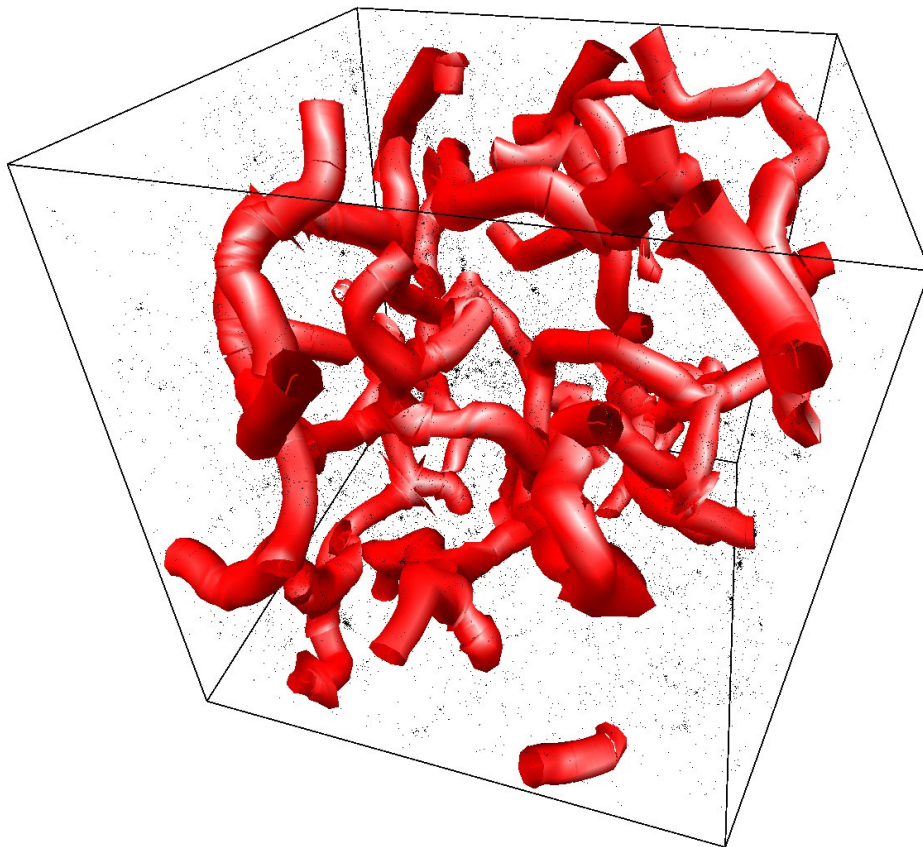
Le principal intérêt de la connaissance de la géométrie tri dimensionnelle des filaments réside sans doute dans la possibilité qu'elle donne de définir une taille caractéristique en chaque point du squelette. Il devient alors possible par exemple d'isoler les galaxies appartenant aux vides de celles des halos et de celles des filaments de manière rigoureuse. La section moyenne (voir la figure 4.22(b)) ou encore l'ellipticité des filaments (voir la figure 4.22(a)) pourrait de plus être caractéristique de différents modèles d'Univers, et c'est une voie que nous comptons explorer dans un avenir proche. L'étude de la section moyenne des filaments semble particulièrement intéressante et apparaît croître comme l'échelle de lissage à la puissance  $2/3$ , les raisons exactes restant encore à déterminer.

#### 4.3.4 Mesures de distance entre deux squelettes

Un des atouts du squelette est de permettre l'extraction des zones d'intérêt de la distribution de matière. Avoir la possibilité de donner une mesure quantitative de la différence entre les réseaux filamentaires de deux distributions pourrait permettre d'en distinguer les propriétés. Nous désirerions par exemple mesurer la différence qu'il existe entre le squelette du gaz et celui de la matière noire dans les simulations hydrodynamiques.

#### Définition d'une distance inter-filaments

Le squelette peut être résumé à une courbe de dimension 1 plongée dans un espace 3D. La question est donc comment définir une distance entre deux courbes qui

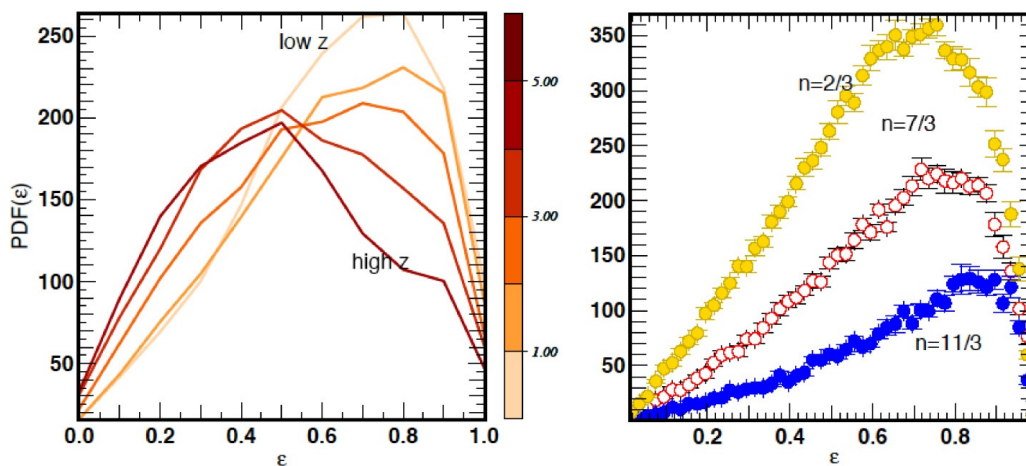


**Fig. 4.21:** Résultat obtenu en appliquant la méthode de recherche du point d'inflexion. Le squelette est calculé à partir de la distribution de galaxies galics3 comprenant seulement 30,765 galaxies dans un volume de  $100h^{-1} \text{ Mpc}^3$ . Malgré la faible définition du champ de densité (lissé sur six pixels d'une grille de  $128^3$  cellules), l'algorithme est suffisamment robuste pour ne pas diverger.

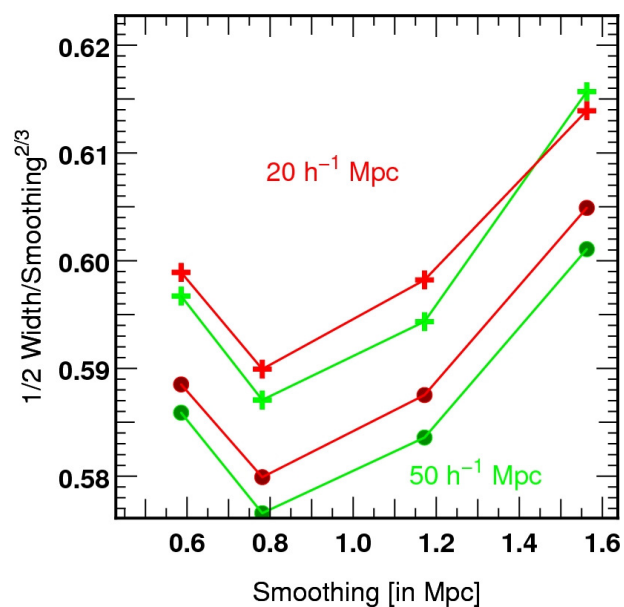
pourrait donner de manière quantitative une mesure de leur différence? Pour une application telle que la caractérisation de la similarité des squelettes du gaz et de la matière noire, il est clair que cette mesure doit refléter deux caractéristiques principales : un filament du squelette de la matière noire à-t-il toujours un homologue dans le squelette du gaz? Et si oui, se trouve-t-il exactement au même endroit ou est-il éloigné?

Les squelettes sont toujours constitués d'un ensemble de segments mais leur nombre varie forcément entre deux squelettes. Pour comparer un squelette  $A$  à un squelette  $B$ , la méthode imaginée consiste donc à considérer chaque segment du squelette  $A$ , à trouver le segment appartenant au squelette  $B$  qui soit le plus proche possible et à en mesurer la distance. L'histogramme  $d(A, B)$  des distances mesurées donnera alors une bonne idée de l'écart entre  $A$  et  $B$ . Afin d'obtenir une distance qui soit symétrique, il faut cependant tenir compte du fait que si un segment  $s_i$  de  $A$  a pour plus proche voisin un segment  $s'_j$  de  $B$ , alors  $s_i$  n'est pas forcément le plus proche voisin de  $s'_j$ . La distance entre  $A$  et  $B$  sera donc donnée par





(a) Mesures d'ellipticité



(b) Mesures de section

**Fig. 4.22:** Influence du redshift (figure 4.22(a), partie gauche) et de l'indice spectral  $n$  (figure 4.22(a), partie droite) sur l'ellipticité des filaments. La figure 4.22(b) montre l'influence de l'échelle de lissage considérée sur la section mesurée des filaments pour deux simulations de  $20h^{-1}$  et  $50h^{-1}$  Mpc respectivement. Sur cette courbe, la largeur de la section apparaît directement proportionnelle à la longueur de lissage à la puissance  $2/3$ .

$$D(A, B) = (d(A, B) + d(B, A))/2.$$

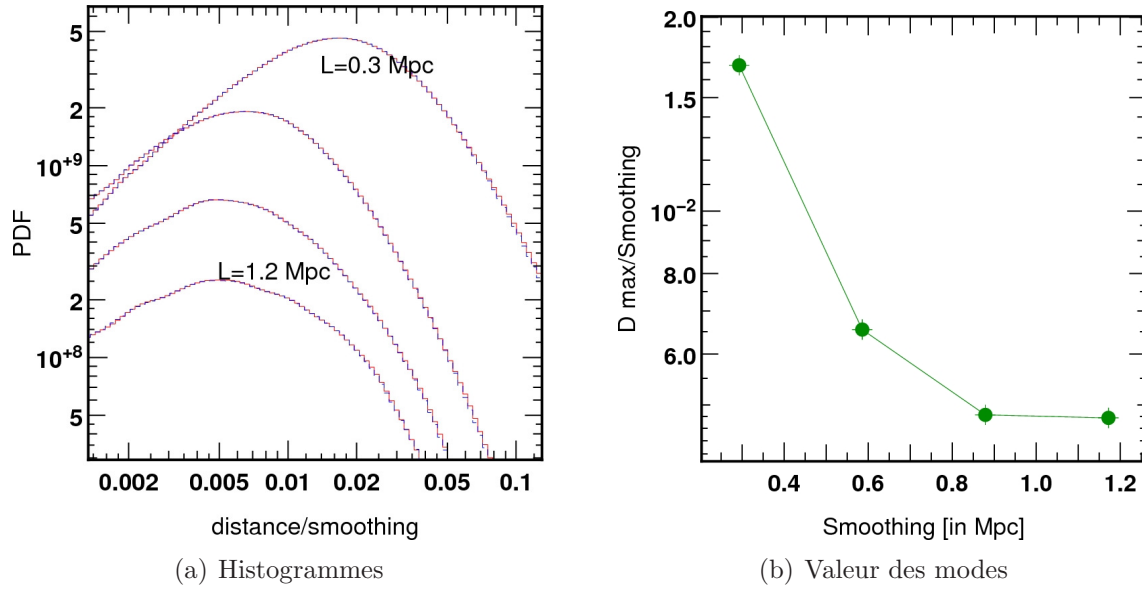
### Comparaison des filaments du gaz et de la matière noire

La simulation que nous allons utiliser est une simulation hydrodynamique réalisée dans le cadre du projet HORIZON<sup>5</sup>. Cette simulation appelée MareNostrum (le nom du calculateur utilisé) compte parmi ce qui se fait de mieux actuellement. Il s'agit d'une simulation AMR de  $50h^{-1}$  Mpc<sup>3</sup> dont la grille initiale comporte  $1024^3$  cellules raffinables 8 fois (soit une résolution équivalente de  $8192^3$ ). Le champ de matière noire est quant-à-lui échantillonné par  $1024^3$  particules. L'intérêt d'utiliser une telle simulation réside bien sûr dans la possibilité qu'elle donne d'explorer une vaste gamme d'échelles différentes tout en conservant une statistique suffisante. Malheureusement, à ce jour, le décalage spectral atteint n'est que de  $z = 4$  du fait de la résolution, de nombreuses régions sont à ce redshift déjà entré dans un mode non-linéaire entraînant ainsi un taux de raffinage élevé et donc un fort ralentissement des calculs). Nous allons donc utiliser la méthode introduite précédemment pour comparer les squelettes du gaz et de la matière noire à  $z = 4$ .

La figure 4.23(a) présente les deux histogrammes des distances entre le squelette du gaz  $S_g$  et celui de la matière noire  $S_{dm}$ . Le fait que les histogrammes  $d(S_g, S_{dm})$  (en rouge) et  $d(S_{dm}, S_g)$  (en bleu) soient superposés nous fournit une information importante : là où la distribution du gaz est filamentaire, il en est de même pour celle de la matière noire. En effet, si par exemple, il existait des filaments dans le gaz à petite échelle qui n'avaient pas d'équivalent dans la distribution de matière noire, une asymétrie serait apparente (une bosse serait visible dans l'histogramme  $d(S_g, S_{dm})$  mais pas dans  $d(S_{dm}, S_g)$ ). Quelque soit l'échelle de lissage (de  $0.3h^{-1}$  à  $1.2h^{-1}$  Mpc), il est de plus clair que les PDF présentent un maximum à une distance de l'ordre du centième de l'échelle de lissage. On peut en conclure que les deux squelettes sont donc toujours proches, confirmant ainsi l'hypothèse selon laquelle la distribution du gaz est principalement influencée par les puits de potentiels créés par la matière noire.

Une analyse de l'évolution de la position des modes en fonction de l'échelle de lissage (figure 4.23(b)) est cependant instructive. Si la distance moyenne par unité d'échelle de lissage entre les deux squelettes est proportionnelle à l'échelle de lissage pour les échelles supérieures à  $\approx 800h^{-1}$  kpc, celle-ci augmente exponentiellement pour des échelles inférieures. Les filaments de la distribution du gaz présentent donc une tendance marquée à se différencier de ceux de la matière noire aux petites échelles. On peut montrer que cette échelle correspond à peu près à la limite du régime linéaire. L'approximation linéaire étant valable pour  $\delta \ll 1$ , on peut en effet désigner comme non linéaires toutes les régions où  $\delta > 1$ . En calculant, sur des simulations, la variance de champs lissés sur différentes échelles, on trouve ainsi que  $\delta \approx 1$  pour un lissage de l'ordre du  $h^{-1}$  Mpc. Nous venons donc de montrer que si

<sup>5</sup><http://www.projet-horizon.fr>



**Fig. 4.23:** Sur la figure 4.23(a), mesure de l'histogramme de la distribution des distances entre les plus proches voisins parmi les segments du squelette des baryons et de la matière noire pour la simulation AMR Mare Nostrum (voir le texte principal). Le tracé de la position des modes par unités de lissage pour des échelles de lissage variant de 0.3 Mpc à 1.2 Mpc (figure 4.23(b)) permet de mettre en évidence un changement de comportement à une échelle de l'ordre de 800 kpc, signifiant qu'en deçà de cette échelle, les squelettes des baryons et de la matière noire divergent.

l'approximation selon laquelle la distribution du gaz suit celle de matière noire aux échelles linéaires est réaliste, ce n'est pas le cas aux plus petites échelles où le gaz a tendance à former ses propres filaments indépendamment de ceux de la matière noire.

---

# Bibliographie

- [1] J. K. & al. ADELMAN-MCCARTHY. The Fourth Data Release of the Sloan Digital Sky Survey. *ApJS*, 162 :38–48, January 2006.
- [2] C. ALCOCK and B. PACZYNSKI. An evolution free test for non-zero cosmological constant. *Nature*, 281 :358–+, October 1979.
- [3] D. AUBERT, C. PICHON, and S. COLOMBI. The origin and implications of dark matter anisotropic cosmic infall on  $L_*$  haloes. *MNRAS*, 352 :376–398, August 2004.
- [4] J. M. BARDEEN, J. R. BOND, N. KAISER, and A. S. SZALAY. The statistics of peaks of Gaussian random fields. *ApJ*, 304 :15–61, May 1986.
- [5] J. BINNEY and J. SILK. Tidal interactions in the expanding universe - The formation of prolate systems. *MNRAS*, 188 :273–283, July 1979.
- [6] J. BLAIZOT, I. SZAPUDI, S. COLOMBI, T. BUDAVÀRI, F. R. BOUCHET, J. E. G. DEVRIENDT, B. GUIDERDONI, J. PAN, and A. SZALAY. GALICS - V : Low- and high-order clustering in mock Sloan Digital Sky Surveys. *MNRAS*, 369 :1009–1020, July 2006.
- [7] J. BLAIZOT, Y. WADADEKAR, B. GUIDERDONI, S. T. COLOMBI, E. BERTIN, F. R. BOUCHET, J. E. G. DEVRIENDT, and S. HATTON. MoMaF : the Mock Map Facility. *MNRAS*, 360 :159–175, June 2005.
- [8] S. COLOMBI, D. POGOSYAN, and T. SOURADEEP. Tree Structure of a Percolating Universe. *Physical Review Letters*, 85 :5515–+, December 2000.
- [9] J. DA ÂNGELA, P. J. OUTRAM, T. SHANKS, B. J. BOYLE, S. M. CROOM, N. S. LOARING, L. MILLER, and R. J. SMITH. The 2dF QSO Redshift Survey- XV. Correlation analysis of redshift-space distortions. *MNRAS*, 360 :1040–1054, July 2005.
- [10] A. FALTENBACHER, S. GOTTLÖBER, M. KERSCHER, and V. MÜLLER. Correlations in the orientations of galaxy clusters. *A&A*, 395 :1–9, November 2002.
- [11] A. J. S. HAMILTON. Measuring Omega and the real correlation function from the redshift correlation function. *ApJ*, 385 :L5–L8, January 1992.

- 
- [12] T. MATSUBARA and A. S. SZALAY. Constraining the Cosmological Constant from Large-Scale Redshift-Space Clustering. *ApJ*, 556 :L67–L70, August 2001.
- [13] T. MATSUBARA and A. S. SZALAY. Cosmological Parameters from Redshift-Space Correlations. *ApJ*, 574 :1–8, July 2002.
- [14] P. MCDONALD and J. MIRALDA-ESCUDE. Measuring the Cosmological Geometry from the LYALPHA Forest along Parallel Lines of Sight. *ApJ*, 518 :24–31, June 1999.
- [15] D. NOVIKOV, S. COLOMBI, and O. DORÉ. Skeleton as a probe of the cosmic web : the two-dimensional case. *MNRAS*, 366 :1201–1216, March 2006.
- [16] B. PANDEY and S. BHARADWAJ. The luminosity, colour and morphology dependence of galaxy filaments in the Sloan Digital Sky Survey Data Release Four. *MNRAS*, 372 :827–838, October 2006.
- [17] P. J. E. PEEBLES. *The large-scale structure of the universe*. Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p., 1980.
- [18] E. ROLLINDE, P. PETITJEAN, C. PICHON, S. COLOMBI, B. ARACIL, V. D’ODORICO, and M. G. HAEHNELT. The correlation of the Lyman  $\alpha$  forest in close pairs and groups of high-redshift quasars : clustering of matter on scales of 1-5 Mpc. *MNRAS*, 341 :1279–1289, June 2003.
- [19] M. P. VAN HAARLEM, C. S. FRENK, and S. D. M. WHITE. Projection effects in cluster catalogues. *MNRAS*, 287 :817–832, June 1997.

# Chapitre 5

## Conclusion

Le but que nous nous étions fixé au commencement de cette thèse était de trouver de nouvelles méthodes d'étude de la formation des grandes structures. L'approche que nous avons utilisée est principalement numérique et repose au final sur le développement de deux logiciels permettant respectivement l'extraction du lieu géométrique des filaments et la création de catalogues de galaxies réalistes à grande échelle (MoLUSC).

Le premier a pour objectif de palier au fait qu'aucune méthode ne permettait jusqu'alors de considérer les filaments de la matière noire ou du gaz comme des objets à part entière, ce qui est assez paradoxal sachant qu'ils sont ce qu'il y a de plus caractéristique de la distribution de matière à grande échelle. MoLUSC a quand à lui été créé dans le but de rendre possible la comparaison des mesures effectuées sur les grandes structures de galaxies dans les catalogues observationnels aux résultats obtenus par simulation numérique. Le développement de ces deux logiciels a certainement été ce qui a pris le plus de temps, et plus particulièrement celui du squelette qui a demandé le plus d'efforts afin de trouver des méthodes originales et efficaces de résolution des équations différentielles ainsi que d'un point de vue algorithmique, afin que le temps de calcul soit minime. Le résultat est qu'il est possible de calculer le squelette sur une grille de  $1024^3$  cellules en quelques minutes seulement (comme par exemple sur la simulation Mare Nostrum utilisée au chapitre 4) et que la méthode est suffisamment robuste pour être applicable à des données observationnelles.

L'ensemble des outils développés devraient être rendu disponible pour la communauté dans un avenir proche et nous espérons qu'ils pourront être utile à un grand nombre de personnes. Parmi eux, outre MoLUSC et le squelette, un outil de reconnaissance de halos de type friend-of-friend autorisant aussi la recherche de voisins ou le calcul de densités SPH (des opérations très souvent utiles dans l'analyse des simulations numériques) ainsi qu'un visualiseur de données utilisant la technologie OpenGL appelé SimuView. La plupart des illustrations de cette thèse ont été réalisées en utilisant ce dernier, qui permet un rendu en temps réel aussi bien que la création de films.

Afin d'essayer d'apporter des éléments de réponse aux questions de la composition de l'Univers ainsi que de la répartition de son contenu en matière, plusieurs applications ont été présentées. Parmi elles, trois semblent particulièrement intéressantes. La mesure de la valeur de  $\Omega_m$  par mesure de la longueur des filaments par unité de volume dans le catalogue SDSS tout d'abord. Si les contraintes obtenues ne sont pas aussi bonnes que celles du CMB ou des supernovæ, elle a au moins le mérite d'être originale et l'on peut imaginer qu'en explorant toutes les autres caractéristiques du squelette on pourrait améliorer les résultats (en mesurant la courbure ou la torsion du squelette par exemple). La deuxième application est le test de Alcock-Paczynski qui est extrêmement prometteur. Combiné aux contraintes de longueur du squelette, ils devraient fournir dans un avenir proche une contrainte bien meilleure sur l'ensemble des paramètres cosmologiques. De plus, on peut encore espérer améliorer ces résultats avec la mise en place de nouveaux catalogues de galaxies à plus grands redshifts. Finalement, la troisième voie explorée est la caractérisation des flux de matière le long des filaments. L'utilisation du squelette rend en effet possible des mesures inédites des propriétés de l'écoulement de la matière. Nous avons présenté ici les résultats obtenus pour la matière noire mais rien n'empêche d'étendre la méthode au gaz ainsi qu'aux galaxies, un des objectifs étant par exemple l'amélioration des modèles semi-analytiques de formation galactique par la prise en compte des propriétés des filaments.

D'autres voies peuvent aussi être explorées. Pour l'étude des propriétés de l'Univers local par exemple (voir l'annexe B). Le squelette pourrait en effet servir à l'identification des structures connues de la distribution des galaxies locales dans le but d'en quantifier les caractéristiques. D'un point de vue plus théorique, l'extension du travail fait dans la section 4.2 à la mesure de propriétés des premiers moments (équations (1.67) ) de l'équation d'Euler (1.68) autour des filaments pourrait aboutir à l'obtention d'une équation de fermeture locale. Il serait alors possible en toute légitimité de résoudre l'équation d'Euler dans les filaments en utilisant seulement les moments d'ordre inférieur tout en démontrant la validité de l'approximation faite. Cette étude pourrait de plus s'appliquer aussi bien au gaz qu'à la matière noire (comme il a été montré avec la simulation Mare Nostrum dans les sections 4.2.6 et 4.3.4).

Le concept de squelette peut de plus être encore développé avec la conception d'algorithmes robustes permettant le calcul du *vrai* squelette qui permettrait par exemple l'étude des propriétés des connections entre halos et filaments. Les travaux effectués par Christophe Pichon sur ce sujet sont déjà bien avancés et une approche probabiliste est développée par Stéphane Colombi. Il est aussi envisagé de développer une version fonctionnant dans un espace de dimension supérieure afin d'étudier le squelette dans l'espace des phases ...

## Annexe A

# Tirage aléatoire selon une loi de probabilité donnée

Dans le chapitre 2.2.1, on désire générer un ensemble de densités  $\{n_i\}$  suivant une distribution  $P(n|\rho) dn$  à partir de nombres aléatoires suivant une loi uniforme. Si l'on prend  $f(x)$  une fonction monotone sur  $x \in [0, 1]$ , alors la probabilité de générer un nombre au hasard entre  $f(x)$  et  $f(x) + dx$  est notée  $Q(f(x)) dx$ . Or, d'après la loi de combinaison des probabilités,

$$Q(f(x)) dx = Q(x) \left| \frac{dx}{df} \right| dx. \quad (\text{A.1})$$

En remplaçant  $x$  par  $a$  et  $f(x)$  par  $n_i(a)$ , on obtient :

$$Q(n_i(a)) da = Q(a) \left| \frac{da}{dn_i} \right| da. \quad (\text{A.2})$$

Soit  $P^*(x) = \int_0^x P(x|\rho_i) dx$ , alors

$$n_i(a) = P^{*-1}(a - P^*(0)) \quad (\text{A.3})$$

avec  $P^{*-1}(P^*(x)) = x$ . Étant donné que  $P^*(0) = 0$ ,

$$Q(n_i(a)) da = Q(a) \frac{dP^*}{da}(a) = Q(a) P(a|\rho_i) da, \quad (\text{A.4})$$

ce qui montre que  $n_i$  a bien la densité de probabilité  $P(a|\rho_i)$ . En effet,  $a$  suivant une loi uniforme,  $Q(a) da = da$  pour  $a \in [0, 1]$  et donc

$$Q(n_i(a)) da = P(a|\rho_i) da. \quad (\text{A.5})$$



---

## Annexe B

# Le squelette et l'Univers local

Une étude est actuellement en cours concernant les propriétés du squelette dans l'Univers local. En effet dans l'Univers Local nous pouvons disposer de mesures de *distances* (par oppositions aux décalages spectraux) pour les galaxies, ce qui nous donne accès aux mouvements propres, hors de l'expansion.

Afin de connaître la position d'une galaxie dans l'Univers on a en général une mesure de son redshift, cependant si l'on dispose également d'une mesure de sa magnitude absolue, connaissant sa magnitude apparente, on obtient une mesure indépendante de sa distance. En combinant ces deux mesures : redshift et distance, on obtient la mesure de la vitesse propre ou "particulière" de la galaxie : hors du flot de Hubble. Cette vitesse propre est alors due uniquement aux champs gravitationnels que subit la galaxie de la part de son environnement.

Si l'on connaît son environnement de masse lumineuse par le squelette, on peut alors comprendre la distribution de la masse totale (masse non lumineuse comprise) environnante. Par exemple, dans l'univers local, plusieurs grands mouvements d'ensemble (des "flots") sont connus. Le flot global vers le grand attracteur est mesuré à environ 600 km/s : le filament dans lequel nous vivons, mais également les autres filaments près de nous, ont ce mouvement d'ensemble dans la direction générale de Centaurus. Ce mouvement a été mentionné depuis une dizaine d'années. Les mesures de distances de galaxies sont parmi les mesures de paramètres astrophysiques les plus difficiles. Les observations sont difficiles car très variées : on doit mesurer des courbes de lumière de Supernovae Ia, de Céphéides, faire de la radioastronomie pour mesurer la rotation du gaz d'hydrogène neutre dans les galaxies spirales, mesurer les étoiles des galaxies proches pour la méthode TRGB ("tip of the red giant branch"), mesurer des élargissements de raies optiques pour les galaxies elliptiques, utiliser les images du Telescope Spatial Hubble, ... Ces observations sont si variées et si difficiles qu'il faut attendre plusieurs dizaine d'années pour que les efforts communs de dizaines d'astronomes permettent de construire un catalogue de distances extragalactiques qui ait un nombre suffisant de mesures de très bonne qualité.

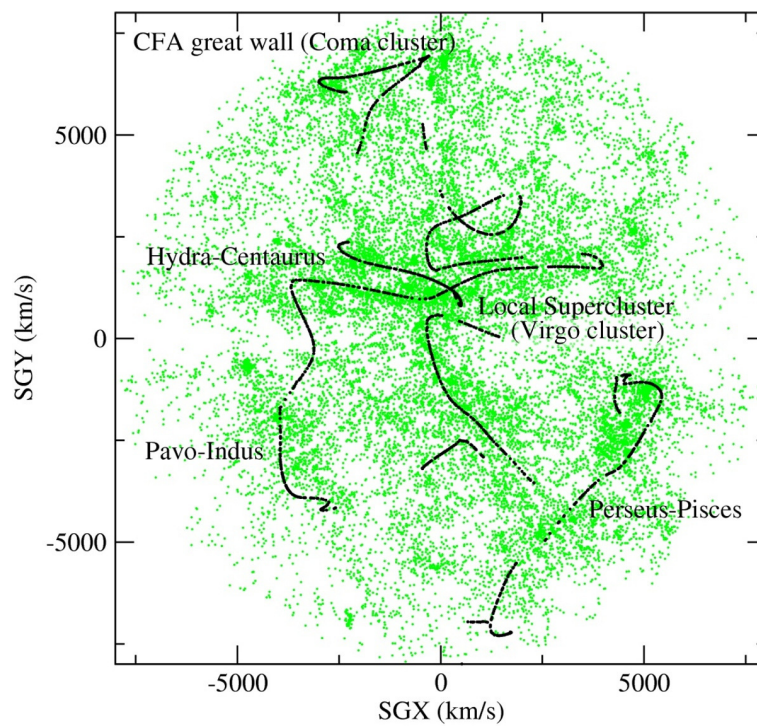
Aujourd'hui ces catalogues commencent à être construits. Le processus est long

et délicat, il faut mettre ensemble des mesures de natures très différentes et qui ont donc des erreurs systématiques, des biais, très différents. Il faut aussi que la calibration de ces méthodes soit ajustées à un même point zéro, avant tout effort de combiner des mesures de Céphéides avec des mesures de HI radio. Dans Tully et al 2007, une première analyse d'un catalogue de 1,485 galaxies mesurées en distances dans un rayon de 3,000 km/s est disponible. Ce catalogue et cette analyse sont focalisées sur l'univers très local, ils mettent en évidence un nouveau mouvement propre : le filament auquel nous appartenons s'éloigne d'un grand vide local à 211 km/s. Simultanément un filament proche du notre se rapproche de notre filament. Ces deux filaments volent ensemble dans le mouvement global vers l'amas Virgo à 200 km/s. A plus grande échelle tout ce volume de 3,000 km/s vole vers le grand attracteur. Cette étude du catalogue observationnel soulève de nombreuses questions car les mesures qui sont accessibles aux astronomes ne sont que les parties radiales des vecteurs.

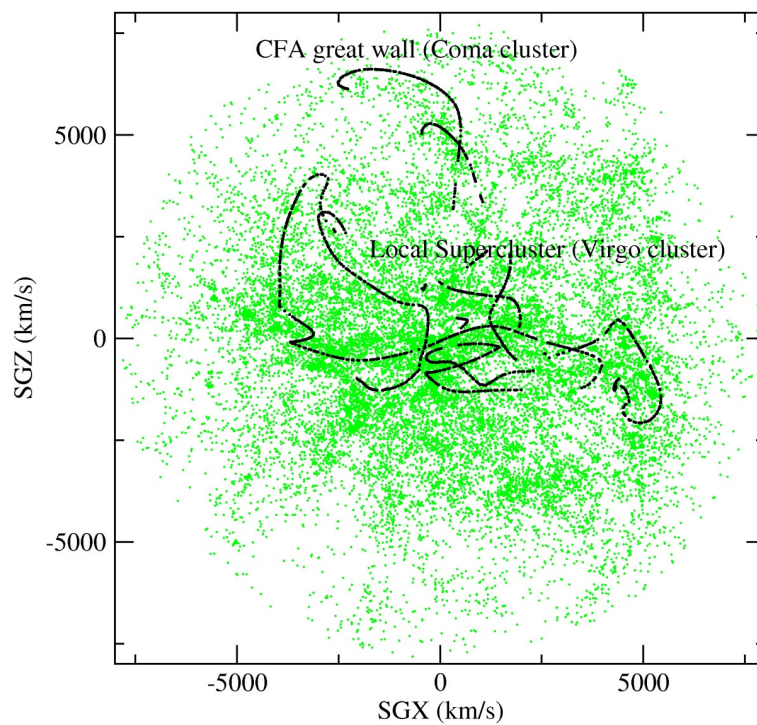
Quelle est la dispersion vraies des vitesses dans le filament local ? Quelle est la vitesse d'ensemble des galaxies parallèlement au filament local en direction de Virgo qui semble être l'amas auxquels viennent se connecter plusieurs filaments ? Les vitesses parallèles au filament sont-elles plus grandes que les vitesses orthogonales comme nous l'avons vu avec le squelette des simulations ? L'image générale que nous avons dégagé avec les simulations : de routes secondaires, venant se connecter à une autoroute principale où la vitesse autorisée est bien supérieure et qui vient nourrir les grandes villes de véhicules roulant tous dans la même direction, est-elle correcte lorsque comparée aux observations ?

Pour cela nous commençons à appliquer le squelette aux catalogues de distance, nous devons tout d'abord identifier les filaments par une méthode qui soit reproductible (actuellement ils sont identifiés à l'oeil et sans cohérence entre la définition des différents filaments), ensuite nous devons analyser les composantes uniquement radiales des vitesses de ces galaxies indépendamment dans chacun des filaments. Nous devons extraire des simulations une région correspondante à l'univers local, extraire seulement les composantes radiales des vitesses et reproduire cette analyse.

C'est un domaine où la méthode du squelette pourrait réellement être utile et servir à prédire des comportements observables pour les distributions de galaxies dans différentes cosmologies, analyser les données observer exactement de la même façon que les simulations et en déduire la cosmologie dans laquelle nous vivons.



(a) Projection XY



(b) Projection XZ

**Fig. B.1:** Identification par le squelette des structures connues de l'Univers local dans le catalogue de galaxies LEDA.

# Table des figures

1.1	Evolution du paramètre d'échelle $a(t)$ . . . . .	7
1.2	Diagramme de Hubble . . . . .	12
1.3	Résultats de WMAP3 . . . . .	15
1.4	Evolution de l'Univers dans l'espace des paramètres cosmologiques . .	17
1.5	Le catalogue LICK . . . . .	29
1.6	Le grand mur du catalogue CfA . . . . .	29
1.7	Couverture angulaire du catalogue SDSS DR4 . . . . .	31
2.1	Analyse du biais dans GaLISC . . . . .	46
2.2	Comparaison des distribution de galaxies obtenues avec GalICS et MoLUSC . . . . .	50
2.3	Comparaison des fonctions de corrélation à deux points de la distribution de matière noire et des distribution générées avec GalICS et MoLUSC . . . . .	51
2.4	Illustration de la technique du pavage aléatoire . . . . .	54
2.5	Projection de Flamsteed des coordonnées angulaires des galaxies de SDSS . . . . .	56
2.6	Calcul de la convolution spectre/filtre . . . . .	58
3.1	Coupe de la projection du champ de densité de matière noire d'une simulation ${}^3\text{LCDM}_{50}^{256}$ . . . . .	61
3.2	Illustration d'un vide et d'un halo dans une simulation de matière noire	62
3.3	Calcul des fonctionnelles de Minkowsky par la technique des sphères .	66
3.4	Une isosurface du champ de densité d'une simulation de matière noire	68
3.5	Arbre à recouvrement minimal . . . . .	71
3.6	Illustration d'un champ de densité gaussien et de son squelette 2D . .	74
3.7	Comportement du gradient et du hessien d'un champ de densité 2D .	75
3.8	Approximation locale du champ autour d'un point selle . . . . .	77
3.9	Représentation d'un champ 2D en fonction du signe des valeurs propres du Hessien $\mathcal{H}$ . . . . .	78
3.10	Choix du système de coordonnées sur une surface d'isodensité . . . .	81
3.11	Illustration d'un champ 3D en fonction du signe des vecteurs propres de $\mathcal{H}$ . . . . .	84
3.12	Notations pour la grille d'échantillonnage . . . . .	87

3.13	Projection d'une tranche extraite d'une simulation lissée à différentes échelles . . . . .	91
3.14	L quinze configurations différentes de l'algorithme "marching cubes" .	94
3.15	Une surface d'isodensité sur une simulation LCDM . . . . .	95
3.16	Cas problématique dans la définition des configurations de cellules pour l'algorithme "Marching cubes" . . . . .	96
3.17	Solution des équations du squelette par calcul d'isosurfaces . . . . .	98
3.18	Les différents problèmes numérique apparaissant lors du calcul du squelette . . . . .	99
3.19	Influence des critères de sélection sur le squelette total . . . . .	102
3.20	La méthode du squelette appliquée à une distribution de galaxies simulées par GalICS . . . . .	104
3.21	Extréma d'un champs de densité détectés par la méthode des isosurfaces	105
3.22	Influence du post-traitement sur le squelette local . . . . .	109
3.23	Le squelette local d'une distribution de galaxies . . . . .	111
4.1	Influence de $\Omega_m$ sur la distribution de la matière noire . . . . .	120
4.2	Comparaisons des catalogues virtuels au SDSS . . . . .	122
4.3	Comparaison des méthodes de dérivation numérique . . . . .	124
4.4	Méthode de calcul du squelette d'un catalogue à géométrie complexe	125
4.5	Le squelette du SDSS . . . . .	127
4.6	Squelette du volume limité de SDSS et évolution avec la taille du lissage.	128
4.7	Variation de la longueur du squelette en fonction de l'échelle de lissage	131
4.8	Contraintes sur $\Omega_m$ . . . . .	132
4.9	Méthode de calcul de moments le long du squelette . . . . .	135
4.10	PDF 1D des champs de vitesse et de dispersion au voisinage des filaments. . . . .	137
4.11	PDF de $\mathbf{V}$ . . . . .	140
4.12	PDF de $\mathbf{P1}$ . . . . .	141
4.13	PDF de $\mathbf{P3}$ . . . . .	142
4.14	PDF de $\mathbf{V}$ et $\mathbf{P1}$ en fonction du redshift . . . . .	144
4.15	Influence des filaments sur le moment cinétique des halos . . . . .	145
4.16	Equation d'état $T = f(\rho)$ dans les filaments . . . . .	147
4.17	Mesure de $C_\theta(z)$ . . . . .	150
4.18	Contours de vraisemblances . . . . .	152
4.19	Longueur différentielle du squelette pour des champs Gaussiens . . .	154
4.20	Méthode de calcul de la section des filaments . . . . .	155
4.21	Le squelette avec calcul de la section des filaments . . . . .	157
4.22	Influence de l'indice spectral et du redshift sur l'ellipticité des filaments et de l'échelle de lissage sur leur section . . . . .	158
4.23	Comparaison des filaments des baryons et de la matière noire . . . . .	160
B.1	Squelette dans l'Univers local . . . . .	168