Topological techniques for inference on periodic functions with phase variation

Wojciech Reise

université PARIS-SACLAY

Topological techniques for inference on periodic functions with phase variation

Techniques d'inférence topologique pour fonctions périodiques avec variation de phase

Thèse de doctorat de l'université Paris-Saclay

École doctorale de mathématiques Hadamard n° 574 (EDMH)
Spécialité de doctorat: mathématiques appliquées
Graduate School : Mathématiques. Référent : Faculté des sciences d'Orsay

Thèse préparée dans le **Laboratoire de Mathématiques d'Orsay (Université Paris-Saclay, CNRS)**, sous la direction de **Frédéric CHAZAL**, Directeur de Recherche, et la co-direction de **Bertrand MICHEL**, Professeur.

Thèse soutenue à Paris-Saclay, le 6 Décembre 2023, par

Wojciech REISE

Composition du jury

Membres du jury avec voix délibérative

Mathilde MOUGEOT Professeure, ENS Paris-Saclay	Présidente
Wolfgang POLONIK Professeur, University of California at Davis	Rapporteur & Examinateur
Bharath SRIPERUMBUDUR Maître de conférences, HDR, Pennsylvania State university	Rapporteur & Examinateur
Paul DOUKHAN Professeur, Cergy-Pontoise Université	Examinateur
Anthea MONOD Maîtresse de conférences, Imperial College London	Examinatrice
Laurent OUDRE Professeur, ENS Paris Saclay	Examinateur

THESE DE DOCTORAT

NNT : 2023UPASM030

université PARIS-SACLAY | ÉCOLE DOCTORALE de mathématiques Hadamard (EDMH)

Titre: Techniques d'inference topologique pour fonctions periodiques avec variation de phase
Mots clés: variation de phase, statistiques non-paramétriques, analyse topologique des données, odométrie

Résumé: Les outils statistiques modernes permettent de comparer des données temporelles comportant des variations de phase, souvent en les alignant. Nous proposons des techniques pour l'analyse d'une observation qui est composée de plusieurs périodes d'une fonction périodique, avec variation de phase. Ce cadre est motivé par un problème de positionnement de voiture. D'abord, nous montrons que des constructions de l'Analyse Topologique des Données contiennent des informations sur la période et le nombre de ses occurrences. Dans une première contribution, nous proposons un descripteur d'une période de la fonction construit à partir d'un signal bruité contenant plusieurs périodes de celles-ci. On montre que ce descripteur converge quand le nombre de périodes observées augmente. Dans une seconde contribution, on estime le nombre de période que contient une observation. On montre que l'estimation est correcte pour une classe de fonctions assez large, généralisant ainsi un estimateur basé sur les lignes de niveau. De notre méthode découle également une segmentation du signal en périodes qui utilise les extrema locaux. Finalement, nous appliquons cette dernière au signal magnétique enregistré dans une voiture en mouvement. Affecté par l'empreinte magnétique du pneu, ce signal contient l'information sur la position angulaire d'une des roue et par conséquent sur la dynamique du véhicule. Nos expériences montrent que la méthode de segmentation proposée permet d'estimer la distance parcourue par le véhicule. De plus, elle localise les révolutions de manière précise, ouvrant ainsi la voie pour une méthode d'odométrie magnétique simple et accessible.

Title: Topological techniques for inference on periodic functions with phase variation
Keywords: phase variation, non-parametric statistics, topological data analysis, odometry

Abstract: Modern statistical techniques allow to compare or summarize temporal data with phase variation, often by aligning the trajectories. We propose techniques to analyse a single observation of a periodic function with phase variations, a setting motivated by a car positioning problem. We show that tools from topological data analysis can capture information about both the periodic function and the number of observed periods, quantifying the repetitive structure. In a first contribution, we propose a descriptor of a single period constructed from a noisy observation of several periods. We prove that it converges as the number of periods in the observation increases. In a second contribution, we estimate the number of periods contained in the observation. Based on similar topological tools, we prove that our estimator is correct for a wide class of functions, generalizing the zero-crossings estimator. We also give a method, guided by the topological estimator, to segment the signal into periods using local extrema. Finally, we apply this method to the magnetic signal recorded in a moving car. Composed of a tyres' magnetic fingerprint, the signal reflects the angular position of the wheel and hence the dynamics of the vehicle. Preliminary experiments show that the proposed segmentation method allows to estimate the distance travelled by the vehicle. More interestingly, it locates the revolutions of that wheel precisely, paving a way to straightforward and accessible odometry.

Acknowledgments

It is said that a doctorate is an individual journey, but it is never made alone. I would like to express my gratitude to those who contributed to the completion of this work.

To Wolfgang Polonik and Bharath Sriperumbudur, who thoroughly read the present manuscript. Your scientific perspective on my work is invaluable to me. I also thank Mathilde Mougeot, Paul Doukhan, Anthea Monod and Laurent Oudre for agreeing to form an examination committee.

À mes directeurs de thèse pour leur soutient infaillible et l'énorme liberté qu'ils m'ont accordé. Et le crédit de confiance que celle-ci a nécessité. Je garde un magnifique souvenir des ces journées passés à Nantes et au LMJL, qui donnaient une cadance à mon travail. Le temps entre l'aube et la nuit rempli de science, discussions et rencontres. L'écoute et l'encouragement que Bertrand m'a accordé ont porté ce travail et sont devenus un modèle pour moi. Merci à Fred pour sa générosité. Je suis reconnaissant pour ces discussions scientifiques, se terminant par des conseils judicieux. Je lui doit également cet environement si confortable et stimulant dans lequel j'ai pu baigner, ainsi que des moments hors du temps à sillonner la côte.

À David, pour m'avoir permis de découvrir l'univers du positionnement et donné les moyens de travailler dessus. À Pierre-Jean, pour son partage d'éxpérience et perspectives sur le sujet. À Mathieu, pour un accompagnement scientifique généreux, son attention et efficacité inégallée. À Florian, pour toute l'aide et la patience qu'à nécessité la conception et la mise en place de l'acquisition de données.

À Thomas, pour une démonstration pratique de la porosité de la membrane entre l'académique et le privé.

Aux collègues de DataShape et de l'IMO, et particulièrement aux (ex-)doctorants, pour leur présence et engagement qui continuent de créer une ambiance de travail chaleureuse et amicale. À tous ceux, qui gardent leur porte et l'esprit toujours ouverts.

Aux grimposiciens, cyclo-touristes et lindyhoppers, pour les nombreuses aventures qui m'ont toujours distraites avec succès. To friends, scattered across Switzerland, Poland and the UK, for inspiring me on a daily basis.

Rodzinie, za dom, niezachwiane wsparcie i wyrozumiałość. Za zaufanie, że to jakoś kiedyś skończe. E anche a l'altra famiglia, per la calorosa accoglienza e l'ospitalità, *in una casa, fuori città*.

A Giulia, che mi ha imparato a scrivera, a leggere, tanta parola. Grazie per la tua fiamma che mi spinge avanti.

Contents

Introduction en français

Dans cette thèse, nous étudions des données temporelles avec une structure récurrente et sujettes à des variations de phase. Les données temporelles prennent souvent la forme d'une fonction continue $S : [0, T] \to \mathbb{R}$, représentant par exemple la taille d'un enfant mesuré au cours du temps, ou les fluctuations du prix d'une action à la bourse. Un des enjeux en Statistique est l'extraction d'information portant sur la distribution sous-jacente des données. Un problème standard est la construction d'intervalles de fluctuation à partir de la donnée de M réalisations $S^1, \dots, S^M$ d'un même processus. Ces intervalles peuvent être construits *ponctuellement*, pour tout $t \in [0, T]$ séparément, en utilisant l'échantillon $(S^m(t))_{m=1}^M$. Une telle procédure ponctuelle attribue cependant toute la variance au sein de la population à la variation d'amplitude, sans prendre en compte la variation de phase. Cependant, une partie de cette variance peut être expliquée en ajustant l'axe du temps des observations, menant à analyser les données temporelles comme trajectoires.

Pour le premier exemple, le but d'une analyse ponctuelle serait de déterminer ce qu'est la taille normale d'un enfant d'âge donné, et parallèlement, ce qu'est une anomalie. Même si plusieurs âges $t_1, \dots t_n$ sont considérés, une estimation ponctuelle ne prend pas en compte les différences entre le développment des individus. En d'autres termes, le fait qu'un enfant soit de taille plus petite que son âge pourrait le prédire ne devrait pas susciter d'inquiétude. En revanche, une différence dans la forme de la courbe de croissance $(t, S(t))$ le devrait, par exemple, si sa taille venait à baisser soudainement.

Cette situation illustre de manière typique l'intêret de prendre en compte les *variations de phase* dans l'étude de données temporelles. Dans ce contexte, on dit souvent que S est une *courbe* et une réalisation est appelée *trajectoire*. Une courbe peut-être à valeur dans $\mathbb{R}^d$ ou sur une variéte (par exemple, des trajectoires d'avions ou des parcours migratoires d'oiseaux, projetés sur la surface terrestre). Les outils statistiques modernes incluent souvent la variation de phase par l'*alignement (de phase)*. Aligner deux trajectoires S^1 et S^2 consiste à trouver un homéomorphisme $\gamma : [0, T] \to [0, T]$ tel que S^1 et $S^2 \circ \gamma$ soient proches. L'alignement entre deux courbes de croissances donne une correspondance entre les maturations des deux enfants, un objet qui pourrait être utile en soi pour la comparaison de l'évolution d'autres quantités biologiques. Lorsque toutes les trajectoires sont alignées, il est possible de calculer une statistique sur cette population, comme une trajectoire moyenne.

Le prix d'une action au cours du temps est un long processus qu'on suppose souvent régie par un modèle paramétrique ou une équation différentielle stochastique. Ici, une analyse statistique visant à estimer les paramètres d'un tel modèle ou même identifier son type utiliserait des portions de la donnée temporelle. Par exemple, un estimateur pour le modèle paramétrique est construit sur $S|_{[t,t+\tau]}$, avec $0 \leq t \leq T - \tau$. Une fois les paramètres identifiés, le modèle peut être utilisé pour quantifier les risques ou gains de certaines positions.

Ces deux exemples montrent que les données temporelles peuvent être analysées avec des points de vue différents. Cela n'est pas surprenant, car les quantités d'intérêt dépendent du cadre statistique choisi. Dans l'exemple des courbes de croissance, les observations S^i, S^j sont vues comme des trajectoires. Pour la fluctuation d'indices boursiers, les propriétés locales du processus sont souvent plus pertinentes que les propriétés de la trajectoire globale.

Dans cette thèse, on s'intéresse aux données temporelles qui ont à la fois une structure interne et une variation de phase, et pour lequelles l'étude des deux est importante. Plus précisémment, on considère une fonction périodique $\phi : \mathbb{R} \to \mathbb{R}$ de période 1, ainsi qu'un homéomorphisme (strictement) croissant $\gamma : [0, T] \to [0, R]$, appelé *reparamétrisation*. On dira que

$$\begin{array}{rccl} S: & [0,T] & \to & \mathbb{R} \\ & t & \mapsto & \phi(\gamma(t)) + W(t), \end{array} \tag{1}$$

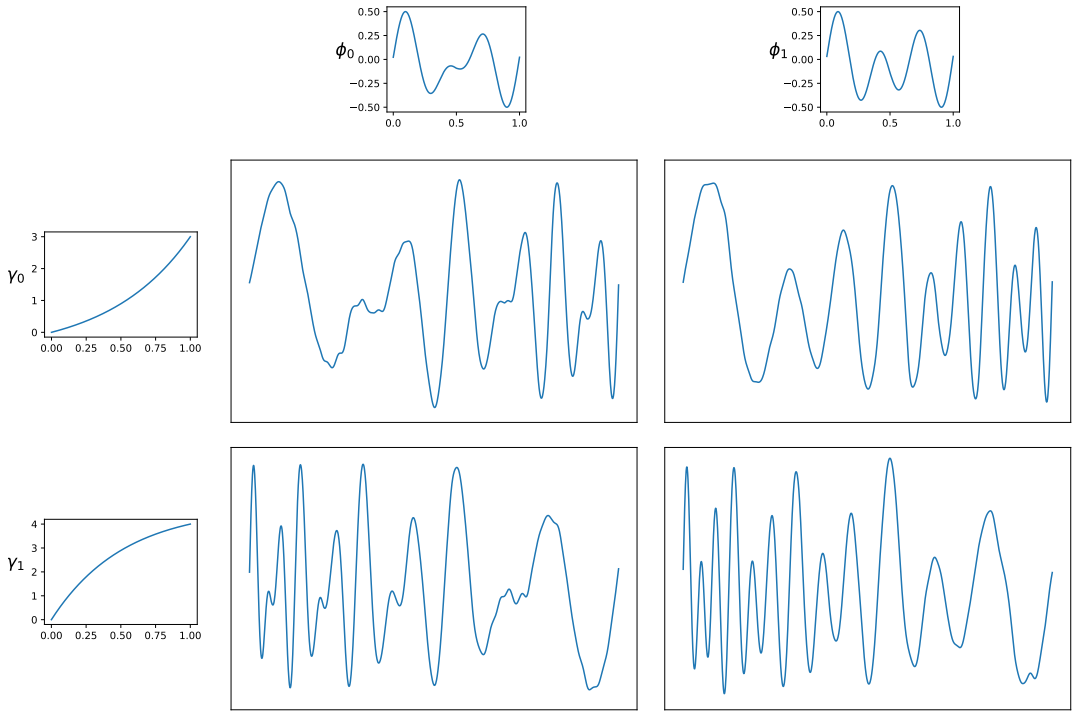


Figure 1: Deux fonctions périodiques, ϕ_0 et ϕ_1, reparamétrisées par γ_0 et γ_1, et perturbées par du bruit additif.

est une *fonction périodique reparamétrisée* corrompue par une fonction continue aléatoire $W : [0,T] \to \mathbb{R}$, qu'on identifiera comme du bruit. On nommera également S *fonction périodique avec variation de phase*. Figure 1 montre quatres exemples, obtenus en composant des fonctions périodiques ϕ_1, ϕ_2 avec des reparamétrisations $\gamma_1 : [0,1] \to [0,3]$ et $\gamma_2 : [0,1] \to [0,4]$ et en ajoutant du bruit.

Dans cette thèse, on suppose que ϕ, γ et W sont inconnues, mais on va toujours s'intéresser au cas où on observe au moins une période ($\gamma(T) \geq 1$) et où $\|W\|_\infty$ est toujours faible devant l'amplitude de ϕ. On étudie deux problème sur des données issues du modèle (1). Bien que simples à formuler, ceux-ci nécessitent d'analyser la structure répétitive dans S.

Problème 1. À partir de données temporelles (1), estimer $\gamma(T) - \gamma(0)$ et γ.

Problème 2. À partir de données temporelles (1), construire une signature de ϕ, telle que pour deux reparamétrisations γ_1, γ_2, les signatures $S^1 = \phi \circ \gamma_1$ et $S^2 = \phi \circ \gamma_2$ soient similaires.

Informellement, le Problème 1 consiste à estimer le nombre de périodes de S. Par exemple, et même sans être expert, on peut intuiter que les signaux de la première et deuxième ligne de la Figure 1 contiennent 3 et 4 périodes respectivement, et de localiser celles-ci sur l'intervalle $[0,1]$. Une segmentation en périodes donne déjà une première estimation de γ. Une solution au Problème 2 permet d'identifier si deux signaux S^1, S^2 proviennent d'observations de la même fonction périodique ϕ. L'enjeu dans le Problème 2 est de construire un descripteur statistique de ϕ à partir de S: en observant un nombre important de périodes, on souhaite être en mesure de caractériser ϕ, au moins grossièrement. Par exemple, ϕ_0 et ϕ_1 représentées en Figure 1 possèdent deux et trois 'grandes variations' par période. Un descripteur qui décrirait l'amplitude et le nombre de ces variations pourrait ainsi distinguer les signaux de la colonne de droite de ceux de la colonne de gauche. Une propriété fondamentale d'un bon descripteur est donc l'invariance par rapport à la reparamétrisation et en particulier, le nombre de périodes observées. Précisons qu'il ne s'agit tout de même pas d'estimer ϕ comme fonction, ce problème n'étant pas bien posé.

Les données de la forme (1) ainsi que le Problème 1 sont motivés par une question de positionnement de véhicule à partir de données inertielles, posée par Sysnav[1], une entreprise française spécialisée dans les techniques de positionnement et navigation. Bien que la technologie de Géolocalisation et Navigation par Système de Satellites (GNSS) soit omniprésente, elle n'est pas

[1] https://www.sysnav.fr/

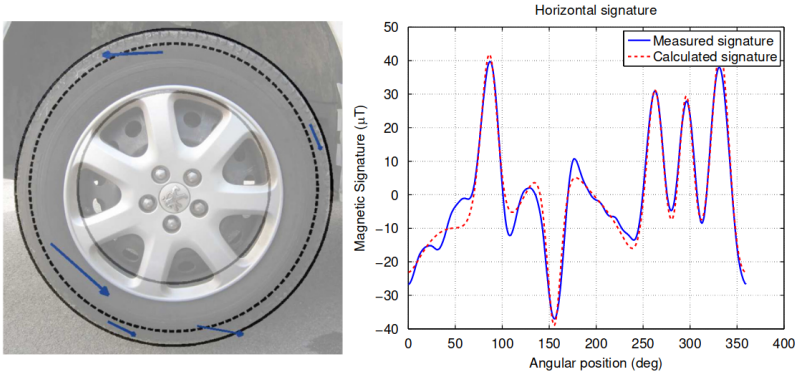


Fig. 2. Results of the optimization. Left: magnetic dipoles found after optimization. Right: measured and calculated signatures.

Figure 2: Reproduction de Le Goff et al. (2012, Figure 2) qui montre un exemple de ϕ, avec de nombreux extrema locaux. Les fonctions périodique qu'on rencontrera seront souvent plus simples.

sans faille. Le signal satellite nécessaire pour le positionnement n'est pas toujours disponible et peut même être facilement bruité. Pour atteindre le niveau de robustesse requis dans des applications sensibles, une idée consiste à développer plusieurs estimateurs de positions et de les agréger. Si les sources de données qui alimentent ces estimateurs sont de natures diverses, leurs failles sont indépendantes, augmentant ainsi la robustesse de l'estimateur agrégé. On peut, par exemple, estimer le déplacement d'une voiture en intégrant les mesures d'accélération et de vitesse angulaire fournies par une centrale à intertie (IMU). Dans ce travail, on construit un odomètre à partir des mesures d'intensité du champ magnétique. En effet, le pneu d'une voiture possède une empreinte magnétique (Le Goff et al., 2012), représentée dans la Figure 2. Un magnétomètre positionné à proximité de la roue d'un véhicule en mouvement enregistre cette empreinte, à une fréquence proportionnelle à la vitesse de rotation de la roue et donc à l'allure du véhicule. Cette dernière étant variable, l'empreinte observée comporte des variations de phase. Par rapport au modèle (1), ϕ représente l'empreinte magnétique de la roue et γ la phase de celle-ci. Une solution au Problème 1 fournit donc une estimation de la distance parcourue par le véhicule.

Analyse de données temporelles avec variation de phase

Nous décrivons quelques méthodes développées spécifiquement pour l'analyse de données temporelles avec variation de phase. Souvent, plusieurs trajectoires $S^1, \ldots, S^M$ d'une même fonction inconnue ϕ sont observées via des reparamétrisations $\gamma_1, \ldots, \gamma_M$ différentes. Notre but est d'illustrer comment le problème de variation de phase est traité ainsi que la particularité du cas où ϕ est périodique.

Une première façon de comparer des courbes avec variation de phase est de les considérer dans un espace quotient par l'action de la reparamétrisation. En définissant les espaces et la métrique, on peut munir cet espace d'une structure suffisante pour y définir des statistiques, comme la moyenne. C'est le point de vue adopté dans Srivastava et al. (2011) et le *Square-root velocity (SRV)*. Les auteurs considèrent deux actions sur l'espace des courbes $\mathcal{C} = \{\alpha : \mathbb{S}^1 \to \mathbb{R}^D\}$: la reparamétrisation $(\alpha, \gamma) \mapsto \alpha \circ \gamma$ et la rotation $(Q, \alpha) \mapsto Q\alpha$, avec $\gamma : \mathbb{S}^1 \to \mathbb{S}^1$ un homéomorphisme et $Q \in \mathrm{SO}(\mathbb{R}^D)$ une matrice de rotation. La fonction $t \mapsto \ln(\|\alpha'(t)\|)$ et la courbe $t \mapsto \alpha'(t)/\sqrt{\|\alpha'(t)\|}$ sont deux coordonnées qui s'avèrent bien décrire ces actions. En particulier, la deuxième coordonnée donne son nom au SRV et permet de montrer que l'espace quotient est une variété Riemannienne. Cette théorie a été étendue aux trajectoires sur des variétés Riemanniennes par Su et al. (2014) et a vu de nombreuses applications.

Certaines propriétés du SRV, intéressantes par natures, sont contraignantes dans notre cas. Par exemple, Srivastava et al. (2011, Figure 4) montre plusieurs courbes en forme d'hélice avec des nombres de spirales différents. La représentation dans l'espace quotient permet de partitionner ces courbes en fonction du nombre de spirales, ce qui montre que la métrique est sensible au nombre de répétitions d'une structure bien plus qu'à la reparamétrisation. Pour autant, elle ne permet pas d'obtenir $\gamma(T) - \gamma(0)$ directement.

Un autre outil qui incorpore les variations de phase dans l'analyse de données temporelles sont

les *path signatures*. Introduits dans le cadre de la théorie des chemins rugueux (Lyons et al., 2007), celles-ci sont également utilisées pour l'apprentissage statistique (Fermanian, 2021). La signature d'une courbe de variation bornée $\alpha = (\alpha_1, \dots, \alpha_D) : [0,1] \to \mathbb{R}^D$ est une suite formelle de tenseurs, $\mathcal{S}(\alpha) = (a_0, a_1, \dots, a_n, \dots)$, où $a_n \in (\mathbb{R}^D)^{\otimes n}$ est obtenu en intégrant différentes coordonnées les unes contre les autres. Par exemple, $a_1 = (\int d\alpha)_{d=1}^D = (\alpha_d(1) - \alpha_d(0))_{d=1}^D$, tandis que $a_2 = (\int_{0<s<t<1} d\alpha_i(s) d\alpha_j(t))_{1\leq i,j\leq D}$. Ainsi, l'invariance $\mathcal{S}(\alpha) = \mathcal{S}(\alpha \circ \gamma)$ par rapport à la reparamétrisation $\gamma : [0,1] \to [0,1]$ découle du changement de variable. De manière encore plus importante dans notre contexte, l'*identité de Chen* exprime la signature d'un signal en fonction des signatures de ses composantes: si on note $\alpha|_I$ la restriction de α à un intervalle $I \subset [0,1]$, alors on a l'égalité $\mathcal{S}(\alpha|_{[0,1]}) = \mathcal{S}(\alpha|_{[0,1/2]}) \otimes \mathcal{S}(\alpha|_{[1/2,1]})$, où $\otimes$ désigne le produit tensoriel. En particulier, cette identité permet de reconstruire une fonction linéaire par morceaux à partir de sa signature (Fermanian, 2021, Chapitre 6). Dans le contexte de (1), elle donne une relation explicite entre la signature de $\phi \circ \gamma$ et celle de $\phi|_{[0,1]}$. Cependant, à notre connaissance, cette identité n'a pas été exploitée pour l'étude de fonctions périodiques.

Kneip and Gasser (1992) adoptent une approche plus directe, en alignant des fonctions ̀partir des 'points de repère'. Ils localisent certains extrema locaux ainsi que des grandes variations de la fonction étudiée via, respectivement, les *e-fonctionnelles* et les *p-fonctionnelles*. Étant donné une collection de $N \in \mathbb{N}$ telles fonctionnelles, un signal $S : [0,T] \to \mathbb{R}$ est encodé par les repères $(T_n(S))_{n=1}^N$, avec $0 \leq T_1(S) \leq \dots \leq T_N(S) \leq T$. Ainsi, pour synchroniser deux fonctions $S_1, S_2 : [0,T] \to \mathbb{R}$, il suffit de trouver une correspondance monotone entre les deux vecteurs correspondants. Dans un modèle semi-paramétrique et sous une hypothèse d'unicité de localisation des points de repère, la procédure est consistente: quand la fréquence d'échantillonnage augmente, les points de repère sont localisés avec toujours plus de précision, permettant d'inférer les paramètres t_0 et t_1 de $\gamma(t) = (t - t_0)/t_1$. Il est à noter que les résultats sont conditionnés au fait que les points de repères ont été au moins partiellement identifiés. Dans le cas des fonctions périodiques, les repères se répètent et les conséquences de cela ne sont pas claires pour nous.

Le Dynamic Time Warping (DTW) (Sakoe and Chiba, 1978) est une mesure de dissimilarité adaptée à la comparaison de courbes avec variations de phase. Elle est définie pour des séries temporelles: si on note $S_n = S(t_n)$ pour $0 \leq t_1 < \dots < t_N \leq T$, alors la DTW est la solution du problème d'optimisation

$$\mathrm{DTW}((S_n^1)_{n=1}^N, (S_m^2)_{m=1}^M) = \min_{i,j} \sum_{k=1^K} |S^1_{i(k)} - S^2_{j(k)}|, \tag{2}$$

où le minimum est pris sur des fonctions $i : \{1, \dots, K\} \to \{1, \dots, N\}$ et $j : \{1, \dots, K\} \to \{1, \dots, M\}$, avec $K \in \mathbb{N}$. Des contraintes telles qu'un module de croissance et la surjectivité sont habituellement imposées sur i et j. De nature explicite et simple à calculer, la DTW est un outil priviligié pour comparer ou aligner des séries temporelles. En incorporant les dérivées de S dans le critère à optimiser, les solutions i, j de (2) alignent S^1 et S^2 sur leurs points de repères donnés par les e- et p-fonctionnelles (Gasser and Wang, 1997).

En traitement du signal, le problème d'estimation de la modulation de la phase d'une fonction périodique est connu sous le nom de *instantaneous phase estimation* (Boashash, 2015, Chapitre 10). Pour des fonctions "simples", comme $\phi(x) = a \sin(2\pi(x - b))$, la solution exacte est $\gamma(t) = \arctan(H(S)(t)/S(t))$, où H est la *transformée de Hilbert* (Boashash et al., 1990). Pour traiter un signal complexe, on analyse sa décomposition temps-fréquence avec des techniques comme la détéction d'extrema (Rankine et al., 2007). De telles procédures sont mêmes extensibles à des signaux qui sont des sommes de composantes avec différentes phases, pourvu que celles-ci soient bien séparées dans le spectre des fréquences (Khan and Boashash, 2016, Hussain and Boashash, 2002). C'est le point de vue adopté dans Bristeau (2012), Zabulon et al. (2019) pour le positionnement de véhicule à partir de signaux magnétiques.

Une autre technique reposant sur des 'repères' est celle des *zero-crossings*. Elle consiste à utiliser $S^{-1}(\alpha)$ pour déterminer le nombre de périodes et la phase, pour $\alpha \in \mathbb{R}$, et souvent $\alpha = 0$. En l'absence de bruit, le nombre de fois que le graphe de S traverse le niveau α est proportionnel au nombre de périodes dans le signal et $S^{-1}(\alpha)$ permet de retrouver une segmentation qui correspond à ces dernières. Cette méthode simple nécessite seulement de choisir α. Elle n'est cependant pas stable par rapport aux perturbations de S par du bruit additif W, ni adaptée à l'étude de fonctions dont la période contient plusieurs passages par le niveau α, comme il en sera discuté dans le Chapitre 5.

Il est alors intéressant de se demander si les outils introduits pour la comparaison de courbes

avec variation de phase sont adaptables au comptage du nombre de périodes ou à la quantification des répétitions. Le zero-crossing n'est qu'un exemple de 'repère'. De telles adaptations nous sont inconnues, sauf dans des rares cas où la fonction est effectivement périodique (Stefan et al., 2013), mais on en mesure pourtant la difficulté sur l'exemple des e-fonctionnelles de Kneip and Gasser (1992), pourtant ci efficaces. La n-ième e-fonctionnelle détecte la position du n-ième plus petit minimum. Pour une fonction périodique reparamétrisée (1), le nombre d'extrema locaux dépend du nombre de périodes, et leur ordre est arbitraire et instable: toute perturbation même minime peut placer le minimum global dans la première ou la dernière période de S.

Dans notre travail, nous proposons des 'repères' et descripteurs issus de l'Analyse Topologique des Données (TDA). Ceux-ci sont stables et plus spécifiques que, par exemple, les zero-crossings et les e-fonctionnelles. Ils permettent aussi de cerner la structure répétitive des signaux observés, avec une complexité algorithmique linéaire en la longueur du signal (Glisse, 2023) et au plus quadratique en le nombre d'extrema locaux.

Analyse topologique des données

L'Analyse Topologique des Données est un domaine récent et désigne également un ensemble de techniques qui décrivent la forme des données par l'extraction de quantités géométriques ou topologiques (Chazal and Michel, 2021). Ces techniques s'appliquent à des données de natures diverses, souvent complexes: des nuages de points, des graphes (Hofer et al., 2020, Lasalle, 2023), des images (Bleile et al., 2021, Garin and Tauzin, 2019) et également des séries temporelles (Fernández and Mateos, 2022, Gidea and Katz, 2018, Gidea et al., 2020, Perea, 2019). Pour un panorama plus complet de ce domaine, nous renvoyons le lecteur à la revue Chazal and Michel (2021).

Un des outils phares de la TDA est l'*homologie persistante*. L'homologie est un invariant topologique qui qualifie rigoureusement la notion intuitive de 'trou': un trou de dimension 1 correspond à un cycle, et de dimension 2 - à une cavité. L'homologie persistante permet d'étudier l'évolution de l'homologie à travers une suite d'espaces : de leur apparition à leur disparition. Ceci définit la notion de *persistance*: savoir à quels moments ce 'même trou'[2] est présent. L'homologie persistante est typiquement vue comme un version multi-échelle de l'homologie. En effet, l'estimation de l'homologie à partir de données discrètes (tel un nuage de points) passe souvent par la construction d'un complexe simplicial, ou l'estimation de sous-niveaux d'une densité; qui nécessitent le choix d'un paramètre d'échelle. Comment choisir cette échelle est compris dans un cadre théorique et pour des modèles géométriques, mais les applications de l'homologie comme descripteur des données vont au-delà. L'introduction de la persistance permet de contourner ces questions, en décrivant l'évolution de l'homologie à travers les échelles.

L'*homologie persistante des sous-niveaux* d'une fonction continue $f : [0, T] \to \mathbb{R}$ décrit les extrema locaux de f et leur ordre. Informellement, dans ce cadre simple, on peut la comprendre par l'analogie suivante illustrée par la Figure 3. Supposons que f représente un profil d'altitude et qu'on examine la formation de réservoirs d'eau au fur et à mesure que le niveau de mer monte. Des réservoirs se forment dans les minima globaux, puis dans d'autres minima locaux, tandis que les maxima locaux deviennent des îlots. Lorsque le niveau d'eau franchit le niveau d'un maximum, l'îlot disparait et deux réservoir adjacents fusionnent pour n'en former plus qu'un: le plus profond des deux absorbe l'autre. L'homologie persistante décrit l'évolution de ces réservoirs, composantes connexes, dans $(f^{-1}(]-\infty, \alpha]))_{\alpha\in\mathbb{R}}$, ici représentés par des intervalles verticaux. Une représentation différente de ces intervalles est le *diagramme de persistance*, un (multi-)ensemble de $\mathbb{R}^2$, où chaque intervalle d'extremités (y_1, y_2) est associé à un point de coordonnées (y_1, y_2).

L'homologie persistante des sous-niveaux associe donc une notion de persistance aux minima locaux. Des intervalles courts, ou de façon équivalente, des points près de la diagonale $\Delta = \{(y, y) \in \mathbb{R}^2 \mid y \in \mathbb{R}\}$ représentent des oscillations de faible amplitude, souvent assimilés à du bruit de haute fréquence ou à un manque de regularité de f. En effet, la *p-persistance totale*, définie comme la somme des $(y_2 - y_1)^p$ avec (y_1, y_2) les points du diagramme, est finie pour les fonctions α-Hölder dès que $p > 1/\alpha$ (Perez, 2023).

[2]Plus rigoureusement, la persistance est une notion attribuée aux générateurs de l'homologie, car les 'troux géometriques' varient au cours dans la suite d'espaces.

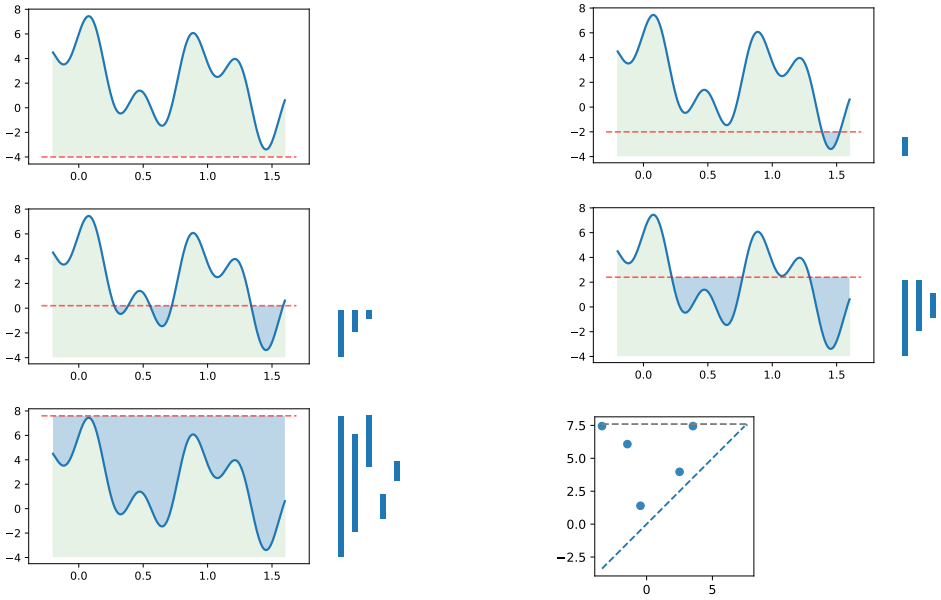


Figure 3: Une représentation graphique d'une séquence de sous-niveaux de f, croissants de gauche à droite et de haut en bas. Les intervalles verticaux à coté de chaque graphique représentent les composantes connexes et la figure en bas à droite - le diagramme de persistance de f.

TDA pour les données temporelles

Il existe de nombreuses techniques d'analyse de données temporelles avec l'homologie persistante. Souvent, une structure combinatoire telle qu'un complexe simplicial est d'abord définie, de sorte à ce que l'homologie de celle-ci soit un invariant intéressant du problème.

La technique sans doute la plus répandue pour l'étude de série temporelles est le *plongement de Takens*. Takens (1981) donne un plongement simple et explicite de l'*attracteur* d'un système dynamique dans $\mathbb{R}^D$, défini en terme d'une trajectoire de ce système, et l'attracteur étant un sous-espace spécifique de l'espace des phases. Les propriétés fractales de l'attracteur qui sont de grand intérêt pour déterminer si le système est chaotique ou périodique peuvent alors être étudiées avec l'homologie. En TDA, cette application de plongement est souvent utilisée pour extraire de l'information sur le comportement d'une série temporelle. On construit $\mathrm{SWE}(S) \in \mathbb{R}^{d\times(N-(d-1)\tau)}$ en faisant glisser une fenêtre de taille d sur la série temporelle $S = (S_n)_{n=1}^N$,

$$\mathrm{SWE}(S)_n = (S_n, S_{n+\tau}, \ldots, S_{n+(d-1)\tau}) \in \mathbb{R}^d, \tag{3}$$

pour un certain $\tau \in \mathbb{N}$. Si S est périodique, l'homologie persistante d'un complexe sur $\mathrm{SWE}(S)$ ressemble à celle du cercle $\mathbb{S}^1$: en particulier, elle contient un générateur persistent en dimension 1. Cette idée a été mise en œuvre pour classifier des systèmes climatiques (Ghil and Sciamarella, 2023), comme un indicateur de volatilité des produits financiers (Gidea et al., 2020, Ruiz-Ortiz et al., 2022), pour détécter des vibrations dans un processus de fabrication (Khasawneh et al., 2018), la reconnaissance de syllabes dans le chant d'oiseaux (Fernández et al., 2023). Cette liste non-exhaustive témoigne du succès que rencontre cette méthode.

Cependant, lorsqu'on introduit la variation de phase, le signal observé n'est plus périodique. La longueur de la structure récurrente de $\phi\circ\gamma$ varie selon γ, ce qui casse la géométrie de SWE, comme illustré par la Figure 4. Les paramètres $d, \tau \in \mathbb{N}$ doivent être choisis en fonction de la période, ce qui n'est pas possible quand la longueur de celle-ci change. Les résultats de Perea (2019), Kim and Jung (2022) reliant la décomposition de Fourier de S à la géométrie de $\mathrm{SWE}(S)$ et son homologie persistante quantifient indirectement ce problème. Dłotko et al. (2019) propose une méthode qui s'adapterait à des variations de phase.

Bien que répandue, la technique du plongement n'est pourtant pas la seule approche topologique pour l'étude de données temporelles périodiques ou récurrentes. Khasawneh and Munch (2018) s'intéressent à des signaux binaires, constants par morceaux, sujets à un bruit qui introduit des transitions parasites entre les deux états. Souhaitant compter le nombre de "vraies" transitions,

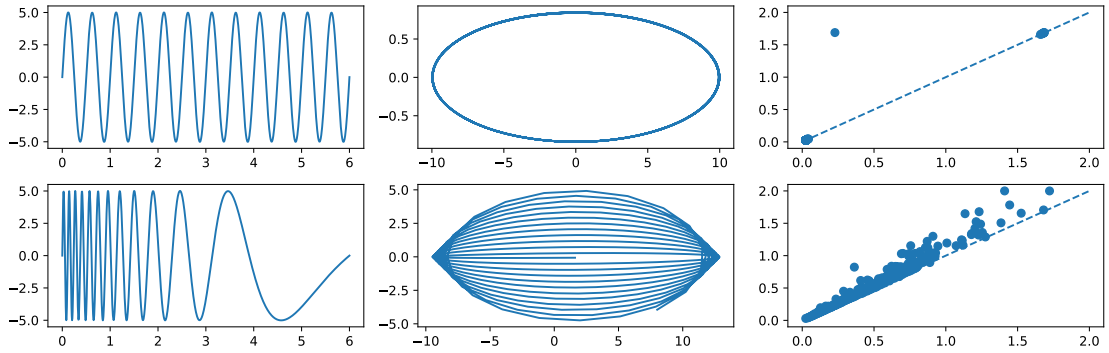


Figure 4: En haut, de gauche à droite, le graphe d'une fonction périodique, la projection de SWE(S) sur deux composantes principales et le diagramme de persistence en dimension 1 du complexe de Vietoris-Rips construit sur SWE(S). En bas, les graphiques correspondants pour une reparamétrisation du signal en haut. La géométrie du plongement est différente pour le signal reparamétrisé, comme en témoigne l'homologie persistante qui ne contient plus de générateur persistant.

qui sont généralement plus longues, les auteurs attribuent une notion de persistance aux transitions, ce qui leur permet d'éliminer celles qui sont parasites. Erden and Cetin (2017) considèrent $(\frac{nT}{N}, S(\frac{nT}{N}))_{n=1}^{N}$ comme un nuage de points. Les oscillations de S donnent lieu à des générateurs de l'homologie en dimension 1, et les auteurs obtiennent une variante plus robuste des zero-crossings. Finalement, Bois et al. (2022) utilisent la construction des sous-niveaux décrite en Figure 3 pour caractériser la marche de patients atteints de maladies neuro-dégénératives. Ils constatent que le nombre de répétitions d'un intervalle reflète le nombre de pas dans l'enregistrement. Après l'avoir formalisé, nous exploitons cette observation.

Analyse topologique des données aléatoires et statistiques

L'homologie et sa version persistante contiennent des informations pertinentes sur des structures complexes, mais en pratique, celles-ci ne sont observées qu'à travers des réalisations aléatoires. Dans le cadre de (1), on dira par exemple qu'on observe ϕ seulement à travers S, avec γ et W aléatoires. Il est donc important de comprendre comment l'homologie persistante se comporte pour des données aléatoires, corrompues ou incomplètes.

Un cadre statistique commun en TDA est celui de l'inférence de l'homologie d'une variété à partir d'un échantillon de points. Une solution consiste à construire un complexe simplicial sur ce nuage de points, à la bonne échelle. Ceci permet aussi d'estimer le diagramme de persistance de façon consistante (Chazal et al., 2015). En pratique, les données sont souvent bruitées (Aamari et al., 2023) ou corrompues (Vishwanath et al., 2022).

Au-delà de l'inférence, l'homologie persistante est également souvent utilisée comme un descripteur des données structurées, même si le modèle n'introduit pas explicitement de géométrie ou de topologie. Dans ce cas, on est souvent intéressé par la distribution des caractéristiques topologiques multi-échelle. Par exemple, en apprentissage statistique, Garin and Tauzin (2019) proposent d'extraire d'une image plusieurs descripteurs topologiques, une procédure qui permet de classifier le jeux de données MNIST.

On distingue deux cadres statistiques. Dans le premier, on observe une collection i.i.d d'un certain type de données, par exemple des images. On décrit la distribution empirique de ces images par la biais de celles des diagrammes de persistance. C'est souvent le cas dans l'apprentissage statistique, et c'est ce cadre qui est à l'origine du développement des méthodes de *représentation des diagrammes de persistance*. En effet, en tant que multi-ensembles, ces derniers ne sont pas adaptés à l'apprentissage et un vecteur leur est souvent associé (Bubenik, 2015, Zielinski et al., 2021, Adams et al., 2017, Reininghaus et al., 2014). Pour la plupart, les constructions proposées convergent asymptotiquement : quand la taille de l'échantillon (nombre d'images) croît, la vectorisation empirique moyenne converge vers son espérance et on peut approcher sa distribution par des techniques de *bootstrap* (Chazal et al., 2014, Berry et al., 2020). Ces résultats sont une conséquence de la théorie générale des processus empiriques (Kosorok, 2008) et reposent sur la

donnée de plusieurs diagrammes de persistance indépendants.

La limite de ces processus est parfois étudié dans un deuxième cadre, qui consiste en l'observation d'une structure qui grandit ou qui est échantillonnée de plus en plus finement. Un exemple typique est la convergence de descripteurs géométriques quand la taille d'un nuage de points augmente. Décrire un diagramme aléatoire qui a un caractère 'global' à travers des interactions et propriétés statistiques locales n'est pas simple. La distribution moyenne des descripteurs topologiques et leur estimation a été abordée dans Kahle (2011), Divol and Polonik (2019). Récemment, Roycraft et al. (2023) proposent des techniques de bootstrap pour construire des intervalles de confiance sur ces descripteurs. Pour les données temporelles, Krebs (2021) établit la convergence (en espérance) de l'homologie persistante du SWE d'une chaîne de Markov, sous certaines hypothèses portant sur le noyaux de transition. Bobrowski and Borman (2012) détérmine l'espérance du nombre de certaines barres pour les (sous-niveaux de) processus Gaussians, Divol and Chazal (2019) pour le mouvement Brownien et Perez (2022b) pour des processus de Markov plus généraux.

Contributions

Dans cette thèse, on propose des techniques de TDA pour l'analyse de fonctions périodiques avec variation de phase (1), par le biais des problèmes 1 et 2. D'une part, on compare deux signaux avec des nombres de périodes différents et de l'autre, on compte le nombre de celles-ci dans le cas où ϕ est inconnue. On applique nos méthodes pour le problème de positionnement décrit ci-dessus.

1. Additivité des diagrammes de fonctions périodiques (Chap. 3)

On montre que l'homologie persistante des sous-niveaux d'un signal de forme (1) est additive. On note $D(\phi|_I)$ le diagramme de persistance de $\phi|_I$, la restriction d'une fonction 1-périodique ϕ à un intervalle $I \subset \mathbb{R}$. Alors, le diagramme de persistance est additif.

1. Si $R \in \mathbb{N}$, alors
$$D(\bar{\phi}\big|_{[0,R]}) = RD(\bar{\phi}\big|_{[0,1]}), \tag{4}$$
où $\bar{\phi}$ désigne une version de ϕ adaptée aux traitement des bords de l'intervalle.

2. Pour tout $R > 1$,
$$D(\phi|_{[0,R]}) = \lfloor R-1 \rfloor D^* + D', \tag{5}$$
où $D^* = D(\bar{\phi}\big|_{[0,1]})$ et D' sont deux diagrammes de persistance. De plus, la persistance totale de D' est majorée par (le double de) celle de D^*, donc D' peut être considéré comme un "reste".

Une identité qui exprime $D(\phi|_{[0,R]})$ à l'aide de $D(\phi|_{[0,1]})$ ne surprendra pas les lecteurs familiers de l'homologie persistante. On prend soin de formaliser celle-ci, avec une attention particulière pour les conditions de bord: même avec $R \in \mathbb{N}^*$, $D(\phi|_{[0,R]}) \neq RD^*$.

Les deux identités (4, 5) sont à la fondation de nos deux contributions principales. Une, qui repose sur le fait que le nombre de périodes apparait explicitement dans (4), et l'autre, où nous caractérisons ϕ via D^*.

2. Segmentation d'un signal périodique avec variation de phase (Chap. 5)

Quand ϕ est une fonction 'simple', les zero-crossings segmentent $[0, T]$ en périodes de ϕ. Par contre, ni le nombre de périodes obtenu ni cette segmentation ne sont stables par rapport aux perturbations de S. On propose une solution en deux étapes au Problème 1 quand le nombre de périodes $N := \gamma(T) - \gamma(0) \in \mathbb{N}$ est un entier.

D'abord, on estime le nombre de périodes avec $\hat{N}(S, \tau)$, qui compte les multiplicités des points dans le diagramme et repose sur (4). Cet estimateur dépend d'un paramètre d'échelle $\tau > 0$, qui détermine un seuil en dessous duquel deux points sont considérés proches. On montre que si cette échelle est bien choisie et que le bruit est d'amplitude suffisamment petite, alors $\hat{N}(S, \tau) = N$. On propose deux adaptations de $\hat{N}$, qui s'avèrent plus robustes que $\hat{N}$ dans nos études numériques, mais pour lesquelles les garanties restent encore à établir. Les estimateurs généralisent l'approche zero-crossings, car $\mathrm{card}(S^{-1}(\alpha))$ se lit dans $D(S)$.

Dans un deuxième temps, on construit une segmentation de $[0, T]$ en périodes de ϕ. Si $\hat{N}(S, \tau) = N$, par construction de $\hat{N}$, on peut extraire exactement NK points du diagramme de S, pour un $K \in \mathbb{N}$. Si on considère $(x_i)_{i=1}^{NK}$ les minima locaux associès à ces NK points, alors $(x_{1+K(n-1)})_{n=1}^{N}$ segmente l'intervalle $[x_1, x_{NK-K+1}] \subset [0, T]$ en périodes de ϕ.

3. Odométrie magnétique (Chap. 7)
En s'appuyant sur l'algorithme topologique de segmentation, nous développons un odomètre magnétique. Pour des trajectoires simples, celui-ci détermine la phase d'une roue de manière plus fiable qu'un estimateur basé sur les zero-crossings, surtout quand l'empreinte magnétique du pneu possède plusieurs extrema locaux. Nous étudions également ses limitations sur des trajectoires et environnements plus complexes. Le bruit de basse ou de haute fréquence casse la structure répétitive, qui néanmoins peut être parfois retrouvée en analysant le champ magnétique selon une bonne direction, ou en utilisant le gradient de ce dernier.

4. Signatures topologiques de fonctions périodiques (Chap. 4)
De nombreuses méthodes existent pour comparer des signaux avec variations de phase. Par contre, n'exploitant pas la structure répétitive des fonctions périodiques, elles ne sont pas adaptées au cas où les signaux contiennent des nombres de périodes différents.

Nous proposons une signature $F(S)$ de S, qui caractérise ϕ. Nous la définissons comme la moyenne de fonctions associées à des points du diagramme de persistance $D(S)$. Grâce à (5), on montre que cette signature converge quand $R \to \infty$ et $W = 0$. De plus, dans un cadre stochastique où γ_1 et γ_2 sont aléatoires avec $\gamma_1(0) = \gamma_2(0)$ et $\gamma_1(T) = \gamma_2(T)$, nous contrôlons la distance entre $F(S_1)$ et $F(S_2)$ par une distance entre les distributions de γ_1 et γ_2. Bien que la nature des descripteurs s'y prête, nous ne démontrons pas la stabilité pour des nombres de périodes différents entre les signaux.

Calculer la signature de $S = (S_n)_{n=1}^N$ demande de connaître sa loi et n'est donc pas possible à partir d'une seule observation. Nous proposons d'utiliser la signature $F(X)$ d'un processus plus court $X = (S_m)_{m=1}^M$, pour $M < N$, qui décrit bien la structure de ϕ également. Nous montrons qu'une méthode de bootstrap permet d'estimer $F(X)$ à partir d'une collection de fenêtres $(X_k)_{k=1}^{N-M+1}$ avec $X_k = (S_{k+m})_{k=0}^{M-1}$, pourvu que la dépendance dans le signal décroit suffisamment vite. Motivé par le cadre industriel où γ représente le déplacement d'un véhicule au cours du temps, nous spécifions γ comme une chaîne de Markov et montrons qu'une mesure de la dépendance décroit suffisamment vite pour guarantir la validité de la méthode bootstrap.

5. Fonctionnelles normalisées et persistance tronquée (Chap. 4)
Les représentations vectorielles et fonctionnelles des diagrammes de persistance sont souvent linéaires et leur continuité est bien comprise. Elle dépend de la p-persistance totale, dont la continuité et convergence sont régies par la régularité de la fonction et la dimension du domaine, pour un diagramme des sous-niveaux d'une fonction.

On introduit des représentations normalisées des diagrammes ainsi qu'une version tronquée de la persistance totale. On montre dans la Section 4.3 que la troncature rend la persistance continue sur une classe de fonctions plus large. Il en est de même pour les représentations normalisées, ce qui est un élément fondamental pour la stabilité des signatures $F(S)$.

Structure

La Figure 5 est une représentation schématique du contenu (contributions) et des liens de dépendance entre les chapitres de cette thèse. Le Chapitre 2 introduit le bagage technique de cette thèse et ne contient pas nos contributions. Dans les Sections 2.1 à 2.4, nous rappelons des notions de mesurabilité de processus stochastiques et des résultats de convergence. La présentation est dense et s'adresse à des experts. La Section 2.5 introduit l'homologie, l'homologie persistante et discute ses propriétés. Notre but principale est de rendre précise l'intuition établie dans la Figure 3 et présenter notre utilisation de la théorie de la persistance. En particulier, cette section contient les détails techniques sollicités dans le Chapitre 3.

Dans le Chapitre 3, nous montrons les propriétés additives des diagrammes de persistance de fonctions périodiques de la contribution 1. Même si les discussions sont principalement algébriques, nous illustrons les parties clés. Le Chapitre 4 introduit les signatures. La première partie de ce chapitre est dédiée à l'étude de leur continuité et de leurs propriétés statistiques. La deuxième partie, en Section 4.3, présente la contribution 5. Dans le Chapitre 5, nous étudions le Problème 1, que nous proposons de résoudre comme décrit dans la contribution 2. Le Chapitre 6 explicite comment appliquer les algorithmes d'estimation du nombre de périodes et de segmentation sur des signaux de dimension 3 - une question venant de l'application étudiée dans le Chapitre 7. Ce dernier présente le contexte du positionnement de véhicule suivi de l'étude de la performance de certaines méthodes du Chapitre 5 appliquées aux données magnétiques.

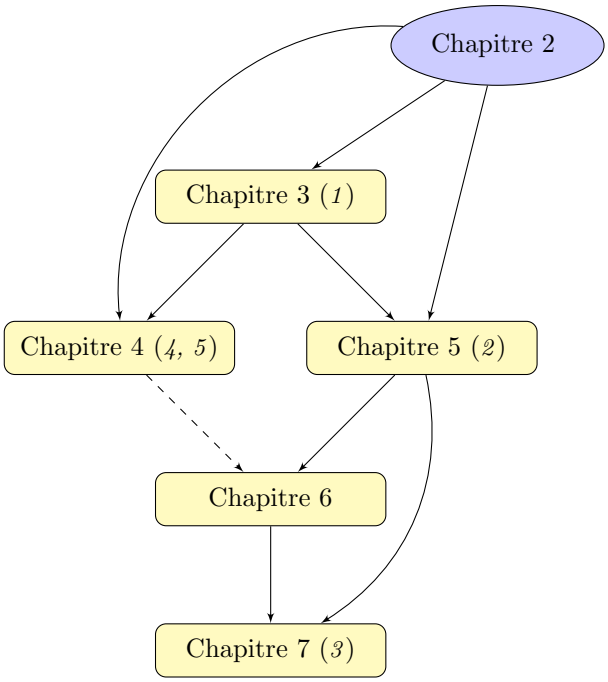


Figure 5: Un schéma de la structure de cette thèse, avec (n) faisant référence à la contribution n et les flèches indiquent la dépendances entre les chapitres. Par exemple, le Chapitre 3 repose sur le bagage introduit dans le Chapitre 2.

Chapter 1

Introduction

In this thesis, we study temporal data with recurrent structure and phase variation. *Temporal data* can be a random continuous function $S : [0, T] \to \mathbb{R}$, for example the height of a child over time, or the price of a stock market share. One of the main concerns in Statistics is extracting information about the underlying distribution modelling a phenomenon, from observed data. A typical problem is the construction of prediction intervals based on M different realizations of the same process, $S^1, \dots, S^M$. They can be constructed *point-wise*, for each $t \in [0, T]$ separately, based on $(S^m(t))_{m=1}^M$. It attributes all variance to amplitude variations among individual realizations, but it does not include *phase-variation*. Some of the variations in a point-wise analysis can indeed be explained by adjusting the time axis of the observation, what leads to the idea of analysing temporal data as a *trajectory*.

Going back to the first example, a point-wise analysis would aim to determine what constitutes a 'normal' height for a child at a given age, and, as a consequence, what is a worrying anomaly. Even if not one, but the heights at a few, fixed time points were considered, it does not take into account the variations in the maturation tempi between individuals, of which age is only a proxy. In laymans' terms, it should not be worrying if an child is too short and falls outside of classically-constructed prediction intervals. However, more abnormal are differences in the 'shape' of the growth curve, $(t, S(t))$. An example would be if its height suddenly started to decrease.

This is the prototypical motivation to introduce *phase variation* in the study of temporal data. In this context, S is sometimes called a *curve* and its realization a *trajectory*: the same limitations of a pointwise, or fixed-time analysis are valid for temporal data with values in $\mathbb{R}^d$ or a manifold (for example, plane trajectories projected on the Earths' surface). Modern statistical methods take phase variation into account, often by *curve (or phase) registration*: 'aligning' the observations by finding a homeomorphism $\gamma : [0, T] \to [0, T]$ such that S^1 and $S^2 \circ \gamma$ are close, in some sense. The interpretation in the example with children is finding a correspondence between the maturation tempi of two children. Once all curves are synchronized, it is for example possible to define and compute an average trajectory: a statistic of a population.

Consider now the example of stock prices. For a given stock, only a single, long process of prices is observed. Often, it is assumed that the price is somehow 'stationary' and that it follows a parametric model, or satisfies a differential equation driven by a stochastic term. A statistical analysis which aims to infer the parameters of that model is conducted on 'some pieces' of the temporal data. For example, a statistic is computed on $S|_{[t,t+\tau]}$ for all $0 \leq t \leq T - \tau$ and some $\tau > 0$, and the values are used to construct an estimator for the parametric model. Practically, once the parameters are known, one can assess the risk, estimate prospective gains, and even provide an interpretation about what drives the price in certain situations.

The two examples show that temporal data can be analysed from different points of view. In a way, our examples state the obvious: the quantities of interest depend on the data model. In the first setting, observations S^i, S^j are analysed as whole trajectories and it makes little sense to 'cut' out a piece of a given length or analyse the 'internal structure'. The opposite can be said of stock market prices, where the stochastic term has some local or regularity properties, but the trajectory is not shared across realizations.

In this thesis, we study temporal data which has both internal structure and phase variation, which we are interested in analysing. Specifically, consider $\phi : \mathbb{R} \to \mathbb{R}$ a periodic function with period 1, and $\gamma : [0, T] \to [0, R]$ an increasing homeomorphism for some $R > 1$, also called a

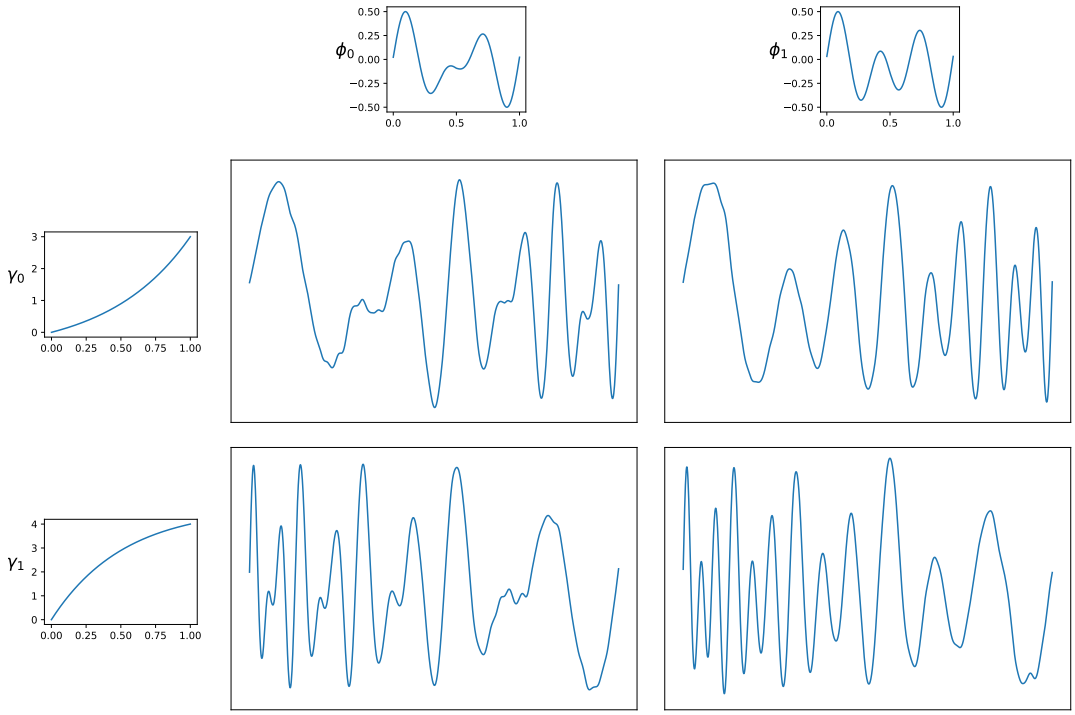


Figure 1.1: Two periodic functions, ϕ_0 and ϕ_1, reparametrised by γ_0 and γ_1, with some additive noise.

reparametrisation. Then, we say that

$$\begin{array}{rccl} S: & [0,T] & \to & \mathbb{R} \\ & t & \mapsto & \phi(\gamma(t)) + W(t), \end{array} \tag{1.1}$$

is a *reparametrized periodic function* corrupted by noise $W : [0,T] \to \mathbb{R}$, a random continuous function. We will also call S a *periodic function with phase variation* and often omit the noise in the terminology. Figure 1.1 shows an example of four different signals S, obtained by composing the periodic functions ϕ_1 and ϕ_2 with two parametrisations $\gamma_1 : [0,1] \to [0,3]$ and $\gamma_2 : [0,1] \to [0,4]$, with some additive noise.

In this thesis, we will assume that ϕ, γ and W are all unknown, although $\|W\|_\infty$ will always be small compared to the variations of ϕ. We focus on two, simple to formulate problems on data of form (1.1), both of which require analysing the "repetitive structure" of S.

Problem 1.1. Given temporal data S of the form (1.1), estimate $\gamma(T) - \gamma(0)$ and construct an estimator of γ.

Problem 1.2. Given temporal data S of the form (1.1), construct a signature of ϕ, such that for any two reparametrisations γ_1, $\gamma_2 : [0,T] \to \mathbb{R}$, the signatures of $S^1 = \phi \circ \gamma_1$ and $S^2 = \phi \circ \gamma_2$ are similar.

Informally, Problem 1.1 consists in counting the number of periods. For example, asked to determine the number of periods in Figure 1.1, a non-expert would clearly identify them as 3 and 4 in the top and bottom rows respectively, and would segment $[0,T]$ into the periodic structure without much difficulty. From such a segmentation, one could construct an estimator of γ.

Roughly speaking, a solution to Problem 1.2 indicates whether two signals S^1 and S^2 are observations with the same underlying periodic function ϕ. It involves constructing a statistical descriptor of ϕ, based on an observation of S: as we observe an increasing number of periods of ϕ, we should be able to characterize ϕ, at least roughly. Clearly, the point is not to estimate ϕ as a function, because estimating the phase in model (1.1) is an ill-posed problem, but only characterize some of its features. For example, ϕ_0 and ϕ_1 from Figure 1.1 have two and three large 'variations' in a single period respectively. A descriptor that captures the number of such variations per period could distinguish the left column from the right one. For a descriptor to characterize ϕ, it needs to be invariant with respect to the reparametrisation, as formalized in Problem 1.2.

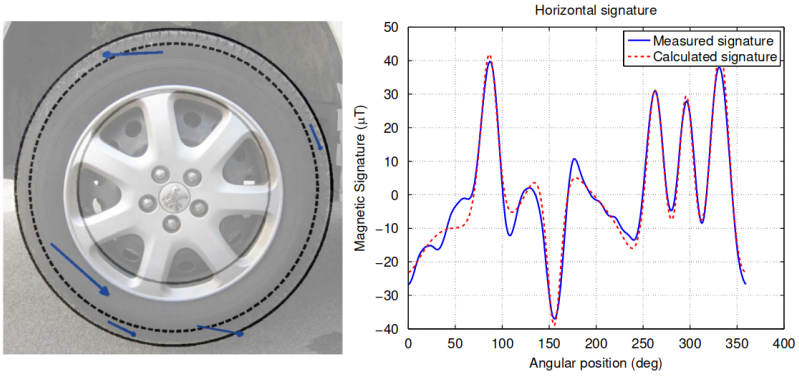


Fig. 2. Results of the optimization. Left: magnetic dipoles found after optimization. Right: measured and calculated signatures.

Figure 1.2: Figure 2 from Le Goff et al. (2012) which shows an example of ϕ with numerous local extrema. It is rather complicated and the periodic functions we encounter in practice have a simpler form.

Data of the form (1.1) and Problem 1.1 in particular are motivated by a car positioning problem, introduced to us by Sysnav[1], a French company specialized in positioning systems. While ubiquitous, the Global Navigation Satellite System (GNSS) is not entirely reliable: certain applications require constant positioning and the satellite signal might be temporarily not available or even jammed. To increase robustness, it is common to augment the Global Navigation Satellite System (GNSS) with other position estimators, all of which are then aggregated to produce a single estimator, whose reliability does not depend on a single source of data. As an example, an estimator of the displacement of a car is constructed by integrating the accelerations and angular velocities, acquired by an Inertial Measurement Unit (IMU). In this thesis, we are interested in exploiting the magnetic signal. A car's tire is magnetised, so its revolution has a magnetic fingerprint (Le Goff et al., 2012), as shown by Figure 1.2. When a magnetometer is placed inside a moving car and near one of its wheels, the fingerprint is observed with frequency corresponding to the speed of that car. The last not being constant, the fingerprint is observed with some variation in the phase. In terms of our model, ϕ is the magnetic fingerprint projected along one direction, and γ represents the angular phase of a wheel in time. A solution to Problem 1.1 is an estimator of the number of revolutions of one of the cars wheels, and hence of its displacement.

1.1 Analysis of temporal data with phase variation

We describe quite extensively a few selected methods developed specifically to analyse temporal data with phase variation. Most often, the data consists of several curves, $S^1, \ldots, S^M$ with common ϕ but different phases γ. Our aim is to show how phase variation is usually circumvented and how it is different for recurrent structures.

One way to compare curves with phase variation is to consider temporal data as a quotient space by the action of reparametrisation. By appropriately defining the considered spaces and the metric, one can push statistical constructions, like that of a mean, on that quotient space. This is the point of view adopted in the popular *Square-root velocity* (SRV) framework (Srivastava et al., 2011). Consider two actions on the space of curves $\mathcal{C} = \{\alpha : \mathbb{S}^1 \to \mathbb{R}^D\}$: that of reparametrisation $(\alpha, \gamma) \mapsto \alpha \circ \gamma$ and that of rotation $(Q, \alpha) \mapsto Q\alpha$, for $\gamma : \mathbb{S}^1 \to \mathbb{S}^1$ a homeomorphism and $Q \in \mathrm{SO}(D)$ a rotation matrix. The authors represent a curve α by two "coordinates": a function $t \mapsto \ln(\|\alpha'(t)\|)$ and a curve $t \mapsto \alpha'(t)/\sqrt{\|\alpha'(t)\|}$, for which it is more natural to describe those actions. The second 'coordinate' gives the framework its name and allows to show that the quotient space has a Riemannian structure. This framework has been extended to trajectories on Riemannian manifolds in Su et al. (2014).

Certain properties, desirable in SRV by design, are not suitable for our setting. Srivastava et al. (2011, Figure 4) shows a collection of helices with varying numbers of spirals. The developed metric allows to cluster these helices based on their winding numbers. It shows that the metric is

[1] https://www.sysnav.fr/

more sensitive to the number of repetitions of a pattern than to the reparametrisation, but it does not allow to explicitly obtain $\gamma(T) - \gamma(0)$.

Another tool for analysis with phase variation are *path signatures*. Introduced in the 1960s and exploited in the rough-path theory (Lyons et al., 2007), signatures are popular in the machine learning community (Fermanian, 2021). The idea is to map a curve $\alpha = (\alpha_1, \dots, \alpha_D) : [0,1] \to \mathbb{R}^D$ (of bounded variation) to a formal sequence of tensors, $\mathcal{S}(\alpha) = (a_0, a_1, \dots, a_n, \dots)$, for $a_n \in (\mathbb{R}^D)^{\otimes n}$, by integrating different coordinates against each other. For example, $a_1 = (\int d\alpha)_{d=1}^D = (\alpha_d(1) - \alpha_d(0))_{d=1}^D$, while $a_2 = (\int_{0<s<t<1} d\alpha_i(s) d\alpha_j(t))_{1\leq i,j\leq D}$. Then, the invariance $\mathcal{S}(\alpha) = \mathcal{S}(\alpha \circ \gamma)$ to reparametrisation $\gamma : [0,1] \to [0,1]$ is a consequence of the change of variables formula. More importantly for our context, the signature satisfies *Chens' identity*: if we denote by $\alpha|_I$ the restriction of α to an interval $I \subset [0,1]$, then $\mathcal{S}(\alpha|_{[0,1]}) = \mathcal{S}(\alpha|_{[0,1/2]}) \otimes \mathcal{S}(\alpha|_{[1/2,1]})$, where $\otimes$ denotes the tensor product. Apart from the numerous theoretical implications, Chens' identity allows to reconstruct piecewise linear curves from their signatures (Fermanian, 2021, Chapter 6). In the context of (1.1), it gives an explicit relation between the signature of $\phi \circ \gamma$ and that of its constituent pieces, ϕ. However, we are not aware of work that would exploit this relation for the analysis of periodic functions.

Kneip and Gasser (1992) adopt a more direct point of view, aligning 'visual features'. Specifically, they define *e-functionals* and *p-functionals* which localize certain local extrema and large variations of the curves, respectively. Considering a collection of, say, $N \in \mathbb{N}$ such functionals, a signal $S : [0,T] \to \mathbb{R}$ is encoded by $0 \leq T_1(S) \leq \ldots \leq T_N(S) \leq T$. This transforms the problem of synchronizing two functions $S_1, S_2 : [0,T] \to \mathbb{R}$ into finding a monotone mapping between their corresponding encodings. In a semi-parametric model and under the assumption that the features are 'uniquely localized', the authors prove consistency of a procedure based on two features. They show that as the sampling (in time) increases, the features are localized increasingly finely, allowing to infer the parameters of a linear reparametrisation, $\gamma(t) = (t - t_0)/t_1$. It is observed that the behavior of the method is unknown if such features were missed. The incidence of potentially repeated features (ubiquitous in periodic functions) is not discussed.

Initially proposed in engineering sciences (Sakoe and Chiba, 1978), Dynamic-Time Warping (DTW) is a dissimilarity measure tailored to comparing time series, samples of curves with phase variation. If we let $S_n = S(t_n)$ for $0 \leq t_1 < \ldots < t_N \leq T$, then the DTW is determined by solving an optimization problem for a reparametrisation,

$$\mathrm{DTW}((S_n^1)_{n=1}^N, (S_m^2)_{m=1}^M) = \min_{i,j} \sum_{k=1}^K |S^1_{i(k)} - S^2_{j(k)}|, \tag{1.2}$$

where the minimum is taken over warping functions of the form $i : \{1, \dots, K\} \to \{1, \dots, N\}$ and $j : \{1, \dots, K\} \to \{1, \dots, M\}$, for varying $K \in \mathbb{N}$, satisfying some constraints. Typically, i, j are required to be non-decreasing and surjective. The explicit nature of the method and its computational simplicity make it very appealing both for comparing time series or phase registration, especially in applications. Changing the cost function (1.2) to interpolate between matching the signals and their derivatives, one can also align on structural characteristics, e- and p-locations from Gasser and Wang (1997).

In signal processing, the problem of estimating the phase, γ, which modulates a periodic function is known as *instantaneous phase estimation* (Boashash, 2015, Chapter 10). Originally studied for periodic functions of a simple form, for example $\phi(x) = a \sin(2\pi(x - b))$, the exact solution is given in terms of the Hilbert transform H, by $\gamma(t) = \arctan(H(S)(t)/S(t))$ (Boashash et al., 1990). Such models were motivated by applications in communications, where the carrier and signal waves have a single sinusoidal component. For more complex signals, one can analyze the time frequency representations, for example with image processing techniques like peak detection and component linking (Rankine et al., 2007). This procedure is even extended to sums of independently-parametrized functions, under hypothesis of separation of different components in the frequency spectrum (Khan and Boashash, 2016, Hussain and Boashash, 2002). In the context of positioning with magnetic signals, it has been the point of view adopted in Bristeau (2012) and Zabulon et al. (2019).

A simple method from signal processing which relies on 'visual' and not spectral characteristics is *zero crossings*. It consists of counting the number of times that the graph of S crosses a certain level $\alpha \in \mathbb{R}$, often zero (Boashash et al., 1990). In noiseless conditions, it is proportional to the number of periods in the signal and the times of the zero-crossings segment the signal. This is a

very straightforward method for which we only need to choose α. As we will see in Chapter 5, it is not stable with respect to perturbations of S by additive noise W.

An interesting question is whether any methods of comparing curves with phase variation can also be used to count or quantify repetition. A zero-crossing is just one specific feature. We do not know of such adaptations, except in the rare cases where the curve is actually periodic, see for example Stefan et al. (2013). We illustrate a difficulty on the e-functionals from Kneip and Gasser (1992). In general, the n-th e-functional detects the position of the n-th smallest local minimum, for $n \in \mathbb{N}^*$. Consider a reparametrized periodic function (1.1) with no noise. The number of local extrema with a given value depends on the number of periods, but their order by value is unstable: a small perturbation can place the global minimum in the first or in the last period of S.

In our work, we use constructions from Topological Data Analysis (TDA) to propose descriptors more specific and stable, while similar in spirit to the visual features like zero-crossings or the e-locations used in analysis of samples of curves. While TDA tools are notorious for high computational complexity, our estimators are computable in time linear in the length of the time series (Glisse, 2023) and at most quadratic in the number of local extrema.

1.2 Studying the shape of data with Topological Data Analysis

Topological Data Analysis (TDA) is a recent field and also an umbrella term used to designate techniques that describe the 'shape of data', often by extracting geometric or topological information from observations (Chazal and Michel, 2021). Such techniques are applicable to a variety of data types: point clouds, graphs (Hofer et al., 2020, Lasalle, 2023), images (Bleile et al., 2021, Garin and Tauzin, 2019) and also time-series (Perea, 2019, Gidea and Katz, 2018, Gidea et al., 2020, Fernández and Mateos, 2022). We recommend Chazal and Michel (2021) for a thorough review.

One of the most popular tools from TDA is *persistent homology*. Homology is a topological invariant, rigorously defining the intuitive notion of holes, in different dimensions: a 0-hole is a connected component, a 1-hole - a loop, a 2-hole - a cavity etc. Persistent homology is associated to a sequence of spaces, allowing the study of the evolution of those holes: when they appear, and when they disappear in that sequence. This also gives the notion of *persistence*: for how long a 'hole'[2] is present. Persistent homology is typically advertised as a multi-scale version of homology for point clouds. Indeed, in a practical setting, estimating the homology of an underlying manifold from a sample passes through level set (or support) estimation. Usually, that involves the choice of a meaningful 'scale' or threshold. Persistent homology allows to circumvent that, by describing a point-cloud by its homology at all scales.

Interestingly, the *persistent homology of sub level sets*[3] of a continuous function $f : [0, T] \to \mathbb{R}$ captures the structure of local extrema of f: their values, but also their order. Very informally, persistent homology in this simple setting can be understood via a 'raising sea level' analogy. If we let f represent a landscape profile as illustrated in Figure 1.3, small reservoirs (connected components) appear in deepest valleys (global minima). As the level raises further, other reservoirs appear in local minima and some merge, at levels of local maxima. When two reservoirs merge, it is the shallower one that ceases to exist. Persistent homology in dimension 0 in this case tracks the evolution of connected components in $(f^{-1}(]-\infty, \alpha]))_{\alpha\in\mathbb{R}}$, here represented by vertical intervals. A different representation of these intervals, the persistence diagram is a (multi-)set of points, where each point with coordinates (y_1, y_2) corresponds to an interval with coordinates (y_1, y_2). The setting of persistent homology is introduced precisely in Section 2.5.

It is clear that persistent homology of sub level sets of a function assigns a notion of 'persistence' to local extrema. Short bars, or equivalently points near the diagonal $\Delta = \{(x, x) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$ represent 'wiggles' of small amplitude, often associated to high frequency noise. The total p-persistence is the sum of p-powers of the lengths of all bars. For persistence diagrams of sublevel sets, whether the total persistence is finite reflects the regularity of the function (Perez, 2023): typically, an α-Hölder function on $[0, T]$ will have finite p-persistence for $p > 1/\alpha$.

[2] Technically, it is the homology generators which are persistent. The geometric 'holes' will inevitably vary with the filtration.

[3] In dimension zero.

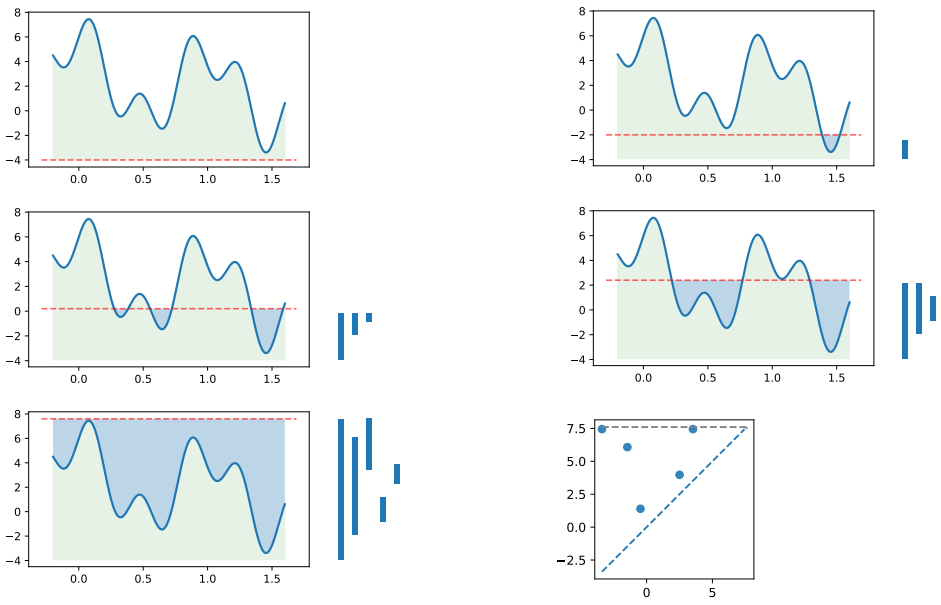


Figure 1.3: A sequence of graphical representations of sub level sets of f, with the level increasing from left to right and top to bottom. Besides each representation is a collection of vertical bars, representing the different connected components.

1.2.1 TDA for temporal data

Persistent homology has been applied in time series analysis in very different ways. A combinatorial structure, usually a weighted simplicial complex, is defined in such a way that its homology is an interesting invariant for a particular problem.

The arguably most popular topological technique for studying temporal data consists of calculating the persistent homology of a Takens Embedding. Takens (1981) showed that an *attractor* of a dynamical system, a specific subspace of the phase-space, can be embedded in $\mathbb{R}^D$ via a simple and explicit map of a trajectory from that system. The fractal properties of the attractor, which determine whether that dynamical system is periodic or chaotic, can be probed with the homology of that embedding. In applications, that embedding is used to extract some information about the phase space, most often its periodicity. To do so, one constructs the *sliding-window embedding* (SWE) $\mathrm{SWE}(S) \in \mathbb{R}^{d\times(N-(d-1)\tau)}$ of the time series $S = (S_n)_{n=1}^N$,

$$\mathrm{SWE}(S)_n = (S_n, S_{n+\tau}, \ldots, S_{n+(d-1)\tau}) \in \mathbb{R}^d, \tag{1.3}$$

for some $\tau, d \in \mathbb{N}$. If S is periodic, the persistent homology (of a complex) on $\mathrm{SWE}(S)$ should contain persistent generators in dimension one. This idea has been used to classify climate systems (Ghil and Sciamarella, 2023), as a volatility indicator in finance (Gidea et al., 2020, Ruiz-Ortiz et al., 2022), chatter detection in an industrial cutting process (Khasawneh et al., 2018), pattern recognition in birdsongs (Fernández et al., 2023) or periodicity detection and quantification (Dłotko et al., 2019). This list is not exhaustive, due to the numerous applications.

Reparametrized periodic functions are not periodic. As it is the case in the industrial context at hand, the length of the periodic structure in $\phi \circ \gamma$ changes according to γ, what in turn breaks the geometry of the embedding (1.3). We illustrate this in Figure 1.4. It can be understood thanks to the relation between the geometry of $\mathrm{SWE}(S)$ and to the Fourier decomposition of S described in Perea (2019), Kim and Jung (2022).

Three topological techniques for the study of 'recurrent' time series have been proposed. In Khasawneh and Munch (2018), authors count the number of transitions in a piecewise constant, binary signal. The noise model meant that several spurious transitions could occur near a true transition and a persistent homology based technique was used to discard those. The second method proposed in Erden and Cetin (2017) consists of analysing the graph $(\frac{nT}{N}, S(\frac{nT}{N}))$ with persistent homology, obtaining a more robust variant of zero-crossings. In Bois et al. (2022), authors use the construction from Figure 1.3 to analyze the gait of patients. They recognize that the multiplicity

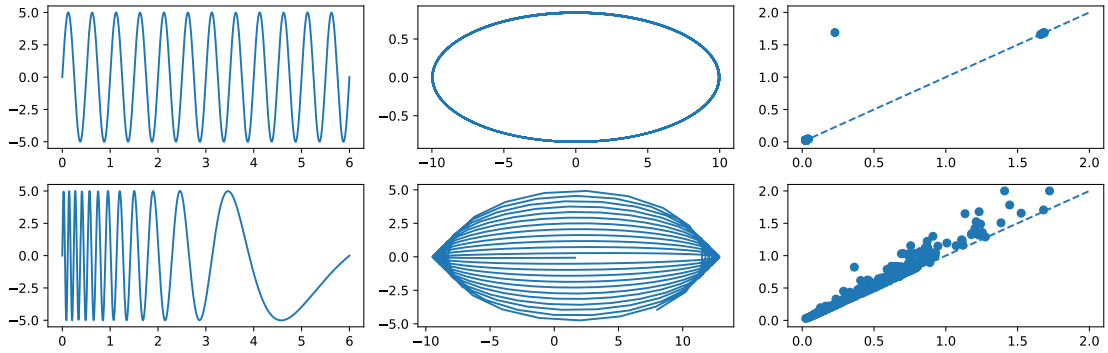


Figure 1.4: On the top row, from left to right, a periodic function, a projection of the $\mathrm{SWE}_{4,1}$ on the first two principal components, and the persistence diagram in dimension 1 of the Vietoris-Rips complex of the embedding. The top row shows that periodic temporal data has a circular embedding (for a suitable choice of parameters), what is reflected in the persistence diagram: there is one, persistent generator in homology of dimension one. The bottom row shows the analogue, for a reparametrisation of the signal in the top row. Due to the homeomorphism, the (projection of the) embedding looks different and this is also reflected in the diagram, which is not that of a closed curve.

of certain bars reflect the number of steps in the recording. We will formalize and exploit this observation.

1.2.2 TDA of random data and Statistics

While homology and its persistent version might contain relevant information about a geometric object of interest, in practical scenarios, that object is only observed through a random realisation. In our case, for example, we can say that we 'observe ϕ only through the temporal data S', with random γ and W. Therefore, it is important to understand how persistent homology behaves for random data, potentially corrupted or incomplete.

A popular TDA setting is data sampled from a certain distribution on a manifold. As mentioned before, one way to infer the homology of that manifold is to build a complex on the sample point cloud, at a right scale. With such a construction, one estimates the persistence diagram consistently (Chazal et al., 2015). In practice, data is often corrupted, so situations with additive noise (Aamari et al., 2023) or outliers/out of distribution samples (Vishwanath et al., 2022) are also studied.

Persistent homology is also often used as a descriptor of random, structured data, even when the statistical model does not introduce geometry or topology explicitly. In such a scenario, one is rather interested in capturing the distribution of geometric features at different scales. An example from statistical learning is classification of digits from Garin and Tauzin (2019), where authors use topological features extracted from images to classify handwritten digits.

We distinguish two statistical settings. In the first, we observe an i.i.d collection of such structures, say images. We describe this empirical distribution by a distribution of persistent diagrams, or features thereof. Common in statistical learning, this setting has fuelled the development of feature-extraction methods from persistence diagrams. Indeed, multi-sets are not suited for statistical learning, so several ways of representing them as vectors (Bubenik, 2015, Zielinski et al., 2021), images (Adams et al., 2017) or elements of a Banach space (Reininghaus et al., 2014) have been proposed. Such constructions converge asymptotically: as the sample size grows, the empirical mean representation of a persistence diagram converges to the its expectation and the distribution of the empirical mean can be approximated by bootstrap techniques (Chazal et al., 2014, Berry et al., 2020). These results follow from the quite general machinery of empirical process theory (Kosorok, 2008) and rely on the asymptotics of many, independent realizations of persistence diagrams.

The limit of these processes is sometimes characterized in a second setting, which consists of the observation of a structure that grows in size or is sampled increasingly-finely. A typical example is a large point-cloud, and the convergence of the topological representation is studied as the size of that point cloud increases. Establishing strong statistical guarantees is then considerably

harder. In this case, the expected number of topological features and its estimation have been characterized (Kahle, 2011, Divol and Polonik, 2019). Recently, Roycraft et al. (2023) construct confidence sets on those descriptors with bootstrap techniques. Overall, characterising a single diagram requires understanding what is the 'distance' at which interactions take place and it is challenging given the global nature of persistent homology. We are aware of several results on temporal data, albeit only in expectation. First, Krebs (2021) shows that the persistent homology of SWE of a Markov chain converges, under some assumptions on the Markov kernel. It is the only work we are aware of that considers estimation guarantees for topological descriptors on time series. For random processes, Bobrowski and Borman (2012) gives an explicit characterisation of persistent homology for Gaussian processes, Divol and Chazal (2019) for Brownian motion and Perez (2022b) for more general Markov processes.

1.3 Contributions

In this thesis, we show how TDA can be used to analyse periodic functions with phase variation (1.1), and in particular, we study Problems 1.1 and 1.2. We introduce tools to compare signals with different number of periods through descriptors thereof and to count the number of periods in an unknown signal. We apply this last to the car positioning problem mentioned above.

1. Additivity of the persistence diagram for periodic functions (Chap. 3)
We show that persistent homology of sub level sets of a signal (1.1) captures its recurrent nature. Specifically, let $\phi : \mathbb{R} \to \mathbb{R}$ be a one-periodic, continuous function. We denote by $D(\phi|_I)$ the persistence diagram of $\phi|_I$, the restriction of ϕ to an interval $I \subset \mathbb{R}$. Then, the persistence diagram is additive in the following sense:

1. if $R \in \mathbb{N}$, then
$$D(\bar{\phi}\big|_{[0,R]}) = RD(\bar{\phi}\big|_{[0,1]}), \tag{1.4}$$
where $\bar{\phi}$ denotes a version of ϕ, adapted to periodic boundary conditions.

2. For any $R > 1$,
$$D(\phi|_{[0,R]}) = \lfloor R - 1 \rfloor D^* + D', \tag{1.5}$$
where $D^* = D(\bar{\phi}\big|_{[0,1]})$ and D' are two persistence diagrams. We also show that D' has total persistence smaller than D^*, so it can be treated as a remainder term.

Additive relations between $D(\phi|_{[0,R]})$ and $D(\phi|_{[0,1]})$ are not surprising for readers familiar with persistent homology. However, we formalize them with particular care for the boundary conditions: even if $R \in \mathbb{N}$, we have in general $D(\phi|_{[0,R]}) \neq RD^*$.

The above are the building blocks of our two main contributions: one, which relies on the fact that the number of periods appears explicitly in (1.4) and the other one, which characterizes ϕ through D^* from (1.5).

2. Segmenting a periodic signal with phase variation (Chap. 5)
When ϕ is a 'simple' function, zero-crossings of $\phi \circ \gamma$ segment $[0, T]$ into periods of ϕ. However, the so-obtained number of periods, or the segmentation are not stable with respect to perturbations of S. We propose a solution to Problem 1.1 with stability guarantees, valid when the number of observed periods $N \coloneqq \gamma(T) - \gamma(0) \in \mathbb{N}$ is an integer. We proceed in two steps.

First, we propose an estimator of the number of periods, $\hat{N}(S, \tau)$, which counts the multiplicity of points in the persistence diagram, relying crucially on (1.4). The estimator depends on a scale parameter $\tau > 0$ that defines a threshold for points to be considered close. We show that if the scale is chosen appropriately and the noise W is small enough, the estimation is correct $\hat{N}(S, \tau) = N$. We also propose two adaptations of $\hat{N}$, which prove more robust in our simulation studies than $\hat{N}$, but for which any guarantee remain to be established.

In the second step, we construct a segmentation of $[0, T]$ into periods of ϕ. If $\hat{N}(S, \tau) = N$ is correct, by construction of $\hat{N}$, the persistence diagram $D(\bar{\phi} \circ \gamma)$ contains NK persistent points, for some $K \in \mathbb{N}$. We can retrieve all local minima corresponding to those points, and by taking every K-th minimum, we obtain the desired segmentation.

Our method generalizes the zero-crossings estimator. Specifically, the number of zero-crossings can be expressed as the number of points in a certain region in the persistence diagram.

3. Magnetic odometry (Chap. 7)
We develop a magnetic odometry method based on our segmentation algorithm. In simple scenarios, it counts revolutions of a wheel more reliably than an estimator based on zero-crossings, especially for non-trivial magnetic fingerprints and it localizes periods more precisely. We study its limitations for more complex trajectories and environments. There, the repetitive structure in the signal is obfuscated by the high- or low-frequency noise, but can sometimes still be detected correctly: by projecting the magnetic signal along a suitable direction or analysing the gradient of the magnetic field.

4. Topological signature of a reparametrized periodic function (Chap. 4)
Several metrics or notions of discrepancy to compare signals up to phase variation have been proposed. However, they are not adapted to comparing two observations of reparametrized periodic functions with different numbers of periods, as they do not exploit the recurrent structure of the signal.

We construct a signature $F(S)$ of S, which is somehow characteristic of ϕ. It is defined by averaging elements of persistence diagrams, in a functional space. Thanks to (1.5), we show that the signature converges as $R \to \infty$ when $W = 0$. Also, when the reparametrisations γ_1, γ_2 are random with $\gamma_1(0) = \gamma_2(0)$ and $\gamma_1(T) = \gamma_2(T)$, a distance between $F(S_1)$ and $F(S_2)$ is controlled by the distance between the distributions of γ_1 and γ_2. Equality of the endpoints of γ_1 and γ_2 implies that the number of periods of ϕ is the same in both signals. We explain what is missing for the general setting of a different number of periods, and we show an upper-bound on the distance, which includes a bias term between the signatures.

The signature of the whole process is not computable from a single observation $S = (S_n)_{n=1}^N$. Instead, we propose to compute a signature of a shorter process $X = (S_m)_{m=1}^M$ of length $M < N$, which still accurately describes the structure of ϕ. We show that $F(X)$ can be estimated from a collection of windows $(X_k)_{k=1}^{N-M+1}$, where $X_k = (S_{k+m})_{m=0}^{M-1}$, provided that the dependence structure in the signal is not too long. Part of the dependence is hidden in the reparametrisation process. We consider a Markov chain model for the raparametrisation γ, motivated by the industrial application where γ represents the displacement of a vehicle in time. We show that the dependence decreases exponentially fast and this justifies a bootstrap procedure.

5. Normalized functionals and truncated total p-persistence (Chap. 4)
Considered most often in the literature are functionals 'linear' in the persistence diagram, and their continuity has been fully characterized. It depends on the total p-persistence of the persistence diagram, which is not necessarily continuous, nor even finite: for a diagram of sub level sets of a function, these properties depend on the dimension of the domain and the regularity of that function.

We introduce normalized functionals of persistence diagrams and a truncated version of total persistence. In Section 4.3, we show that it is lower- and upper-bounded, and, more importantly, enjoys better continuity properties than the non-truncated counter-part. With this notion of truncated persistence, the normalized functionals are continuous. Overall, this contribution is crucial in establishing the stability of the signatures.

1.4 Structure

An overview of the dependence between chapters and their content in terms of contributions is provided in Figure 1.5. Chapter 2 introduces the technical tools used throughout this thesis and does not contain our contributions. The aim of Sections 2.1 to 2.4 is to introduce measurability and convergence results from stochastic processes and time series. The presentation is rather dense and oriented to an expert reader. Section 2.5 retraces the construction of homology, persistent homology and its resulting properties. While not fully accessible to a non-expert audience, our goal is to make precise the intuition conveyed with Figure 1.3 and how we use persistent homology. It also provides technical details necessary for Chapter 3.

In Chapter 3, we show the additive properties of persistence diagrams for periodic functions. We state (1.4) and (1.5) in more precise terms and give the corresponding proofs. While the technical discussion is algebraic, we illustrate the key parts. In Chapter 4, we present our contributions on signatures. The first part of the chapter is devoted to stability and statistical properties (contribution 4). The second part, Section 4.3, presents contribution 5. In Chapter 5, we present the estimators and segmentation algorithm mentioned in contribution 2. In Chapter 6, we consider

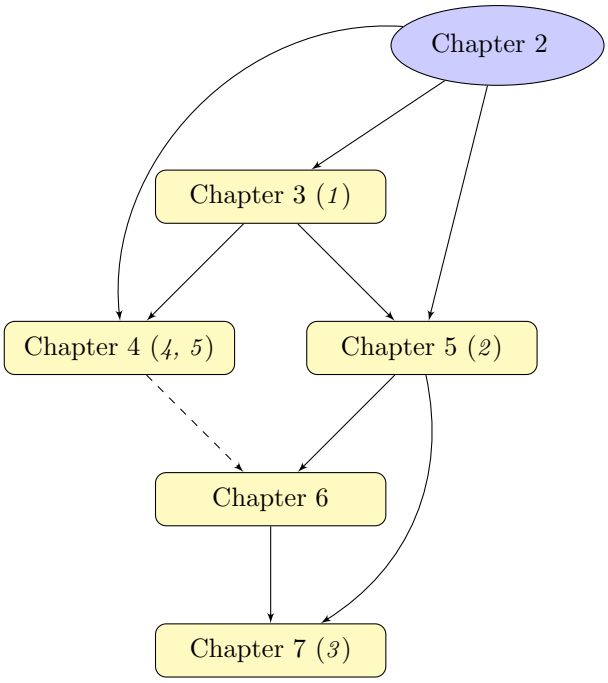


Figure 1.5: A schematic illustration of the structure of this thesis. The numbers in parentheses refer to the contribution. The arrows outline the dependence structure between the chapters: the solid arrow between Chapters 2 and 3 indicates that Chapter 3 crucially relies on the results or methods introduced in Chapter 2.

how to apply those estimators to three-dimensional signals - a question stemming from the application. Chapter 7 presents the background on positioning and motivates the need for another position estimator. Crucially, it contains the application of the methods developed in Chapters 5 and 6.

Chapter 2

Background

2.1 Measure theory and stochastic processes

We call a stochastic process $S = (S_t)_{t\in I} : (\Omega, \mathcal{A}) \to (\mathbb{R}^I, \sigma(\mathbb{R}^I)))$ a random variable on a measure space $(\Omega, \mathcal{A}, \eta)$, for $I \subset \mathbb{R}$ an interval or a subset of $\mathbb{Z}$. A major appeal of stochastic processes is that they model temporal data, but can be also thought of as random functions. To justify this point of view, we explicit the notions of measurability.

We denote by $\sigma(\mathbb{R}^I)$ the σ-algebra generated by the *product topology on* $\mathbb{R}^I$. Specifically, it is the topology which makes the projections $S \mapsto S_t$ measurable, for all $t \in I$. In this manuscript, we are mostly interested in the case where S is $C(I)$-valued, that is, for all $\omega \in \Omega$, $t \mapsto S_t(\omega)$ is a continuous function. For this reason, we will think of S as a 'function-valued' random variable, what we justify that in this section.

The space of continuous functions $(C(I), \|\cdot\|_\infty)$, equipped with the supremum norm is a Banach space when I is a compact interval. The Borel σ-algebra $\mathcal{B}(\|\cdot\|_\infty)$, is the smallest σ-algebra containing all the open sets for the topology generated by that norm. The product topology on $\mathbb{R}^I$ is coarser than that of $\|\cdot\|_\infty$, and therefore, $\sigma(\mathbb{R}^I) \subset \mathcal{B}(\|\cdot\|_\infty)$. While S is a $(\mathbb{R}^I, \sigma(\mathbb{R}^I)$-valued random variable, it is not clear that it is measurable with respect to $\mathcal{B}(\|\cdot\|_\infty)$. Theorem 2.1 tells us that it is, justifying considering S a 'function-valued' random variable.

Theorem 2.1. *Assume that $I \subset \mathbb{R}$ is a compact interval, $t \mapsto S_t(\omega)$ is continuous for all $\omega \in \Omega$ and $S_t : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a random variable for all $t \in I$. Then, $S = (S_t)_{t\in I}$ is $\mathcal{B}(\|\cdot\|_\infty)$-measurable.*

Proof (taken from Steinwart (2022)). By definition, S is a *Carathéodory function*, that is, $S.(\omega) : I \to \mathbb{R}$ is continuous, and $S_t : \Omega \to \mathbb{R}$ is measurable. If S is a Carathéodory function, then $S : (\Omega \times I, \mathcal{A} \otimes \mathcal{B}(\|\cdot\|_\infty)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable (Aliprantis and Border, 2006, Lemma 4.51). By Steinwart (2022, Lemma 5.1.5), S is *weakly-measurable*. That is, for any $x' \in C(I)'$, the dual of $C(I)$, $x'S : \Omega \to \mathbb{R}$ is measurable. We conclude the proof by applying Pettis' measurability theorem.

Lemma 2.2 (Pettis' measurability theorem)**.** *Consider $h : \Omega \to E$, where (E, d_E) is a Banach space. If E is separable as a metric space and h is weakly-measurable, then h is measurable with respect to the Borel σ-algebra induced by d_E.*

Because I is compact, $(C(I), \|\cdot\|_\infty)$ is complete and separable. We can therefore apply Pettis' theorem to S weakly-measurable. □

Gaussian processes are an important example of stochastic processes. We say that S is a *Gaussian process* if $(S_t)_{t\in K}$ is a Gaussian random variable, for every finite subset $K \subset I$. A centered Gaussian process is entirely characterized by its covariance functions, $(s, t) \mapsto \mathbb{E}[S_s S_t]$ (Steinwart, 2022, Theorem 4.4.5). Indeed, for a Gaussian process S and $x' \in C(I)'$, $x'S$ is also Gaussian and it is determined by its mean and variance.

Definition 2.3. We say that a sequence of stochastic processes $(S_n)_{n\in\mathbb{N}}$ *converges weakly* to a stochastic process S if for all $x' \in C(I)'$, $x'S_n \to x'S$ in $\mathbb{R}$.

If the distribution of $x'S^1$ and $x'S^2$ is the same for all x', then S^1 and S^2 are the same as weakly-measurable variables. By Kosorok (2008, Lemma 7.1), two Borel-measurable processes equal as weakly-measurable processes, induce the same measures on the measure-space, $(C(I), \mathcal{B}(|\cdot|_\infty))$.

As we will see in Section 2.5.3, the descriptors that we will consider are defined for continuous functions. When evaluating the descriptors on random function, to show their continuity, boundedness or discuss their stability, we might need to control two key quantities of the stochastic process: its supremum and its regularity.

The supremum $M_I \coloneqq \sup_{t\in I} S_t$ is a random variable. For a Gaussian process, we can bound the deviations of M_I using its covariance function. We will use the following result in Chapter 5.

Proposition 2.4 (Azäis and Wschebor (2009, Proposition 4.1))**.** *Consider a centred Gaussian process S with covariance* $\mathrm{Cov}(s,t)$ *such that* $\mathrm{Cov}(t,t) = 1$*, for all $t \in I$. Then,*

$$P(M_I > u) \leq \frac{\exp(-u^2/2)}{2\pi} \int_I \sqrt{\frac{\partial^2 \mathrm{Cov}}{\partial s \partial t}(r,r)}dr + 1 - \phi(u),$$

where ϕ is the cumulative distribution function of a standard normal random variable.

For topological descriptors to be continuous or bounded, it is necessary for the functions to be more regular than just continuous. We recommend Azäis and Wschebor (2009, Sections 4.2 and 4.3), which contains results on the regularity of sample paths. As an example, for Gaussian processes, it is enough for the mean to be C^k and the covariance to be C^{2k} for the paths to be (almost-surely) C^k. Our work is not restricted to Gaussian processes and we impose weaker regularity conditions, typically Lipschitz or Hölder regularity.

Definition 2.5. We say that S is *α-Hölder* if there is $\Lambda : \Omega \to \mathbb{R}$, such that

$$|S_s(\omega) - S_t(\omega)| \leq \Lambda(\omega)|s-t|^\alpha, \qquad \text{for all } s,\, t \in I,$$

and we say that S is *Lipschitz* if it is 1-Hölder.

Notice that we do not require Λ to be uniformly bounded. In fact, it is not immediately clear that it is a random variable. For it to be the case, it suffices for the increments to be bounded. Specifically, we assume that for some $0 < r_1 < r_2$, there exists $K = K_{r_2,r_1}$ such that for all $s, t \in I$,

$$\mathbb{E}[|S_t - S_s|^{r_2}] \leq K_{r_2,r_1}|t-s|^{1+r_1}. \tag{2.1}$$

The following proposition is known as the *Kolmogorov continuity theorem.*

Proposition 2.6 (Azäis and Wschebor (2009, Proposition 1.11))**.** *Suppose S satisfies* (2.1) *with K_{r_2,r_1} and let $\alpha \in]0, \frac{r_1}{r_2}[$. Then, there exists a modification $(X_t)_{t\in[0,1]}$ of S and a random variable $\Lambda_{X,\alpha} : \Omega \to \mathbb{R}_+^*$, such that for all $s, t \in [0,1]$,*

$$P(|X_t - X_s| \leq \Lambda_{X,\alpha}|t-s|^\alpha) = 1.$$

Recall that a process X is said to be a *modification* of S, if $P(S_t = X_t) = 1$, for all $t \in I$. A modification therefore prescribes the evaluations, but does not directly imply equality in terms of paths. We say that X and S are *indistinguishable*, if $P(\{\omega \mid S_t(\omega) = X_t(\omega), \forall t \in I\}) = 1$. If S and X are modifications of each other and are continuous (almost everywhere), then they are indistinguishable, so there is a unique 'modification' of S which has continuous paths. In Section 4.H, we will show that we can upper-bound the moments of $\Lambda_{X,\alpha}$.

We show an example of an explicit upper-bound of the constant in (2.1) for processes other than Gaussian. Instead, we require the increments of the process to be sub-Gaussian, with a controlled constant.

Proposition 2.7. *Assume that the increments of S are sub-Gaussian,*

$$P(|S_t - S_s| > u) \leq 2\exp\left(-\tfrac{u^2}{2d^2(s,t)}\right). \tag{2.2}$$

Then, for any $0 < r_1$ and $r_1 + 1 < r_2$, we have

$$K_{r_2,r_1} \leq C_{r_2}\mathrm{diam}(I)^{r_2-1-r_1}.$$

Proof. Recall that for a positive random variable X with cumulative distribution function F_X, we have $\mathbb{E}[X^k] = \int_0^\infty u^{k-1}(1-F_X(u))du$. For $X = |S_s - S_t|$, we use that F_X is lower-bounded by (2.2) to obtain that

$$\begin{aligned}\mathbb{E}[|S_t - S_s|^{r_2}] &\leq \int_0^\infty u^{r_2-1} P(|S_t - S_s| > u)du \\ &\leq 2\int_0^\infty u^{r_2-1} \exp\left(-\tfrac{u^2}{2d^2(s,t)}\right) du \\ &= 2\tfrac{d(s,t)\sqrt{\pi}}{\sqrt{2}} \int_0^\infty u^{r_2-1} \sqrt{\tfrac{2}{d(s,t)\pi}} \exp\left(-\tfrac{u^2}{2d^2(s,t)}\right) du.\end{aligned}$$

We recognize that the integral is the (r_2-1)-th moment of a half-normal random variable with variance $d(s,t)^2$, which is $C_{r_2} d(s,t)^{r_2-1}$, where

$$C_{r_2} = \begin{cases} \frac{r_2!}{(\frac{r_2}{2})!} 2^{-r_2/2}, & \text{if } r_2 \text{ is even} \\ 2^{(r_2-1)/2} \frac{r_2-1}{2}! \sqrt{\frac{2}{\pi}}, & \text{if } r_2 \text{ is odd.} \end{cases}$$

Hence,

$$\mathbb{E}[|S_t - S_s|^{r_2}] \leq \sqrt{2\pi} \cdot C_{r_2} d(s,t)^{r_2}.$$

□

2.2 Markov chains

Time series, a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ on a common measureable space, are a stochastic process indexed by a discrete set. Time series could be thought of as a discretization of a random function, but this does not reflect the more fine structure of the σ-algebra permitted by a countable index set, nor the breadth of applications. A common structure is that of a system or phenomenon represented using a state and a law governing transitions to the next state. In a *Markov chain*, the transitions depend only on the previous state and such models have found numerous applications (storage models, population growth, exchange rates). The theory of Markov chains is most widely known for finite or countable state-space $\mathcal{X}$. However, the intuitions and developments in the countable setting have been adapted to general state spaces, where $\mathcal{X}$ is only assumed measurable and sometimes topological. In this section, we introduce elements from the general state space theory, because, in this thesis, we use Markov chains on real intervals to model the displacement of a vehicle over time. We rely on Meyn and Tweedie (1993).

Definition 2.8. A *Markov chain* is a time series $(X_n)_{n\in\mathbb{N}}$ which satisfies the *Markov property*,

$$P(X_n \in A \mid X_j,\ j \leq n-1) = P(X_n \in A \mid X_{n-1}).$$

In other words, the future value of the process is independent on the past, conditioned on the present value. We will be only interested in *homogeneous* processes, where these probabilities do not depend on the time index n. A homogeneous Markov chain on a space $(\mathcal{X}, \sigma(\mathcal{X}))$ is specified by an initial distribution and, thanks to the Markov property, the transition probabilities.

Definition 2.9. The *initial distribution* is simply a probability measure on $\mathcal{X}$, μ. The *transition probability kernel*, is a map $P : \mathcal{X} \times \sigma(\mathcal{X}) \to \mathbb{R}^+$, such that

1. for each $A \in \sigma(\mathcal{X})$, $P(\cdot, A)$ is $\sigma(\mathcal{X})$-measurable,
2. for each $x \in \mathcal{X}$, $P(x, \cdot)$ is a probability measure on $(\mathcal{X}, \sigma(\mathcal{X}))$.

The n-transition kernel is defined recursively as a product: for $n = 0$, $P^0(x, A) = 1_x(A)$ and for $n \geq 1$,

$$P^n(x, A) = \int_{\mathcal{X}} P(x, dy) P^{n-1}(y, A). \tag{2.3}$$

The transition probability kernel can be thought of as the probability of $X_1 \in A$, conditional on $X_0 = x$. With the initial distribution and the transitions, we can characterize the distribution of the Markov chain, $\mathbb{P}(X_n \in A) = \int_{\mathcal{X}} \mu(dx) P^n(x, A)$.

Because Markov chains are used to model an extremely wide range of phenomena, they naturally exhibit different behaviors. We introduce three different properties: stationarity, irreducibility and recurrence, each corresponding to a different aspect of the behavior of a Markov chain.

Recall that a time series is called *stationary* if, for any $k \in \mathbb{N}$, $(X_n, \dots X_{n+k})$ has the same distribution for all $n \in \mathbb{N}$. The distribution of X_n is μP^n, so the time series is stationary if and only if μP^n does not change with n. An *invariant measure* $\pi : \mathcal{A} \to \mathbb{R}$ is a σ-finite measure such that $\pi(A) = \int_{\mathcal{X}} \pi(dx) P(x, A)$. An important question is thus whether the transition probability kernel admits an invariant measure.

Theorem 2.11 relates the concept of invariance to that of recurrence, but before we state it, we need to introduce irreducible chains. The chain is said to be *ρ-irreducible*, if for any $x \in \mathcal{X}$ and $A \in \sigma(\mathcal{X})$ with $\rho(A) > 0$, the probability of $(X_n)_{n\in\mathbb{N}}$ ever falling in A, having started from x, is positive, $\mathbb{P}_x(\exists n, X_n \in A) > 0$. Among ρ-irreducible measures, Meyn and Tweedie (1993, Proposition 4.2.2) allows to choose the maximal one, ψ, for which (X_n) is said to be *ψ-irreducible*.

Definition 2.10. We say that (X_n) is *recurrent*, if it is ψ-irreducible, and, for every $x \in \mathcal{X}$ and $A \in \sigma(\mathcal{X})$ such that $\psi(A) > 0$, we have $\sum_{n=1}^{\infty} P^n(x, A) = \infty$.

On finite state spaces, a chain is irreducible if we cannot partition the states in such a way that the chain does not transition between different sets of the partition. It is recurrent if a typical realization passes infinitely often by any particular state. On general spaces, the notion of a 'state' is replaced by that of a set of positive ψ-measure: the chain is recurrent if, from every initial condition, the expected number of visits to any set A of positive ψ-measure is unbounded. Example 2.12 shows a simple, recurrent process.

Theorem 2.11 (Meyn and Tweedie (1993, Theorem 10.0.1))**.** *A recurrent chain $(X_n)_{n\in\mathbb{N}}$ admits an invariant measure, π.*

Note that a ψ-irreducible chain is either recurrent or transient (Meyn and Tweedie, 1993, Theorem 8.3.4). A chain is *transient* if there is a finite cover, $\mathcal{X}_1, \dots \mathcal{X}_D$ of $\mathcal{X}$, such that for all $d = 1, \dots, D$, there is $0 < M_d < \infty$ such that the expected number of returns to $\mathcal{X}_d$ is finite, ie $\sum_{n\geq 1} \mathbb{P}^n(x, \mathcal{X}_d) < \infty$, for every $x \in \mathcal{X}_d$. An standard example of a transient Markov chain is a random walk on $\mathbb{Z}$, with biased transitions.

Example 2.12 (Random walk on $\mathbb{R}^+$)**.** Consider a chain defined inductively by $X_n = \max(X_{n-1} + W_n, 0)$, where $W_1, \dots, W_n$ are all iid distributed as W. The chain X is called a *random walk on* $\mathbb{R}^+$ and it is recurrent if $\mathbb{E}[W] < 0$ or if both $\mathbb{E}[W] = 0$ and $\mathbb{E}[W^2] < \infty$ (Meyn and Tweedie, 1993, Propositions 8.5.1 and 8.5.4).

For a stationary process, convergence to an invariant measure is implied by the *Doeblin condition*.

Definition 2.13. A stationary process X with TPK P satisfies the *Doeblin condition* if there exists $r \in \mathbb{N}^*$ and a measure ν (non-trivial, $\nu(\mathcal{X}) > 0$) such that

$$P^r(x, A) \geq \nu(A), \qquad \text{for any } x \in \mathcal{X}, \text{ and } A \in \sigma(\mathcal{X}). \tag{2.4}$$

A weaker version of (2.4) is a consequence of ϕ-irreducibility. By Meyn and Tweedie (1993, Theorem 5.22), any ϕ-irreducible chain admits a *small set*: a set C, for which (2.4) is satisfied for $x \in C$. The Doeblin condition is thus the small-set condition for $C = \mathcal{X}$.

2.3 Measures of dependence

We can now model complex phenomena or systems, by introducing dependence between random variables. In this section, we present ways to measure that dependence. These notions allow to generalize the estimation guarantees from Section 2.4 from independent, to depend data. For further reading, we recommend Doukhan (1995), Dedecker et al. (2007) and the review Bradley (2005).

Consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and recall that two σ-algebras $\mathcal{A}_1$, $\mathcal{A}_2 \subset \mathcal{A}$ are independent if

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2), \qquad \text{for all } A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2. \tag{2.5}$$

Definition 2.14. Denoted by $\sigma(X)$, the σ-algebra generated by a random variable $X : (\Omega, \mathcal{A}) \to (\mathbb{X}, \mathcal{B})$ is the smallest one which contains $\{X^{-1}(B) \mid B \in \mathcal{B}\}$. Two random variables X_1, X_2 defined on a common probability space are independent, if $\sigma(X_1)$ and $\sigma(X_2)$ are independent.

For $(X_n)_{n\in\mathbb{Z}}$ a sequence of random variables on $(\Omega, \mathcal{A})$, we denote by $\sigma^X_{a,b} = \sigma(X_a, \dots, X_b)$ the σ-algebra generated by $X_a, \dots X_b$, for some $a \leq b \in \mathbb{Z}$. For simplicity, we will focus on stationary sequences $(X_n)_{n\in\mathbb{Z}}$.

Definition 2.15. A time series is *m-dependent* if there is $m \in \mathbb{N}$, such that $\sigma^X_{-\infty,k}$ and $\sigma^X_{k+m,\infty}$ are independent, for all $k \in \mathbb{Z}$.

Common examples are processes of the form $X_n = f(Z_n, \dots, Z_{n+m})$, for $(Z_n)_{n\in\mathbb{Z}}$ independent and are often obtained through transformations, like renormalization or centering, of an otherwise independent process. However, m-dependence is a very strong notion. Mixing coefficients quantifying how far certain equations, usually satisfied by independent σ-algebras, are from being satisfied. We will work with β-mixing coefficients which quantify the deficit in (2.5). In Chapter 4, we will also be interested in how fast these coefficients decay.

Definition 2.16 (Dedecker et al. (2007, Section 1.2))**.** The k-th *β-mixing coefficient* of $(X_n)_{n\in\mathbb{N}}$ is defined as

$$\beta_X(k) \coloneqq \sup_{\mathcal{A}_1,\mathcal{A}_2} \sum_{A_1\in\mathcal{A}_1, A_2\in\mathcal{A}_2} |\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)|, \tag{2.6}$$

where the supremum is taken over all finite partitions $\mathcal{A}_1 \subset \sigma^X_{-\infty,0}$ and $\mathcal{A}_2 \subset \sigma^X_{k,\infty}$ of Ω. We will say that the process is *β-mixing* or *absolutely regular* if $\lim_{k\to\infty} \beta_X(k) = 0$.

Trivially, an m-dependent sequence is absolutely regular, as $\beta(k) = 0$ for $k \geq m + 1$. Let us compare the two notions for stationary Gaussian sequences. Then, m-dependence is equivalent to ϕ-mixing, where the ϕ-coefficients are analogue to $\beta_X(k)$, except for a normalization, $|\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)\mathbb{P}(B)} - 1|$. Recall that the spectral density, f, of a covariance function is $f(\lambda) = \sum_{t\in\mathbb{Z}} \mathrm{Cov}(0, t)e^{i\lambda t}$, for $\lambda \in [0, 2\pi]$. A sequence of stationary, Gaussian random variables is absolutely regular if and only if the spectral density of the covariance, can be written as $f(\lambda) = |p(e^{i\lambda})|^2 \exp(\sum_{j\in\mathbb{Z}} a_j e^{i\lambda j})$, with $\sum_{j\in\mathbb{Z}} |j||a_j|^2 < \infty$ and the roots of the polynomial p all lie on the unit circle (Bradley, 2005, Theorem 7.1). As mentioned in the remark after Doukhan (1995, Theorem 3), $\beta_X(n) = \mathcal{O}(n^{2-\alpha})$ if $\mathrm{Cov}(0, t) = \mathcal{O}(t^{-\alpha})$.

Another example of absolutely regular processes are Markov chains, which satisfy the Doeblin condition. We will use this characterization in Chapter 4.

Theorem 2.17 (Doukhan (1995, Section 2.4, Theorem 1))**.** *If a stationary Markov chain X satisfies* (2.4)*, then X is absolutely regular and β_X decays exponentially fast.*

A big benefit of working with mixing coefficients is that they are defined through σ-algebras. This means that any point-wise mapping of the chain cannot increase the dependence between variables. We will call such propeties *mixing-preservation* and, as we will see, it will be particularly important in Chapter 4, where $h = \phi$ is a continuous, periodic function.

Proposition 2.18. *For any $h : \mathcal{X} \to \mathcal{Y}$ measurable and $h(X) \coloneqq (h(X_n))_{n\in\mathbb{N}}$, we have $\beta_{h(X)}(k) \leq \beta_X(k)$*

Proof. By definition, $\mathbb{P}(h(X) \in A) = \mathbb{P}(X \in h^{-1}(A))$. Any partition $\mathcal{A}$ of $\sigma(\mathcal{Y})$ gives a partition of $h_*\mathcal{A} \coloneqq \{h^{-1}(A) \mid A \in \mathcal{A}\} \subset \sigma(\mathcal{X})$. So,

$$\begin{aligned}
\beta_{h(X)}(k) &= \sup_{\substack{\mathcal{A}_1,\mathcal{A}_2 \\ \text{partitions of } \sigma(\mathcal{Y})}} \sum_{A_1\in\mathcal{A}_1, A_2\in\mathcal{A}_2} |\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)| \\
&= \sup_{\substack{\mathcal{A}_1,\mathcal{A}_2 \\ \text{partitions of } \sigma(\mathcal{Y})}} \sum_{B_1\in h_*\mathcal{A}_1, B_2\in h_*\mathcal{A}_2} |\mathbb{P}(B_1 \cap B_2) - \mathbb{P}(B_1)\mathbb{P}(B_2)| \\
&\leq \sup_{\substack{\mathcal{A}_1,\mathcal{A}_2 \\ \text{partitions of } \sigma(\mathcal{X})}} \sum_{A_1\in\mathcal{A}_1, A_2\in\mathcal{A}_2} |\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)| \\
&= \beta_X(k).
\end{aligned}$$

□

Another way to quantify dependence is via *weak-dependence* coefficients. In the same way as mixing is related to a lack of independence (2.5), these coefficients quantify a deficit in *orthogonality* between the random variables. Even though we do not use this framework further, we give a general outline, which we hope motivates working with mixing coefficients instead. We refer to Dedecker et al. (2007) for a thorough presentation of the theory and examples. see Bobbia et al. (2022) for a recent review. Consider $\mathcal{F}$, $\mathcal{G}$ two families of functions and c which maps two functions to a positive scalar. A sequence X is $(c, \mathcal{F}, \mathcal{G}, (\epsilon_k)_{k\in\mathbb{N}})$-weakly-dependent, if for any $k \in \mathbb{N}$, d_f, $d_g \in \mathbb{N}_+$, $f : \mathbb{R}^{d_f} \to \mathbb{R}$, $g : \mathbb{R}^{d_g} \to \mathbb{R}$ and $i_1, \leq \ldots \leq i_{d_f} \leq j_1 - k \leq \ldots \leq j_{d_g}$, we have

$$|Cov(f(X_{i_1}, \ldots X_{i_{d_f}}), g(X_{j_1}, \ldots X_{j_{d_g}}))| \leq c(f, g)\epsilon_k, \tag{2.7}$$

and $\lim_{k\to\infty} \epsilon_k = 0$. For example, η-dependence corresponds to $\mathcal{F} = \mathcal{G}$ being L-Lipschitz functions with $L < 1$ and sup-norm 1, while $c(f, g) = d_f Lip(f) + d_g Lip(g)$. Clearly, the 'preservation properties' depend on the regularity of the function h.

2.4 Empirical processes

Consider the standard statistical problem of determining whether and how an estimator converges as the sample size increases. When the estimated quantity is finite-dimensional, the law of large numbers (LLN) or the central limit theorem (CLT) provide answers. Empirical process theory allows us to answer similar questions when the estimated quantities are functions. We first recall the general setting for indepedent data, and we provide two results for dependent sequences. For a detailed introduction to empirical process theory, we refer to Kosorok (2008).

Consider a probability measure $\mathbb{P}$ on $(\mathcal{X}, \sigma(\mathcal{X}))$ and let $X_1, \ldots, X_N$ be an i.i.d N-sample from $\mathbb{P}$. We call $\mathbb{P}_N \coloneqq \frac{1}{N}\sum_{n=1}^N \delta_{X_n}$ the *empirical distribution* associated to that sample. For $f : \mathcal{X} \to \mathbb{R}$ a measurable function, we denote by $\mathbb{P}_N f \coloneqq \frac{1}{N}\sum_{n=1}^N f(X_n)$ the *empirical mean* of f over the sample and we assume that the *mean* $\mathbb{P}f \coloneqq \mathbb{E}_{X\sim\mathbb{P}}[f(X)]$ exists. As the sample size increases, the empirical mean converges to the true mean. Specifically, when $\mathbb{P}f^2 < \infty$, the LLN implies that $\mathbb{P}_N f \to \mathbb{P}f$ in probability and the CLT states that $\sqrt{N}(\mathbb{P}_N f - \mathbb{P}f) \to \mathcal{N}(0, \mathbb{P}f^2)$. Analogue conclusions also hold in $\mathbb{R}^D$, for a function $f = (f_1, \ldots, f_D)$, for $f_d : \mathcal{X} \to \mathbb{R}$ measurable and $1 \leq d \leq D$.

If we now have a family of functions, $\mathcal{F}$, the collection of random variables $(\mathbb{P}_N f)_{f\in\mathcal{F}}$ forms a stochastic process, which we call the *empirical process*. A realization of the empirical measure $\hat{P}_N$ (that is, of the random variables $X_1, \ldots, X_N$) gives a *path* $(\mathbb{P}_N f)_{f\in\mathcal{F}}$. The question of convergence of that process as a stochastic process is central in empirical process theory, motivated initially by the following example.

Example 2.19. Consider the classic question of determining whether the empirical distribution function of a real-valued random variable, $F_N(t) \coloneqq \frac{1}{N}\sum_{n=1}^N 1_{X_n\leq t}$ converges to the distribution function F. In the language of empirical process theory introduced above, $\mathcal{X} = \mathbb{R}$, $f_t = 1_{]-\infty,t]}$ and $\mathcal{F} = \{f_t \mid t \in \mathbb{R}\}$. The LLN and the CLT imply convergence of $F_N(t)$ to $F(t)$ for any $t \in \mathbb{R}$ (or any finite collection of evaluations points), but do not specify whether the convergence is uniform over $t \in \mathbb{R}$.

Definition 2.20. The functional family $\mathcal{F}$ is $\mathbb{P}$-*Glivenko-Cantelli* if $\sup_{f\in\mathcal{F}} |\mathbb{P}_N f - \mathbb{P}f| \to 0$ almost surely. The functional family is said to be $\mathbb{P}$-*Donsker* if $\sqrt{N}(\mathbb{P}_N f - \mathbb{P}f) \to \mathbb{G}$ converges weakly to a Gaussian process $\mathbb{G}$ with mean 0 and covariance $\mathrm{Cov}(f, g) = \mathbb{P}(fg) - (\mathbb{P}f)(\mathbb{P}g)$.

Being Glivenko-Cantelli (resp. Donsker) is analogous to satisfying the LLN (resp. CLT) in the finite-dimensional setting. To pass from the finite-dimensional convergence to the functional one, it is common to quantify the 'complexity' of $\mathcal{F}$ using one the two notions of entropy, uniform and bracketing entropy. We focus on the latter in this manuscript.

Definition 2.21. We consider $L_p(\mathbb{P}) = \{f : \mathcal{X} \to \mathbb{R} \mid \mathbb{P}|f|^p < \infty\}$, for some $1 \leq p < \infty$. A pair of functions l, $u \in L_p(\mathbb{P})$ is an ϵ-*bracket*, $[l, u]$, if $\mathbb{P}(l(X) \leq u(X)) = 1$ and $(\mathbb{P}|l - u|^p)^{1/p} \leq \epsilon$. We write $f \in [l, u]$ if $\mathbb{P}(l(X) \leq f(X) \leq u(X)) = 1$. The *bracketing number* of $\mathcal{F}$, $\mathcal{N}_{[]}(\epsilon, \mathcal{F}, L_p(\mathbb{P}))$ denotes the least number of ϵ-brackets which cover $\mathcal{F}$. The *bracketing integral* is $\mathcal{J}_{[]}(\mathcal{F}, L_p(\mathbb{P})) = \int_0^\infty \sqrt{\mathcal{N}_{[]}(\epsilon, \mathcal{F}, L_p(\mathbb{P}))}d\epsilon$.

By Kosorok (2008, Theorems 2.2 and 2.3), a family of measurable functions $\mathcal{F}$ is Glivenko-Cantelli if its ϵ-bracketing number is finite for every $\epsilon > 0$ and it is Donsker if its bracketing integral is finite. The family $(1_{]-\infty,t]})_{t\in\mathbb{R}}$ of point-wise evaluations from Example 2.19 has bounded bracketing integral, so it is Donsker. Many other, natural classes of functions have finite bracketing numbers or even bracketing integrals. In Chapter 4, we study a particular class of empirical processes and we show in Proposition 4.21 that it has a finite bracketing integral.

The Donsker property is related to weak-convergence, a mode particularly adapted to empirical processes with continuous paths. Consider functionals of the form $f_t : \mathcal{X} \to \mathbb{R}$, for $t \in I \subset \mathbb{R}$. Assume that $t \mapsto f_t(x)$ is continuous for each $x \in \mathcal{X}$ and that f_t is measurable. By Theorem 2.1, the empirical process $(f_t)_{t\in I}$ is Borel-measurable and, moreover, equality of processes $(f_t(X))_{t\in I}$ in the weak sense implies that the measures they induce on $C(I, \mathbb{R})$ are the same.

A motivation for establishing Gaussian approximation for empirical processes was justifying the construction of confidence bands and hypothesis testing. One technique to construct such intervals is called *bootstrapping*. We construct the *bootstrap sample* $X_1^b, \dots, X_N^b$, by sampling from $X_1, \dots, X_N$ uniformly and with replacement. Let $\hat{\mathbb{P}}_N \coloneqq \sum_{n=1}^N \delta_{X_n^b}$ be the empirical measure associated with the bootstrap sample. The bootstrap is automatically consistent, in that it provides an approximation of $\mathbb{P}_N f$: if $\mathcal{F}$ is Donsker, then $\sqrt{N}(\hat{\mathbb{P}}_N f - \mathbb{P}_N f)$ converges weakly in probability to the Gaussian process $\mathbb{G}$ (Kosorok, 2008, Theorem 2.6). In such a case, we will say the bootstrap is *valid*. The 'weak convergence' here is with respect to the resampling, $\lim_{N\to\infty} \mathbb{P}^{\otimes N}(\sup_{h\in C_b(\mathbb{X})} \sqrt{N}|\mathbb{E}_B h(\hat{\mathbb{P}}_N f) - h(\mathbb{P}_N f)| > \epsilon) = 0$, where $\mathbb{E}_B$ denotes the expectation taken with respect to the sampling, with initial data $X_1, \dots, X_N$ fixed.

Suppose that we want to construct a confidence interval for $(\mathbb{P}f)_{f\in\mathcal{F}}$, at level $\alpha \in]0, 1[$. The Gaussian approximation implies that $[\mathbb{P}_N f - c_\alpha, \mathbb{P}_N f + c_\alpha]$ is such an interval, where c_α is the α-quantile of $\sup_{f\in\mathcal{F}} \sqrt{N}(\mathbb{P}_N f - \mathbb{P}f)$. Naturally, the distribution of that last is not directly accessible, but, by validity of the bootstrap, we can approximate that distribution, and in particular, c_α. Specifically, consider $\hat{c}_\alpha$, the empirical α-quantile of a large number of realisations of $\sup_{f\in\mathcal{F}} \sqrt{N}(\hat{\mathbb{P}}_N f - \mathbb{P}_N f)$. Then, $\left[\mathbb{P}_N f - \frac{\hat{c}_\alpha}{\sqrt{N}}, \mathbb{P}_N f + \frac{\hat{c}_\alpha}{\sqrt{N}}\right]$ is a valid confidence interval, in that

$$\mathbb{P}^{\otimes N}\left(\mathbb{P}f \in \left[\mathbb{P}_N f - \tfrac{\hat{c}_\alpha}{\sqrt{N}}, \mathbb{P}_N f + \tfrac{\hat{c}_\alpha}{\sqrt{N}}\right]\right) \geq \alpha + o(1).$$

In Chapter 4, we will study $X_1, \dots, X_N$ which are not independent. Under additional assumptions, analogue convergence results hold and bootstrap techniques can be used. The covariance of the limiting Gaussian process has a different form, which takes into account the dependence,

$$(f, g) \mapsto \lim_{k\to\infty} \sum_{n=1}^{\infty} \operatorname{cov}(f(X_k), g(X_n)). \tag{2.8}$$

If the chain X is absolutely regular, with coefficients decreasing sufficiently fast, it is Donsker.

Theorem 2.22 (Kosorok (2008, Theorem 11.22))**.** *Let X be a stationary sequence and $\mathcal{F}$ with finite bracketing entropy. Suppose there exists $r \in]2, \infty[$, such that*

$$\sum_{k=1}^{\infty} k^{\frac{2}{r-2}} \beta_X(k) < \infty, \tag{2.9}$$

Then, $\sqrt{N}(\mathbb{P}_N f - \mathbb{P}f)$ converges to a tight, zero–mean Gaussian $\mathbb{G}$ process with covariance (2.8)*.*

However, the bootstrap technique needs to be adapted, as the resampling in the standard bootstrap breaks the dependence structure. We will use the *moving block bootstrap* (MBB) (Bühlmann, 2002), which we now describe. Like in the standard bootstrap, we sample a total N values, but we do so by sampling $B = N/L$ blocks, where $L = L(N) \in \mathbb{N}$ is the number of variables in each block. Each block is composed of L consecutive elements from X: that is, $(X_n, \dots, X_{n+L-1})$, for $n \in \{1, \dots N - L + 1\}$. The MBB sample is then

$$X_1^*, \dots, X_N^* = X_{n_1}, \dots, X_{n_1+L}, X_{n_2}, \dots, X_{n_2+L}, \ \dots, X_{n_B}, \dots X_{n_B+L},$$

where $n_1, \dots n_B \sim \mathcal{U}(1, \dots, N - L + 1)$ are independent and we define

$$\mathbb{P}_N^* f = \frac{1}{N} \sum_{n=1}^{N} f(X_n^*).$$

Theorem 2.23 (Bühlmann (1995, Theorem1))**.** *In addition to assumptions from Theorem 2.22, suppose that* $\beta_X(k) \xrightarrow{k\to\infty} 0$ *exponentially, i.e.*

$$\beta_X(k) \leq a\exp(-ck), \qquad \text{for some } a,\, c > 0.$$

Let the block size $L(n)$ *in the MBB satisfy* $L(n) \to \infty$ *and* $L(n) = \mathcal{O}(n^{1/2-\epsilon})$*, for some* $0 < \epsilon < \frac{1}{2}$*. Then,*

$$\sqrt{N}(\mathbb{P}_N^* f - \mathbb{P}_N f) \to \mathbb{G}, \qquad \mathbb{P}\text{-almost surely,}$$

where $\mathbb{G}$ *is the zero-mean Gaussian Process with the covariance* (2.8)*.*

Theorem 2.23 gives conditions under which MBB is a valid technique to construct confidence intervals or perform statistical tests using functionals of dependent data.

2.5 Persistent homology and its functional representations

Homology is a topological invariant and formalizes the intuitive notion of a "hole", defining it as "a cycle which is not a boundary" through an algebraic construction. An extension of homology, persistent homology encodes the evolution of those "holes", as the underlying topological space changes. We first introduce singular homology, a flexible and general theory which we will use throughout this work. Then, we introduce simplicial homology, a more computable version, which coincides with singular homology in numerical settings. The exposition of the persistence theory is standard, although we emphasize the importance of persistence diagrams constructed from measures, as we will use this in Chapter 3. Section 2.5.4 points to the link between the theory from Sections 2.5.1-2.5.3 and the intuitive picture of persistence from Figure 1.3. We recommend Hatcher (2002) for an introduction to general algebraic topology, Chazal et al. (2016) for the theory of persistence modules, and finally Chazal and Michel (2021) for a broader picture of the developments and applications of persistent homology.

2.5.1 Singular homology

For $n \in \mathbb{N}^*$, the standard n-simplex is $\Delta^n = \{(t_0, \ldots, t_n) \mid 0 \leq t_k \leq 1,\ \sum_{k=0}^n t_k = 1\}$, with the standard topology induced by $\mathbb{R}^{n+1}$. The standard 0-simplex is a point $\Delta^0 = \{1\}$. A *singular n-simplex* in a topological space X is a continuous map $\sigma : \Delta^n \to X$. For example, a singular 1-simplex is a map $\sigma : [0,1] \to \mathbb{R}$ and a singular 0-simplex, $\sigma : \Delta^0 = \{1\} \to X$ can be identified with a point, $\sigma(1) \in \mathbb{X}$.

Definition 2.24. The *boundary of* σ is the alternating formal sum of restriction maps

$$\partial_n \sigma = \sum_{k=0}^{n} (-1)^k \sigma^k, \tag{2.10}$$

where $\sigma^k : \Delta^{n-1} \to X$ is the restriction to the k-th face,

$$\sigma^k(t_0, \ldots, t_{n-1}) = \sigma(t_0, \ldots, t_{k-1}, 0, t_{k+1}, \ldots, t_{n-1}).$$

The interpretation of this construction is that σ is a (possibly degenerate) realization of an n-dimensional structure in X. Since Δ^{n-1} is a face of Δ^n, σ^i is the realisation of the restriction of σ to that face.

Example 2.25. By (2.10), the boundary of a 1-simplex σ is a sum of 0-simplices $\partial_1\sigma = \sigma^0 - \sigma^1$. Because the domain of σ^0 and σ^1 is a point, we can identify the boundary with a formal sum of points from X,

$$\partial_1\sigma = \sigma^0 - \sigma^1 = \sigma(0,1) - \sigma(1,0).$$

Let now $X = \mathbb{S}^1$ be the unit circle seen in the complex plane and consider the following 1-simplices

$$\sigma_1(t_0, t_1) = \exp(\pi i), \qquad \sigma_2(t_0, t_1) = \exp(\pi i t_1), \qquad \sigma_3(t_0, t_1) = \exp(2\pi i t_1).$$

The map σ_1 is constant, so $\partial_1\sigma_1 = 0$. The function σ_2 maps the unit interval to the upper-half circle. Its boundary is $\partial_1\sigma_2 = \sigma_2(0,1) - \sigma_2(1,0) = [-1] - [1]$, where the notation $[x]$ refers to the algebraic element (0-simplex) corresponding to a point $x \in \mathbb{C}$. So, the formal sum $\partial_1\sigma_2$ is not trivial. The simplex σ_3 represents a full revolution around $\mathbb{S}^1$. Even though it is not constant, we have $\sigma_3(0,1) = \sigma_3(1,0)$, so its boundary is trivial.

We can now introduce singular homology. Let $\mathbb{F}$ be a field. We call *n-chains* the elements of C_n, the free abelian $\mathbb{F}$-module generated by the singular n-simplices of X. We extend the definition of the boundary ∂_n from the basis to the whole module by linearity and we call $\partial_n : C_n \to C_{n-1}$ the *boundary map*, where we set $C_{-1} \coloneqq 0$.

Definition 2.26. A n-cycle is an element of $\ker \partial_n$ and a n-boundary an element of $\mathrm{im}\partial_{n+1}$.

From (2.10), it is clear that $\partial_n \circ \partial_{n+1} = 0$, so $\mathrm{im}\partial_{n+1} \subset \ker \partial_n$ and a n-boundary is a n-cycle. The sequence $((C_n)_{n\geq -1}, (\partial_n)_{n\geq 0})$ is a *chain complex*. So, we can define the *n-th homology group of* X as the quotient module

$$H_n(X) \coloneqq \ker \partial_n / \mathrm{im}\partial_{n+1}. \tag{2.11}$$

Homology is a topological invariant, in that it satisfies several, natural properties and reflects the relationships between topological spaces. First, a continuous function $f : X_1 \to X_2$ induces a morphisms between homology groups of those spaces. Indeed, for any $n \in \mathbb{N}$ and any n-simplex σ in X_1, $f \circ \sigma : \Delta^n \to X_2$ is an n-simplex of X_2. We can extend f to $C_n(X_1) \to C_n(X_2)$ by linearity. By definition, it commutes with the boundary maps $\partial_n(f \circ \sigma) = f(\partial_n \sigma)$, so, in particular, it maps $\ker \partial_n(X_1)$ to $\ker \partial_n(X_2)$ and $\mathrm{im}\partial_{n+1}(X_1)$ to $\mathrm{im}\partial_{n+1}(X_2)$. It therefore extends to a morphism $f : H_n(X_1) \to H_n(X_2)$. This is one of numerous, natural properties satisfied by singular homology. Singular homology satisfies the Eilenberg-Steenrod axioms, some of which we recall in Proposition 2.27.

Proposition 2.27.

2.27a. *If $f : X_1 \to X_2$ is a homeomorphism (bijective, continuous function, with a continuous inverse), then the homology groups of X_1 and X_2 are isomorphic $H_n(X_1) \simeq H_n(X_2)$.*

2.27b. *If $X = \bigsqcup_{k=1}^{K} X_k$, then $H_n(X) \simeq \bigoplus_{k=1}^{K} H_n(X_k)$.*

2.27c. *The homology of a point is trivial: for $X = \{x\}$, we have $H_n(X) = 0$ for $n \geq 1$ and $H_0(X) = \mathbb{F}$.*

Take X and the simplices from Example 2.25. We have established that σ_1, $\sigma_3 \in \ker(\partial_1)$ and it is clear that $\sigma_1 \in \mathrm{im}\partial_2$, so its image in the homology group $H_1(\mathbb{S}^1)$ is trivial. However, $\sigma_3 \in \ker \partial_1$ and it is known that $\sigma_3 \notin \mathrm{im}\partial_2$. It is, in fact, a generator of $H_1(\mathbb{S}^1)$, which is of rank 1. The fact that σ_3 is not a boundary is not direct from the definition of singular homology groups, but it can be shown if we admit the remaining Eilenberg-Steenrod axioms or using homotopy theory[1] One can also use other homology theories (cellular or simplicial) to study the homology of $\mathbb{S}^1$.

2.5.2 Simplicial homology

Apart from some particular cases (like finite spaces or small topologies), there are uncountably many distinct n-singular simplices so $C_n(X)$ is huge, making direct computations of homology prohibitive. Some natural spaces (like a sphere or a torus) have a decomposition (for example, CW-complexes), and then theoretical machinery can be used to determine homology groups. The spaces that arise in practical and numerical applications are often discrete and represented in such a decomposition: graphs, polyhedra, or their generalizations, simplicial complex of finite size. This reduces the computation of homology to linear algebra.

Definition 2.28. Consider a finite set S. A *simplex* σ is a tuple (ordered set) from S. We say that another simplex σ' is a *face* of σ if $\sigma' \subseteq \sigma$. An *abstract simplicial complex* $\mathcal{K}$ is a family of simplices such that

- $\mathcal{K}$ contains all the singletons: for all $x \in S$, $[x] \in \mathcal{K}$,
- all the faces of a simplex from $\mathcal{K}$ are also in $\mathcal{K}$: ($\sigma \in \mathcal{K}$ and σ' is a face of σ) $\implies \sigma' \in \mathcal{K}$.

If we denote by $|\cdot|$ the cardinality of a set, we call $|\sigma| - 1$ the *dimension* of σ and the *dimension of the complex* is the largest dimension of a simplex of $\mathcal{K}$.

[1] It suffices to show that $\sigma_3([1-t, t])$ is not homotopy equivalent to a constant map.

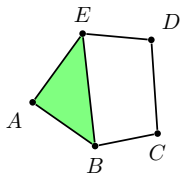


Figure 2.1: Simplicial complex $\mathcal{K}$ from Example 2.30.

Example 2.29. A finite graph $G = (V, E)$ is an example of an abstract simplicial complex of dimension one. Indeed, let $S = V$, so that the 0-simplices correspond to the graphs' vertices and E are precisely the 1-simplices, and that $\mathcal{K} = S \cup E$. The faces of an edge $[v_1, v_2]$ are the vertices $[v_1]$, $[v_2]$, so $\mathcal{K}$ trivially verifies the two properties and is indeed an abstract simplicial complex. Example 2.30 shows a simplicial complex of dimension 2.

The definition of the boundary of a simplex, chains, cycles, boundaries and homology groups is analogue to the singular setting. For $\sigma = [x_0, \dots x_n]$, we define its boundary as the alternating sum of faces,

$$\partial_n \sigma = \sum_{k=0}^{n} (-1)^k [x_0, \dots, x_{k-1}, x_{k+1}, \dots, x_n],$$

similarly to the singular case. By linearity, we can extend this boundary operator to n-chains C_n, the $\mathbb{F}$-module generated by n-simplices of $\mathcal{K}$. We also obtain a chain complex of which the homology is defined as in (2.11).

Singular and simplicial homologies capture the same information, although via different constructions. For abstract simplicial complexes and their geometric realizations, the two constructions coincide (Hatcher, 2002, Theorem 2.27). However, the notable advantage of simplicial homology is that, for a finite set S, the chain complex is finite-dimensional. Hence, we can represent ∂_n as a matrix using the n- and $(n-1)$-simplices as bases and use linear algebra to calculate the homology of $\mathcal{K}$. In addition, the field is often set to $\mathbb{F} = \mathbb{Z}_2$ and in that case, a n-chain can be interpreted as a collection of n-simplices and a sum of two n-chains as their symmetric difference.

Example 2.30. Consider the following simplicial complex from Figure 2.1

$$\mathcal{K} = \{A,\ B,\ C,\ D,\ E, [A,B],\ [B,C],\ [C,D],\ [D,E],\ [E,A],\ [B,E],\ [A,B,E]\}.$$

We can represent the boundary operators as matrices

$$\partial_1 = \begin{array}{c} \begin{matrix} [A,B] & [B,C] & [C,D] & [D,E] & [E,A] & [B,E] \end{matrix} \\ \begin{pmatrix} -1 & 0 & 0 & 1 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 \end{pmatrix} \end{array} \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix}, \qquad \partial_2 = \begin{array}{c} \begin{matrix} [A,B,E] \end{matrix} \\ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \end{array} \begin{matrix} [A,B] \\ [B,C] \\ [C,D] \\ [D,E] \\ [E,A] \\ [B,E] \end{matrix}$$

From the reduced form of ∂_1, we can see that $\operatorname{rank}(\partial_1) = 4$ and that $\ker \partial_1 = \operatorname{Span}(c_1, c_2)$, with $c_1 = [A,B]+[B,C]+[C,D]+[D,E]+[A,E]$ and $c_2 = [A,B]+[B,E]+[E,A]$. However, $c_2 \in \operatorname{im}\partial_2$. So,

$$H_0(\mathcal{K}) = \mathbb{F}, \qquad H_1(\mathcal{K}) = \mathbb{F}, \qquad H_n(\mathcal{K}) = 0 \text{ for } n \geq 2, \tag{2.12}$$

which gives the same homology groups as those of $\mathbb{S}^1$.

2.5.3 Persistent homology

In applications, the data often consists of an isolated set of points, or a graph, and its 'topological features' (number of connected components, cycles) depend on the scale at which it is considered. Persistent homology encodes how the topology changes as the scale varies. Specifically, it allows to study the evolution of homology groups for a parametrized family of spaces. Part of the appeal of

that theory resides in the celebrated decomposition theorem, which gives the interpretation that the module is characterized by some persistent homological generators. We present the theory from Chazal et al. (2016).

A *persistence module* over $\mathbb{R}$ is a collection of vector spaces $(V_t)_{t\in\mathbb{R}}$ along with morphisms $(\iota_s^t : V_s \to V_t)_{s\leq t\in\mathbb{R}}$, such that $\iota_s^t \circ \iota_r^s = \iota_r^t$ and $\iota_t^t = \mathrm{id}_{V_t}$. The interpretation of the persistence module is that it allows us to see the evolution of an element of $v \in V_s$ through $\mathbb{R}$, by looking at its image $\iota_s^t(v)$ for any $t \geq s$.

A standard example of a persistence module from topological data analysis is that of homology of sub level sets of some continuous function, $f : X \to \mathbb{R}$. For any $t \in \mathbb{R}$, we consider the closed sub level sets $X_t = f^{-1}(]-\infty, t])$ and we call $(X_t)_{t\in\mathbb{R}}$ the *sub level set filtration of* X. We will sometimes call f the *filter function*. For each level, we define $V_t \coloneqq H_n(X_t)$, where H_n denotes n-dimensional singular homology. A natural map between any two level sets is the inclusion $X_s \hookrightarrow X_t$, for $s \leq t$. By the functorial properties of homology mentioned in Section 2.5.1, this induces morphisms ι_s^t at the homology level, so $((V_t)_{t\in\mathbb{R}}, (\iota_s^t)_{s\leq t\in\mathbb{R}})$ is a persistence module. We elaborate on this example in Section 2.5.4. Another closely related example is the terminated persistence module.

Definition 2.31. The *terminated persistence module* of f, denoted $\bar{\mathbb{V}}(f)$ is defined by

$$\bar{V}_t(f) = \begin{cases} H_0(X_t), & \text{if } t < \max f \\ 0, & \text{otherwise.} \end{cases} \tag{2.13}$$

The morphisms ι_s^t are induced by inclusions for $s \leq t < \max f$ and are 0 for $t \geq \max f$.

The terminated persistence module differs from the homology module of sub level sets in that it is 0 once the maximum of f is reached. We will argue that it shares many properties of the non-terminated module, with the notable exception that all generators disappear eventually.

In general, persistence modules can have complicated structure, but, in certain cases of particular interest, they are remarkably simple: they decompose as a sum of interval modules

$$\mathbb{V} = \sum_{I\in\mathcal{B}} 1_I(t), \tag{2.14}$$

where $\mathcal{B}$ is a collection of intervals and the interval module 1_I is defined as

$$1_I(t) = \begin{cases} \mathbb{F} & \text{if } t \in I \\ 0 & \text{otherwise,} \end{cases} \qquad \iota_s^t = \begin{cases} \mathrm{id} & \text{if } s, t \in I \\ 0 & \text{otherwise.} \end{cases}$$

When such a decomposition exists, it is unique up to isomorphism of persistence modules (Chazal et al., 2016, Theorem 2.7), so we can use the intervals as a representation of a module.

Definition 2.32. A morphism between two persistence modules $\mathbb{V}, \mathbb{W}$ is a collection of maps $\rho_t : V_t \to \mathbb{W}_t$ such that the following diagram commutes

$$\begin{array}{ccc} V_s & \xrightarrow{\iota_s^t} & V_t \\ {\scriptstyle\rho_s}\downarrow & & \downarrow{\scriptstyle\rho_t} \\ W_s & \xrightarrow{\iota_s^t} & W_t \end{array}$$

and it is an isomorphism if ρ_t is an isomorphism for all t.

It is particularly important in the scope of this thesis that the intervals in $\mathcal{B}$ are not necessarily unique, so we treat this collection as a multiset. There are two graphical representations of $\mathcal{B}$ in the plane. One is a vertical stack of intervals in $\mathbb{R}^2$, called *the persistence barcode*

$$\{I \times \{n\} \mid n \in \mathbb{N},\, I \in \mathcal{B}\},$$

and the other, the *persistence diagram* is a multiset of points in $\mathbb{R}^2$,

$$\{(b, d) \text{ the endpoints of } I \mid I \in \mathcal{B}\}.$$

Remark 2.33. Note that the persistence diagram as defined above is not a complete invariant of the module, as it does not capture whether the endpoints are included or not in the interval. The *decorated persistence diagram* is a version of the diagram which contains that information Chazal et al. (2016).

We discuss the interpretation of decomposition (2.14) in the context of a sub level set filtration. If f is bounded, there is $t_0 \in \mathbb{R}$ sufficiently small such that $X_t = \emptyset$ for any $t \leq t_0$ and $t_1 \in \mathbb{R}$ such that $X_{t_1} = X$ for any $t \geq t_1$. So, persistent homology interpolates between $H_n(\emptyset) = 0$ and $H_n(X)$, and what happens in between is specified by f. Typically, if X is an interval and the interval decomposition exists, the homology of X_t changes at t only when t is a critical value of f, so the endpoints of the intervals correspond to critical values of f. Consider an interval $I = [b, d[$. It corresponds to a homological feature which appears for the first time in X_b, exists for all $t < d$ and which disappears in X_d. It is particularly important that it does not exist after d. We say that the generator is *born* at time b, *dies* at time d, and that b and d are its *birth* and *death* times.

The interval decomposition of a module exists when the module is indexed by a finite set (instead of $\mathbb{R}$), or, when V_t is of finite dimension for all $t \in \mathbb{R}$ (Chazal et al., 2016, Theorem 2.8). We mentioned already in Section 2.5.2 that in case we have a finite simplicial complex, the space of chains (and therefore also the homology groups) is of finite dimension. Therefore, in practical computations, the persistence modules are decomposable, what justifies the above definition of persistence diagrams. Algorithmically, the persistence diagrams can be computed using linear algebra, not unlike in the case of non-persistence case, although the reduction algorithm has to be adapted to respect the order in which the simplices appear in the filtration (Herbert Edelsbrunner and John Harer, 2010).

The vector spaces V_t are not finite-dimensional in general, and the persistence diagram of a general function f is constructed using the theory developed in Chazal et al. (2016). The idea is that we can associate to a persistence module $\mathbb{V}$ a measure $\mu_{\mathbb{V}}$ on rectangles in $\mathbb{R}^2$, in such a way that it characterizes the number of generators with life-span constricted by that rectangle. Specifically, by Chazal et al. (2016, Proposition 3.8), the *persistence measure* $\mu_{\mathbb{V}}$ is

$$\mu_{\mathbb{V}}([s_1, s_2] \times [t_1, t_2]) = \dim \left(\frac{\operatorname{im}(\iota_{s_2}^{t_1}) \cap \ker(\iota_{t_1}^{t_2})}{\operatorname{im}(\iota_{s_1}^{s_2}) \cap \ker(\iota_{t_1}^{t_2})} \right). \tag{2.15}$$

The numerator represents the non-trivial generators which exist already at s_2 (were born earlier than that) and die after t_1, but before t_2. The denominator[2] cancels those generators which satisfy the criterion above, but which were actually born even before s_1. With that interpretation, we can see that the persistence measure $\mu_{\mathbb{V}}$ of an interval decomposable module $\mathbb{V}$ counts the number of points in the persistence diagram which are contained in a given rectangle.

Chazal et al. (2016, Theorem 3.12) show the equivalence between persistence measures and multi-sets in the plane, as long as both are 'finite[3] and Lemma 3.16 gives an explicit way to associate a persistence diagram D to a persistence module $\mathbb{V}$, which we adapt for our purposes of undecorated diagrams. For any non-increasing sequences $(\epsilon_k)_{\in\mathbb{N}}$, $(\eta_k)_{\in\mathbb{N}}$ converging to 0, the multiplicity of (b, d) in the persistence diagram D of $\mathbb{V}$ is

$$\mathrm{m}_D(b, d) = \mathrm{m}_{\mathbb{V}}(b, d) = \lim_{k \to \infty} \mu_{\mathbb{V}}([b - \epsilon_k, b + \epsilon_k] \times [b - \eta_k, b + \eta_k]). \tag{2.16}$$

The finiteness conditions are satisfied as soon as the module is *q-tame*, that is, when $\operatorname{rank}(V_s \to V_t)$ is finite for any $s < t \in \mathbb{R}$. The sub level set persistence module is q-tame when X is compact and f continuous (Chazal et al., 2016, Theorem 3.33), what guarantees that the persistence diagram is well-defined. The same is true for the terminated persistence module. The persistence diagram of the persistence measure associated to an interval-decomposable module coincides with the diagram as defined from the decomposition.

2.5.4 Persistent homology of a filtered interval or circle

The persistent homology of a function on an interval is a simple illustration of the theory introduced in the previous section. In fact, in this thesis, we will almost exclusively be interest in the filtration of either the circle or of a compact interval and homology dimension $n = 0$ (for $n \geq 1$, the homology

[2] Note that the quotient is algebraic, so it is a difference between dimensions of vector spaces.

[3] In the sense of 'finite measure' for the measure and 'locally finite' for the multiset

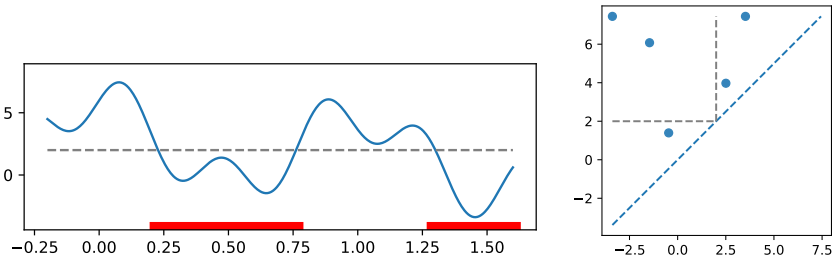


Figure 2.2: On the left, the graph of a function, with sub level sets in red, for the level t given by the dashed gray line. On the right, the persistence diagram of that function. Delimited by the gray dashed lines and with lower-right vertex (t, t), the unbounded region contains points which correspond to the connected components visible in red.

is trivial). Chapter 3 is dedicated to studying the structure of the persistence module when f is a periodic function.

Let X be a circle or a compact interval. The persistence diagram of $f : X \to \mathbb{R}$, contains information about the values and order of local extrema of f. The morphism ι_s^t in homology is an isomorphism if f has not critical points in $[s, t]$. If $x \in X$ is an isolated local minimum, $f(x)$ is a critical value and $\sigma \equiv x \in C_0(X_{f(x)})$ generates a non-trivial homology class. Hence, a local minimum of f leads to a new connected component. Analogously, a local maximum leads to the death of one. A point in the persistence diagram can be interpreted as a pair of a local minimum and a local maximum. Figure 2.2 depicts sub level sets and the corresponding diagram.

As observed in the previous section, $H_0(X_t) = H_0(X)$ for any $t \geq \max f$. In particular, for an interval and the circle, this implies that there is one connected component that has no death time. In the decomposition (2.14), it corresponds to an unbounded, half-open interval $[b, \infty[$, and in the persistence diagram, it is often represented as a point (b, ∞). For practical purposes, it is often convenient to either discard such points, or to assign them a finite death value. The terminated persistence module (2.13) is a way to do so algebraically. The persistence diagrams of so-modified modules are contained in a bounded region, as formalized in Proposition 2.34.

Proposition 2.34. *Let $M \coloneqq \max f$, $m \coloneqq \min f$ and denote by $D(f)$ the persitence diagram of $\bar{\mathbb{V}}(f)$. Then, the birth (resp. death) of any point is lower (resp. upper) bounded by m (resp. M),*

$$D(f) \subset \{(b, d) \mid m \leq b,\, d \leq M\}.$$

This modification does not impact the invariance properties. In particular, the persistent homology is invariant with respect to homeomorphisms.

Proposition 2.35 (Invariance to reparametrisation)**.** *Consider a continuous function $f : X \to \mathbb{R}$ and let $\gamma_1, \gamma_2 : [0, T] \to X$ be a homeomorphism such that $\gamma_1(0) = \gamma_2(0)$ and $\gamma_1(T) = \gamma_2(T)$. Then,*

$$D(f \circ \gamma_1) = D(f \circ \gamma_2).$$

Proof. For any $t \in \mathbb{R}$, the homeomorphism $g \coloneqq (\gamma_1^{-1} \circ \gamma_2) : [0, T] \to [0, T]$ maps the t-sublevel set of $f \circ \gamma_2$ to that of $f \circ \gamma_1$. Indeed,

$$\begin{aligned}(f \circ \gamma_1)^{-1}(]-\infty, t]) &= \{y \in [0, T] \mid (f \circ \gamma_1)(y) \leq t\} \\ &= \{y = g(x) \mid (f \circ \gamma_1)(g(x)) = (f \circ \gamma_2)(y) \leq t\} \\ &= g(\{y \in [0, T] \mid (f \circ \gamma_2)(y) \leq t\}).\end{aligned}$$

By Proposition 2.27a, g induces an isomorphism between the two corresponding persistence modules. So, the corresponding persistence diagrams are the same (as well as any invariants thereof). □

2.5.5 Persistent homology of a time series

When the data is not a continuous function but a vector $S \in \mathbb{R}^N$, a time series of length $N \in \mathbb{N}$, we can define its diagram in two ways which can be shown to coincide.

First, we define a continuous function f by discretizing $[0,1]$, prescribing the values at the nodes and linearly interpolating between,

$$f(t) = \sum_{k=0}^{N-1} 1_{\left[\frac{k}{N}, \frac{k+1}{N}\right[}(t) N \left(\left(t - \tfrac{k}{N}\right) S_{k+1} + \left(\tfrac{k+1}{N} - t\right) S_k\right).$$

Then, we can define the persistence module of S as the singular homology of the sub level sets of f. Because every V_t is of finite dimension, that persistence module has an interval decomposition and thus, a persistence diagram.

The second construction consists in defining a simplicial complex, $\mathcal{K} = \{(k), (k, k+1) \mid k = 1, \dots N\}$, and a function $f : \mathcal{K} \to \mathbb{R}$ by $f(\sigma) = \max_{k \in \sigma} S_k$. Then, we define the persistence module of S as the simplicial homology module of $(f^{-1}(]-\infty, t]))_{t \in \mathbb{R}}$. Each space is finite dimensional, so it also admits an interval decomposition.

The two construction coincide, which allows us to use them interchangeably. The singular point of view allows us to show that the diagram of a function and of a time series of samples thereof, are close in bottleneck distance (defined in Section 2.5.6). The latter point of view allows us to actually compute the persistence diagram: Algorithm 1 is an example. In general, persistent homology is notorious for having high computational complexity. However, it is not the case for a time series, as the algorithm recently proposed by Glisse (2023) to calculate the persistence diagram has time-complexity $\mathcal{O}(N)$.

Algorithm 1: Computing the persistence diagram of dimension 0 of a sublevel set filtration of f

Data: A time series or a function f with a finite number of local extrema

$D \leftarrow \{\}$;
for *c a local extremum of f, ordered by ascending value of f* **do**
  if *c is a local min* **then**
    Push $(f(c), *)$ to D with key c.;
  end
  else
    `/* c is a local max */`
    $I \leftarrow$ connected component of c in $f^{-1}(]-\infty, f(c)])$.;
    $c_1, c_2 \leftarrow \operatorname{argmin}_{x<c} f(I), \operatorname{argmin}_{x>c} f(I)$.;
    $i \leftarrow \operatorname{argmax}\{f(c_1), f(c_2)\}$;
    Set the second coordinate of $D(c_i)$ to $f(c)$.;
  end
end
Set the second coordinate of $D(\operatorname{argmax}(s))$ to $\max(f)$.;
return D

2.5.6 Stability of persistence diagrams

The persistent homology module of sub level sets is stable with respect to perturbations of a function. It justifies the use of persistence as an invariant in practical scenarios, where only a noisy or partial observation (for example, on a fine mesh) of that function is available. In this section, we recall the celebrated bottleneck-stability result and also the algebraic ingredients of the proof: the interleaving distance, to justify the terminated modules' stability.

Definition 2.36. We call a *ϵ-matching* between two persistence diagrams D and D' a bijection $\Gamma : A \to A'$ between some subsets of $A \subset D$ and $A' \subset D'$, considered with multiplicity, if

- all matched points are ϵ-close: for any $a \in A$, $d_\infty(a, \Gamma(a)) \leq \epsilon$,
- all non-matched points are ϵ-close to the diagonal: for any $a \in (D \backslash A) \cup (D' \backslash A')$, $d_\infty(a, \Delta) \leq \epsilon$,

where $\Delta = \{(x, x) \in \mathbb{R}^2\}$ denotes the diagonal. The *bottleneck distance* (Herbert Edelsbrunner and John Harer, 2010, VIII.2), $\mathrm{d_b}(D, D')$ is the infimum $\epsilon > 0$ such that there exists and ϵ-matching between D and D'.

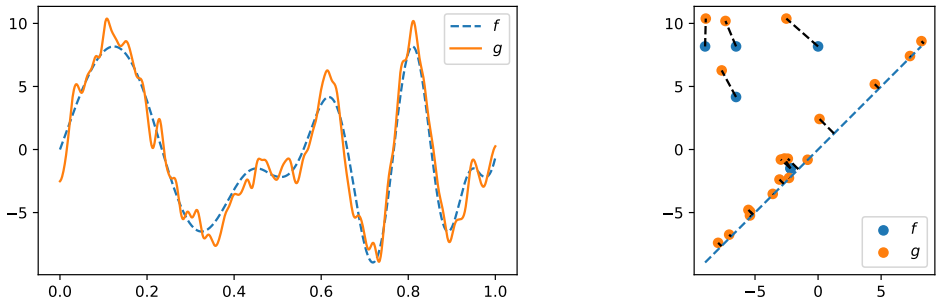


Figure 2.3: On the left, the graph of a function f and a noisy version, g. On the right, the persistence diagrams. The dashed lines between points represent a matching Γ that realizes the bottleneck distance.

Theorem 2.37 (Bottleneck stability of diagrams)**.** *Let $f, g : X \to \mathbb{R}$ be two continuous functions on a compact space X. Then,*

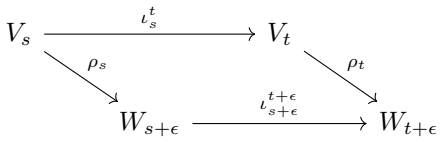
$$\mathrm{d_b}(D(f), D(g)) \leq \|f - g\|_\infty.$$

Note that we will often abuse notation and see an ϵ-matching as defined on the whole diagram $\Gamma : D \cup \Delta \to D' \cup \Delta$, with the diagonal with infinite multiplicity. In that case, we require that $\Gamma^{-1}\big|_{\Delta \setminus \Gamma(D)}$ and $\Gamma|_{\Delta \setminus \Gamma^{-1}(D')}$ are both the identity. Figure 2.3 shows an example of a function f and a noisy version g, their respective persistence diagrams and a grapical representation of a matching. In Chapters 3 and 4, we will use the terminated persistence module whose diagram is also stable.

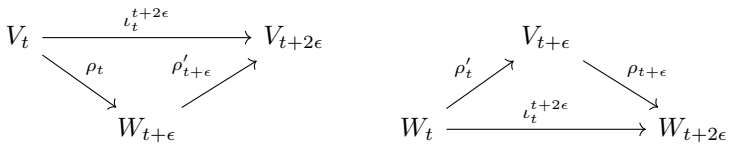
Proposition 2.38 (Stability of the terminated persistence module)**.** *Let $f, g : X \to \mathbb{R}$ be two continuous functions on a compact space X. If we denote by $D(f)$ and $D(g)$ the persistence diagrams of $\mu_{\bar{\mathbb{V}}f}$ and $\mu_{\bar{\mathbb{V}}g}$ respectively, then*

$$\mathrm{d_b}(D(f), D(g)) \leq \|f - g\|_\infty.$$

To show that the terminated module is also stable, we need to understand where the stability in Theorem 2.37 comes from. As such, it is a consequence of two elements: interleavings and the isometry theorem. We call $\rho = (\rho_t : V_t \to W_{t+\epsilon})_{t\in\mathbb{R}}$ an *ϵ-morphism* between $\mathbb{V}$ and $\mathbb{W}$ if it is a morphism between persistence modules $(V_t)_{t\in\mathbb{R}}$ and $(W_{t+\epsilon})_{t\in\mathbb{R}}$. In other words, we require the following diagram to commute



An ϵ-interleaving is a tuple of ϵ-morphisms, $((\rho_t : V_t \to W_{t+\epsilon})_{t\in\mathbb{R}}, (\rho' : W_t \to V_{t+\epsilon})_{t\in\mathbb{R}}))$ which are inverse of one another up to a 2ϵ-shift: for all $s \leq t$, also the following diagrams commute



Using interleavings, we define the *interleaving distance* $\mathrm{d_i}$,

$$\mathrm{d_i}(\mathbb{V}, \mathbb{W}) := \inf\{\epsilon > 0 \mid \text{there exists an } \epsilon\text{-interleaving between } \mathbb{V} \text{ and } \mathbb{W}\}. \tag{2.17}$$

The isometry theorem relates the interleaving and bottleneck distances (Chazal et al., 2016, Theorem 5.14).

Theorem 2.39 (Isometry theorem)**.** *Let $\mathbb{V}$, $\mathbb{W}$ be q-tame modules. Then,*

$$\mathrm{d_i}(\mathbb{V}, \mathbb{W}) = \mathrm{d_b}(D(\mu_{\mathbb{V}}), D(\mu_{\mathbb{W}})). \tag{2.18}$$

Proof of Proposition 2.38. The proof follows the standard proof of Theorem 2.37. Let $\epsilon \coloneqq \|f - g\|_\infty$. For $t < \min(\max f, \max g - \epsilon)$, the inclusion $f^{-1}(]-\infty, t[) \subset g^{-1}(]-\infty, t+\epsilon[)$ induces a morphism $\rho_t : \bar{V}(f)_t = V(f)_t \to V(g)_{t+\epsilon} = \bar{V}(g)_{t+\epsilon}$. If we set $\rho_t \equiv 0$ for $t \geq \min(\max f, \max g - \epsilon)$, then $(\rho_t)_{t\in\mathbb{R}}$ is an ϵ-morphism. We proceed similarly for ρ'. The two morphisms are an ϵ-interleaving. Since $\mathbb{V}(f)$ and $\mathbb{V}(g)$ are q-tame, so are $\bar{\mathbb{V}}(f)$ and $\bar{\mathbb{V}}(g)$ and we conclude by Theorem 2.39. □

While the bottleneck distance is natural because of its relation with the algebraic interleaving distance, another popular distance is the p-Wasserstein distance,

$$d_p(D, D') = \inf_{\Gamma : A \to A'} \left[\sum_{a \in A} d(a, \Gamma(a))^p + \sum_{a \in D' \setminus A' \cup D \setminus A} d(a, \Delta)^p \right]^{1/p}, \tag{2.19}$$

where the infimum is taken over matchings of D and D' as above. It is considered a natural generalization of $\mathrm{d_b}$, which is d_∞. However, it is not always well-defined.

2.5.7 Total persistence and functional representations of persistence diagrams

For $p \in \mathbb{N}^*$, the *total p-persistence* of a persistence diagram D is defined as $\mathrm{pers}_p(D) = (\sum_{(b,d)\in D}(d-b)^p)^{1/p}$. The space $\mathcal{D}^p = \{D \mid \mathrm{pers}_p(D) < \infty\}$ of persistence diagrams with the total p-persistence is finite is a Polish metric space for the p-Wasserstein metric (Divol and Lacombe, 2021, Proposition 3.3). For the persistence diagram of a function, $D = D(f)$, the total p-persistence is related to the regularity of that function and to its total variation (Perez, 2022a, Theorem 3.7 and Proposition 3.6): typically, if f defined on a subset of $\mathbb{R}^d$ is α-Hölder, its total p-persistence is finite, for any $p > d/\alpha$. Continuity of total persistence was initially obtained in Cohen-Steiner et al. (2010) for Lipschitz functions. In Chapter 4, we introduce a weaker variant of total persistence, the ϵ-truncated persistence, which we show to be continuous for less regular functions.

The space of persistence diagrams is not a vector space and is ill-suited for statistical learning. It is common to map diagrams to a functional space, often Banach (Chazal and Michel, 2021).

Definition 2.40. Consider $(\mathbb{T}, d)$ a Euclidean space and let $\mathcal{H} \subset \mathbb{R}^{\mathbb{T}}$ be a functional Banach space. A *functional representation* of persistence diagrams with finite total p-persistence is a function $\rho : \mathcal{D}^p \to \mathcal{H}$.

Numerous functionals have been proposed in the literature (Carrière et al., 2020, Bubenik, 2015, Adams et al., 2017, Chung and Lawson, 2022). Many functionals are of the form

$$\rho(D)(t) = \sum_{(b,d)\in D} (d-b)^p k(b,d)(t),$$

where $k : \mathbb{R}^2 \to \mathcal{H}$ is a map, which to a point (b, d) in the plane associates a function $k(b, d) : \mathbb{T} \to \mathbb{R}$. They are called 'linear', as $\rho(D_1 \sqcup D_2) = \rho(D_1) + \rho(D_2)$, where $\sqcup$ denotes the union of multisets. We provide Examples 2.41 and 2.42, to which we will come back to in Chapter 4 and of which Figure 2.4 is a graphical representation.

Example 2.41 (Persistence Silhouette)**.** The persistence silhouette (Chazal et al., 2014) is a weighted sum of 'tent' functions $\Lambda_{(b,d)}(t) = \left(\frac{d-b}{2} - |t - \frac{b+d}{2}|\right)_+$, for $(\mathbb{T}, d) = (\mathbb{R}, |\cdot|)$. The kernel $t \mapsto \Lambda_{b,d}(t)$ is piecewise linear and 1-Lipschitz.

Example 2.42 (Persistence Image)**.** The kernel that corresponds to the persistence image (Adams et al., 2017) is $k^{pi}(b,d)(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(b-x)^2+(d-y)^2}{2\sigma^2}\right)$, for some $\sigma > 0$ and $(\mathbb{T}, d) = (\mathbb{R}^2, \|\cdot\|_2)$ The function $(x, y) \mapsto \exp(-(x^2 + y^2))$ is $(4/e)$-Lipschitz.

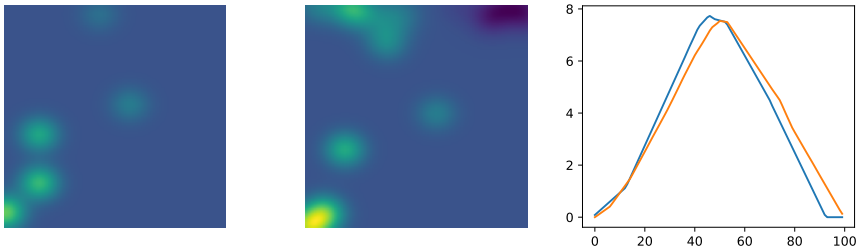


Figure 2.4: Illustration of functionals for diagrams of the two functions from Figure 2.3. On the left, persistence images with $p = 1$ and $\sigma = 1$. On the right, the persistence silhouettes.

The properties of certain classes of functionals have been fully characterized (Divol and Polonik, 2019, Divol and Lacombe, 2021). Their convergence can be studied using empirical process theory, as introduced in Section 2.4. We particularly emphasize the regularity of the kernels, which is often used to bound the covering number of the functional family (Chazal et al., 2014, Berry et al., 2020). Asymptotic results are available also for other topological descriptors. Stabilization theory have been leveraged to show convergence of topological descriptors: Betti numbers of simplicial complexes over point processes in $\mathbb{R}^d$ (Roycraft et al., 2023). Functional representations are also flexible, in that it is also possible to optimize them for a statistical task at hand (Carrière et al., 2020, Hacquard, 2023).

Chapter 3

Persistent homology of periodic functions

Consider a signal $S : [0, T] \to \mathbb{R}$ of the form (1.1), that is $S(t) = (\phi \circ \gamma)(t)$, where $\phi : \mathbb{R} \to \mathbb{R}$ is a 1-periodic continuous function and $\gamma : [0, T] \to [0, R]$ is a continuous and increasing bijection. We say that S is a *reparametrized periodic function*. A way to extract from S information about ϕ or γ is to look at level sets. For example, zero-crossings is the cardinality of $S^{-1}(\{0\})$. It is clear that this number reflects the number of periods of ϕ in S if 0 is in the range of ϕ.

It is also clear that a similar relation generalizes to connected components in the sub level sets, and so, to persistent homology. In contrast to cardinalities of level sets, persistence is stable with respect to perturbations, so we expect to obtain a descriptor with similar properties but more robust and versatile. Describing a signal with a recurrent structure by its persistence diagram has already been used in a medical context, to characterize the gait (Bois et al., 2022).

In this section, we formalize this intuition and we show two additivity properties of persistence modules and diagrams. Specifically, in Section 3.1, we deal with the case when the number of observed periods is exact, in which case the diagram is exactly additive. We will see how to exploit that further in Chapter 5. In Section 3.2, we consider signals with a non-integral number of periods. We show that the diagram satisfies the desired property, up to a remainder which has small persistence. In both cases, we adapt the standard persistence modules slightly.

We also present some perspectives on further results. Notably, generalizations of the properties mentioned above to functions on higher-dimensional domains (Section 3.3.1) and refining stability to perturbations (Section 3.3.2), which leads us to discuss a different potential characterization of total persistence.

3.1 Filtering the circle: no boundary effects

The signal S contains an integral number of periods if $\gamma(T) - \gamma(0) \in \mathbb{N}^*$. In that case, we can see S as defined on the circle $\mathbb{S}^1 \simeq [0, T]/(0 \sim T)$. We will denote by $\bar{\phi} : \mathbb{S}^1 \to \mathbb{R}$ a version of ϕ defined on the circle, via the canonical projection $\mathbb{R} \to \mathbb{S}^1 \simeq \mathbb{R}/\mathbb{Z}$, $t \mapsto (\cos(2\pi t), \sin(2\pi t))$. We denote by $D(\bar{\phi})$ the persistence diagram (in 0-homology) of $\mathbb{S}^1$ filtered by $\bar{\phi}$, where we have set the death of the essential component to $\max \phi$. For a multiset D and $N \in \mathbb{N}$, we define ND to be a multi-set with the same points as D, except with all multiplicities increased by a factor N, $\mathrm{m}_{ND}(b, d) \coloneqq N\mathrm{m}_D(b, d)$.

Proposition 3.1. *For $N \in \mathbb{N}$, consider ϕ_N which realizes N periods of ϕ,*

$$\phi_N : [0, 1] \to \mathbb{R}, \qquad t \mapsto \phi(Nt).$$

If ϕ_1 has a finite number of local extrema, then

$$D(\bar{\phi_N}) = ND(\bar{\phi_1}). \tag{3.1}$$

Proof. Consider $\pi : \mathbb{S}^1 \to \mathbb{S}^1$ defined by $z \mapsto z^N$. It defines an N-cover of $\mathbb{S}^1$ by itself[1] and satisfies $\bar{\phi_N} = \bar{\phi} \circ \pi$. In the proof, we characterize the persistence module $\left(H_0\left((\bar{\phi} \circ \pi)^{-1}(]-\infty, t]\right)\right)_{t \in \mathbb{R}}$, $(\iota_s^{t'})_{s \le t'})$.

[1] For $z \in \mathbb{S}^1$, $\mathrm{card}(\pi^{-1}(z)) = N$.

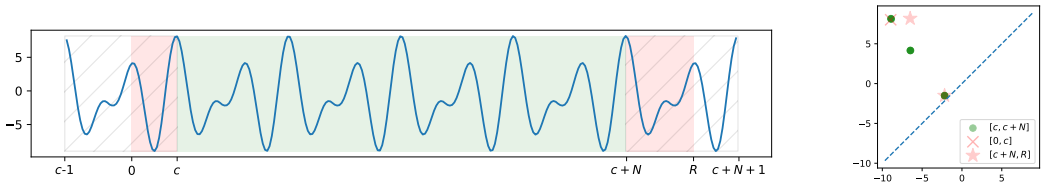


Figure 3.1: A graph of several periods a periodic function. The signal is observed on $[0, R]$, but only $N = 4$ periods are fully included and are marked in green on the graph. The red parts correspond to the remainder. The intervals $[c-1, c]$ and $[R, c+N+1]$ used in the proof of Proposition 3.3 are included as hatched regions.

Let $M = \max(\bar{\phi})$ and $k = \left| \bar{\phi}^{-1}(M) \right|$. Since $\bar{\phi}$ has a finite number of critical points, we can choose a small constant $\epsilon > 0$ such that $\bar{\phi}^{-1}([M - 2\epsilon, M[)$ contains no critical points. The space $\bar{\phi}^{-1}(]-\infty, M - 2\epsilon])$ has k connected components. Because π is an N-cover, we have that $|(\bar{\phi} \circ \pi)^{-1}(M)| = Nk$, so that $(\bar{\phi} \circ \pi)^{-1}(]-\infty, M - 2\epsilon])$ has Nk connected components. Therefore, for any $t < M$,

$$H_0((\bar{\phi} \circ \pi)^{-1}(]-\infty, t])) = \oplus_{n=1}^{N} H_0(\bar{\phi}^{-1}(]-\infty, t])),$$

and the morphisms $(\iota_s^t)_{s \leq t' < M}$ are induced by direct sums. In particular, at the level of persistence measures, it implies that $\mu_{\bar{\phi}\circ\pi}(A) = \sum_{n=1}^{N} \mu_{\bar{\phi}}(A)$, for any rectangle $A \subset \Omega_1 =]-\infty, M[\times]-\infty, M[$.

The measures of $\mu_{\bar{\phi}\circ\pi}$ and $\mu_{\bar{\phi}}$ are supported on $\Omega_1 \sqcup \Omega_2$, where $\Omega_2 = \{(u, M) \mid u \in [-\infty, M[\}$, so it remains to consider $A \subset \Omega_2$. The circle $\mathbb{S}^1 = (\bar{\phi} \circ \pi)^{-1}(]-\infty, M])$ has a single connected component, so that

$$\dim(\ker \iota_{M-\epsilon}^{M}) = Nk - 1.$$

In addition, since the death of the essential component is set to M, at the level of persistence measures, we have $\mu_{\bar{\phi}\circ\pi}(\Omega_2) = (Nk - 1) + 1 = NK$. Since the local minimal values of $\bar{\phi}$ and $\bar{\phi} \circ \pi$ are the same, by the additivity of the persistence measure,

$$\begin{aligned} \mu_{\bar{\phi}\circ\pi}(A) &= \mu_{\bar{\phi}\circ\pi}(A \cap \Omega_1) + \mu_{\bar{\phi}\circ\pi}(A \cap \Omega_2) \\ &= N\mu_{\bar{\phi}}(A \cap \Omega_1) + N\mu_{\bar{\phi}}(A \cap \Omega_2) \\ &= N\mu_{\bar{\phi}}(A). \end{aligned}$$

Because the persistence measures are equal, so are the persistence diagrams, by the characterisation (2.16). □

3.2 Filtering an interval: boundary effects

In this section, we use the terminated persistence module (2.13), because it helps us to systematically deal with the essential generators. Here, $D(\phi)$ will refer to the persistence diagram of the persistence measure $\mu_{\bar{\mathbb{V}}(\phi)}$. That diagram exists, is stable by Proposition 2.38, and Proposition 2.35 also holds. It differs from the diagram of sub level sets of ϕ except that the point $(\min \phi, \infty)$ has now coordinates $(\min \phi, \max \phi)$.

We are now ready to study the structure of the persistence diagram of a periodic function, evaluated on a compact interval $[0, R]$. Consider $\phi : \mathbb{R} \to \mathbb{R}$ a 1-periodic and continuous function. We denote by $\phi|_A$ the restriction of ϕ to $A \subset \mathbb{R}$ and by $D \sqcup D'$ the union of two multisets: for each $(b, d) \in D \cup D'$, $m_{D \sqcup D'}(b, d) = m_D(b, d) + m_{D'}(b, d)$.

Proposition 3.2 (Additivity of diagrams)**.** *For any $R > 1$, there exist persistence diagrams D_1 and D', such that*

$$D(\phi|_{[0,R]}) = (\lfloor R - 1 \rfloor D_1) \sqcup D'. \tag{3.2}$$

The idea of the proof consists in first choosing $c \in [0, 1]$ a global maximum of ϕ and defining "the period" to be $\phi|_{[c,c+1]}$. This allows us to decompose the diagram as a sum of diagrams of individual periods. Thanks to the periodicity of ϕ, these diagrams are the same and we obtain (3.2). Figure 3.1 illustrates the proof.

Proof. Let $M \coloneqq \max \phi$ and $c \coloneqq \inf\{x \in [0,1] \mid \phi(x) = M\}$. Since $\phi^{-1}(M)$ is closed, then of course $\phi(c) = M$. This defines "the start of a period" and also the cutting point. Now, let $N = \max\{n \in \mathbb{N} \mid c + n \leq R\}$, which corresponds to the number of whole periods in the signal.

Let us first fix $t \in \mathbb{R}$ and consider $\bar{V}_t(\phi|_{[0,c]})$, $\bar{V}_t(\phi|_{[c,c+N]})$ and $\bar{V}_t(\phi|_{[c+N,R]})$. If $t < M$, then

$$\phi|_{[0,c]}{}^{-1}(]-\infty,t]) \cap \phi|_{[c,c+N]}{}^{-1}(]-\infty,t]) \subset \{c\}$$

and $\phi(c) = M$, so that intersection is empty. The same holds for $\phi|_{[c+N,R]}$ and $\phi|_{[c,c+N]}$. Therefore, by Proposition 2.27b

$$\begin{aligned}\bar{V}_t(\phi|_{[0,R]}) = H_0(\phi|_{[0,R]}{}^{-1}(]-\infty,t])) \simeq & H_0(\phi|_{[0,c]}{}^{-1}(]-\infty,t])) \oplus H_0(\phi|_{[c,c+N]}{}^{-1}(]-\infty,t])) \\ & \oplus H_0(\phi|_{[c+N,R]}{}^{-1}(]-\infty,t])) \\ = & \bar{V}_t(\phi|_{[0,c]}) \oplus \bar{V}_t(\phi|_{[c,c+N]}) \oplus \bar{V}_t(\phi|_{[c+N,R]}).\end{aligned} \tag{3.3}$$

Since the isomorphism is induced by inclusions, it is an isomorphism between the persistence modules restricted to $t \in]-\infty, M[$. By (2.13), the persistence modules are all 0 for $t \geq M$, so $\bar{V}_t(\phi|_{[0,R]})$ and $\bar{V}_t(\phi|_{[0,c]}) \oplus \bar{V}_t(\phi|_{[c,c+N]}) \oplus \bar{V}_t(\phi|_{[c+N,R]})$ are trivially isomorphic for $t \geq M$ and the morphisms commute. Therefore, the two persistence modules are isomorphic (on $t \in \mathbb{R}$).

By repeating the same argument as above, we can decompose $\bar{\mathbb{V}}(\phi|_{[c,c+N]})$ as the direct sum of $(\bar{\mathbb{V}}(\phi|_{[c+n,c+n+1]}))_{n=0}^{N-1}$. For any $n = 0, \dots, N-1$, $g_n : x \mapsto x + n$ is a homeomorphism between the sub level set of $\phi|_{[c,c+1]}$ and $\phi|_{[c+n,c+n+1]}$, so $\bar{\mathbb{V}}(\phi|_{[c,c+N]}) \simeq \bigoplus_{n=0}^{N-1} \bar{\mathbb{V}}(\phi|_{[c,c+1]})$. Thus, (3.3) becomes

$$\bar{\mathbb{V}}(\phi|_{[0,R]}) \simeq \bar{\mathbb{V}}(\phi|_{[0,c]}) \oplus \bigoplus_{n=0}^{N-1} \bar{\mathbb{V}}(\phi|_{[c,c+1]}) \oplus \bar{\mathbb{V}}(\phi|_{[c+N,R]})$$

The second crucial observation is that the diagram of a direct sum of two persistence modules is the union of diagrams. For decomposable modules, since the decomposition is unique, it is straightforward (Chazal et al., 2016, Proposition 2.16). The persistence modules that we consider are not-necessarily interval decomposable, so the diagrams are defined through persistence measures (2.16). For any two persistence modules $\mathbb{V} = (V_t)_{t\in\mathbb{R}}$, $\mathbb{W} = (W_t)_{t\in\mathbb{R}}$ and any $s, t \in \mathbb{R}$, we have that

$$\operatorname{im}{}_{(\mathbb{V}\oplus\mathbb{W})}\iota_s^t = (\operatorname{im}{}_{\mathbb{V}}\iota_s^t) \oplus (\operatorname{im}{}_{\mathbb{W}}\iota_s^t), \qquad \ker{}_{(\mathbb{V}\oplus\mathbb{W})}\iota_s^t = \ker{}_{\mathbb{V}}\iota_s^t \oplus \ker{}_{\mathbb{W}}\iota_s^t.$$

Hence, the persistence measures $(\mu_{\mathbb{V}} + \mu_{\mathbb{W}})$ and $\mu_{\mathbb{V}\oplus\mathbb{W}}$ are equal, as they are characterized by ranks of certain maps (2.15). Their persistence diagrams, defined from persistence measures (2.16), are therefore equal. If we denote by $D_1 \coloneqq D(\phi|_{[c+n,c+n+1]})$ and by $D' = D(\mu_{\bar{\mathbb{V}}(\phi|_{[c+N,R]})} + \mu_{\bar{\mathbb{V}}(\phi|_{[0,c]})})$ the diagram of the sum of the rectangle measures of the $\phi|_{[0,c]}$ and $\phi|_{[c+N,R]}$, then (3.2) follows. □

From the proof, we immediately see that $D_1 = D(\phi|_{[c,c+1]})$: it is realized as the diagram of "the period" of ϕ. We can further characterize the remainder: we show that it can be truly considered a remainder, as each of the left and right parts have persistence less than that of D_1.

Proposition 3.3. *With the same notation as in Proposition 3.2, for any $p \geq 1$,*

$$\operatorname{pers}_p(D') \leq 2\operatorname{pers}_p(D_1),$$

where $\operatorname{pers}_p(D) = \left(\sum_{(y_1,y_2)\in D}(y_2 - y_1)^p\right)^{1/p}$.

Proof. The period $\phi|_{[0,c]}$ is part of that of $\phi|_{[c-1,c]}$, so, for any $t \in \mathbb{R}$, $\phi|_{[0,c]}{}^{-1}(]-\infty,t]) \subset \phi|_{[c-1,c]}{}^{-1}(]-\infty,t])$ induces a map $\bar{V}_t(\phi|_{[0,c]}) \to \bar{V}_t(\phi|_{[c-1,c]})$. We claim that it is an injective morphism between persistence modules. Hence, $\operatorname{rank}(\bar{V}_s(\phi|_{[0,c]}) \to \bar{V}_t(\phi|_{[0,c]})) \leq \operatorname{rank}(\bar{V}_s(\phi|_{[c-1,c]}) \to \bar{V}_t(\phi|_{[c-1,c]}))$ for any $s < t \in \mathbb{R}$ and both are finite because the modules are q-tame. Hence, to every point $(b,d) \in D(\phi|_{[0,c]})$ with $0 < t - s < d - b$, we can assign a point $(b',d') \in D(\phi|_{[-1+c,c]})$ in such a way that this assignment is injective (considered with multiplicity) and such that $b' \leq b < d \leq d'$. So, $\operatorname{pers}^p_{p,\epsilon}(D(\phi|_{[0,c]})) \leq \operatorname{pers}^p_{p,\epsilon}(D(\phi|_{[-1+c,c]}))$. A similar argument shows that $\operatorname{pers}^p_{p,\epsilon}(D(\phi|_{[c+N,R]})) \leq \operatorname{pers}^p_{p,\epsilon}(D(\phi|_{[c+N,c+N+1]}))$. □

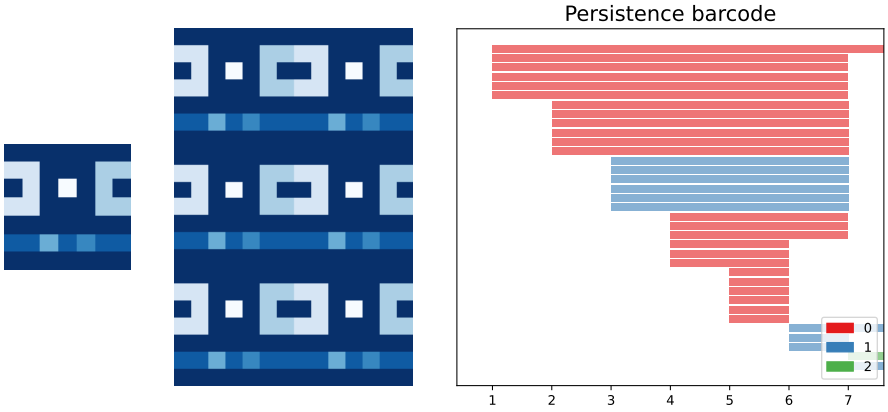


Figure 3.2: On the left, an image representing a filtration on a torus. The lighter colors correspond to smaller values. In the center, an image with $n_1 = 2$ and $n_2 = 3$. On the right, the persistence barcode associated with the image in the center.

3.3 Perspectives

There are several natural questions about extensions of the results presented above. Proposition 3.1 describes the structure of a diagram of $\phi : \mathbb{S}^D \to \mathbb{R}$ for $D = 1$, but can we devise similar properties for $D \geq 2$? Second, the noiseless results trivially extend to the noisy case by bottleneck stability Theorem 2.37. However, bottleneck stability does not reflect that a localized perturbation leaves several periods intact and that the corresponding points might still appear in the diagram. This section presents partial results and observations which are not used in the remaining of this thesis. In particular, its content is not a prerequisite for other sections and will not be referred to, except for Figure 3.3.

3.3.1 Periodic functions of several variables

It is convenient to consider periodic functions in $D \geq 1$ variables as defined on $(\mathbb{S}^1)^D$. For $\phi : (\mathbb{S}^1)^D \to \mathbb{R}$ and $\gamma : (\gamma_d : \mathbb{S}^1 \to \mathbb{S}^1)_{1 \leq d \leq D}$, each a n_d-cover of $\mathbb{S}^1$, is $D(\phi \circ \gamma)$ a $(\Pi_{d=1}^D n_d)$-multiple of $D(\phi)$? We start the section by the analysis of an example for $D = 2$, which shows that the answer to the above is negative[2]. Then, we point to the Mayer-Vietoris sequence in persistent homology, which we believe to be the right tool to study the problem further. Finally, we provide a characterization for functions with a particular structure.

Example 3.4. Consider the function whose representation is shown in the left panel of Figure 3.2. The central panel shows a function which features $n_1 = 2$ horizontal and $n_2 = 3$ vertical repetitions. We can read from the barcode that all the bars have multiplicity $n_1 n_2 = 6$, until $t = 6$, after which $n_2 = 3$ additionnal bars in dimension 1 appear. Those bars correspond to the 3 dark-blue, horizontal lines, which create cycles (due to periodic boundary conditions). The multiplicity of these bars reflect only n_2, but not n_1. Crucially, one of those 3 generators is one of the two generators of $H_1(\mathbb{S}^1 \times \mathbb{S}^1)$ that appear as the semi-finite bars.

Conjecture. Let $f(x, y) = \phi(\gamma_1(x), \gamma_2(y))$ and denote by $\mathbb{U}^n = (H_n(f^{-1}(]-\infty, t])))_{t \in \mathbb{R}}$ and $\mathbb{V}^n = (H_n(\phi^{-1}(]-\infty, t])))_{t \in \mathbb{R}}$ the singular persistence homology modules for $n = 0, 1$. If we let $t_1 \coloneqq \min_s \operatorname{rank}(\mathbb{V}(\phi)_s \to \mathbb{V}(\phi)_{\max \phi + 1}) \geq 1$, then for any $s \leq t < t_1$, we have

$$\operatorname{rank}(\mathbb{U}^n_s \to \mathbb{U}^n_t) = n_1 n_2 \operatorname{rank}(\mathbb{V}^n_s \to \mathbb{V}^n_t).$$

The value t_1 is the birth of the first essential generator of $\mathbb{S}^1 \times \mathbb{S}^1$. Based on Example 3.4, we believe that the additive property expressed above holds until t_1. In that case, a reasoning analogue to that for $D = 1$ should be valid for $n = 0$. We believe that a proof could be established

[2] We use cubical complexes and cubical homology for illustration purposes; see Kaczynski et al. (2011, Section 2) for an introduction.

for $n = 1$ using the Mayer-Vietoris sequence for persistent homology and Di Fabio and Landi (2011, Theorem 3.7). Characterising the module beyond t_1 seems more complex.

We give a more complete characterization in case ϕ has more structure. Consider two continuous functions $\phi_1, \phi_2 : \mathbb{S}^1 \to \mathbb{R}$ and let

$$\begin{array}{rccl} \phi: & \mathbb{S}^1 \times \mathbb{S}^1 & \to & \mathbb{R} \\ & (x, y) & \mapsto & \max(\phi_1(x), \phi_2(y)). \end{array} \tag{3.4}$$

We will also need to consider $f_d \coloneqq \phi_d \circ \gamma_d$ and the modules $\mathbb{V}^{n,d}$, $\mathbb{W}^{n,d}$ for f_d and ϕ_d respectively.

Proposition 3.5. *Assume that* $\max \phi_1 \leq \max \phi_2$. *Then,*

$$\operatorname{rank}(\mathbb{U}_s^0 \to \mathbb{U}_t^0) = \begin{cases} n_1 n_2 \operatorname{rank}(\mathbb{W}_s^{0,1} \to \mathbb{W}_t^{0,1}) \operatorname{rank}(\mathbb{W}_s^{0,2} \to \mathbb{W}_t^{0,2}) & \text{if } t < \max \phi_1 \\ n_2 \operatorname{rank}(\mathbb{W}_s^{0,2} \to \mathbb{W}_t^{0,2}) & \text{if } \max \phi_1 \leq t < \max \phi_2 \\ 1 & \text{if } \max \phi_2 \leq t, \end{cases}$$

$$\operatorname{rank}(\mathbb{U}_s^1 \to \mathbb{U}_t^1) = \begin{cases} 0 & \text{if } s < \max \phi_1 \\ n_2 \operatorname{rank}(\mathbb{W}_s^{0,2} \to \mathbb{W}_t^{0,2}) & \text{if } \max \phi_1 \leq s \leq t < \max \phi_2 \\ 1 & \text{if } \max \phi_1 \leq s < \max \phi_2 \leq t \\ 2 & \text{if } \max \phi_2 \leq s < t. \end{cases}$$

$$\operatorname{rank}(\mathbb{U}_s^2 \to \mathbb{U}_t^2) = \begin{cases} 1 & \text{if } \max \phi_2 \leq s, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. The sub level sets of f are the product of those of f_1 and f_2: for any $t \in \mathbb{R}$,

$$(x, y) \in f^{-1}(]-\infty, t) \iff \max(f_1(x), f_2(y)) \leq t \iff (f_1(x) \leq t) \text{ and } (f_2(y) \leq t).$$

Then, the Künneth formula for persistent homology (Gakhar and Perea, 2019, Proposition 4.3) shows that for any $s, t \in \mathbb{R}$,

$$\operatorname{rank}(\mathbb{U}_s^n \to \mathbb{U}_t^n) = \sum_{i+j=n} \operatorname{rank}(\mathbb{V}_s^{n,1} \to \mathbb{V}_t^{n,1}) \operatorname{rank}(\mathbb{V}_s^{n,2} \to \mathbb{V}_t^{n,2}).$$

For $n = 0$, we have simply

$$\operatorname{rank}(\mathbb{U}_s^0 \to \mathbb{U}_t^0) = \operatorname{rank}(\mathbb{V}_s^{0,1} \to \mathbb{V}_t^{0,1}) \operatorname{rank}(\mathbb{V}_s^{0,2} \to \mathbb{V}_t^{0,2}).$$

Fix $d = 1, 2$. For $t \leq \max_{x \in \mathbb{S}^1} \phi_d(x)$, Proposition 3.1 implies that

$$\operatorname{rank}(\mathbb{V}_s^{0,d} \to \mathbb{V}_t^{0,d}) = n_d \operatorname{rank}(\mathbb{W}_s^{0,d} \to \mathbb{W}_t^{0,d}).$$

Hence, for $t < \min_{d=1,2} \max_{x \in \mathbb{S}^1} \phi_d(x)$, we recover an additive property

$$\operatorname{rank}(\mathbb{U}_s^0 \to \mathbb{U}_t^0) = n_1 n_2 \operatorname{rank}(\mathbb{W}_s^{0,1} \to \mathbb{W}_t^{0,1}) \operatorname{rank}(\mathbb{W}_s^{0,2} \to \mathbb{W}_t^{0,2}).$$

For $\max_{x \in \mathbb{S}^1} \phi_1(x) \leq t < \max_{x \in \mathbb{S}^1} \phi_2(x)$, we have $\operatorname{rank}(\mathbb{V}_s^{0,1} \to \mathbb{V}_t^{0,1}) = 1_{[\min \phi_1, \infty[}(s) = \operatorname{rank}(\mathbb{W}_s^{0,1} \to \mathbb{W}_t^{0,1})$, so

$$\operatorname{rank}(\mathbb{U}_s^0 \to \mathbb{U}_t^0) = n_2 \operatorname{rank}(\mathbb{W}_s^{0,1} \to \mathbb{W}_t^{0,1}) \operatorname{rank}(\mathbb{W}_s^{0,2} \to \mathbb{W}_t^{0,2}).$$

Finally, for $\max_{x \in \mathbb{S}^1} \phi_2(x) \leq t$, $\operatorname{rank}(\mathbb{U}_s^0 \to \mathbb{U}_t^0) = 1$.

Consider higher homology dimensions. Recall that $\operatorname{rank}(\mathbb{W}_s^{1,d} \to \mathbb{W}_t^{1,d}) = 1_{[\max \phi_d, \infty[}(s) = \operatorname{rank}(\mathbb{V}_s^{1,d} \to \mathbb{V}_t^{1,d})$. Assuming $\max_{x \in \mathbb{S}^1} \phi_1(x) < \max_{x \in \mathbb{S}^1} \phi_2(x)$, the Künneth formula implies that for $n = 1$,

$$\begin{aligned} \operatorname{rank}(\mathbb{U}_s^1 \to \mathbb{U}_t^1) &= \operatorname{rank}(\mathbb{V}_s^{1,1} \to \mathbb{V}_t^{1,1}) \operatorname{rank}(\mathbb{V}_s^{0,2} \to \mathbb{V}_t^{0,2}) + \operatorname{rank}(\mathbb{V}_s^{0,1} \to \mathbb{V}_t^{0,1}) \operatorname{rank}(\mathbb{V}_s^{1,2} \to \mathbb{V}_t^{1,2}) \\ &= 1_{[\max \phi_1, \infty[}(s) \operatorname{rank}(\mathbb{V}_s^{0,2} \to \mathbb{V}_t^{0,2}) + \operatorname{rank}(\mathbb{V}_s^{0,1} \to \mathbb{V}_t^{0,1}) 1_{[\max \phi_2, \infty[}(s) \\ &= 1_{[\max \phi_2, \infty[}(s) + \left(n_2 \operatorname{rank}(\mathbb{W}_s^{0,2} \to \mathbb{W}_t^{0,2}) 1_{]-\infty, \max \phi_2[}(t) + 1_{[\max \phi_2, \infty[}(t) \right) 1_{[\max \phi_1, \infty[}(s) \\ &= \begin{cases} 0 & \text{if } s < \max \phi_1 \\ n_2 \operatorname{rank}(\mathbb{W}_s^{0,2} \to \mathbb{W}_t^{0,2}) & \text{if } \max \phi_1 \leq s \leq t < \max \phi_2 \\ 1 & \text{if } \max \phi_1 \leq s < \max \phi_2 \leq t \\ 2 & \text{if } \max \phi_2 \leq s < t. \end{cases} \end{aligned}$$

For $n = 2$, we have $\operatorname{rank}(\mathbb{V}_s^{1,2} \to \mathbb{V}_t^{1,2}) = 0 = \operatorname{rank}(\mathbb{V}_s^{2,2} \to \mathbb{V}_t^{2,2})$ for any $s < t$, so

$$\begin{aligned}\operatorname{rank}(\mathbb{U}_s^2 \to \mathbb{U}_t^2) =& \operatorname{rank}(\mathbb{V}_s^{0,1} \to \mathbb{V}_t^{0,1})\operatorname{rank}(\mathbb{V}_s^{2,2} \to \mathbb{V}_t^{2,2}) + \operatorname{rank}(\mathbb{V}_s^{2,1} \to \mathbb{V}_t^{2,1})\operatorname{rank}(\mathbb{V}_s^{0,2} \to \mathbb{V}_t^{0,2}) \\ &+ \operatorname{rank}(\mathbb{V}_s^{1,1} \to \mathbb{V}_t^{1,1})\operatorname{rank}(\mathbb{V}_s^{1,2} \to \mathbb{V}_t^{1,2}) \\ =& 1_{[\max\phi_1,\infty[}(s) 1_{[\max\phi_2,\infty[}(s) \\ =& 1_{[\max\phi_2,\infty[}(s).\end{aligned}$$

□

This simple case shows that the vertical and horizontal multiplicities appear, but at times which are characteristic to functions f_1 and f_2: there might exist bars with multiplicity $n_1 n_2$ if $[\min\phi_1, \max\phi_1] \cap [\min\phi_2, \max\phi_2]$ has non-empty interior. However, if $\max\phi_1 \leq \min\phi_2$, the function f is degenerate in that $f(x, y) = f_2(y)$.

Proposition 3.5 shows that for functions of the form (3.4), homology in dimension $n = 1$ is sensitive to multiplicity only in one of the directions. In particular, the image in Figure 3.2 has $n_1 n_2$ bars $[3, 7[$, which shows that it cannot be of the form (3.4).

3.3.2 Persistent homology of noisy signals: an alternative characterization of total persistence

We numerically illustrate how persistence diagrams of sub level sets change when the periodic signal is perturbed in different ways. One of those observations leads us to conjecture a characterization of total persistence, which might provide an alternative bound to that of Proposition 4.13.

We illustrate how different types of noise change the structure of the persistence diagram. The first row in Figure 3.3 shows ϕ with Gaussian noise with square-exponential covariance with time scale 0.15 and standard deviation 0.1. The noise has a higher frequency than the oscillations in ϕ, what produces a persistence diagram with points clustered around those of $D(\phi)$. In the second and third rows, we show ϕ corrupted by punctual noise. The comparison between the second and third diagrams shows that the location of the perturbation matters little, confirming the 'global' character of persistent homology. The interesting observation is that such a 'sparse' perturbation produces outliers in the persistence diagram, but numerous points corresponding to the periodicity of ϕ are left un-perturbed. While the amplitude of the noise is high, the Wasserstein distance (2.19) between the diagrams is small. The fourth figure shows ϕ superimposed with a drift. The drift is slow enough for local minima of ϕ to appear, and in the persistence diagram, almost all points lie on segments parallel to the diagonal. The exception is the pair of global extrema, which is far from all points. Finally, we picture an amplitude-modulated signal. Points in the persistence diagram lie on segments, whose extension intersects the diagonal at $(0, 0)$.

Remark 3.6. Amplitude modulation occurs as a consequence of frequency modulation and a fixed sampling frequency. Because the extrema are never attained when sampling, the signal is attenuated: for a signal with a highest frequency component $f_{\max}$ sampled at rate f_s, the attenuation is of the order $\cos(f_{\max}/f_s)$. Hence, while the persistence diagram of a function is invariant to reparametrisation, it is not true for the diagram of its sampled version. We quantify that distance in Proposition 5.12.

The observations from the localized perturbation scenario remind of the sought-after Wasserstein stability of persistence diagrams. Here, $\|W\|_{L^1}$ is much smaller than $\|W\|_\infty$, leading to a poor bottleneck stability bound. The total p-persistence is the p-Wasserstein distance to the empty diagram. We now present a different notion of "variation", which we believe is related to the total p-persistence: an observation that we do not prove.

Definition 3.7. Let $a \leq b \in \mathbb{R}$ and $a \leq a_k \leq b_k \leq b \in \mathbb{R}$, for $k = 1, \ldots, K \in \mathbb{N}$. A collection of tuples $\mathcal{I} = ((a_k, b_k))_{k=1}^K$ is a *nested sequence for* $[a, b]$ if for any k, k', intervals $]a_k, b_k[$ and $]a_{k'}, b_{k'}[\subset [a, b]$ are disjoint or nested (one is included in the other).

In the standard notions of p-variation, the considered sequences are partitions, whereas here, the intervals can be nested. To ensure that the variation does not increase as we keep adding nesting elements, we orient the intervals.

Definition 3.8. An oriented pair (a, b, v) is a pair (a, b) with an orientation $v \in \{0, 1\}$. We call a pair $(a_{k'}, b_{k'})$ *a maximal element of* (a_k, b_k) if it is maximal with respect to the order in $\mathcal{I}$. A

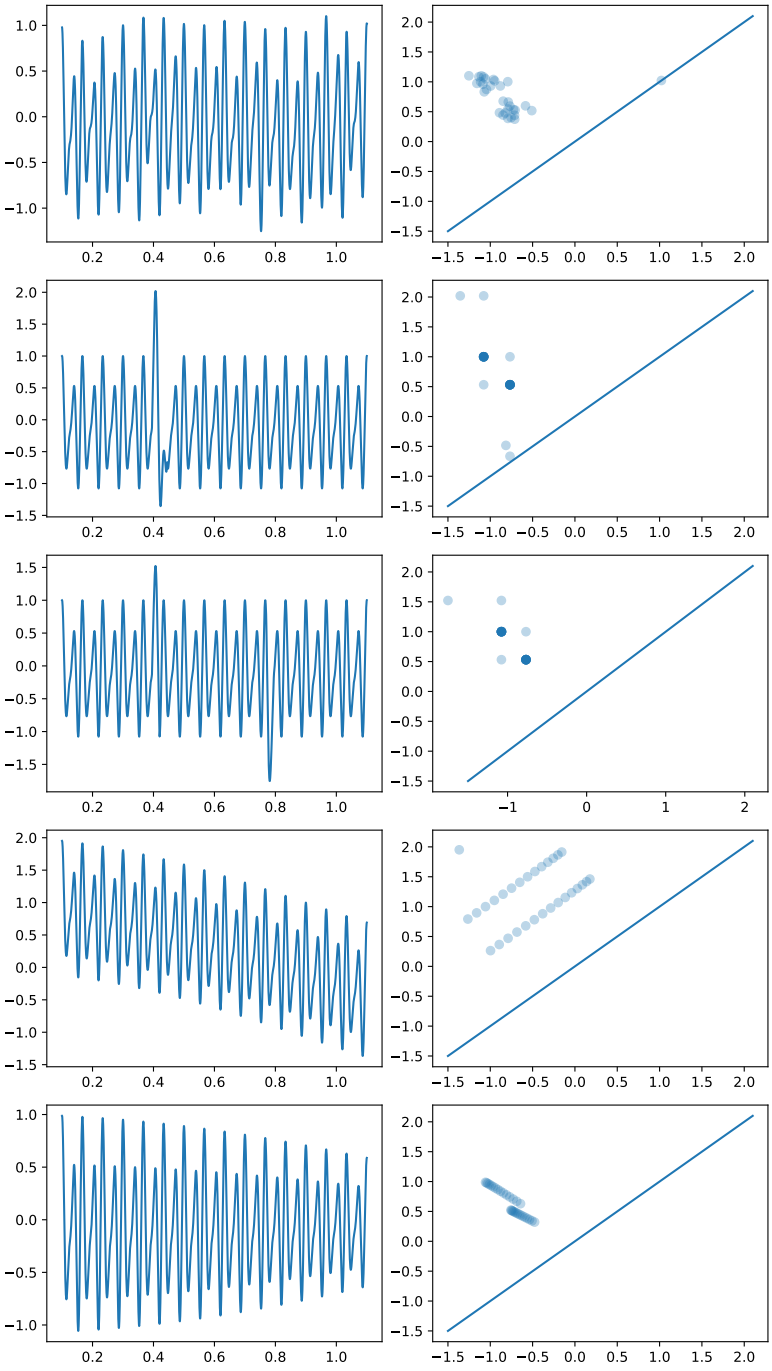


Figure 3.3: On the left, a periodic function, perturbed using several kinds of noise. On the right, the corresponding persistence diagrams.

nested, alternating sequence for $[a, b]$ is a nested sequence of oriented pairs, $\bar{\mathcal{I}} = ((a_k, b_k, v_k))_{k=1}^K$, such that if $(a_{k'}, b_{k'})$ is a maximal element of (a_k, b_k), then $v_{k'} = 1 - v_k$.

Example 3.9. Consider the sequence $\bar{\mathcal{I}} = ((1, 8, 0), (2, 3, 1), (4, 7, 1), (5, 6, 0))$ for the interval $[1, 8]$. It is a nested sequence, because $]5, 6[\subset]4, 7[$ and $]2, 3[, \ ]4, 7[\subset]1, 8[$. It is also alternating: pairs $(2, 3)$ and $(4, 7)$ are maximal elements of $(1, 8)$, and they have orientations opposite to that last.

Definition 3.10. Let $S : [a, b] \to \mathbb{R}$ be a continuous function and $\bar{\mathcal{I}}$ a nested alternating sequence for $[a, b]$. For $p \in \mathbb{N}^*$, the *p-variation of* S *associated to* $\bar{\mathcal{I}}$ is

$$V(S, \bar{\mathcal{I}}) \coloneqq \sum_{k=0}^{K} w(S(b_k) - S(a_k), v_k), \qquad \text{where } w(x, v) \coloneqq |x|^p (-1)^v (-1)^{x<0},$$

and the *p-variation of* S is

$$V(S) \coloneqq \max_{\bar{\mathcal{I}}} V(S, \bar{\mathcal{I}}),$$

where the maximum is taken over all finite, nested and alternating sequences $\bar{\mathcal{I}}$.

Example 3.11. Consider $S : [1, 8] \to \mathbb{R}$ a continuous, piecewise linear function, defined by prescribing $(S(k))_{k=1}^8 = (0, 4, 1, 6, 3, 5, 2, 7)$, and the nested alternating sequence shown at the bottom of Figure 3.4. Its p-variation is

$$V(S, ((1, 7, 1), (3, 4, 0), (5, 6, 0))) = w(-2, 1) + w(5, 0) + w(2, 0) = 2^p + 2^p + 5^p.$$

Notice that the p-variation of a nested alternating sequence is not necessarily positive, as $w(x, 0)$ has the same sign as x. The p-variation of S defined above is certainly not the same as the notions of p-variation from probability or the one studied in persistence theory. If we denote that last by p-var, then Perez (2022b, Proposition 3.6) states that p-var$(S) \leq 2\text{pers}_p^p(S)$. This shows that persistence is 'more-efficient' at capturing variation than p-var[3]. The nested alternating sequences are more flexible than p-var, so we believe that the following relation might hold.

Conjecture. Let $S : [0, T] \to \mathbb{R}$ be a continuous function, such that $S(0) = \min S$ and $S(T) = \max S$. If S has a finite number of critical points, then $\text{pers}_p^p(S) = V(S)$.

Example 3.12 motivates that statement. The pairing of critical points from the persistence module gives a nested alternating sequence. However, it is not immediately clear to us why that particular sequence maximizes the p-variation.

Example 3.12. Let S be as in Example 3.11, illustrated on Figure 3.4. The persistence diagram defined in Section 2.5.4 in the simplicial setting is $D(S) = \{(1, 4), (3, 5), (2, 6), (0, 7)\}$. We notice that the same pairs are realized by the sequence $\bar{\mathcal{I}}$ from Example 3.9, so

$$V(S, \bar{\mathcal{I}}) = w(7, 0) + w(-3, 1) + w(-4, 1) + w(2, 0) = 7^p + 3^p + 4^p + 2^p = \text{pers}_p^p(S).$$

Note that p-var$(S) = 4^p + 3^p + 5^p + 3^p + 2^p + 3^p + 5^p \leq 2 \cdot (7^p + 3^p) + 5^p < 2\text{pers}_p^p(S)$.

The nested alternating sequences introduced above are similar to "windows with wave", presented in Sebastiano Cultrera di Montesano et al. (2023), Biswas et al. (2023). There are two main differences. The windows are two-dimensional, $[a_k, b_k] \times [\min S|_{[a_k, b_k]}, \max S|_{[a_k, b_k]}]$, with the second coordinate taking into account the values of the function on $[a_k, b_k]$. The second difference is that the windows are partitioned into 3 sub-windows, which allow to characterize their mutual relations and nesting. In particular, the set of windows for piecewise linear functions is shown to contain information to recover the diagrams. The framework of windows and windows with waves might be adapted to showing the conjecture, and also allow to remove the assumption on the maxima of S being attained on the boundaries of the interval.

The fact that such a relation might hold came to us when working on a lower bound for $\text{pers}_p^p(S)$, alternative to Proposition 4.13. In that lower–bound, we rely on the bottleneck stability of the persistence diagram, which is most pessimistic. In particular, in the case of high amplitude and localized noise (see bottom graphs in Figure 3.3 for an example), the bound is trivial. If the characterization conjectured above was true in a more general form, we hoped to lower bound $\text{pers}_p^p(S)$ with a term that would feature $\text{pers}_p^p(W)$ instead.

[3]The factor 2 is natural, as the variation is counted 'once' in total persistence, but twice in p-var.

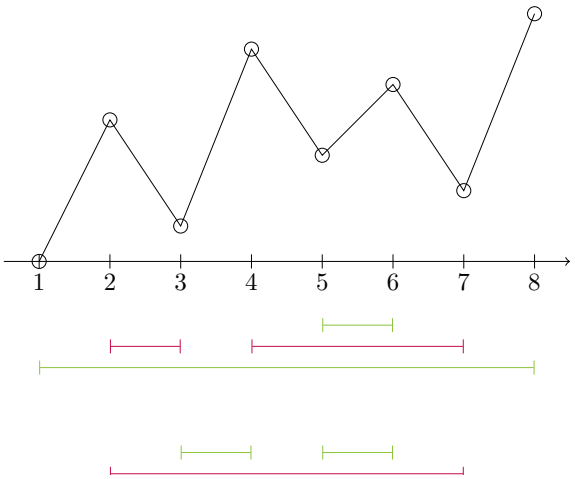


Figure 3.4: A graphical representation of S from Example 3.12, and of two alternating sequences. Green and red segments represent orientations 0 and 1 respectively.

Chapter 4

Topological signatures

In Chapter 3, we have established that the persistence diagram of a periodic signal with phase variation contains repeated points. The coordinates of those points are certain local extrema of the signal and the number of repetitions is proportional to the number of observed periods. In the present chapter, we leverage the repetitive character (Proposition 3.2) and stability (Proposition 2.38) of persistence diagrams to propose a descriptor of a periodic function, calculated on a noisy observation of several periods. Specifically, we assume that ϕ is 1-periodic and we let $\gamma : [0, T] \to [0, R]$ be an increasing bijection, $W : [0, 1] \to \mathbb{R}$ a continuous noise process. We consider an observation S of the form

$$S : [0, T] \to \mathbb{R}, \qquad t \mapsto (\phi \circ \gamma)(t) + W(t). \tag{4.1}$$

Our aim is to construct a signature $F : S \mapsto F(S)$ which contains information about ϕ while remaining somehow robust to W and to changes in γ, both assumed random. The signature we propose is an average of topological features of S, putting weight on repeated and persistent pairs of local extrema.

Related literature

Time series or functional observations of the form (4.1) appear in many applications, where ϕ is somehow characteristic of a population: child growth dynamics (Ramsay and Silverman, 2002), physiological signals (Goldberger et al., 2000), bird migration curves (Su et al., 2014). The reparametrisation γ is the main source of variability in the pointwise evaluations of the signals, as in the 'phase variation' model in Functional data analysis (FDA), see Marron et al. (2015) for a review. The problems typically considered in FDA consist in aligning a population of curves or computing a representative curve, for which methods with guarantees have been proposed (Gasser and Wang, 1997, Khorram et al., 2019, Tang and Muller, 2008). Underlying most of the models is the assumption that the start and end points ($\gamma(0)$ and $\gamma(T)$ here) are common for all curves.

In gait analysis (Bois et al., 2022) or in the positioning problem motivating this thesis, a single observation is composed of several periods of ϕ and the number of periods varies across observations. In the latter, there is little reason for two observations to have the same number of periods, unless the initial angular position of the wheel and the trajectory are exactly the same across those two observations. Therefore, in contrast with FDA, the assumption of common endpoints is not satisfied. In fact, the problem changes from describing the whole signal, to that of describing its constituent parts, that is, the periods of ϕ.

Techniques from topological data analysis (TDA) have been increasingly used to extract geometric or topological information from observations (Chazal and Michel, 2021). The arguably most popular TDA technique for analyzing a time series consists in computing the homology of the sliding-window embedding (SWE) of the time series, and indicates whether the underlying phenomenon is periodic or not (Perea, 2019). It has been used in many applications (Ghil and Sciamarella, 2023, Gidea and Katz, 2018, Fernández and Mateos, 2022). The SWE of a time series $(S_n)_{n=1}^N$ is a point cloud in $\mathbb{R}^d$, where each point is of the form $(S_n, S_{n+\tau}, \ldots, S_{n+(d-1)\tau})$ for parameters $d, \tau \in \mathbb{N}$. If S is periodic, a simplicial complex constructed on the SWE at the right scale will have a non-trivial homology group in dimension one, as illustrated in the top row in Figure 1.4. In signals with phase variation however, the length of the periodic structure changes and

so does the geometry of the SWE, as shown in the bottom row in Figure 1.4. This is corroborated by the fact that the geometry of the delay embedding contains information about the frequencies supporting the signal (Perea, 2019, section 5).

Techniques other than the SWE have been proposed to extract topological information from time series, but they either already assume that the signal is of a specific form (Khasawneh and Munch, 2016), that it is periodic (Erden and Cetin, 2017), or are not concerned with extracting a signature thereof (Kennedy et al., 2018).

Building on the properties of persistent homology of periodic function proved in Chapter 3, we propose to use the persistent homology of sublevel sets of the signal to describe this last. This descriptor summarizes the height, order and number of local extrema. The idea of quantifying the shape of the curve is not new: for example, the landmark method extracts visual features like local extrema or inflection points (Perng et al., 2000).

In many statistical applications, it is convenient to map a persistence diagram to a vector or a function, via a functional representation (Chazal and Michel, 2021). Numerous functionals (Carrière et al., 2020, Adams et al., 2017) are 'linear in the diagram' and their properties have been well-studied (Divol and Polonik, 2019). In our case, it seems natural to renormalize the functionals by the total persistence of the diagram, a proxy for the number of periods. Building on Divol and Polonik (2019) and a recent characterization of the stability of total persistence for Hölder regular processes (Perez, 2022a), we study the robustness of the signatures we propose.

Guarantees on the estimation of functionals of persistence diagrams, in both asymptotic and non-asymptotic cases, have been provided in Chazal et al. (2014), Berry et al. (2020), under the assumption that the persistence diagrams (or functionals thereof) in the collection are all independent. In a setting motivated by the industrial context of this thesis, we have a single time series of which we would like to estimate the signature. The natural procedure is to construct a sample by taking contiguous vectors from that observation, what leads to a collection of shorter and dependent observations. We study two reparametrisation models inspired by Marron et al. (2015) and, building on the theory of strong mixing (Doukhan, 1995, Dedecker et al., 2007), we show that the dependence between observations decreases. When the β-mixing coefficients decrease sufficiently fast, the estimators of the functionals also converge in the dependent setting (Radulović, 1996, Bühlmann, 1995, Kosorok, 2008), not unlike in the independent setting (Chazal et al., 2014). So far, estimation of topological signatures from dependent data has been less explored: Krebs (2021) gives a concentration inequality for persistent Betti numbers from dependent data.

Contributions and structure

In Section 4.1, we concisely introduce the signature and show its key invariance properties. In Section 4.2, we introduce models for reparametrisations and we discuss the guarantees of estimation of the signatures defined for time-series. Section 4.3 contains additional results on persistent homology, essential for the previous sections. The main contributions can be summarized as the following.

1. We demonstrate that the signature converges as the number of observed periods grows, in case there is no additive noise (Theorem 4.2). We rely on Propositions 3.2 and 3.3, which characterize the persistence diagram of sublevel sets of several periods of a function.

2. We show that the signature is invariant under changes of the distribution of γ for fixed endpoints $\gamma(0)$ and $\gamma(T)$ (Theorem 4.3). Recent results on regularity of total persistence allow us to obtain quantitative stability bounds.

3. We provide a technique to estimate the signature from a single time–series observation (Theorem 4.7). We show that the moving block bootstrap technique allows us to construct confidence intervals around the the signature, for two reparametrisation models.

Finally, Section 4.4 provides a simple numerical illustration of the signature and its invariance properties.

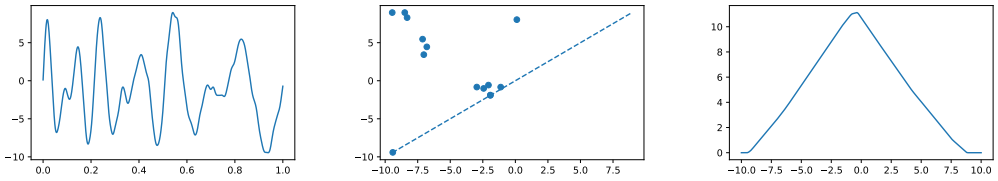


Figure 4.1: An example of a noisy observation of a reparametrised periodic function (left), its persistence diagram (center) and a functional summary (right): the persistence silhouette.

4.1 Signatures of reparametrized periodic functions and their properties

The signatures we propose are functions constructed using the local minima and maxima of the signal. We define those signatures with persistent homology of the sublevel sets of the signal and its functional representations. For the sake of readability, we defer the justification of some properties of persistence diagrams and the corresponding functionals to Sections 4.3 and 4.3.2. We include an illustration of these concepts in Figure 4.1.

4.1.1 Normalized functionals of truncated persistence

Let $D(S)$ denote the persistence diagram of the truncated module (2.13) of $S \in C([0,T],\mathbb{R})$. It is a multiset of points in $\mathbb{R}^2$, where the coordinates are the values of local extrema of S. Each point can be interpreted as a local minimum paired with a local maximum. That pairing is constructed by tracking the evolution of connected components in sublevel sets $S^{-1}(]-\infty,t])$ as t changes. We use the terminated persistence modules from Section 2.13. The pairing is consistent in that if S is reparametrised, the persistence diagram remains unchanged, see Proposition 2.35. In addition, if S is periodic, the multiplicity of any point in the diagram reflects the number of periods in S, except for some extra points due to incomplete periods close to the lower- and upper- endpoints of the domain $[0,T]$, see Proposition 3.2. Therefore, for a periodic function ϕ, the only component of the parametrisation that the persistence diagram depends on is the starting point, $\gamma(0)$, and the number of observed periods, $\gamma(T)-\gamma(0)$.

To gain algebraic and statistical properties, it is often convenient to map the diagram to a functional representation, as motivated in Section 2.5.7. In such a representation, we typically associate to each point (y_1,y_2) from the persistence diagram a function, $k_{y_1,y_2}:\mathbb{T}\to\mathbb{R}$, for some metric space $\mathbb{T}$. A functional representation is then a weighted sum of such functions, where the weights are commensurate with a measure of importance of each point.

Definition 4.1. The *ϵ-truncated persistence* of a point $(y_1,y_2)\in\mathbb{R}^2$ is $w_\epsilon(y_1,y_2)=\max(y_2-y_1-\epsilon,0)$ for some $\epsilon>0$. The *normalized functional of ϵ-truncated p-persistence*, for some $p>1$ and for any $t\in\mathbb{T}$, is

$$\bar{\rho}_{k,\epsilon,p}(S)(t)=\frac{\sum_{(y_1,y_2)\in D(S)} w_\epsilon(y_1,y_2)^p k_{y_1,y_2}(t)}{\sum_{(y_1,y_2)\in D(S)} w_\epsilon(y_1,y_2)^p}, \tag{4.2}$$

if the denominator is positive and $\bar{\rho}_{k,\epsilon,p}(S)(t)=0$ otherwise. We omit the dependence of $\bar{\rho}$ on k,ϵ,p, writing $\bar{\rho}=\bar{\rho}_{k,\epsilon,p}$.

We defer the proof of the fact that $\bar{\rho}$ is well-defined and we now introduce the topological signature. When $\gamma\sim\mu$ and $W\sim\nu$ are independent random variables, S is also random. For each path and $t\in\mathbb{T}$, we can calculate $\bar{\rho}(S)(t)\in\mathbb{R}$. We define the *signature of S* point-wise as

$$F(S)(t) := \mathbb{E}[\bar{\rho}(S)(t)], \tag{4.3}$$

where the expectation is taken with respect to the law of the process, induced by the product measure of μ and ν.

It is clear that $\bar{\rho}(S)(t)$ is a real-valued random variable, but it is less clear whether $\bar{\rho}(S)\in C(\mathbb{T},\mathbb{R})$ can be seen as a random variable: is it $\mathcal{B}(\|\cdot\|_\infty)$-measurable? Appendix 4.A and results from Section 2.1 give a positive answer. We will see in Sections 4.1.2 and 4.1.3 that the normalization of the functional makes it robust to the number of periods of ϕ in S to a certain extent. The

details of the construction of the persistence diagram, examples of $\bar{\rho}$ and a study of the properties of the truncated persistence $\mathrm{pers}_{p,\epsilon}(D) \coloneqq \left(\sum_{(y_1,y_2)\in D} w_\epsilon(y_1,y_2)^p\right)^{1/p}$ are included in Section 4.3.

4.1.2 Properties of functionals of a periodic function

We examine the consistency of the signature (4.3) and its invariance with respect to reparametrisations for noiseless observations. That is, we consider the case $W = 0$, so that (4.1) becomes $S(t) = \phi(\gamma(t))$. Recall that $\gamma : [0,T] \to \mathbb{R}$ is a continuous and increasing function.

For consistency, normalizing the functional by the total truncated p-persistence is akin to normalizing by the number of periods. As $\gamma(T) - \gamma(0)$ increases, the contribution of the boundary effects becomes less significant and we gain invariance to the number of observed periods. Theorem 4.2 is in fact a corollary of Proposition 3.2. It also justifies calling the limit the "signature of a periodic function".

Theorem 4.2 (Consistency)**.** *Assume that k satisfies* (4.16) *and* (4.17)*. Then, as $R \to \infty$,*

$$\bar{\rho}_{\epsilon,p,k}(D(\phi|_{[0,R]})) \xrightarrow{\|\cdot\|_\infty} \bar{\rho}_{\epsilon,p,k}(D(\phi|_{[c,c+1]})).$$

Proof. We have fixed ϵ, p and k and we will denote $\bar{\rho}_{\epsilon,p,k}$ by $\bar{\rho}$. Let $D_1 = D(\phi|_{[c,c+1]})$, D' be given by Proposition 3.2 and let $D_R = D(\phi|_{[0,R]})$. In addition, we will write $\rho(D) = \rho_{k,\epsilon,p}(D) = \sum_{x\in D} w_\epsilon(x)^p k_x(t)$ for the linear (non-normalized) version of the functional $\bar{\rho}$. Then, for any $t \in \mathbb{T}$,

$$\left|\tfrac{\rho((R-1)D_1)+\rho(D')}{\mathrm{pers}^p_{p,\epsilon}(D_R)} - \tfrac{\rho(D_1)}{\mathrm{pers}^p_{p,\epsilon}(D_1)}\right| \le \left|\tfrac{\rho(D')}{\mathrm{pers}^p_{p,\epsilon}(D_R)}\right| + \left|\tfrac{\rho((R-1)D_1)}{\mathrm{pers}^p_{p,\epsilon}(D_R)} - \tfrac{\rho(D_1)}{\mathrm{pers}^p_{p,\epsilon}(D_1)}\right|,$$

and

$$\begin{aligned}\left|\tfrac{\rho((R-1)D_1)}{\mathrm{pers}^p_{p,\epsilon}(D_R)} - \tfrac{\rho(D_1)}{\mathrm{pers}^p_{p,\epsilon}(D_1)}\right| &\le \left|\tfrac{\mathrm{pers}^p_{p,\epsilon}(D_1)\rho((R-1)D_1) - (\mathrm{pers}^p_{p,\epsilon}((R-1)D_1)+\mathrm{pers}^p_{p,\epsilon}(D'))\rho(D_1)}{\mathrm{pers}^p_{p,\epsilon}(D_R)\mathrm{pers}^p_{p,\epsilon}(D_1)}\right| \\ &\le \tfrac{|\mathrm{pers}^p_{p,\epsilon}(D')\rho(D_1)|}{\mathrm{pers}^p_{p,\epsilon}(D_R)\mathrm{pers}^p_{p,\epsilon}(D_1)},\end{aligned}$$

where we have used that for any $N \in \mathbb{N}$,

$$\mathrm{pers}^p_{p,\epsilon}(ND_1)\rho(D_1) = N\mathrm{pers}^p_{p,\epsilon}(D_1)\rho(D_1) = \mathrm{pers}^p_{p,\epsilon}(D_1)\rho(ND_1).$$

Now, we observe that $\mathrm{pers}^p_{p,\epsilon}(D_R) = \mathrm{pers}^p_{p,\epsilon}(\lfloor R-1\rfloor D_1) + \mathrm{pers}^p_{p,\epsilon}(D') \ge (R-2)\mathrm{pers}^p_{p,\epsilon}(D_1)$ and $\mathrm{pers}^p_{p,\epsilon}(D') \le 2\mathrm{pers}^p_{p,\epsilon}(D_1)$ to obtain that

$$\|\bar{\rho}(D(\phi|_{[0,R]})) - \bar{\rho}(D(\phi|_{[c,c+1]}))\|_\infty \le \tfrac{|\rho(D')|}{\mathrm{pers}^p_{p,\epsilon}(D_R)} + \tfrac{|\mathrm{pers}^p_{p,\epsilon}(D')\rho(D_1)|}{\mathrm{pers}^p_{p,\epsilon}(D_R)\mathrm{pers}^p_{p,\epsilon}(D_1)} \le \tfrac{|\rho(D')|+2|\rho(D_1)|}{(R-2)\mathrm{pers}^p_{p,\epsilon}(D_1)} \quad (4.4)$$

Using the Minkowski inequality,

$$|\rho_t(D')| = |\sum_{x\in D'} w_\epsilon(x)^p k_x(t)| \le \sum_{x\in D'} |w_\epsilon(x)^p| \max_{x\in D'} |k_x(t)| \le \mathrm{pers}^p_{p,\epsilon}(D') \max_{x\in D'} \|k_x\|_\infty.$$

Because k is L_k-Lipschitz by (4.16), for any $x \in D'$, we have $\|k_x\| \le L_k\|x - \pi(x)\| + \|k_{\pi(x)}\|$, where $\pi(b,d) = (\frac{b+d}{2}, \frac{b+d}{2})$. Using (4.17) on one hand, and the fact that the distance of any point in the diagram to Δ is bounded by A_ϕ, we obtain $\|k_x\| \le \frac{L_k A_\phi}{2} + C$. A similar bound holds for $\rho_t(D_1)$. Going back to (4.4), we have that

$$\begin{aligned}\|\bar{\rho}(D(\phi|_{[0,R]})) - \bar{\rho}(D(\phi|_{[c,c+1]}))\|_\infty &\le \frac{(2|\mathrm{pers}^p_{p,\epsilon}(D_1)| + |\mathrm{pers}^p_{p,\epsilon}(D_1)|)\max_{x\in D'}\|k_x\|_\infty}{(R-2)\mathrm{pers}^p_{p,\epsilon}(D_1)} \\ &\le \frac{4(C + L_k A_\phi)}{R-2},\end{aligned}$$

what converges uniformly to 0 as $R \to \infty$. □

Without noise, Proposition 3.2 implies that the functional depends only on the number of periods and $\gamma(0)$. As a consequence, the signature F is also robust to the distribution of reparametrisations, but only to a certain extent. Consider $\gamma_1 \sim \mu_1$ and $\gamma_2 \sim \mu_2$ such that the distributions of

endpoints $(\gamma_1(0), \gamma_1(T))$ and $(\gamma_2(0), \gamma_2(T))$ are the same. When μ_1 and μ_2 are such that we can condition on the endpoints, then

$$F(\phi \circ \gamma_1) = F(\phi \circ \gamma_2). \tag{4.5}$$

In light of Proposition 2.35, (4.5) is not surprising, but we cannot show it without a strong disintegration condition. That condition holds when μ_1, μ_2 are measures on a closed subspace of $(C([0,T]), \|\cdot\|_\infty)$. An example of such a space is

$$\Gamma_{v_{\min}} = \{\gamma \in C([0,T], \mathbb{R}) \mid \gamma(s) - \gamma(t) \geq v_{\min}(s-t), \text{ for all } s \geq t\}, \tag{4.6}$$

for any $v_{\min} > 0$. We give more details in Appendix 4.B, notably, we restate (4.5) in more precise terms in Proposition 4.23.

We stress that relaxing the assumption on the equality of distributions is not straightforward. In short, the main problem lies in obtaining a fine control on the persistence diagram when 'cutting' a domain, $[0, T_2]$, into $[0, T_1]$ and $[T_1, T_2]$, for any $0 < T_1, T_2$. Specifically, we need to consider the difference between $D(\phi|_{[0,T_2]})$ and $D(\phi|_{[0,T_1]}) \cup D(\phi|_{[T_1,T_2]})$. When T_1 is a global maximum of ϕ, we can reason as in the proof of Proposition 3.2. However, this is far from the general situation, in which case the cut at T_1 might induce some spurious points in the diagram.

4.1.3 Properties of functionals of noisy periodic functions

Consider now the noisy observations as in (4.1). As explained in Section 3.3.2, we loose the structural property Proposition 3.2 and the invariance with respect to γ. We now obtain two bounds using different strategies. For fixed endpoints, we control the differences produced by the noise (Theorem 4.3). Otherwise, more generally, we can compare the functionals of noisy observations with the signature of the periodic function (Proposition 4.5). Let us detail the assumptions on W and γ.

We impose three conditions on the noise W, whose distribution we will denote by ν. First, we assume that $\|W\|_\infty$ is almost–surely bounded by a constant smaller than the amplitude of the signal: there is $q > 0$ such that $\|W\|_\infty \leq (A(\phi) - \epsilon - q)/2$, where $A(\phi) = \max\phi - \min\phi$. Second, we assume that W is independent of γ. It implies that the law of S is the image measure of the product of μ and ν by the map $(x, y) \mapsto \phi(x) + y$. Finally, we assume that W satisfies the regularity condition (2.1). Difficulties in treating W come both from controlling its amplitude and the regularity. The tools that we use are sensitive to many, small fluctuations. Condition (2.1) allows us to control the regularity, without imposing a uniform Hölder constant on all paths. We refer to Section 2.1 for a discussion of this condition and examples of processes which satisfy it.

For Theorem 4.3, we assume that γ has a lower-bounded modulus of variation and fixed endpoints. Specifically, let $0 < T, R$ and consider

$$\Gamma_{T,R,v_{\min}} \coloneqq \{\gamma \in C([0,T],[0,R]) \mid \gamma(0) = 0, \gamma(T) = R,\ 0 \leq v_{\min}(t-s) \leq \gamma(t) - \gamma(s), \forall s \leq t\}.$$

The set $\Gamma_{T,R,v_{\min}}$ is convex. It is also included in $C([0,T], \mathbb{R})$, so it can be naturally endowed with the sup-norm $\|\cdot\|_\infty$, for which it is a closed, complete and separable space. In particular, it is a Radon space, so that all measures on $(\Gamma_{T,R,v_{\min}}, \mathcal{B}(\Gamma_{T,R,v_{\min}}))$ are inner–regular and locally-finite. Hence, we can equip the space of probability measures on $(\Gamma_{T,R,v_{\min}}, \mathcal{B}(\Gamma_{T,R,v_{\min}}))$ with the Wasserstein distance W_1 (Panaretos and Zemel, 2020). Another reason for working with $\Gamma_{T,R,v_{\min}}$ is that γ^{-1} all have the same domain, what allows us to take full advantage of the invariance properties of homology. Finally, the lower–bound on the modulus provides a relation between $\|\gamma_1^{-1} - \gamma_2^{-1}\|_\infty$ and $\|\gamma_1 - \gamma_2\|_\infty$. The proof of Theorem 4.3 is deferred to Appendix 4.C.

Theorem 4.3 (Stability)**.** *Let μ_1, μ_2 be two probability measures on $\Gamma_{T,R,v_{\min}}$ and let $\gamma_k \sim \mu_k$, for $k = 1, 2$. If $p \geq 1 + \max(r_2, r_2/(r_1 - 1))$, $\bar{\rho} = \bar{\rho}_{\epsilon,p,k}$,*

$$\|\mathbb{E}[\bar{\rho}(\phi \circ \gamma_1 + W)] - \mathbb{E}[\bar{\rho}(\phi \circ \gamma_2 + W)]\|_\infty \leq \frac{\tilde{C}(K_{r_2,r_1})}{v_{\min}^\alpha} W_1(\mu_1, \mu_2)^\alpha,$$

where $\tilde{C}(x) = \mathcal{O}(x^{1/r_2}(1 + x^{1/(r_1-1)}))$ depends on ϕ, ϵ, p, q, k and K_{r_1,r_2}.

Two cases show that the control in Theorem 4.3 is useful. First, suppose that $\mu_k = \delta_{\gamma_k}$ for $k = 1, 2$, for some fixed $\gamma_1, \gamma_2 \in \Gamma_{T,R,v_{\min}}$. Then, we obtain that the normalized truncated silhouette is Hölder, with respect to the distance $\|\gamma_1 - \gamma_2\|_\infty$. It is expected that we do not have

complete invariance: for a fixed path W, the reparametrisation γ can influence how the points in the persistence diagram are displaced, as shows Figure 3.3. Consider now the case of vanishing noise. If K_{r_2,r_1} decreases to zero, then so does the Hölder constant Λ_W and we have indeed that the right-hand side vanishes.

Note that controlling $\|W\|_\infty$ is not sufficient for the stability. When $A(W) < \epsilon$, the constant factor in $\tilde{C}(x)$ is $C_{\Lambda_W} = L_k(1+\frac{8p^2A(\phi)(A(\phi)-\epsilon)\mathrm{pers}_{p-2,\epsilon}^{p-2}(\phi)}{(R-2)q^p})$. We can take the truncation parameter ϵ small, in which case $q = (A(\phi)-\epsilon)$ and so, for ϕ with a single maximum and minimum per period, we have $C_{\Lambda_W} \approx L_k(1+8p^2) > 0$, which is not zero. Even though the amplitude of the noise is smaller than the cut-off ϵ, it still has an influence on the signature. Therefore, it is important that as the amplitude decreases, the noise does not become increasingly irregular: it is the case of aW, with $a \to 0^+$. We require the almost-sure bound on $\|W\|_\infty$ for a different reason: it gives us the lower–bound on $\mathrm{pers}_{p,\epsilon}^p(\phi \circ \gamma + W)$, which appears in the denominator of $\bar{\rho}$.

For processes of decreasing amplitude but increasing irregularity, it is more advantageous to bound $\|W_{\gamma_1^{-1}} - W_{\gamma_2^{-1}}\|_\infty \leq 2\|W\|_\infty$ in the proof. In such a scenario however, we ignore the reparametrisations so the distance $\|\gamma_1^{-1} - \gamma_2^{-1}\|_\infty$ disappears from the bound.

Remark 4.4. When both endpoints are fixed and common to all reparametrisations, there is no reason to normalize by the total persistence. The stability comes from the continuity of the functional, not the renormalisation. Proposition 4.19 states that linear functionals of the form $\sum_{x\in D} w_\epsilon(x)^p k_x$ are also continuous for Hölder functions, so a statement analogue to Theorem 4.3 also holds for such un-normalized functionals.

We now discuss relaxing some assumptions in Theorem 4.3. First, note that the lower–bound on the modulus of continuity ($v_{\min} > 0$) allows us to upper–bound $\|\gamma_1^{-1} - \gamma_2^{-1}\|_\infty$ by $\frac{1}{v_{\min}}\|\gamma_1 - \gamma_2\|_\infty$. But, if we remove this assumption ($v_{\min} = 0$), it is not clear whether $\Gamma_{T,R,0}$ is a complete space for $\|\gamma_1^{-1} - \gamma_2^{-1}\|_\infty$.

Second, we could also allow R to vary. A simple example is to let $\gamma_k = R\tilde{\gamma}_k$, where R is a random variable on a compact set of $]0,\infty[$ and $\tilde{\gamma}_k \sim \tilde{\mu}_k$ is a random element of $\Gamma_{T,1,v_{\min}}$, with $\tilde{\gamma}$ independent of R. In that case, we do obtain the distance $W_{1,\|\cdot\|_\infty}(\tilde{\mu}_1, \tilde{\mu}_2)$ in the bound, but it is not clear that it lower–bounds $W_{1,\|\cdot\|_\infty}(\mu_1, \mu_2)$. For it to be the case, a 'good' coupling of μ_1 and μ_2 would need to be approximated using ν and a coupling of $\tilde{\mu}_1$ and $\tilde{\mu}_2$: we do not know if such an approximation can be constructed.

The final extension is robustness in the case where the distributions of $\gamma_k(T) - \gamma_k(0)$ are not the same for $k = 1, 2$. However, we are short of understanding it already in the noiseless case, as stated in Section 4.1.2 and Appendix 4.B.

Proposition 4.5. *Let $(\gamma_k : [0,T] \to [0,R_k])_{k=1,2}$ be two fixed reparametrisations, for $R_k > 2$. Consider perturbations $W_1, W_2 \in C_\Lambda^\alpha([0,T],\mathbb{R})$, with $\|W_k\|_\infty < A(\phi)/2$. Then,*

$$\|\bar{\rho}(\phi\circ\gamma_1 + W_1) - \bar{\rho}(\phi\circ\gamma_2 + W_2)\| \leq L_k\left(\tfrac{4A(\phi)}{\min(R_1,R_2)-2} + P(\max(\|W_1\|_\infty, \|W_2\|_\infty))\right),$$

where the expression of $P(x) = \mathcal{O}(x)$ is given explicitly in Lemma 4.24.

Proposition 4.5 is much weaker and deterministic, but valid under milder hypotheses. The proof is in Appendix 4.D. Note that the right–hand side is strictly positive, even in the noiseless case $W = 0$ and $\mu_1 = \mu_2$. It is not surprising, because the bounds we use are very crude: we remove the noise and we compare the respective functionals to the limit object $\bar{\rho}(\phi)$. The (uniform) Hölder regularity assumption on the noise stems from the statement being deterministic and pathwise: a similar proof could be carried out for signatures (in expectation), using regularity assumption (2.1).

4.2 Statistical inference for signatures from time–series

We have defined the signature and studied its properties for continuous observations. In practical applications, we do not have access to S, but to observations in the form of a time–series $(S_n)_{n=1}^N$. In our case, this time series is composed of samples from a continuous process. In some situations, it is reasonable to assume that we have access to a collection of independent time–series from the same model, what allows to conveniently estimate F. We consider the case in which we observe a single time series. The purpose of this section is to show asymptotic statistical guarantees for signatures of windows of a discretized signal.

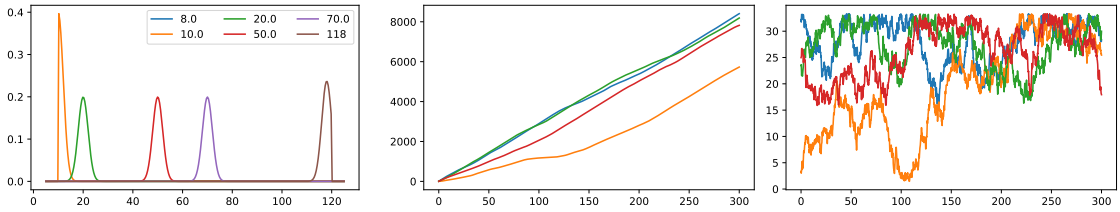


Figure 4.2: On the left, the truncated Gaussian Kernel centered at different point in I, with $\eta = 2$. In the center, several reparametrisation paths, integrated from the Markov chain realisations on the right, generated as specified in Example 4.6 with $\eta = 1.1$.

4.2.1 Time series model

Similarly to the continuous model (4.1), the observations are a reparametrisation of a 1-periodic function ϕ

$$S_n = \phi(\gamma_n) + W_n \in \mathbb{R}, \qquad n = 1, \dots, N, \tag{4.7}$$

where $(\gamma_n)_{n=1}^N$ is a strictly increasing time series and $(W_n)_{n\in\mathbb{N}}$ is a stationary noise time series satisfying $\mathbb{E}[W_n] = 0$. It is also convenient and straightforward to consider the limit of a time series of infinite length, $(S_n)_{n\in\mathbb{N}}$.

In the industrial setting motivating this thesis, γ represents the displacement of a vehicle. We introduce two reparametrisation processes, defined as discrete integrals of another, positive time series V_n. Specifically, let

$$\gamma_{n+1} = \gamma_n + hV_n = \gamma_0 + h\sum_{k=0}^{n} V_k, \tag{4.8}$$

where $(V_n)_{n=0}^N$ is a sequence of random variables in $I \coloneqq [v_{\min}, v_{\max}]$, independent of γ_0 and $0 < h$ is a time step. This model is inspired by dynamics, where the sequence $(\gamma_n)_{n\in\mathbb{N}}$ could model the displacement of a body over time and V_n should be thought of as the instantaneous speed and $h = \frac{T}{N}$. We will consider two models for $(V_n)_{n\in\mathbb{N}}$.

Model 1 ($(V_n)_{n\in\mathbb{N}}$ i.i.d)**.** We assume that V_n are independent and follow the same, unknown distribution on $\mathbb{R}_+^*$, which satisfies the following property: there exists $0 < a, b, c$ such that, for all $A \in \mathcal{B}(]a, b[)$ measurable, $P(V_k \in A) \geq c\mu(A)$, where μ is the Lebesgue measure.

Model 2 ($(V_n)_{n\in\mathbb{N}}$ a Markov Chain)**.** Let $(V_n)_n$ be a Markov Chain with transition probability kernel P (Definition 2.9). We assume that $P(x, \cdot)$ is a probability measure that has a density f_x with respect to μ and that

1. the density is lower–bounded in a small neighborhood: there exists $\eta, \mu_0 > 0$, such that

$$f_v|_{[v-\eta, v+\eta]\cap I} \geq \mu_0, \tag{4.9}$$

2. $v \mapsto f_v(x)$ is continuous for any $x \in I$.

Note that if $f_x = f$, for all $x \in I$, model 2 reduces to a particular case of model 1, where P has density f, $a = v_{\min}, b = v_{\max}$ and $c = \mu_0$.

Example 4.6. Set $V_0 \sim \mathcal{U}(I)$ and let $0 < \eta < \frac{v_{\max} - v_{\min}}{4}$. An example of a kernel satisfying the above assumption is a truncated Gaussian kernel. The truncation is such that the support is I and $\sigma = \eta$. In Figure 4.2, we show the kernel and several sample trajectories from this model.

4.2.2 Estimation of signatures

Consider the situation where we observe a single time–series $(S_n)_{n=1}^N$, $S_n \in \mathbb{R}$. On a technical level, we can calculate the persistence diagram of a time series as explained in Section 2.5.5 and the resulting functional in the same way, $\bar{\rho}_t((S_n)_n) = \bar{\rho}_t(D((S_n)_{n=1}^N)) = \frac{\sum_{x\in D} w_\epsilon(x)^p k_x(t)}{\sum_{x\in D} w_\epsilon(x)^p}$. However, with a single observation, we cannot expect to reliably estimate $\bar{\rho}(S)$. Persistent homology is a global descriptor, which can link two events, even if they happen far in time. Even though the

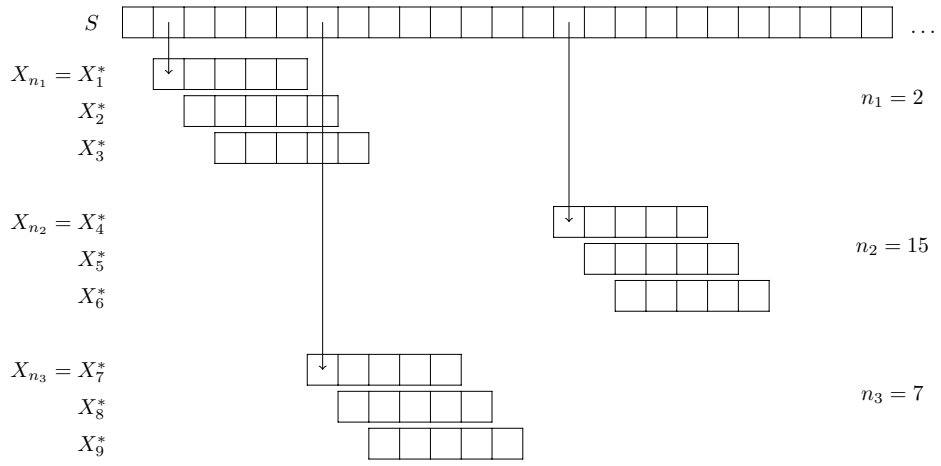


Figure 4.3: A schematic representation of the MBB, for $M = 5$ and $L = 3$.

descriptor $\bar{\rho}$ effectively represents the average homological feature, it is not immediately clear that it benefits from the same properties as the empirical mean. Understanding this poses the same challenges as those explained at the end of Section 4.1.2 and Appendix 4.B.

We fix a window length $M \in \mathbb{N}$ and we use as a signature $F_M(S) \coloneqq F(X) = \mathbb{E}[\bar{\rho}(X)]$, where $X = (S_1, \dots S_M)$. It is a quantity which we can estimate with an empirical mean, whose distribution we can also characterize by bootstrap techniques. In the noiseless case, Theorem 4.2 shows that $F_M(S)$ converges to $F(S)$, as $M \to \infty$. This justifies replacing the 'global' signature with a local version, when the number of periods observed in each window is sufficiently large.

From $(S_n)_{n=1}^N$, we construct the empirical measure

$$\mathbb{P}_{N-M+1} = \tfrac{1}{N-M+1} \sum_{n=1}^{N-M+1} \delta_{X_n}, \qquad \text{where } X_n = (S_n, \dots, S_{n+M-1}). \tag{4.10}$$

The empirical counter-part of $F_M(S)$ is the empirical mean

$$\hat{F}_M(S) \coloneqq \mathbb{P}_{N-M+1}\bar{\rho} = \tfrac{1}{N-M+1} \sum_{n=1}^{N-M+1} \bar{\rho}(X_n).$$

We will estimate its distribution by Moving Block Bootstrap (MBB), as described in Section 2.4. It is applied here to windows $(X_1, \dots, X_{N-M+1})$ as illustrated in Figure 4.3. If we denote by $L = L(N-M+1) \in \mathbb{N}$ the block length, the MBB sample is then

$$X_1^*, \dots, X_{N-M+1}^* = X_{n_1}, \dots, X_{n_1+L}, X_{n_2}, \dots, X_{n_2+L}, \ \dots, X_{n_B}, \dots X_{n_B+L},$$

where $n_1, \dots n_B \sim \mathcal{U}(1, \dots, N-M+1)$ are independent and we define

$$F_M^*(S) = \hat{\mathbb{P}}_{N-M+1}\bar{\rho} = \tfrac{1}{N-M+1} \sum_{n=1}^{N-M+1} \bar{\rho}(X_n^*).$$

The purpose of this section is to prove that the empirical mean $\hat{F}_M(S)$ converges to $F_M(S)$ and that we can approximate the distribution of $\hat{F}_M$ by that of F_M^*, as $N \to \infty$. The core idea is to control how the dependence between X_1 and X_{1+k} changes as k increases. We do so using the β-mixing coefficients (Definition 2.16).

Theorem 4.7. *Consider $(\gamma_n)_{n=1}^N$ as in (4.8) with $(V_n)_{n=1}^N$ from Model 1 or 2. Assume that W is exponentially β-mixing. Then,*

$$\sqrt{N-M+1}(\hat{F}_M - F_M) \to G \tag{4.11}$$

where G is a zero–mean Gaussian process with covariance

$$(s,t) \mapsto \lim_{k\to\infty} \sum_{n=1}^{\infty} \mathrm{cov}(\bar{\rho}(X_k)(s), \bar{\rho}(X_n)(t)).$$

Then, if $L(N) \to \infty$ and $L(N) = \mathcal{O}(N^{1/2-\epsilon})$ for some $\epsilon > 0$ as $N \to \infty$, then

$$\sqrt{N-M+1}(\hat{F}^* - \hat{F}) \to^* G \qquad \text{in probability,} \tag{4.12}$$

This result is a functional central limit theorem, similar to many in the literature of topological data analysis, see for example Chazal et al. (2014) and Berry et al. (2020, Proposition 2 and 3), except that the samples are not independent. For i.i.d data, it is sufficient to control the complexity of the functional family. Since this aspect has been covered extensively, we only recall Proposition 4.21 for completeness. The novel aspect of Theorem 4.7 is the consideration of dependence and it is what we treat with more care. The rest of this section is devoted to a proof of Theorem 4.7, with details being deferred to appendices.

Sketch of proof of Theorem 4.7. Consider $(\gamma_n)_{n\in\mathbb{N}}$ as in (4.8) with model 1 or 2. Notice that this series is not stationary, because $P(\gamma_n < \gamma_{n+1}) = 1$. Crucially though, the composition $(\phi(\gamma_n))_{n\in\mathbb{N}}$ is stationary. In fact, it can be written as $\phi(x) = \phi(\mathrm{frac}(x))$, through $\mathrm{frac}(x) = x - \lfloor x \rfloor$ the fractional part of a real number. In Appendix 4.E, we show that $(\mathrm{frac}(\gamma_n))_{n\in\mathbb{N}}$ is exponentially β-mixing (Proposition 4.25). While this is not a surprising result, it is the most technical part of the proof. We show the Doeblin condition (2.4): in model 1, γ is a random walk, so the distribution of P^n can be written as a convolution. For model 2, we need to consider the joint process $(\mathrm{frac}(\gamma_n), V_n)$, which is a Markov chain. We show that for a sufficiently large $n \in \mathbb{N}$, we can lower-bound P^n by a measure uniform on $[x, x+1]$ and $[x, x+1] \times I$ for models 1 and 2 respectively. The push-forward by frac of that measure gives a lower–bound for the distribution of $(\frac{(}{\gamma}_n))_{n\in\mathbb{N}})$ on $[0,1]$ and for the distribution of $(\frac{(}{\gamma}_n), V_n)_{n\in\mathbb{N}}$ on $[0,1] \times I$ respectively, uniform with respect to the initial conditions.

Next, we analyze how the dependence of $(\phi(\gamma_n))_{n\in\mathbb{N}}$ and $(W_n)_{n\in\mathbb{N}}$ shapes the dependence of $(S_n)_{n\in\mathbb{N}}$ and that between the windows $X_1, \ldots, X_{N-M+1}$. Specifically, Appendix 4.F contains a proof of the following inequality

$$\beta_X(k) \le \beta_S(k-(M+1)) \le \beta_{\mathrm{frac}(\gamma)}(k-(M+1)) + \beta_W(k-(M+1)), \qquad \text{for } k \ge M+1.$$

Since $(W_n)_{n\in\mathbb{N}}$ is exponentially mixing by assumption, $(X_n)_{n\in\mathbb{N}}$ is exponentially-mixing.

The Gaussian approximation (4.11) is a consequence of Theorem 2.22. By the arguments above, the mixing condition (2.9) is verified for X for any $r > 2$. It remains to verify that the bracketing entropy of the functional family $\{\bar{\rho}_t\}_{t\in\mathbb{T}}$ is controlled. This is done in Proposition 4.21.

The approximation of the distribution of the empirical mean by the bootstrap distribution (4.12) is a consequence of Theorem 2.23, for which we only need the aforementioned results. □

We do not give a procedure to choose the block length L, although it has been observed that, in general, $L \simeq (N-M+1)^{1/3}$ works well in practice (Bühlmann and Künsch, 1999).

Remark 4.8. The literature of functional central limit theorems for dependent data is rich in results for various functional classes and dependence assumptions. We believe it might be possible to use more recent and stronger results than Bühlmann (1995, Theorem1), stated here Theorem 2.23. This would allow us to relax the decay of β_W from an exponential to a polynomial one. For instance, (Radulović, 1996, Theorem 1) is written for VC-classes functionals, but the proof seems to rely on the bracketing entropy bound that the functionals considered in the present work also satisfy.

4.2.3 Discussion

Theorem 4.7 motivates the use of $\bar{\rho}$ as a descriptor of a phase–modulated, periodic signal. A possible application of the asymptotic guarantees is the construction of asymptoticlly valid confidence intervals. Consider a situation when S_1 and S_2 are observed, $S_k = \phi_k(\gamma_k) + W_k$, where $\gamma_k \sim \mu_k$. It is tempting to use the proposed framework to test for $\phi_1 = \phi_2$ based on observations S_1 and S_2. Such a test is justified if $\mu_1 = \mu_2$, but we do not provide theoretical guidance on how to calibrate such a test otherwise.

We do not present a theory for the choice of the window length M. Increasing M reduces the probability of not capturing a whole period in a window of length M and the impact of potential boundary effects. On the other hand, the amplitude of a noisy signal is increasing as the window size increases, and Figure 3.3 shows how sensitive persistence diagrams are.

Model 2 introduces dependence between consecutive velocities $(V_n)_{n\in\mathbb{N}}$, but it is a very simple model. For applications where γ_n represents a position in time, a more realistic model would assume that $(V_n)_{n\in\mathbb{N}}$ is itself an integrated process, for example, $V_{n+1} = V_n + ha_n$ with $(a_n)_{n\in\mathbb{N}}$ modelling the acceleration. A reasonable assumption would be that the acceleration is a hidden Markov chain. Under ergodicity assumptions akin to (4.9), we believe that arguments from Appendix 4.E could be used to show that the reparametrisation sequence $(\mathrm{frac}(\gamma_n))_n$ is mixing.

4.3 Total truncated persistence and functional representations of persistence diagrams

To show the invariance and statistical properties outlined in Sections 4.1 and 4.2 respectively, we have used a certain number of assumptions on the regularity of the normalized functionals, both as functions $t \mapsto \bar{\rho}_t(S)$ for a fixed signal S, as well as $S \mapsto \bar{\rho}_t(S)$ for a fixed t. In this section, we introduce a continuous variant of the total p-persistence, namely the ϵ-truncated p-persistence, and show some of its continuity properties. Then, we show what effect using it as weights in ρ and $\bar{\rho}$ has on the properties of those functionals. We rely on the background provided for persistent homology in Section 2.5 and on the terminated persistence modules from Section 3.2.

4.3.1 Total truncated persistence

The total p-persistence of the persistence diagram of sub level sets of a function quantifies the oscillations of that function. It is similar to total variation for functions on the interval (Plonka and Zheng, 2016). For $p \in \mathbb{R}^+$, the *total p-persistence of a persistence diagram D* is the sum of p-powers of the lifetimes of points, $\mathrm{pers}_p(D) = \left(\sum_{(x,y)\in D} w(b,d)^p\right)^{1/p}$, where $w(b,d) \coloneqq d - b$ is the *persistence of a point* $(b,d) \in \mathbb{R}^2$.

In the case of sublevel set persistence, points with small persistence might be attributed to noise and quantify the regularity of the function, while the more persistent ones capture the biggest oscillations of ϕ. The functionals we propose in Section 4.3.2 use persistence and total persistence to give different weights to certain features, reflecting the intuition given above. The stability of the signatures with respect to the generating process depends on the stability of the weights with respect to the input functions, as well as lower- and upper-bounds thereof.

Continuous functions on compact domains are bounded and attain their extremal values, but, similarly to total variation, that is not enough to bound their total persistence because of possible small oscillations. An $\alpha-$Hölder function has total p-persistence bounded by a constant, for $p > 1/\alpha$ (Perez, 2022a), but there are functions for which it is not finite for $p < 1/\alpha$. For example, consider the function with Hödler constant Λ composed of 'teeth' with height ϵ and width $2(\epsilon/\Lambda)^{1/\alpha}$ each. The interval $[0,T]$ can be covered with $T(\Lambda/\epsilon)^{1/\alpha}/2$ such tents, so that the total p-persistence of this function is $T\Lambda^{1/\alpha}\epsilon^{p-1/\alpha}$, which is unbounded as $\epsilon \to 0$ when $\alpha < 1/p$. As a remedy, Perez (2022a) suggests to consider the persistence of the truncated diagram $D \cap \Delta_\epsilon$, for some $\epsilon > 0$, where $\Delta_\epsilon \coloneqq \{(b,d) \in D \mid d - b \geq \epsilon\}$. This leads to a bounded total persistence, but continuity with respect to the sup norm is lost, even on very regular functions. Namely, consider $f_{\epsilon'}$ the Lipschitz function composed of $T\Lambda/\epsilon'$ "teeth", each of height ϵ'. Then,

$$\mathrm{pers}_p^p(D_{f_{\epsilon'}} \cap \Delta_\epsilon) = \begin{cases} T\Lambda\epsilon'^{p-1} & \text{if } \epsilon \leq \epsilon', \\ 0 & \text{if } 0 < \epsilon < \epsilon, \end{cases}$$

since $D_{f_{\epsilon'}} \cap \Delta_\epsilon$ is empty in the latter case. The total persistence is continuous for Lipschitz functions, with the modulus of continuity depending on the Lipschitz constant (Cohen-Steiner et al., 2010, Total Persistence Stability Theorem), but this truncation prevents continuity.

To guarantee both boundedness and continuity, we introduce the ϵ-truncated p-persistence. We

define the ϵ-*truncated total p-persistence* as

$$\mathrm{pers}_{p,\epsilon}(D) = \left(\sum_{(x,y)\in D} w_\epsilon(b,d)^p \right)^{1/p}$$

and $\mathrm{supp}(w_\epsilon) \subset \Delta_\epsilon \coloneqq \{(b,d) \in \mathbb{R}^2 \mid d-b \geq \epsilon\}$.

The ϵ-truncated persistence is the persistence of a diagram shifted by $(\epsilon/2, -\epsilon/2)$, $\mathrm{pers}^p_{p,\epsilon}(D) = \mathrm{pers}^p_{p,0}(\{(b+\epsilon/2, d-\epsilon/2) \mid (b,d) \in D\})$. Alternatively, we can think of shifting the diagonal, as illustrated in Figure 4.4. Proposition 4.9 shows that truncated persistence is continuous in the bottleneck distance between diagrams. Note that the modulus of continuity in the proof is not uniform, since it depends on the number of points and the maximal persistence of a point in the diagram.

Proposition 4.9. *The ϵ-truncated p−persistence is continuous with respect to the bottleneck distance, for every persistence diagram of a continuous function $h : [0,T] \to \mathbb{R}$.*

Proof. The persistence diagram $D_1 = D(h)$ is contained in $[\min h, \max h] \times [\min h, \max h]$. Since the module is q-tame, $\mathrm{rank}(\mathbb{V}(h)_s \to \mathbb{V}(h)_{s+\epsilon/8}) < \infty$, so $M \coloneqq \mathrm{card}(D_1 \cap \Delta_{\epsilon/4}) < \infty$ and $\max_{(b,d)\in D_1} d-b \leq U < \infty$, for some $U \in \mathbb{R}$. Let D_2 be such that $d_B(D_1, D_2) < \epsilon/4$. Then, $\mathrm{card}(D_2 \cap \Delta_\epsilon) \leq \mathrm{card}(D_1 \cap \Delta_{\epsilon/2}) \leq M$ and the persistence of a point in D_2 is bounded by $U + \epsilon/2$. Trivially, the truncated persistence of a point is 2-Lipschitz,

$$|w_\epsilon(b,d) - w_\epsilon(b',d')| \leq (d-b-\epsilon)_+ - (d'-b'-\epsilon)_+ \leq |d-b-(d'-b')| \leq 2\|(b,d)-(b',d')\|_\infty.$$

Then, we use the technique from the proof of the (Cohen-Steiner et al., 2010, Total Persistence Stability Theorem): writing $|x_2^p - x_1^p| = |p\int_{x_1}^{x_2} t^{p-1}dt| \leq p|x_2 - x_1| \max(x_1^{p-1}, x_2^{p-1})$, we get

$$\begin{aligned}\left| \sum_{x\in D_1} w_\epsilon(x)^p - w_\epsilon(\Gamma(x))^p \right| \leq & p \sum_{D_1} |w_\epsilon(x) - w_\epsilon(\Gamma(x))|(w_\epsilon(x)^{p-1} + w_\epsilon(\Gamma(x))^{p-1}) \\ \leq & 4pM(U+\epsilon/2)^{p-1} d_B(D_1, D_2).\end{aligned}$$

□

By abuse of notation, we will define $\mathrm{pers}_{p,\epsilon}(h) \coloneqq \mathrm{pers}_{p,\epsilon}(D_h)$. Unsurprisingly, the ϵ-truncated total p-persistence of a Hölder function on $[0,T]$ is bounded.

Proposition 4.10. *Let $h \in C^\alpha_\Lambda([0,T], \mathbb{R})$. For $p > 1 + 1/\alpha$,*

$$\mathrm{pers}^p_{p,\epsilon}(h) \leq (A(h)-\epsilon)^p \left(1 + pT\left(\tfrac{2\Lambda}{\epsilon}\right)^{1/\alpha}\right) =: C_{p,\Lambda,\alpha,T},$$

where $A(h) \coloneqq \max h - \min h$ is the amplitude of h.

Proof of Proposition 4.10. We first note that when $A(h) \leq \epsilon$, then $\mathrm{pers}^p_{p,\epsilon}(h) = 0$. For the non-trivial case, we follow Perez (2022a, proof of Theorem 4.13). An upper-bound of the covering number of the image of h, at radius $\tau > 0$ is $T(2\Lambda/\tau)^{1/\alpha} + 1$, so that

$$\begin{aligned}\mathrm{pers}^p_{p,\epsilon}(h) &\leq p\int_\epsilon^{A(h)} \left(T\left(\frac{2\Lambda}{\tau}\right)^{1/\alpha} + 1\right)(\tau-\epsilon)^{p-1}d\tau \\ &= (A(h)-\epsilon)^p + pT(2\Lambda)^{1/\alpha}\int_\epsilon^{A(h)} \frac{(\tau-\epsilon)^{p-1}}{\tau^{1/\alpha}}d\tau\end{aligned}$$

We recall that since $\frac{A(h)}{\tau} \geq 1$ and $\frac{1}{\alpha} \leq p-1$, $(\frac{A(h)}{\tau})^{1/\alpha} \leq (\frac{A(h)}{\tau})^{p-1}$, so

$$\frac{(\tau-\epsilon)^{p-1}}{\tau^{1/\alpha}} = \frac{1}{A(h)^{1/\alpha}}\left(\frac{A(h)}{\tau}\right)^{1/\alpha}(\tau-\epsilon)^{p-1} \leq A(h)^{p-1-1/\alpha}\left(1-\frac{\epsilon}{\tau}\right)^{p-1}.$$

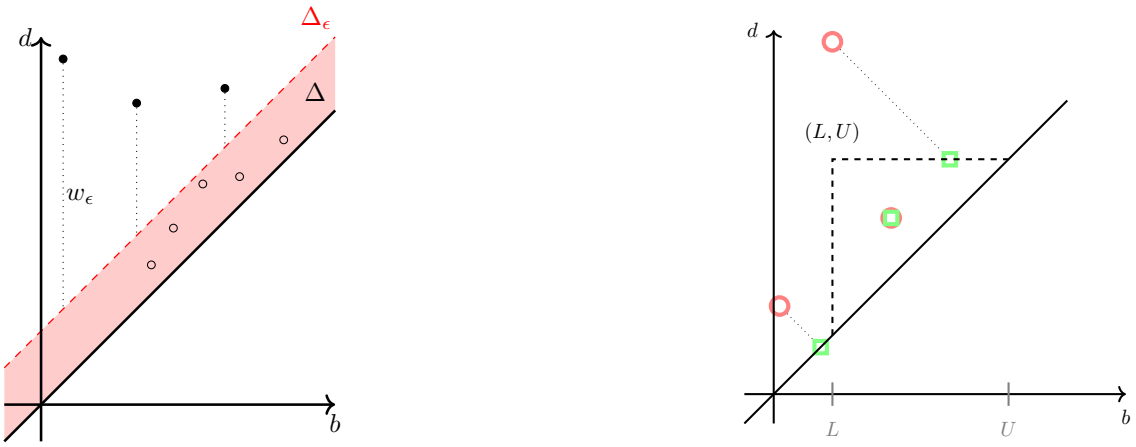


Figure 4.4: On the left, a persistence diagram and Δ_ϵ, the diagonal shifted by $(-\epsilon/2, \epsilon/2)$. The truncated persistence w_ϵ is the distance to Δ_ϵ. The points in the shaded region have zero truncated persistence. On the right, illustration of the projection in (4.18). Points from a diagram and their projections are illustrated with circular and square markers.

Finally, by recognizing that $1 - \epsilon/\tau \leq 1 - \epsilon/A(h)$, we obtain

$$\begin{aligned}
\operatorname{pers}^p_{p,\epsilon}(h) &\leq (A(h) - \epsilon)^p + pT(2\Lambda)^{1/\alpha} A(h)^{p-1-1/\alpha}(1 - \epsilon/A(h))^{p-1}(A(h) - \epsilon) \\
&\leq (A(h) - \epsilon)(1 - \epsilon/A(h))^{p-1}[A(h)^{p-1} + pT(2\Lambda)^{1/\alpha} A(h)^{p-1-1/\alpha}] \\
&\leq (A(h) - \epsilon)^p \left(1 + pT\left(\tfrac{2\Lambda}{A(h)}\right)^{1/\alpha}\right) \\
&\leq (A(h) - \epsilon)^p \left(1 + pT\left(\tfrac{2\Lambda}{\epsilon}\right)^{1/\alpha}\right),
\end{aligned}$$

where we have used that $\epsilon^{1/\alpha} \leq A(h)^{1/\alpha}$ □

By Hölder continuity, $A(h) \leq T^\alpha \Lambda$, so the ratio $\frac{T\Lambda^{1/\alpha}}{A(h)^{1/\alpha}} \geq 1$ denotes how small the amplitude of h is relative to what it could be, under the Hölder assumption. Interestingly, that term increases as $A(h)$ gets smaller, but the whole bound is indeed increasing in $A(h)$, which is of the order of $A(h)^p + A(h)^{2p-1/\alpha}$.

Remark 4.11. Using the proof of Proposition 4.10, we can also upper-bound the total ϵ-truncated p-persistence by $\operatorname{pers}^p_{p,\epsilon}(h) \leq T^{\alpha p}\Lambda^p(1 + p2^{1/\alpha})$.

With $A(h)$ finite and using Proposition 4.9, we could immediately show that total p-persistence is also continuous with respect to the input function. However, we can even show that it is Lipschitz, following the proof of (Perez, 2022a, Lemma 3.20). The argument is based on a Hölders' inequality and the uniform upper-bound on persistence from Proposition 4.10.

Proposition 4.12 (Continuity of truncated p–persistence)**.** *The total ϵ-truncated p-persistence $\operatorname{pers}^p_{p,\epsilon} : C([0,T],\mathbb{R}) \to \mathbb{R}$ is continuous. In addition, $\operatorname{pers}^p_{p,\epsilon}$ is Lipschitz over Hölder functions: for any $f, g \in C^\alpha_\Lambda([0,T])$ such that $p - 2 > \frac{1}{\alpha}$,*

$$\begin{aligned}
|\operatorname{pers}^p_{p,\epsilon}(f) - \operatorname{pers}^p_{p,\epsilon}(g)| &\leq p\|f - g\|_\infty \left(\operatorname{pers}^{p-1}_{p-1,\epsilon}(f) + \operatorname{pers}^{p-1}_{p-1,\epsilon}(g)\right) \\
&\leq C_{p-1,\Lambda,\alpha,T}\|f - g\|_\infty.
\end{aligned}$$

Finally, we give a lower–bound for the truncated persistence. Such a result will be necessary to show the continuity of the class of normalized functionals $\bar{\rho}$, which we define in Section 4.3.2.

Proposition 4.13 (Lower-bound on $p-$persistence)**.** *For continuous functions $f, W : [0,T] \to \mathbb{R}$,*

$$\operatorname{pers}^p_{p,\epsilon}(f + W) \geq \operatorname{pers}^p_{p,\epsilon + A(W)}(f).$$

Proof. Since pers is translation-invariant ($\operatorname{pers}_{p,\epsilon}(f + c) = \operatorname{pers}_{p,\epsilon}(f)$, for any constant $c > 0$), we can assume that $A(W) = 2\|W\|_\infty$. Let $\Gamma : D(f) \to D(f + W)$ be a matching between

the diagrams and denote by $c(\Gamma)$ the associated cost. Thanks to the bottleneck stability theorem (Proposition 2.38), $\inf_\Gamma c(\Gamma) \leq \|W\|_\infty$. Then, for any $(b,d) \in D(f)$ and $(b',d') = \Gamma((b,d)) \in D(f+W)$, we have $d'-b' \geq d-b-2c(\Gamma)$ and, for any $\delta > 0$, $D(f) \cap \Delta_{2c(\Gamma)+\delta} \subset \Gamma^{-1}(D(f+W) \cap \Delta_\delta)$. Then,

$$\begin{aligned}
\mathrm{pers}^p_{p,\epsilon}(f+W) &= \sum_{(b',d')\in D(f+W)} w_\epsilon(b',d')^p \\
&\geq \sum_{(b',d')\in D(f+W)\cap\Delta_\delta} w_\epsilon(b',d')^p \\
&\geq \sum_{(b,d)\in\Gamma^{-1}(D(f+W)\cap\Delta_\delta)} w_\epsilon((b,d)-c(\Gamma)(-1,1))^p \\
&\geq \sum_{(b,d)\in D(f)\cap\Delta_{2c(\Gamma)+\delta}} w_{\epsilon+2c(\Gamma)}(b,d)^p.
\end{aligned}$$

For $\delta = \epsilon$, the last quantity is equal to $\mathrm{pers}^p_{p,\epsilon+2c(\Gamma)}(f)$. By taking the infimum over all matchings Γ, we obtain $\mathrm{pers}^p_{p,\epsilon}(f+W) \geq \mathrm{pers}^p_{p,\epsilon+2\|W\|_\infty}(f)$. □

The result is very weak, but it is tight. If we take f such that $\max f - \min f = 2\|f\|_\infty$ and $W = -\alpha f$, then $f+W = (1-\alpha)f$ and $\|W\|_\infty = \alpha\|f\|_\infty$. Then, $\mathrm{pers}^p_{p,\epsilon}((1-\alpha)f) = \mathrm{pers}^p_{p,\epsilon+2\alpha}(f)$.

Proof of Proposition 4.12. Let $f, g \in C([0,T])$ such that $\|f-g\|_\infty < \epsilon/4$. Let $\Gamma : D(f) \to D(g)$ be a matching. Recall that $|w_\epsilon(b,d) - w_\epsilon(\eta_b,\eta_d)| \leq |b-\eta_b| + |d-\eta_d| \leq 2\|(b,d)-(\eta_b,\eta_d)\|_\infty$. In addition, if $d-b < \epsilon/2$, then both $w_\epsilon(b,d) = 0 = w_\epsilon(\Gamma(b,d))$. Using the bound on the difference of p-powers as in the proof of Proposition 4.9,

$$\begin{aligned}
\left|\sum_{(b,d)\in D(f)} w_\epsilon(b,d)^p - \sum_{(b',d')\in D(g)} w_\epsilon(b',d')^p\right| &\leq p\sum_{(b,d)\in D(f)} |w_\epsilon(b,d) - w_\epsilon(\Gamma(b,d))| \max_{x\in\{(b,d),\Gamma(b,d)\}} w_\epsilon(x)^{p-1} \\
&\leq 2p\|f-g\|_\infty \sum_{\substack{(b,d)\in D(f)\\ d-b\geq\epsilon/2}} \max_{x\in\{(b,d),\Gamma(b,d)\}} w_\epsilon(x)^{p-1} \\
&\leq p\|f-g\|_\infty \sum_{\substack{(b,d)\in D(f)\\ d-b\geq\epsilon/2}} (w_\epsilon(b,d)+2\epsilon/4)^{p-1}.
\end{aligned}$$

Since f is continuous on a compact domain, it is uniformly continuous, so the right-hand side is finite and depends only on f.

For the Lipschitz character, we follow the proof of (Perez, 2022a, Lemma 3.20). For $f, g \in C^\alpha_\Lambda([0,T])$,

$$\begin{aligned}
\left|\sum_{(b,d)\in D(f)} w_\epsilon(b,d)^p - \sum_{(b',d')\in D(g)} w_\epsilon(b',d')^p\right| &\leq p\sum_{(b,d)\in D(f)} |w_\epsilon(b,d) - w_\epsilon(\Gamma(b,d))| \max_{x\in\{(b,d),\Gamma(b,d)\}} w_\epsilon(x)^{p-1} \\
&\leq 2p\|f-g\|_\infty \left(\sum_{(b,d)\in D(f)} w_\epsilon(b,d)^{p-1} + \sum_{(b',d')\in D(g)} w_\epsilon(b',d')^{p-1}\right) \\
&= 2p(\mathrm{pers}^{p-1}_{p-1,\epsilon}(D(f)) + \mathrm{pers}^{p-1}_{p-1,\epsilon}(D(g))\|f-g\|_\infty.
\end{aligned}$$

By Proposition 4.10, $\mathrm{pers}^{p-1}_{p-1,\epsilon}(D(f)) \leq T^{\alpha(p-1)}\Lambda^{p-1}(1+(p-1)2^{1/\alpha})$, so that

$$|\mathrm{pers}^p_{p,\epsilon}(D(f)) - \mathrm{pers}^p_{p,\epsilon}(D(g))| \leq 4pT^{\alpha(p-1)}\Lambda^{p-1}(1+(p-1)2^{1/\alpha})\|f-g\|_\infty.$$

□

4.3.2 Linear and normalized functionals

We have discussed functional representations of persistence diagrams in Section 2.5.7. Many popular functionals are linear, while in the context of signatures of periodic functions, it makes sense

to normalize by the total persistence. As it is usually the case with functionals of persistence, we present a general set of assumptions and we show examples of functionals from the literature (or of their adaptation) which fit within the prescribed framework.

Consider $(\mathbb{T}, d)$ a Euclidean space and let $\mathcal{H}$ be a Banach space of functions $\mathbb{T} \to \mathbb{R}$. Finally, let $k : \mathbb{R}^2 \to \mathcal{H}$ be a map, often called 'kernel', which to a point (b, d) in the plane associates a function $k(b, d)$. For a persistence diagram D with $\text{pers}_{p,\epsilon}(D) > 0$, the linear and the normalized functionals are of the form

$$\rho(D) = \sum_{x \in D} w_\epsilon(x)^p k(x), \qquad \bar{\rho}(D) = \frac{\rho(D)}{\sum_{x \in D} w_\epsilon(x)^p}. \tag{4.13}$$

Otherwise, $\rho(D) = 0 = \bar{\rho}(D)$. We will abuse notation and write $\rho(f) \coloneqq \rho(D(f))$. Compared to how the functionals are usually introduced, we use the ϵ-truncated p-persistence instead of p-persistence. As shown below in Proposition 4.19, this guarantees their continuity but leads to some problems, notably because truncation can make non–empty diagrams empty.

Proposition 4.14. *For any continuous function $f : [0, T] \to \mathbb{R}$ with $\max f - \min f > \epsilon$, the linear and normalized functionals are well-defined.*

Proof. Since f is continuous on a compact domain, it is also uniformly continuous and bounded. Let $\delta > 0$ be such that $|f(t) - f(s)| < \epsilon$, whenever $|s - t| < \delta$. By the reasoning of the proof of (Cohen-Steiner et al., 2010, Persistence Cycle Lemma), $|\omega^{-1}(]\epsilon, \infty[) \cap D(f)| \leq \frac{T}{\delta} + 1$. Let $M_f = \max(f)$, $m_f = \min(f)$. Then,

$$\begin{aligned} \|\rho(D(f))\|_\infty &\leq \sum_{(b,d) \in D(f)} w_\epsilon(d - b)^p \|k(b, d)\|_\infty \\ &\leq (\tfrac{T}{\delta} + 1) \cdot w_\epsilon(M_f - m_f)^p \max_{(b,d) \in D(f) \cap \Delta^+_\epsilon} \|k(b, d)\|_\infty. \end{aligned}$$

As stated above, the number of points is bounded from above, and so is the total persistence. For the normalized functional,

$$\|\bar{\rho}(D)\| \leq (\tfrac{T}{\delta} + 1) \max_{(b,d) \in D(f) \cap \Delta^+_\epsilon} \|k(b, d)\|_\infty \overbrace{\frac{w_\epsilon(M_f - m_f)^p}{\sum_{x \in D} w_\epsilon(x)^p}}^{\leq 1}.$$

□

We consider functionals $\bar{\rho}$ with a kernel k that satisfies the following assumptions:

1. $k(x)$ has a uniformly bounded support, for all $x \in \mathbb{R}^2$

$$\exists K \subset \mathbb{T} \text{ compact}, k(x)|_{\mathbb{T} \setminus K} \equiv 0, \qquad \forall x. \tag{4.14}$$

2. $k(x)$ is Lipschitz, uniformly over $x \in \mathbb{R}^2$

$$\exists L > 0,\ |k(x)(t) - k(x)(s)| \leq L d(s, t), \qquad \forall x \in \mathbb{R}^2, \forall s, t \in \mathbb{T}. \tag{4.15}$$

3. $x \mapsto k(x)$ is Lipschitz

$$\exists L_k > 0, \|k(x) - k(x')\|_{\mathcal{H}} \leq L_k \|x - x'\|_\infty, \qquad \forall\ x, x' \in \mathbb{R}^2. \tag{4.16}$$

4. $k(x)$ is uniformly-bounded on the diagonal

$$\exists C \geq 0,\ \|k|_\Delta\|_\infty \leq C. \tag{4.17}$$

Hypotheses (4.16,4.17) ensure continuity of the functional, while (4.14,4.15) control the bracketing entropy of the family of functionals.

Many functionals proposed in the literature do not satisfy (4.14) as is. To adapt them to this assumption, we precompose the usual kernels with a projection, which we illustrate in Figure 4.4. Specifically, let $L < U \in \mathbb{R}$ and consider $\pi_{L,U} : \Delta_{\geq 0} \to \Delta_{\geq 0}$ the operator which maps points above the diagonal, onto the upper triangle with corner at (L, U)

$$\begin{array}{rccl} \pi_{L,U} : & \Delta_{\geq 0} & \to & \Delta_{\geq 0} \\ & (b, d) & \mapsto & (b, d) + (1, -1) \min(\max(d - U, L - b, 0), \frac{d-b}{2}). \end{array} \tag{4.18}$$

Remark 4.15. It is common in the topological data analysis literature that the proposed functionals do not satisfy (4.14). Instead, it is assumed that all realizations of functionals have a compact support (Berry et al., 2020) or that all diagrams have uniformly bounded birth and death values (Chazal et al., 2014). While in some cases, such assumptions are compatible with the model for the data, we have not made such assumptions in Section 4.2.

We go back to Examples 2.41 and 2.42, which we adapt so that they satisfy the hypotheses above. The calculations of the Lipschitz constants are carried out in Appendix 4.G.

Example 4.16 (Persistence Silhouette)**.** The persistence silhouette (Chazal et al., 2014) is a weighted sum of landscape functions $\Lambda_{(b,d)}(t) = \left(\frac{d-b}{2} - |t - \frac{b+d}{2}|\right)_+$, for $\mathbb{T} = \mathbb{R}$. We set $k^s(b,d)(t) = \Lambda_{(\pi_{L,U}(b,d))}(t)$, so that $\mathrm{supp}(k^s(b,d)) \subset \mathrm{supp}(\Lambda^s(L,U)) = [L,U]$. Since $t \mapsto k^s(b,d)$ is piece–wise linear with slopes 0, 1 and -1, we have $L = 2 = L_k$. The kernel is zero on the diagonal, so $C = 0$ is enough to satisfy (4.17).

Example 4.17 (Persistence Image)**.** The kernel that would correspond to the persistence image (Adams et al., 2017) is $k^{pi}(b,d)(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(b-x)^2+(d-y)^2}{2\sigma^2}\right)$, for some $\sigma > 0$. We propose to modify this functional to have bounded support. Unfortunately, a simple truncation is not enough, because the kernel would not be continuous at the truncation interface. In order to preserve the Lipschitz character, we propose to multiply by the distance to a square of size 2σ to (b,d), namely, for some $r > 1$, set

$$k^{pi,r}(b,d)(x,y) = \left(2 - \frac{\|\pi_{L,U}(b,d) - (x,y)\|_\infty}{\sigma}\right)_+^r k^{pi}(\pi_{L,U}(b,d))(x,y)$$

Thus, the original persistence image kernel corresponds to $r = 0$ and $L = \infty, U = \infty$. The function $(x,y) \mapsto \exp(-(x^2+y^2))$ is $(4/e)$-Lipschitz and $(x,y) \mapsto \left(2 - \frac{\|(b,d)-(x,y)\|_\infty}{\sigma}\right)_+^r$ is $(r2^r/\sigma)$-Lipschitz, for the Minkowski distance. Hence, $L_{k^{pi,r}} = \frac{2^{r-1}}{\pi\sigma^3}(r+2)$ and $L = \frac{2^{r+1}}{\pi e\sigma^3}$.

Remark 4.18. We note a few differences with PersistenceCurves introduced in Chung and Lawson (2022). In that article, the aggregation operator can be different from the sum used here. However, the vectorizations are only curves, i.e $\mathbb{T} = \mathbb{R}$. In addition, for normalized functionals, the authors restrict themselves to kernels of the form $k(b,d)(t) = c1_{[b,d]}(t)$, for some $c > 0$.

Continuity of functionals has been studied, notably in Divol and Polonik (2019) and Chung and Lawson (2022). In the first, it was fully characterized, but only for linear functionals. In the latter, functionals were considered under the L^1 metric. Due to the nature of the statistical results in Section 4.2, we are particularly interested in $\|\cdot\|_\infty$, so we repeat the proof of (Divol and Polonik, 2019, Theorem 3) for linear functionals ρ and we derive results for normalized functionals $\bar{\rho}$.

Proposition 4.19. *Suppose that the persistence of any point in D_1 and D_2 is bounded by a uniform constant U and that k satisfies* (4.14)*,* (4.16) *and* (4.17)*. Then,*

$$\|\rho(D_1) - \rho(D_2)\|_\infty \leq \left(L_k \mathrm{pers}_{p,\epsilon}^p(D_1) + p(L_kU + C)\sum_{k=1,2}\mathrm{pers}_{p-1,\epsilon}^{p-1}(D_k)\right) d_B(D_1,D_2), \quad (4.19)$$

$$\|\bar{\rho}(D_1) - \bar{\rho}(D_2)\|_\infty \leq \left(L_k + 2p(L_kU + C)\frac{\mathrm{pers}_{p-1,\epsilon}^{p-1}(D_1) + \mathrm{pers}_{p-1,\epsilon}^{p-1}(D_2)}{\mathrm{pers}_{p,\epsilon}^p(D_1)}\right) d_B(D_1,D_2). \quad (4.20)$$

Proof. Let $\Gamma : D_1 \to D_2$ be a matching between the two diagrams. For any $t \in \mathbb{T}$,

$$\begin{aligned}
|\rho(D_1)(t) - \rho(D_2)(t)| &\leq \sum_{x\in D_1} w_\epsilon(x)^p |k(x)(t) - k(\Gamma(x))(t)| + k(\Gamma(x))(t)|w_\epsilon(x)^p - w_\epsilon(\Gamma(x))^p| \\
&\leq \sup_{x\in D_1}|k(x)(t) - k(\Gamma(x))(t)| \sum_{x\in D_1} w_\epsilon(x)^p \\
&\quad + \sup_{x\in D_1}|k(\Gamma(x))(t)| \sum_{x\in D_1}|w_\epsilon(x)^p - w_\epsilon(\Gamma(x))^p| \\
&\leq L_k d_B(D_1,D_2)\mathrm{pers}_{p,\epsilon}^p(D_1) \\
&\quad + p(L_kU + C)\sum_{x\in D_1}|w_\epsilon(x) - w_\epsilon(\Gamma(x))|(w_\epsilon(x)^{p-1} + w_\epsilon(\Gamma(x))^{p-1}),
\end{aligned}$$

where in the last inequality, we used that

$$\|k(\Gamma(x))\|_\infty \leq L_k \|(x_1, x_2) - (\tfrac{x_1+x_2}{2}, \tfrac{x_1+x_2}{2})\|_\infty + \|k\left(\tfrac{x_1+x_2}{2}, \tfrac{x_1+x_2}{2}\right)\|_\infty = L_k \tfrac{x_2-x_1}{2} + C.$$

The sum in the second term can be bounded from above by

$$\max_{x\in D_1} |w_\epsilon(x) - w_\epsilon(\Gamma(x))| \sum_{x\in D_1} (w_\epsilon(x)^{p-1} + w_\epsilon(\Gamma(x))^{p-1}) \leq d_B(D_1, D_2)(\mathrm{pers}^{p-1}_{p-1,\epsilon}(D_1) + \mathrm{pers}^{p-1}_{p-1,\epsilon}(D_2)).$$

Consider now the normalized version.

$$\begin{aligned}
|\bar\rho(D_1)(t) - \bar\rho(D_2)(t)| \leq & \frac{|\rho(D_1)(t) - \rho(D_2)(t)|}{\sum_{x\in D_1} w_\epsilon(x)^p} + \bar\rho(D_2) \frac{|\sum_{x\in D_1} w_\epsilon(x)^p - \sum_{y\in D_2} w_\epsilon(y)^p|}{\sum_{x\in D_1} w_\epsilon(x)^p} \\
\leq & d_B(D_1, D_2) \left(L_k + p(L_k U + C) \frac{\mathrm{pers}^{p-1}_{p-1,\epsilon}(D_1) + \mathrm{pers}^{p-1}_{p-1,\epsilon}(D_2)}{\mathrm{pers}^{p}_{p,\epsilon}(D_1)} \right) \\
& + p(L_k U + C) d_B(D_1, D_2) \frac{\mathrm{pers}^{p-1}_{p-1,\epsilon}(D_1) + \mathrm{pers}^{p-1}_{p-1,\epsilon}(D_2)}{\mathrm{pers}^{p}_{p,\epsilon}(D_1)} \\
\leq & \left(L_k + 2p(L_k U + C) \frac{\mathrm{pers}^{p-1}_{p-1,\epsilon}(D_1) + \mathrm{pers}^{p-1}_{p-1,\epsilon}(D_2)}{\mathrm{pers}^{p}_{p,\epsilon}(D_1)} \right) d_B(D_1, D_2).
\end{aligned}$$

Combine $\mathrm{pers}^{p-1}_{p-1,\epsilon}(D_1) + \mathrm{pers}^{p-1}_{p-1,\epsilon}(D_2) \leq 2\max_{k=1,2} \mathrm{pers}^{p-1}_{p-1,\epsilon}(D_k)$ with the observation that the bound is symmetric so that we can have $\mathrm{pers}^{p}_{p,\epsilon}(D_2)$ in the denominator. □

Remark 4.20. The result we give for ρ is a special case of (Divol and Polonik, 2019, Theorem 3). To see this, notice that using the notations of that article, $Lip(\phi) = L_k$, $A = p$, and $\alpha = p$, where 'p' is from our work. In their article, $p = \infty$ and $a = 1$. In particular, we see exactly that $\|k\|_\infty \leq L_k U + C$.

By Proposition 2.34, all points in the persistence diagram of a function f have birth value at least $\min f$ and a death value of at most $\max f$, so that $U = 2A(f)$ is sufficient. Using Proposition 4.9, we can conclude from (4.19) (resp. (4.20)) that ρ (resp. $\bar\rho$) is continuous with respect to the bottleneck distance, on the space of persistence diagrams. Via stability of the diagram with respect to the input from Proposition 2.38, it translates to continuity with respect to the input function.

In Section 4.2, we show convergence of functionals of time series. These results are a consequence of the bracketing entropy of $\mathcal{F} \coloneqq (\bar\rho_t)_{t\in K}, \bar\rho_t : \mathbb{R}^M \to \mathbb{R}$ being finite. It is a well-known result and consequence of (4.14, 4.15).

Proposition 4.21. *Let $N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$ denote the bracketing number of $\mathcal{F}$, with brackets $[u, l]$ of size $\|u - l\| \leq \epsilon$. Consider $\bar\rho$ as in* (4.2) *with k satisfying (4.14, 4.15). Then, for any probability measure P on $\mathbb{R}^M$ and $r \geq 1$,*

$$N_{[]}(\epsilon, \{\bar\rho_t\}_{t\in\mathbb{T}}, \|\cdot\|_{L_r(P)}) \leq \frac{2^{D+1} L^D \operatorname{diam}(K)}{\epsilon^D},$$

where D is the dimension of $\mathbb{T}$. As a consequence, the bracketing entropy $J_{[]}(\infty, \mathcal{F}, \|\cdot\|_{L_r(P)})$ is finite

$$J_{[]}(\infty, \mathcal{F}, \|\cdot\|_{L_r(P)}) \coloneqq \int_0^\infty \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_{L_r(P)})} d\epsilon < \infty.$$

Proof. First, since P is a probability measure, $\|\bar\rho_t\|_{L_r(P)} = \left(\int |\bar\rho_t|^r dP\right)^{1/r} \leq \|\bar\rho_t\|_\infty \int dP = \|\bar\rho_t\|_\infty$, so $N_{[]}(\epsilon, \{\bar\rho_t\}_{t\in\mathbb{T}}, \|\cdot\|_{L_r(P)}) \leq N_{[]}(\epsilon, \{\bar\rho_t\}_{t\in\mathbb{T}}, \|\cdot\|_\infty)$. Combining (4.15) with the fact that $\bar\rho(x)$ is a weighted average of k, for any $S \in \mathbb{R}^M$ and $s, t \in \mathbb{T}$, the normalized functional is L-Lipschitz in the following sense

$$|\bar\rho_t(S) - \bar\rho_s(S)| \leq L d(t, s).$$

Let K be given by (4.14). Then, (Kosorok, 2008, Theorem 9.22) states that

$$N_{[]}(2\epsilon L, \{\bar\rho_t\}_{t\in K}, \|\cdot\|_\infty) \leq N(\epsilon, K, d),$$

where $N(\epsilon, K, d)$ is the covering ϵ-number of (K, d). By assumption, $\mathbb{T}$ is of finite dimension that we will denote by D. By compacity of K, it has a finite diameter, say U. Therefore, $N(\epsilon, K, d) \leq \max(1, \frac{U}{\epsilon^D})$.

Let $t_0 \notin K$, $t_1 \in K$. We have that $\bar{\rho}_{t_1}$ is uniformly bounded,

$$|\bar{\rho}_{t_1}(S)| \leq |\bar{\rho}_{t_0}(S)| + Ld(t_0, t_1) = Ld(t_0, t_1),$$

so that $\bar{\rho}_{t_0} = 0 \in [\bar{\rho}_{t_1} - \epsilon L, \bar{\rho}_{t_1} - \epsilon L]$, for $\epsilon > d(t_0, t_1)$. The brackets in the proof of (Kosorok, 2008, Theorem 9.22) are of the form $[\bar{\rho}_t - \epsilon L, \bar{\rho}_t + \epsilon L]$, so that $N_{[]}(2\epsilon L, \{\bar{\rho}_t\}_{t\in\mathbb{T}}, \|\cdot\|_\infty) \leq N(\epsilon, K, d)$ for $\epsilon > d(t_0, t_1)$. In particular, one bracket is enough for $\epsilon > \max(U^{1/D}, d(t_0, t_1))$, while, for $\epsilon \leq \max(U^{1/D}, d(t_0, t_1))$, we have $N_{[]}(2\epsilon L, \{\bar{\rho}_t\}_{t\in\mathbb{T}}, \|\cdot\|_\infty) \leq 1 + N_{[]}(2\epsilon L, \{\bar{\rho}_t\}_{t\in K}, \|\cdot\|_\infty) \leq 1 + N(\epsilon, K, d) \leq 2N(\epsilon, K, d)$.

Finally, since $L_r(P)$ is dominated by $\|\cdot\|_\infty$ for any probability measure P,

$$\begin{aligned}
J_{[]}(\delta, \{\bar{\rho}_t\}_{t\in\mathbb{T}}, L_r(P)) &= \int_0^\delta \sqrt{\log(N_{[]}(\epsilon, \{\bar{\rho}_t\}_{t\in\mathbb{T}}, L_r(P)))}d\epsilon \\
&\leq \int_0^\delta \sqrt{\log(N_{[]}(\epsilon, \{\bar{\rho}_t\}_{t\in\mathbb{T}}, \|\cdot\|_\infty))}d\epsilon \\
&\leq \int_0^{\min(\delta, 2L\max(U^{1/D}, d(t_0,t_1)))} \sqrt{\log(N(\tfrac{\epsilon}{2L}, K, d))}d\epsilon \\
&\leq \int_0^{\min(\delta, 2L\max(U^{1/D}, d(t_0,t_1)))} \sqrt{\log\left(2^{D+1}UL^D\right) - \frac{1}{D}\log(\epsilon)}d\epsilon.
\end{aligned}$$

As $\lim_{\delta\to 0}\int_\delta^1 \sqrt{-\log(\epsilon)}d\epsilon < \infty$, we conclude that $J_{[]}(\delta, \{\bar{\rho}_t\}_{t\in\mathbb{T}}, L_r(P)) < \infty$. □

4.4 Numerical illustration

To illustrate the signatures and their stability, we propose to estimate the signatures of processes with different periodic functions. Then, we compare the estimate to the signature of a process with a different reparametrisation.

We will consider periodic functions ϕ_1 and ϕ_4 defined by

$$\phi_\theta = \theta(\sin(6\pi t) + |t - \lfloor t \rfloor - \tfrac{1}{2}| - \tfrac{1}{2}) + 5\sin(4\pi t), \qquad \text{for } \theta \in \mathbb{R}.$$

The observed signal follows the discrete model (4.7), with $T = 30$ and a sampling rate of 50Hz. The reparametrisations are generated by integrating twice a Markov chain of accelerations, with a truncated Gaussian transition kernel. The noise is a Gaussian process with covariance

$$\Gamma(s, t) = \sigma^2 \exp\left(-\frac{(s-t)^2}{2\tau^2}\right).$$

We fix the temporal scale τ, but we vary $\sigma = 0.1, 0.5, 2.$ to illustrate the impact of noise on the signature.

For $\bar{\rho}$, we take the silhouette introduced in Example 4.16, where the weights are the 0.2-truncated 1-persistence ($\epsilon = 0.2$, $p = 1$) and we use the projection $\pi_{-9,9}$ as in (4.18). We infer the signatures on 3-second windows ($M = 3 \cdot 50$). We construct the 1%-confidence intervals by resampling 200 times, with block lengths of 2 seconds ($L = 2 \cdot 50$).

In Figure 4.5, for the same random realization γ_1, we calculate the empirical signature $\hat{F}$ for ϕ_1 and ϕ_4, and estimate the corresponding confidence intervals for F. For low noise levels, the variance due to the number of observations and the variability in the endpoints is small, compared to the difference between the functionals. As the noise level increases, the observed function looses its recurrent appearance and the signatures become dominated by the noise.

Consider now two observations with the same periodic function ϕ_1, but different reparametrisations γ_1, γ_2. In Figure 4.6, we can see that for small values of noise, the signatures are close, what confirms their invariance to reparametrisation. It is worth noting that the signals contain different numbers of periods. For more noisy observations, the signatures lose the robustness.

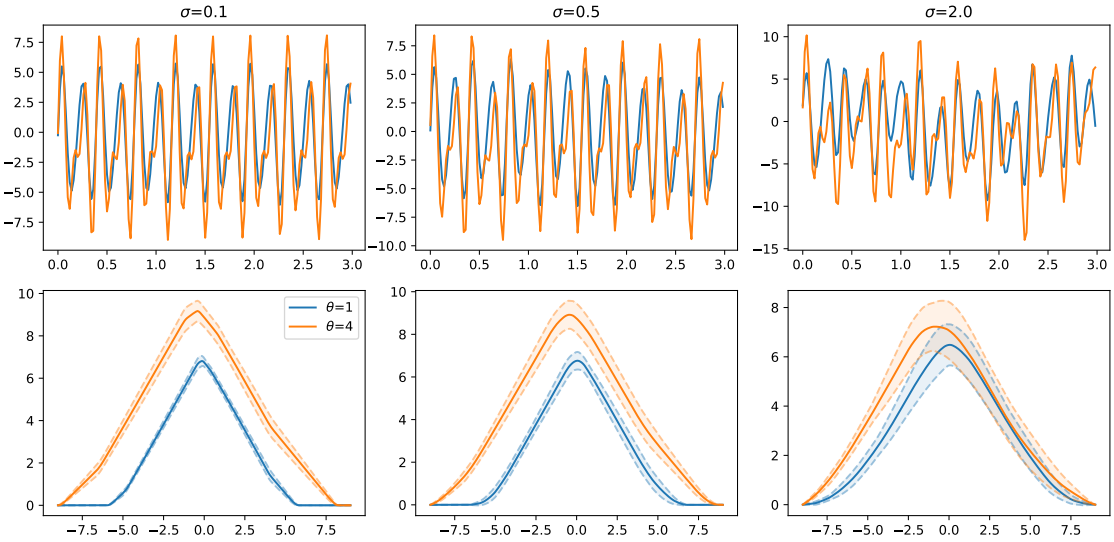


Figure 4.5: Signatures of ϕ_1 and ϕ_4, estimated on reparametrized signals described above. The top row shows the first 3-second window from the 30-second signal, for both functions. The bottom row shows the estimated signatures and the confidence intervals.

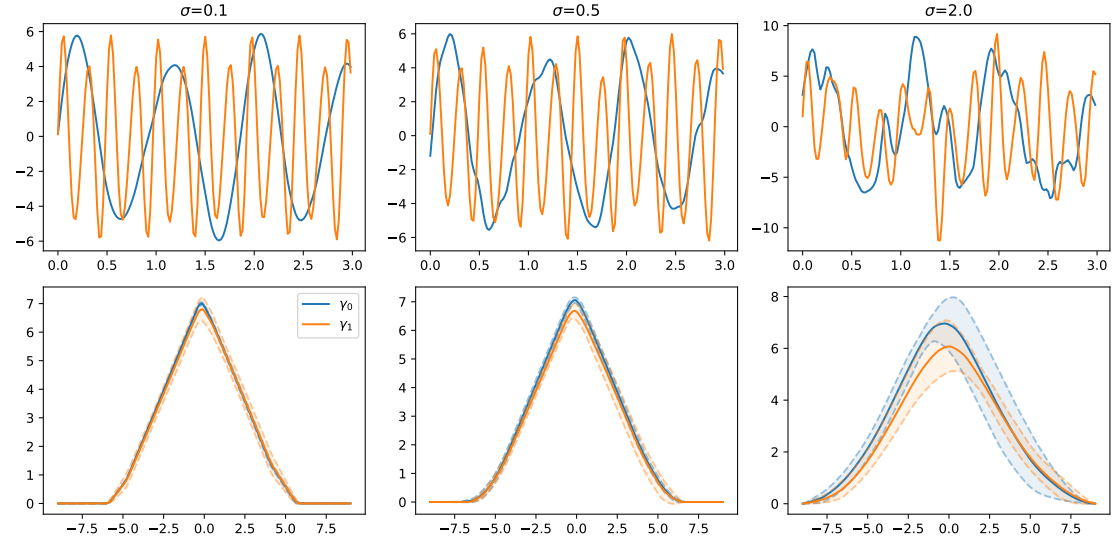


Figure 4.6: Signatures of ϕ_1, estimated on two different reparametrized observations. The top row shows the first 3-second window from the two observed signals. The bottom row shows the estimated signatures and the confidence intervals.

Appendix

4.A Measurability of functionals

The functionals are of the form $\bar{\rho}_t : C([0,T],\mathbb{R}) \to \mathbb{R}$, $t \in \mathbb{T}$, where the index set $\mathbb{T}$ is a (compact) metric space. Then, we will apply these functions pathwise and study the random variable $\bar{\rho}(S)$, where $\bar{\rho}$ is seen as a map $C([0,T],\mathbb{R})$ to $\mathbb{R}^{\mathbb{T}}$. Since $\bar{\rho}_t$ is applied pathwise, it is not obvious under what conditions $\bar{\rho}(S)$ is a random variable. Such considerations could be circumvented by using outer probabilities (Radulović, 1996, Kosorok, 2008), but we address them in Proposition 4.22.

As a stochastic process, $S : (\Omega, \mathcal{A}) \to (C([0,T], \sigma(\mathbb{R}^{[0,T]})))$ is a random variable on the measured space $(\Omega, \mathcal{A}, \eta)$, where $\sigma(\mathbb{R}^{[0,T]})$ is the σ-algebra generated by the product topology on $\mathbb{R}^{[0,T]}$ and η is the law of S. In our model, η is determined by ϕ, μ and ν.

Proposition 4.22. *Consider two independent random variables*

$$\gamma : (\Omega_r, \mathcal{A}_r) \to (C([0,T],\mathbb{R}), \sigma(\|\cdot\|_\infty)), \qquad W : (\Omega_n, \mathcal{A}_n) \to \Big(C([0,T],\mathbb{R}), \sigma(\mathbb{R}^{[0,T]})\Big),$$

and $S = \phi \circ \gamma + W$ as in (4.1). If $f : C([0,T],\mathbb{R}) \to C(\mathbb{T},\mathbb{R})$ is continuous and $C(\mathbb{T},\mathbb{R})$ is $\|\cdot\|_\infty$-separable, then $f(S)$ is $(C(\mathbb{T},\mathbb{R}), \|\cdot\|_\infty)$-measurable.

Proof (proposition 4.22). First, assume that S is weakly-measurable on $E = C([0,T],\mathbb{R})$ and that $(C([0,T],\mathbb{R}), \|\cdot\|_\infty)$ is separable. Using lemma 2.2, we get that S is $\sigma(\|\cdot\|_\infty)$-measurable. Because $f : C([0,T],\mathbb{R}) \to C(\mathbb{T},\mathbb{R})$ is continuous, it is measurable for the two σ-algebras on the domain and co-domain. This allows us to conclude that $f(S)$ is $(C(\mathbb{T},\mathbb{R}), \sigma(\|\cdot\|_\infty))$-measurable.

Let us now verify the assumptions of Lemma 2.2. By continuity of ϕ, the composition $\phi \circ \gamma$ is $\sigma(\mathbb{R}^{[0,T]})$-measurable. As a sum of two (independent) random variables, $S = \phi \circ \gamma + W$ is $\sigma(\mathbb{R}^{[0,T]})$-measurable for $(\Omega, \mathcal{A})$, where $\Omega = \Omega_r \times \Omega_n$ and $\mathcal{A} = \mathcal{A}_r \otimes \mathcal{A}_n$. The product σ-algebra $\sigma(\mathbb{R}^{[0,T]})$ coincides with that of weak measurability on $\mathbb{R}^{[0,T]}$. The space $C([0,T],\mathbb{R})$ with the topology induced by $\|f\|_\infty := \sup_{x\in[0,T]} |f(x)|$ is a Banach, separable space. Any subspace of a separable metric space is separable, so $S(\Omega)$ is also separable. □

4.B Invariance of the signature to reparametrisation

Consider $(C([0,T],\mathbb{R}), \|\cdot\|_\infty)$ with the Borel σ-algebra. We assume that μ_1, μ_2 are Borel measures on the restriction of that σ-algebra to a closed subspace $\Gamma \subset C([0,T])$. We denote by $\delta_t : \gamma \mapsto \gamma(t)$ the evaluation map, we let $(\delta_t)_\star \mu_1 = \mu_1 \circ (\delta_t)^{-1}$ be the measure which characterizes the marginal distribution of $\gamma_1(t)$ and we proceed similarly for μ_2. Note that the evaluation is measurable, as it corresponds to weak-measurability. Similarly, we denote $\delta_{0,T} = (\delta_0, \delta_T) : \gamma \mapsto (\gamma(0), \gamma(T)) \in \mathbb{R}^2$.

Proposition 4.23. *If the marginals $(\delta_{0,T})_\star \mu_1$ and $(\delta_{0,T})_\star \mu_2$ are equal, then*

$$F(\phi \circ \gamma_1) = F(\phi \circ \gamma_2).$$

Proof. We first need to show that we can condition on $(\gamma(0), \gamma(T))$. The space of continuous functions $C([0,T],\mathbb{R})$ is Polish, and so is Γ, because it is a closed subspace. Let $\mathcal{A} = \sigma(\delta_{0,T})$ be the σ-algebra generated by the evaluations. By (Bogachev, 2007, Corollary 10.4.6), there is a regular conditional measure $((\mu_1)_x(d\gamma))_{x\in\mathbb{R}^2}$.

Proposition 2.35 implies that $\gamma \mapsto \bar{\rho}_t(\phi \circ \gamma)$ is constant on $\delta_{0,T}^{-1}(x)$, for any $x = (s, r) \in \mathbb{R}^2$. For any $t \in \mathbb{T}$, using the regular conditional measure property (Bogachev, 2007, Definition 10.4.1),

$$\begin{aligned} F_t(\phi \circ \gamma_1) &= \int_\Gamma \bar{\rho}_t(\phi \circ \gamma)\mu_1(d\gamma) \\ &= \int_{\mathbb{R}^2} \int_{\delta_{0,T}^{-1}(x)} \bar{\rho}_t(\phi \circ \gamma)(\mu_1)_x(d\gamma)(\delta_{0,T})_\star \mu_1(dx) \\ &= \int_{\mathbb{R}^2} \int_{\delta_{0,T}^{-1}(x)} \bar{\rho}_t(\phi \circ \gamma)(\mu_2)_x(d\gamma)(\delta_{0,T})_\star \mu_2(dx) \\ &= F_t(\phi \circ \gamma_2). \end{aligned}$$

□

Since γ_1, γ_2 are reparametrisations, we require Γ to be included in the space of injective functions. An example is given in (4.6).

While it is disappointing to require equality of the marginals $(\delta_{0,T})_\star \mu_1$ and $(\delta_{0,T})_\star \mu_2$ in Proposition 4.23, removing this assumption poses a difficulty which we now discuss. Consider γ_1 and γ_2 fixed, assume that $R \coloneqq R_1 < R_2$ and let $T_1 = \gamma_2^{-1}(R)$. For continuous functions on an interval, we can only control the stability of the persistence diagram in the bottleneck distance d_B (see Proposition 2.38).

As an example, consider the case when $R_2 = R+1$. Then, the distance between the persistence diagrams $d_B(D(\phi \circ \gamma_1), D(\phi \circ \gamma_2))$ is of the order of the amplitude $A(\phi) \coloneqq \max \phi - \min \phi$, as the multiplicity of the point $(\min \phi, \max \phi)$ differs by at least one between both diagrams. The term $\frac{\mathrm{pers}_{p,\epsilon}^p(\phi\circ\gamma_1)+\mathrm{pers}_{p,\epsilon}^p(\phi\circ\gamma_2)}{\mathrm{pers}_{p,\epsilon}^p(\phi\circ\gamma_1)}$ is roughly constant ($1 \leq \frac{R_1+R_2}{R_1} \leq 3$). Therefore, the fact that the difference between $\bar{\rho}(\phi \circ \gamma_1)$ and $\bar{\rho}(\phi \circ \gamma_2)$ will be small is not reflected by Proposition 4.19 which gives a trivial bound. Instead, let $D_1 = D((\phi \circ \gamma_2)|_{[0,T_1]})$, $D_2 = D((\phi \circ \gamma_2)|_{[T_1,T]})$ and consider

$$\|\bar{\rho}(\phi \circ \gamma_1) - \bar{\rho}(\phi \circ \gamma_2)\|_\infty \leq \|\bar{\rho}(D_1) - \bar{\rho}(D_1 \sqcup D_2)\|_\infty + \|\bar{\rho}(D_1 \sqcup D_2) - \bar{\rho}(\phi \circ \gamma_2)\|_\infty. \tag{4.21}$$

Conveniently, a normalized functional of a union of diagrams is a weighted average of the normalized functionals of the individual diagrams

$$\bar{\rho}(D_1 \sqcup D_2)(t) = \bar{\rho}(D_1)(t)\frac{\mathrm{pers}_{p,\epsilon}^p(D_1)}{\mathrm{pers}_{p,\epsilon}^p(D_1 \sqcup D_2)} + \bar{\rho}(D_2)(t)\frac{\mathrm{pers}_{p,\epsilon}^p(D_2)}{\mathrm{pers}_{p,\epsilon}^p(D_1 \sqcup D_2)},$$

so that

$$\begin{aligned} \left|\bar{\rho}(D_1)(t) - \bar{\rho}(D_1 \sqcup D_2)(t)\right| &= \left|\bar{\rho}(D_1)(t)\left(\frac{\mathrm{pers}_{p,\epsilon}^p(D_1)}{\mathrm{pers}_{p,\epsilon}^p(D_1 \sqcup D_2)} - 1\right) + \bar{\rho}(D_2)(t)\frac{\mathrm{pers}_{p,\epsilon}^p(D_2)}{\mathrm{pers}_{p,\epsilon}^p(D_1 \sqcup D_2)}\right| \\ &= |\bar{\rho}(D_1)(t) - \bar{\rho}(D_2)(t)|\frac{\mathrm{pers}_{p,\epsilon}^p(D_2)}{\mathrm{pers}_{p,\epsilon}^p(D_1 \sqcup D_2)} \\ &\leq (L_k A_\phi + C)\frac{\mathrm{pers}_{p,\epsilon}^p(D_2)}{\mathrm{pers}_{p,\epsilon}^p(D_1 \sqcup D_2)} \\ &\leq (L_k A_\phi + C)\frac{\mathrm{pers}_{p,\epsilon}^p(D_2)}{\mathrm{pers}_{p,\epsilon}^p(D_1)}, \end{aligned}$$

what is on the order of $\mathcal{O}(R_2/R - 1)$.

The second term in (4.21) is more problematic. Thanks to Proposition 4.19, it is the error made when approximating the diagram of $\phi|_{[0,R_2]}$ by the union of diagrams of $\phi|_{[0,R]}$ and $\phi|_{[R,R_2]}$. For a particularly good cutting point R, that is, when R is a global maximum, $D_1 \sqcup D_2 = D(\phi \circ \gamma_2)$. However, in general, the support of the union of diagrams differs from the diagram of the whole interval. To show stability, we miss the study of $d_B(D_1 \sqcup D_2, D(\phi \circ \gamma_2))$. A possible avenue is given by the tools introduced in Herbert Edelsbrunner et al. (2023).

4.C Proof of Theorem 4.3

We start by treating S path–wise. Using Proposition 4.19 and the bottleneck stability of persistence diagrams,

$$\begin{aligned}\|\bar{\rho}(\phi\circ\gamma_1+W)-\bar{\rho}(\phi\circ\gamma_2+W)\| &= \|\bar{\rho}(\phi+W_{\gamma_1^{-1}})-\bar{\rho}(\phi+W_{\gamma_1^{-1}})\| \\ &\leq L_k\left(1+4pU\max_{k=1,2}\frac{\mathrm{pers}^{p-1}_{p-1,\epsilon}(\phi+W_{\gamma_k^{-1}})}{\mathrm{pers}^{p}_{p,\epsilon}(\phi+W_{\gamma_k^{-1}})}\right)\|W_{\gamma_1^{-1}}-W_{\gamma_2^{-1}}\|_\infty,\end{aligned} \tag{4.22}$$

where L_k is a regularity constant of the kernel and U is an upper-bound on the persistence of any point in both diagrams. The persistence of any point in the diagram $D(h)$ of a function h is bounded by A_h. Hence, the persistence of a point in $D(\phi+W)$ is bounded by $U=A_{\phi+W}\leq A_\phi+A_W\leq A_\phi+(A_\phi-\epsilon-q)\leq 2A_\phi$.

Next, we obtain an upper–bound of $\max_{k=1,2}\frac{\mathrm{pers}^{p-1}_{p-1,\epsilon}(\phi+W_{\gamma_k^{-1}})}{\mathrm{pers}^{p}_{p,\epsilon}(\phi+W_{\gamma_k^{-1}})}$. By Proposition 2.6, we can assume that W has α-Hölder paths with a (random) constant Λ_W, for $\alpha\coloneqq\frac{\min(1,r_1-1)}{r_2}$. This implies that $\frac{1}{\alpha}+1<p$ and we use the continuity of truncated persistence from Proposition 4.12 to obtain

$$\mathrm{pers}^{p-1}_{p-1,\epsilon}(\phi+W_{\gamma_k^{-1}})\leq \mathrm{pers}^{p-1}_{p-1,\epsilon}(\phi|_{[0,T]})+(p-1)\|W\|_\infty(\mathrm{pers}^{p-2}_{p-2,\epsilon}(\phi|_{[0,T]})+\mathrm{pers}^{p-2}_{p-2,\epsilon}(W_{\gamma_k^{-1}})). \tag{4.23}$$

For any $x\in[0,1]$ and $p\geq 0$, the function $p\mapsto x^p$ is decreasing, so that

$$\begin{aligned}\mathrm{pers}^{p-1}_{p-1,\epsilon}(\phi|_{[0,T]}) &= (A_\phi-\epsilon)^{p-1}\sum_{(b,d)\in D}\max\left(\tfrac{d-b-\epsilon}{A_\phi-\epsilon},0\right)^{p-1} \\ &\leq (A_\phi-\epsilon)^{p-1}\sum_{(b,d)\in D}\max\left(\tfrac{d-b-\epsilon}{A_\phi-\epsilon},0\right)^{p-2} \\ &= (A_\phi-\epsilon)\mathrm{pers}^{p-2}_{p-2,\epsilon}(\phi).\end{aligned}$$

Since $\|W\|_\infty<(A_\phi-\epsilon)/2$ and the persistence does not depend on the parametrisation, equation (4.23) becomes

$$\begin{aligned}\mathrm{pers}^{p-1}_{p-1,\epsilon}(\phi+W_{\gamma_k^{-1}}) &\leq (A_\phi-\epsilon)\mathrm{pers}^{p-2}_{p-2,\epsilon}(\phi)\left(1+\tfrac{p-1}{2}\left(1+\frac{\mathrm{pers}^{p-2}_{p-2,\epsilon}(W)}{\mathrm{pers}^{p-2}_{p-2,\epsilon}(\phi)}\right)\right) \\ &\leq p(A_\phi-\epsilon)\mathrm{pers}^{p-2}_{p-2,\epsilon}(\phi)\left(1+\tfrac{1}{2}\frac{\mathrm{pers}^{p-2}_{p-2,\epsilon}(W)}{\mathrm{pers}^{p-2}_{p-2,\epsilon}(\phi)}\right).\end{aligned}$$

An upper–bound for the persistence of W is given in Proposition 4.10

$$\mathrm{pers}^{p}_{p,\epsilon}(W)\leq(A_W-\epsilon)^p\left(1+pT\left(\tfrac{2\Lambda_W}{\epsilon}\right)^{1/\alpha}\right),$$

where Λ_W is the path–wise Hölder constant of W. The amplitude A_ϕ upper–bounds the persistence of a point and it is also realized as the persistence of a pair of a global minimum and a global maximum, so $\mathrm{pers}^{p-2}_{p-2,\epsilon}(\phi|_{[0,R]})\geq(R-2)(A_\phi-\epsilon)^{p-2}$ and hence

$$\frac{\mathrm{pers}^{p}_{p,\epsilon}(W)}{\mathrm{pers}^{p-2}_{p-2,\epsilon}(\phi)}\leq\left(\tfrac{A_W-\epsilon}{A_\phi-\epsilon}\right)^{p-2}(A_W-\epsilon)^2\tfrac{T}{R-2}\left(1+p\left(\tfrac{2\Lambda_W}{\epsilon}\right)^{1/\alpha}\right).$$

Putting the above together, with $p\geq 2$,

$$\begin{aligned}\mathrm{pers}^{p-1}_{p-1,\epsilon}(\phi+W_{\gamma_k^{-1}}) &\leq p(A_\phi-\epsilon)\mathrm{pers}^{p-2}_{p-2,\epsilon}(\phi) \\ &\quad\times\left(1+\left(\tfrac{A_W-\epsilon}{A_\phi-\epsilon}\right)^{p-2}(A_W-\epsilon)^2\tfrac{T}{R-2}\max\left(1,p\left(\tfrac{2\Lambda_W}{\epsilon}\right)^{1/\alpha}\right)\right).\end{aligned}$$

We have therefore an upper–bound for the numerator. To lower–bound the denominator, we use Proposition 4.13:

$$\begin{aligned}\mathrm{pers}^p_{p,\epsilon}(\phi + W_{\gamma_k^{-1}}) &\geq \mathrm{pers}^p_{p,\epsilon+A_W}(\phi)\\ &\geq (R-2)(A_\phi - (A_W+\epsilon))^p\\ &\geq (R-2)(A_\phi - (A_\phi - \epsilon + q + \epsilon))^p = (R-2)q^p.\end{aligned}$$

We conclude that C_{Λ_W} upper-bounds $\max_k \frac{\mathrm{pers}^{p-1}_{p-1,\epsilon}(\phi+W_{\gamma_k^{-1}})}{\mathrm{pers}^p_{p,\epsilon}(\phi+W_{\gamma_k^{-1}})}$,

$$\begin{aligned}C_{\Lambda_W} \coloneqq L_k \Big(1 + \tfrac{8p^2A_\phi}{(R-2)q^p}(A_\phi-\epsilon)\mathrm{pers}^{p-2}_{p-2,\epsilon}(\phi)\\ \times \left(1 + \left(\tfrac{A_W-\epsilon}{A_\phi-\epsilon}\right)^{p-2}(A_W-\epsilon)^2\tfrac{T}{R-2}\max\left(1, p\left(\tfrac{2\Lambda_W}{\epsilon}\right)^{1/\alpha}\right)\right)\Big).\end{aligned}$$

As $A_W \leq A_\phi - \epsilon - q$, the only remaining stochastic term in C_{Λ_W} is $\Lambda_W^{1/\alpha}$. Also, the bound only depends on R (which is fixed), but not on γ itself.

Let $\pi : \mathcal{A}_{r,1} \times \mathcal{A}_{r,2} \to \mathbb{R}$ be a coupling of μ_1 and μ_2. Specifically, π is a measure on the product space $(\mathcal{G}\times\mathcal{G}, \mathcal{A}_{r,1}\otimes\mathcal{A}_{r,2})$, such that $\pi(A,\mathcal{G}) = \mu_1(A)$ and $\pi(\mathcal{G},A) = \mu_2(A)$, for all $A \in \mathcal{A}$. Then, $\pi\otimes\nu : ((A_1,B_1),(A_2,B_2)) \mapsto \pi(A_1,A_2)\nu(B_1\cap B_2)$ is a coupling of $\mu_1\otimes\nu$ and $\mu_2\otimes\nu$. Using the coupling and (4.22),

$$\begin{aligned}|\mathbb{E}[\bar\rho(\phi\circ\gamma_1 + W) \mid W] - \mathbb{E}[\bar\rho(\phi\circ\gamma_2 + W)\mid W]| &= |\mathbb{E}_\pi[\bar\rho(\phi\circ\gamma_1+W) - \bar\rho(\phi\circ\gamma_2+W)\mid W]|\\ &\leq \mathbb{E}_\pi\left[|\bar\rho(\phi\circ\gamma_1+W) - \bar\rho(\phi\circ\gamma_2+W)| \mid W\right]\\ &\leq C_{\Lambda_W}\mathbb{E}[\|W_{\gamma_1^{-1}} - W_{\gamma_2^{-1}}\|_\infty \mid W],\\ &\leq C_{\Lambda_W}\Lambda_W\mathbb{E}[\|\gamma_1^{-1} - \gamma_2^{-1}\|_\infty^\alpha].\end{aligned}$$

We have thus completely separated the bound into a product, with terms depending on ν and (μ_1,μ_2).

On one hand, it remains to take the expectation with respect to W. We bound the moments of Λ_W using Theorem 4.32, obtaining

$$\begin{aligned}\mathbb{E}[\Lambda_W] &\leq 16\tfrac{\alpha+1}{\alpha}(K_{r_2,r_1})^{1/r_2}\\ \mathbb{E}[\Lambda_W^{1+1/\alpha}] &\leq 6^{r_2+2}K_{r_2,r_1}^{(1/r_2+1/(r_1-1))}.\end{aligned}$$

On the other hand, by Jensens' inequality, $\mathbb{E}[\|\gamma_1^{-1}-\gamma_2^{-1}\|_\infty^\alpha] \leq \mathbb{E}[\|\gamma_1^{-1}-\gamma_2^{-1}\|_\infty]^\alpha$. Using the lower–bound on the modulus of continuity,

$$\sup_{r\in[0,R]}|\gamma_1^{-1}(r) - \gamma_2^{-1}(r)| = \sup_{t\in[0,T]}|t - \gamma_2^{-1}(\gamma_1(t))| \leq \sup_{t\in[0,T]}\frac{1}{v_{\min}}|\gamma_2(t)-\gamma_1(t)|.$$

Taking the infimum over couplings, we obtain the 1-Wasserstein distance $W_1(\mu_1,\mu_2)$.

4.D Proof of Proposition 4.5

We start by proving a lemma.

Lemma 4.24 (Perturbed, path–wise version)**.** *Consider $W \in C^\alpha_\Lambda([0,T],\mathbb{R})$ and set $\delta \coloneqq \|W\|_\infty$. If $2\delta \leq \max\phi - \min\phi$, then*

$$\|\bar\rho(\phi+W) - \bar\rho(\phi)\|_\infty \leq L_k(P_1\delta + P_2\delta^2 + P_3\delta^3) =: L_kP(\delta),$$

where

$$\begin{aligned}P_1 &= 1 + 4A_\phi C_T C^\epsilon_{p-1,p}(\phi),\\ P_2 &= 8C_TC^\epsilon_{p-1,p}(\phi) + 4pA_\phi(C_TC^\epsilon_{p-2,p}(\phi) + \tfrac{C_{p-3,\Lambda,\alpha,T}}{\mathrm{pers}^p_{p,\epsilon}(\phi)}),\\ P_3 &= 4p\left(C_TC^\epsilon_{p-2,p}(\phi) + \tfrac{C_{p-3,\Lambda,\alpha,T}}{\mathrm{pers}^p_{p,\epsilon}(\phi)}\right),\end{aligned}$$

and

$$C_T = \frac{\lceil T\rceil}{\lfloor T\rfloor - 2}, \qquad C^\epsilon_{p,p'}(\phi) = \frac{\mathrm{pers}^p_{p,\epsilon}(\phi)}{\mathrm{pers}^{p'}_{p',\epsilon}(\phi)}, \qquad A_\phi = \|\phi\|_\infty.$$

Proof. By the diagram stability theorem, $d_B(D(\phi+W), D(\phi)) \leq \|W\|_\infty \leq \delta$. The persistence of a point in $D(\phi)$ and $D(\phi+W)$ is bounded by $2A_\phi$ and $2A_{\phi+W} \leq 2(A_\phi+\delta)$ respectively. Using Proposition 4.12, we also bound $\mathrm{pers}^{p-1}_{p-1,\epsilon}(\phi+W) \leq \mathrm{pers}^{p-1}_{p-1,\epsilon}(\phi)+p\delta(\mathrm{pers}^{p-2}_{p-2,\epsilon}(\phi)+\mathrm{pers}^{p-2}_{p-2,\epsilon}(W))$. Using the uniform bound on persistence from Proposition 4.10, $\mathrm{pers}^{p-2}_{p-2,\epsilon}(W) \leq C_{p-3,\Lambda,\alpha,T}$. Finally, putting these together with Proposition 4.19, we obtain:

$$\begin{aligned}
\|\bar\rho(\phi) - \bar\rho(\phi+W)\| &\leq L_k \left(1 + 2pU \frac{\mathrm{pers}^{p-1}_{p-1,\epsilon}(\phi) + \mathrm{pers}^{p-1}_{p-1,\epsilon}(\phi+W)}{\mathrm{pers}^{p}_{p,\epsilon}(\phi)}\right) d_B(D(\phi), D(\phi+W)) \\
&\leq \delta L_k \left(1 + 4p(\|\phi\|_\infty+\delta)\frac{2\lceil T\rceil \mathrm{pers}^{p-1}_{p-1,\epsilon}(\phi|_{[c,c+1]}) + p\delta(\mathrm{pers}^{p-2}_{p-2,\epsilon}(\phi)+C_{p-3,\Lambda,\alpha,T})}{(\lfloor T\rfloor -2)\mathrm{pers}^{p}_{p,\epsilon}(\phi)}\right) \\
&\leq \delta L_k \Big[\big(1 + 4A_\phi C_T C^\epsilon_{p-1,p}(\phi)\big) + \\
&\qquad \Big(8C_T C^\epsilon_{p-1,p}(\phi) + 4pA_\phi(C_T C^\epsilon_{p-2,p}(\phi) + \tfrac{C_{p-3,\Lambda,\alpha,T}}{\mathrm{pers}^{p}_{p,\epsilon}(\phi)})\Big)\delta^1 + \\
&\qquad 4p\left(C_T C^\epsilon_{p-2,p}(\phi) + \tfrac{C_{p-3,\Lambda,\alpha,T}}{\mathrm{pers}^{p}_{p,\epsilon}(\phi)}\right)\delta^2\Big].
\end{aligned}$$

□

Proof of Proposition 4.5. Combining lemma 4.24 and theorem 4.2,

$$\begin{aligned}
\|\bar\rho(\phi\circ\gamma_1 + W_1) - \bar\rho(\phi\circ\gamma_2+W_2)\| \;&\leq\; \|\bar\rho(\phi + (W_1)_{\gamma_1^{-1}}) - \bar\rho(\phi|_{[0,R_1]})\| \\
&+\; \|\bar\rho(\phi|_{[0,R_1]}) - \bar\rho(\phi|_{[0,R_2]})\| \\
&+\; \|\bar\rho(\phi|_{[0,R_2]}) - \bar\rho(\phi + (W_2)_{\gamma_2^{-2}})\| \\
&\leq\; L_k(P(\delta_1)+P(\delta_2)+2\tfrac{4}{\min(R_1,R_2)}\bar\rho(\phi|_{[c,c+1]})) \\
&\leq\; L_k\left(P(\delta_1)+P(\delta_2)+\tfrac{8}{\min(R_1,R_2)-2}\tfrac{A_\phi}{2}\right) \\
&\leq\; L_k\left(P(\max(\delta_1,\delta_2))+\tfrac{4A_\phi}{\min(R_1,R_2)-2}\right).
\end{aligned}$$

□

4.E Exponential mixing of the reparametrisation process

Proposition 4.25. *Consider $(\gamma_n)_{n=1}^N$ as in (4.8) with $(V_n)_{n=1}^N$ as in Model 1 or 2. Then, $\beta_{\mathrm{frac}(\gamma)}(k) \to 0$ exponentially fast.*

The proof of this proposition relies on the continuity of the transition kernel with respect to the Lebesgue measure and the use of Theorem 2.17. In order to apply it, we need to verify (2.4) called a Doeblin condition. It consists in providing a non-trivial lower–bound on the family of measures $(P^r(z,A))_z$. We first treat the case where $(V_n)_{n\in\mathbb{N}}$ are all i.i.d. The case where $(V_n)_{n\in\mathbb{N}}$ is a Markov Chain is similar, but technically more difficult.

4.E.1 Model 1

In Model 1, $(\gamma_n)_{n\in\mathbb{N}}$ is a Markov chain. Indeed, $\gamma_n = \gamma_{n-1} + V_{n-1}$, where V_n is independent from $(V_k)_{k<n}$ and γ_0. We will now verify (2.4). Let $r \coloneqq \lceil 2/(b-a)\rceil$ and $\epsilon = \lfloor \frac{b-a}{r}\rfloor$.

Lemma 4.26. *Consider two measures μ_1, μ_2 such that $\mu_k(A) \geq c_k\mu(A)$, for $A \in \mathcal{B}([a_k,b_k])$. Then, for any $0<\epsilon<\min(b_1-a_1, b_2-a_2)$, we have that $(\mu_1\star\mu_2)(A) \geq c_1c_2\epsilon\mu(A)$, for any $A \in \mathcal{B}([a_1+a_2+\epsilon, b_1+b_2-\epsilon])$.*

We now apply this Lemma 4.26 inductively to μ_1 and μ_2 the measures of $\sum_{n=1}^{r_1} V_n$ and V_{r_1+1} respectively, for $1 \leq r_1 \leq r-1$. We conclude that $P(\sum_{n=1}^r V_n \in A) \geq c\mu(A)$ for all $A \in \mathcal{B}(B)$, where $B \coloneqq [r(a+\epsilon)-\epsilon, r(b-\epsilon)+\epsilon]$ and $c \coloneqq c_1c_2\epsilon^{r-1}$. Thanks to our choice of r and ϵ, B is an interval of length at least 1.

Let $x_0 \in [0,1[$ and $A \in \mathcal{B}([0,1])$. We write $\mathrm{frac}^{-1}(A) = \cup_{k\in\mathbb{Z}} A+k$, where $A+k = \{a+k \mid a \in A\}$. Then,

$$\begin{aligned}
P(\mathrm{frac}(\gamma_r) \in A \mid \gamma_0 = x_0) &= P\left(x_0 + \sum_{n=0}^{r} V_n \in \mathrm{frac}^{-1}(A)\right) \\
&= P\left(\sum_{n=0}^{r} V_n \in \bigcup_{k\in\mathbb{Z}} (A+k) - x_0\right) \\
&\geq P\left(\sum_{n=0}^{r} V_n \in \bigcup_{k\in\mathbb{Z}} (A+k-x_0) \cap B\right) \\
&\geq c\mu\left(\bigcup_{k\in\mathbb{Z}} (A+k-x_0) \cap B\right) \\
&= c\sum_k \mu(A+k-x_0 \cap B),
\end{aligned}$$

where the last equality follows from the fact that $\mu(A \cap (A+1)) = 0$, because $A \subset [0,1]$. Notice that for any set $A + z \cap B = (A \cap (B-z)) + z$ and that $\mu(A+z) = \mu(A)$, for any $z \in \mathbb{R}$. Hence, for any $k \in \mathbb{Z}$,

$$\mu\left(A+k-x_0 \cap B\right) = \mu(k - x_0 + (A \cap (B-k+x_0))) = \mu(A \cap (B-k-x_0)).$$

Recall that B is an interval of length greater than 1, so $(B-k-x_0)_{k\in\mathbb{Z}}$ is a cover of $\mathbb{R}$. Hence,

$$\begin{aligned}
P(\mathrm{frac}(\gamma_r) \in A \mid \gamma_0 = x_0) &\geq c\sum_k \mu(A \cap (B-k-x_0)) \\
&\geq c\mu\left(A \cap \bigcup_k (B-k-x_0)\right) \\
&= c\mu(A).
\end{aligned}$$

We can therefore set $\mu \coloneqq c\mu$. The measure does not depend on x_0 and it has total mass $c > 0$.

We now show that $(\mathrm{frac}(\gamma_n))_{n\in\mathbb{N}}$ is strictly stationary: for any $K \in \mathbb{N}^*$, $\tau \in \mathbb{N}$ and $n_1, \dots n_K$, the vectors $(\mathrm{frac}(\gamma_{n_1}), \dots, \mathrm{frac}(\gamma_{n_K})) \sim (\mathrm{frac}(\gamma_{n_1+\tau}), \dots, \mathrm{frac}(\gamma_{n_K+\tau}))$, where $X \sim Y$ is a shorthand notation for "X and Y have the same distribution". It is enough to show that for any $K \geq 1$, $(\mathrm{frac}(\gamma_0), \dots, \mathrm{frac}(\gamma_K)) \sim (\mathrm{frac}(\gamma_n), \dots \mathrm{frac}(\gamma_{n+K}))$, for any $n \geq 0$. We write

$$(\mathrm{frac}(\gamma_n), \dots \mathrm{frac}(\gamma_{n+K})) = \mathrm{frac}(\mathrm{frac}(\gamma_0 + \sum_{r=0}^{n-1} V_r) + \mathrm{frac}(0, V_n, \dots, \sum_{r=n}^{n+K-1} V_r)),$$

and we analyze the two terms separately. Here, frac is applied component–wise. First, because $(V_n)_{n\in\mathbb{N}}$ are i.i.d, $\left(\sum_{r=0}^{k} V_r\right) \sim \left(\sum_{r=n}^{n+k} V_r\right)$, for any $n, k \in \mathbb{N}$. Therefore, $(0, V_0, \dots, \sum_{r=0}^{n-1} V_r) \sim (0, V_n, \dots, \sum_{r=n}^{n+K-1} V_r)$. It also remains true when we apply frac component–wise, because it is a measurable mapping $\mathbb{R}^{K+1} \to \mathbb{R}^{K+1}$. Second, we claim the following lemma on the sum of two random variables, one of which is uniform.

Lemma 4.27. *If $U \sim \mathcal{U}([0,1])$ and Z is a real–valued random variable independent of U, then* $\mathrm{frac}(U+Z) \sim \mathrm{frac}(U) \sim U$.

Before showing Lemma 4.27, we conclude the proof by applying it to $U = \gamma_0$ and $Z = \sum_{r=0}^{n-1} V_r$. Indeed, γ_0 is independent from $(V_r)_{r=0}^{n-1}$, so we obtain that $\mathrm{frac}(\gamma_0) \sim \mathrm{frac}(\gamma_0 + \sum_{r=0}^{n-1} V_r)$. Finally, combining the above with $\mathrm{frac}((0, V_0, \dots, \sum_{r=0}^{n-1} V_r)) \sim \mathrm{frac}((0, V_n, \dots, \sum_{r=n}^{n+K-1} V_r))$, we have that $\mathrm{frac}(\gamma_0, \dots, \gamma_K) \sim \mathrm{frac}(\gamma_n, \dots, \gamma_{n+K})$.

Proof of Lemma 4.27. First, it is clear that for $s \leq 0$, $P(\mathrm{frac}(U+Z) < s) = 0$ and that for $s > 1$, $1 \geq P(\mathrm{frac}(U+Z) < s) \geq P(\mathrm{frac}(U+Z) \leq 1) = 1$. For $0 < s < 1$,

$$P(\mathrm{frac}(U+Z) \leq s) = P\left(U + Z \in \bigcup_{k\in\mathbb{Z}} [k, k+s]\right) = \sum_{k\in\mathbb{Z}} P(U+Z \in [k, k+s]). \tag{4.24}$$

Because U and Z are independent, $P(U+Z \in [k,k+s]) = (\mu_U \star \mu_Z)([k,k+s])$, where μ_U and μ_Z are the probability measures of U and Z respectively and $\star$ denotes their convolution. Note that since μ is translation–invariant,

$$\begin{aligned}(\mu_U \star \mu_Z)([k,k+s]) &= \int_{\mathbb{R}} \int_0^1 1_{[k,k+s]}(z+u) du d\mu_Z(z) \\ &= \int_{\mathbb{R}} \mu([0,1] \cap [k-z, k+s-z]) d\mu_Z(z) \\ &= \int_{\mathbb{R}} \mu([-k,-k+1] \cap [-z,-z+s]) d\mu_Z(z) \\ &= \int_{\mathbb{R}} \mu([-k,-k+1[\cap [-z,-z+s]) d\mu_Z(z)\end{aligned}$$

Going back to (4.24),

$$\begin{aligned}P(\mathrm{frac}(U+Z) \leq s) &= \sum_{k\in\mathbb{Z}} \int_{\mathbb{R}} \mu([-k,-k+1[\cap [-z,-z+s]) d\mu_Z(z) \\ &= \int_{\mathbb{R}} \sum_{k\in\mathbb{Z}} \mu([-k,-k+1[\cap [-z,-z+s]) d\mu_Z(z) \\ &= \int_{\mathbb{R}} \mu([-z,-z+s]) d\mu_Z(z) \\ &= \mu([0,s]) \int_{\mathbb{R}} d\mu_Z(z). \\ &= s.\end{aligned}$$

Therefore, the distribution function of $\mathrm{frac}(U+Z)$ is uniform on $[0,1]$ and therefore also equal to that of $\mathrm{frac}(U)$. □

4.E.2 Model 2

The process $(\mathrm{frac}(\gamma_n))_{n\in\mathbb{N}}$ is defined in (4.8), via the Markov chain $(V_n)_{n\in\mathbb{N}}$. Recall that this Markov chain has a transition probability kernel P, with support included in $I = [v_{\min}, v_{\max}]$. Therefore, $(\mathrm{frac}(\gamma_n))_{n\in\mathbb{N}}$ is not itself a Markov Chain (of order 1). However, the process $((\gamma_n, V_n))_{n\in\mathbb{N}}$ is a Markov Chain. We characterize its distribution and we verify that it satisfies the Doeblin condition (2.4) and this takes the remaining of this section.

Consider now $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and let $(x,A) \mapsto 1_A(x)$, which is also a transition probability kernel. We define a product kernel on $R \coloneqq \mathbb{R} \times I$, where $I = [v_{\min}, v_{\max}]$. It is characterised by the following measure on rectangles

$$((y,v),(A\times B)) \mapsto 1_A(y) P(v,B).$$

More generally, it extends to any set $A \in \mathcal{B}(R)$ as $((y,v),(A\times B)) \mapsto P(v,A_y)$, where

$$A_y = \{v \in I \mid (y,v) \in A\} \tag{4.25}$$

is the projection of $A \cap \{x = y\}$ onto the second coordinate. We define the map T

$$\begin{array}{rccc} T: & \mathbb{R}^2 & \to & \mathbb{R}^2 \\ & (x,v) & \mapsto & (x+hv, v), \end{array}$$

and we let $\tilde{P}$ be the pull-back of the product kernel P by this map. Explicitly, for $A \in \mathcal{B}(R)$,

$$\tilde{P}((u,v),A) = P(v, A_{u+hv}). \tag{4.26}$$

In what follows, we show (2.4) for the Markov chain $((\mathrm{frac}\gamma_n, V_n))_{n\in\mathbb{N}}$, which has transition probability kernel $\mathrm{frac}_\star \tilde{P}$. Figure 4.E.1 illustrates the proof. For $(u,v) \in R$, we show that $\tilde{P}^n((u,v),\cdot))$ is lower–bounded by a uniform measure of which we carefully characterise the support, $\Omega^n_{u,v}$. In Steps 1-6, we show that for a certain $n_{u,v} \in \mathbb{N}$, the support of this uniform measure, $\Omega^{n_{u,v}}_{u,v}$, is large enough. In 7, we show that $n_{u,v} \leq N \in \mathbb{N}$, for all $(u,v) \in R$. We conclude in Step 8 by showing (2.4). Compared with the i.i.d case treated in Section 4.E.1, Step 1 is the analogue of Lemma 4.26, except that the iteration requires the additional Steps 2-5.

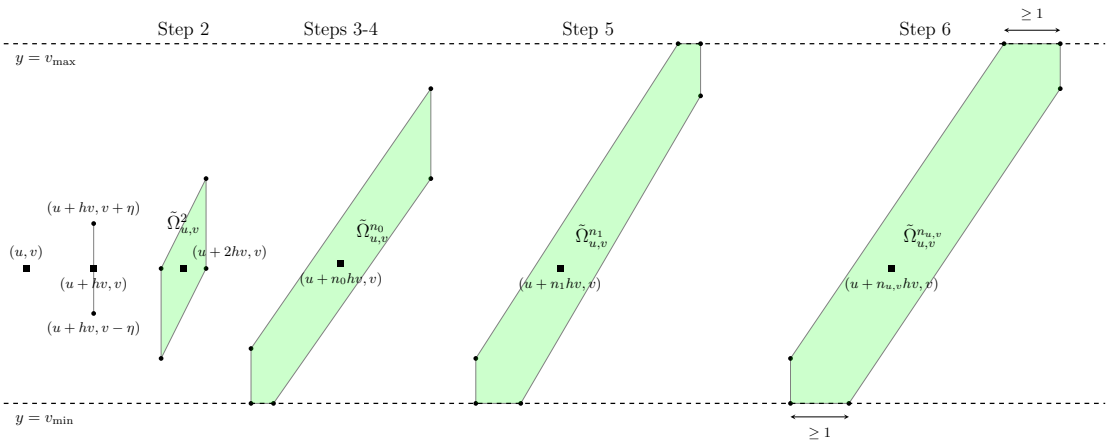


Figure 4.E.1: A schematic illustration of the form of a density's support. The density lower–bounds $\tilde{P}^n((u,v),\cdot))$.

Step 1 (lower–bound for $\tilde{P}^2((u,v),\cdot))$. For $(u,v) \in R$ and $(z_1,z_2) \in R$, according to (4.25),

$$([0,z_1] \times [v_{\min}, z_2])_{u+h(v+y)} = \begin{cases} [v_{\min}, z_2], & \text{if } u + h(v+y) \in [0,z_1], \\ \emptyset, & \text{otherwise.} \end{cases}$$

In (4.26), we observe that integrating with respect to $\tilde{P}^2$ amounts to integrating P along a vertical strip, so marginalizing with respect to (γ_1, V_1),

$$\begin{aligned} \tilde{P}^2((u,v),]-\infty,z_1] \times [v_{\min},z_2]) &= \int_R \tilde{P}((u,v),dxdy))\tilde{P}((x,y),]-\infty,z_1] \times [v_{\min},z_2]) \\ &= \int_I P(v,dy)P(y,(]-\infty,z_1] \times [v_{\min},z_2])_{u+h(v+y)}) \\ &= \int_{v_{\min}}^{\max((z_1-u)/h-v, v_{\max})} P(v,dy)P(y,[v_{\min},z_2]) \end{aligned}$$

Differentiating the above expression with respect to z_1 and then z_2, for $z_1 \leq u + h(v + v_{\max})$, we get

$$\begin{aligned} \frac{\partial \tilde{P}^2((u,v),]-\infty,z_1] \times [v_{\min},z_2])}{\partial z_1} &= f_v\left(\tfrac{z_1-u}{h} - v\right) P\left(\tfrac{z_1-u}{h} - v,]v_{\min}, z_2]\right) \\ f^{\star 2}_{(u,v)}(z_1,z_2) = \frac{\partial^2 \tilde{P}^2((u,v),]-\infty,z_1] \times [0,z_2])}{\partial z_1 \partial z_2} &= f_v\left(\tfrac{z_1-u}{h} - v\right) f_{\frac{z_1-u}{h}-v}(z_2). \end{aligned}$$

As $f_v(y) \geq \mu_0 1_{[v-\eta, v+\eta]}(y)$, we have $f^{\star 2}_{(u,v)}(z_1,z_2) \geq \mu_0^2$, if

$$\begin{cases} \dfrac{z_1-u}{h} - v \in [\max(v_{\min}, v-\eta), \min(v_{\max}, v+\eta)], \\ \qquad z_2 \in \left[\max\left(v_{\min}, \dfrac{z_1-u}{h} - v - \eta\right), \min\left(v_{\max}, \dfrac{z_1-u}{h} - v + \eta\right)\right]. \end{cases}$$

The above is equivalent to

$$\begin{cases} z_1 = u + 2hv + kh\eta \\ z_2 = v + (k+l)\eta, \end{cases} \qquad \text{for some } l \in [-1,1], k \in [-1,1] \cap \left[\tfrac{v_{\min}-v}{\eta}, \tfrac{v_{\max}-v}{\eta}\right]. \tag{4.27}$$

So, $\tilde{P}^2((u,v),\cdot)$ has a density $f^{\star 2}_{(u,v)}$ with respect to the Borel measure on $\mathbb{R}^2$. That density is lower–bounded: for $(z_1,z_2) \in R \cap \Omega^2_{(u,v)}$, we have $f^{\star 2}_{(u,v)}(z_1,z_2) \geq \mu_0^2$, where

$$\Omega^2_{(u,v)} \coloneqq \{(u + 2hv, v) + k(h\eta, \eta) + l(0,\eta) \mid k,l \in [-1,1]\}. \tag{4.28}$$

When $\frac{v_{\min}-v}{\eta} < -1$ and $1 < \frac{v_{\max}-v}{\eta}$, then $\Omega^2_{(u,v)} \subset R$ and we carry on with the induction to 2. Otherwise, we go directly to 3 as $\Omega^{n+1} \cap R^c \neq \emptyset$.

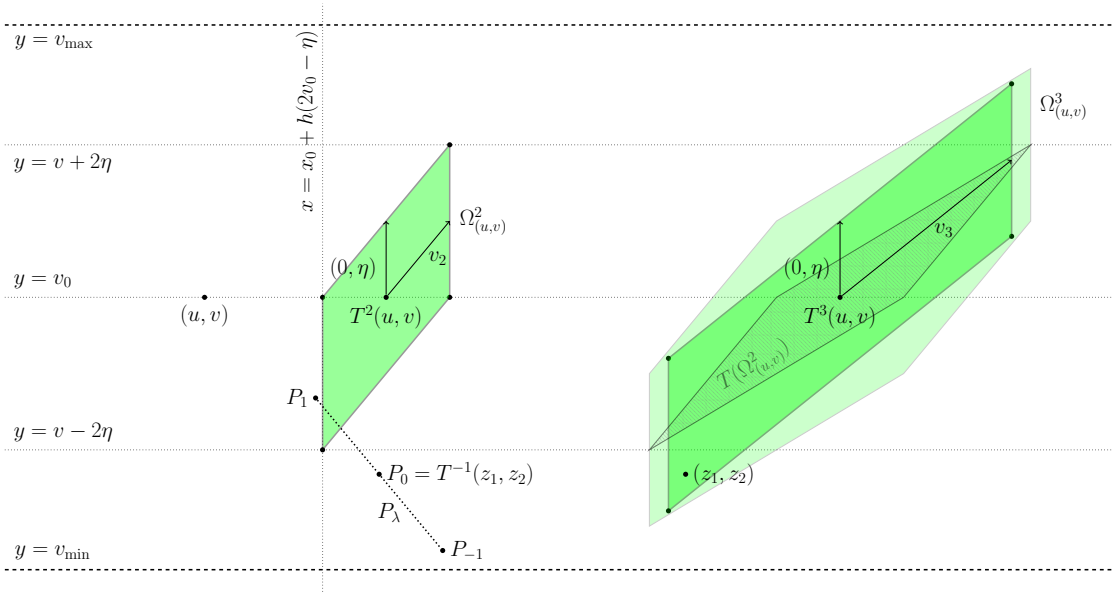


Figure 4.E.2: Illustration of $\Omega^n_{(u,v)}$ for $n = 2, 3$ and of the segment P_λ. Our argument consists in showing that for any $(z_1, z_2) \in \Omega^{n+1}_{(u,v)}$, the length of the intersection of P_λ with $\Omega^n_{(u,v)}$ is at least $\eta\epsilon/2$. While the dark green region is Ω^3, the lighter colour shows a larger region where the lower–bound is valid.

Step 2 (Lower–bound for $n \geq 3$, while $\Omega^n_{(u,v)} \cap R^c = \emptyset$). We start by defining the parallelograms $\Omega^n_{(u,v)}$ and showing some properties of the vectors that generate them. Then, by induction for $n \geq 2$, we will show the following statement:

$$\begin{aligned} &\text{For } 0 < \epsilon < \min\left(\tfrac{1}{4}, \tfrac{\eta}{2}(v_{\max} - v_{\min})\right) \text{ and } \eta < (v_{\max} - v_{\min})/2, \\ &\tilde{P}^n \text{ has a density } f^{\star n} \text{ lower–bounded by } \left(\tfrac{\eta\epsilon}{2}\right)^{n-2} \mu_0^n \text{ on } \Omega^n_{(u,v)}. \end{aligned} \tag{4.29}$$

Our induction is valid while $\Omega^n_{u,v} \subset R$ and Step 3 shows how to modify it when it ceases to be the case. Our arguments become progressively more geometric, for what we find the illustration of the proof in Figure 4.E.2 helpful.

To define $\Omega^n_{(u,v)}$, let $v_2 \coloneqq T(0,\eta) = (h\eta, \eta)$ and for $n \geq 3$,

$$v_n = (1-\epsilon)\left(T(0,\eta) + T(v_{n-1})\right) \in \mathbb{R}^2. \tag{4.30}$$

For $n \geq 3$, we define

$$\Omega^n_{(u,v)} \coloneqq \left\{T^n(u,v) + l(0,\eta) + kv_n \mid l, k \in [-1,1]\right\}. \tag{4.31}$$

Notice that if we take $n = 2$ in (4.31), we get $\Omega^2_{(u,v)}$ from as defined in (4.28).

While one can obtain an explicit expression of v_n, it is of little pratical interest: we only need to ensure that the horizontal component of v_n remains sufficiently large. This is detailed in the proof of Lemma 4.28.

Since we have shown the statement (4.29) for $n = 2$, we proceed with the induction step. For $(z_1, z_2) \in \Omega^{n+1}_{(u,v)} \cap R$, we calculate

$$\begin{aligned} \tilde{P}^{n+1}((u,v),]-\infty, z_1] \times [v_{\min}, z_2]) &= \int_{R \cap \{x+yh \leq z_1\}} \tilde{P}^n((u,v), dxdy) P(y, [v_{\min}, z_2]) \\ &= \int_{R \cap \{x+yh \leq z_1\}} f^{\star n}_{(u,v)}(x,y) P(y, [v_{\min}, z_2]) dxdy. \end{aligned}$$

We can rewrite $R \cap \{x + yh \leq z_1\} = \{(x,y) \in \mathbb{R} \mid y \in I,\, x \leq z_1 - yh\}$. Differentiating with respect to z_1, we obtain

$$\frac{\partial \tilde{P}^{n+1}((u,v),]-\infty, z_1] \times [v_{\min}, z_2])}{\partial z_1} = \int_y f^{\star n}_{(u,v)}(z_1 - yh, y) P(y, [v_{\min}, z_2]) dxdy,$$

where f_v is defined in (4.9). For $z_2 \geq v_{\min}$, we get

$$f^{\star n+1}_{(u,v)}(z_1, z_2) = \frac{\partial^2 \tilde{P}^{n+1}((u,v),]-\infty, z_1] \times [v_{\min}, z_2])}{\partial z_1 \partial z_2} = \int_I f^{\star n}_{(u,v)}(z_1 - yh, y) f_y(z_2) dy. \tag{4.32}$$

The expression in (4.32) is similar to that from Step 1, except that it is integrated over I. We can lower–bound the integrand: $f^{\star n}_{(u,v)}$ is lower–bounded by $\left(\frac{\eta\epsilon}{2}\right)^{n-2} \mu_0^n$ on $\Omega^n_{(u,v)}$ for $n \geq 2$ and f_y by μ_0 on $[y-\eta, y+\eta] \cap I$. To lower-bound $f^{\star n+1}_{(u,v)}$, it remains to lower–bound the length of the integration domain. For the calculations, we take the following parametrisation of $[z_2 - \eta, z_2 + \eta] \cap I$,

$$\left\{ \lambda \in [-1, 1] \mid P_\lambda \coloneqq P_0 + \lambda\eta(-h, 1) \in \Omega^n_{(u,v)} \right\}, \qquad \text{where } P_0 = T^{-1}(z_1, z_2). \tag{4.33}$$

Lemma 4.28. *The length of the segment* (4.33) *is at least* $\frac{\epsilon}{2}$.

For the sake of readablity, we differ the proof of Lemma 4.28 to Section 4.E.2. Finally, going back to (4.32), we have the desired lower–bound

$$f^{\star(n+1)}_{(u,v)}(z_1, z_2) \geq \left(\tfrac{\eta\epsilon}{2}\right)^{n-2} \mu_0^n \times \mu_0 \times \left(\tfrac{\eta\epsilon}{2}\right) = \left(\tfrac{\eta\epsilon}{2}\right)^{(n+1)-2} \mu_0^{n+1}, \qquad \text{for } (z_1,\, z_2) \in \Omega^{n+1}_{u,v}.$$

In addition, for $\epsilon < 1 \Big/ \left(1 + \frac{3(v_{\max} - v_{\min})}{2\eta}\right)$,

$$(0,1) \cdot (v_{n+1} - v_n) = (1-\epsilon)\eta - \epsilon(0,1) \cdot v_n > (1-\epsilon)\eta - \epsilon(v_{\max} - v_{\min}) > \epsilon \tfrac{v_{\max} - v_{\min}}{2}.$$

The height of $\Omega^n_{u,v}$ grows with n, by at least a constant, positive term. Hence, it eventually reaches $v_{\max} - v_{\min}$, in which case $\Omega^n \cap R^c \neq \emptyset$.

Step 3 (First non–empty intersection with the boundary). Let $n_0 \coloneqq \min\{n \in \mathbb{N} \mid \mu(\Omega^n \cap R^c) > 0\}$. Without loss of generality, Ω^{n_0} extends beyond $v_{\min}$. We will now construct a region $\tilde{\Omega}^{n_0+1} \subset R$ such that $f^{\star(n_0+1)}_{(u,v)} \geq \left(\frac{\eta\epsilon}{2}\right)^{n_0-2} \mu_0^{n_0}$ on $\tilde{\Omega}^{n_0+1}$ and for which $\tilde{\Omega}^{n_0+1} \cap (\mathbb{R} \times \{v_{\min}\})$ is lower–bounded. Since we can choose η arbitrarily small, we can treat the lower and upper boundaries independently, so we focus on the construction of $\tilde{\Omega}^{n_0+1}_{u,v}$ on the boundary $\mathbb{R} \times \{v_{\min}\}$ first.

For $P \in \mathbb{R} \times \{v_{\min}\}$, we consider P_λ as in (4.33), under the constraint that the integration segment lies within R, that is, $\{\lambda \in [0,1] \mid P_\lambda \in \Omega^{n_0}\}$. We denote the length of this segment by $L(P)$,

$$L(P) \coloneqq \left|\{\lambda \in [-1,1] \mid P + \eta\lambda\,(-h, 1) \in \Omega^{n_0} \cap R\}\right|,$$

and we let A, B be the endpoints of $\Omega^{n_0} \cap (\mathbb{R} \times \{v_{\min}\})$. We rely on the following claim, whose proof is in Section 4.E.2. Figure 4.E.3 illustrates the situation.

> The set $\{P \in \Omega^{n_0} \cap \mathbb{R} \times \{v_{\min}\} \mid L(P) \geq \frac{\epsilon}{2}\}$ is not empty. If we denote by $D = A + (x_D, 0)$ and $E = A + (x_E, 0)$ its right- and left-endpoints, then for some $c_0 > 0$ independent of (u, v), we have
>
> $$x_B + c_0 \leq x_D - x_E. \tag{4.34}$$

In particular, $L(P) \geq \frac{\epsilon}{2}$ implies that $f^{\star n_0+1}_{(u,v)}(P) \geq \left(\frac{\eta\epsilon}{2}\right)^{n_0-1} \mu_0^{n_0+1}$. By convexity of $\Omega^{n_0} \cap R$, we have $L(P) \geq \frac{\epsilon}{2}$ for P on the segment $T(E)T(D)$, so we can include that segment in $\tilde{\Omega}^{n_0+1}$. As $L(E) = L(P)$, where $P = E + (1 - \epsilon/2)k\eta\,(-h, 1)$, for $k \in [0,1]$, we have the same lower–bound on the density holds on $T(P)$. So, we can include a segment of height $\eta(1 - \epsilon/2)$ above $T(E)$ in $\tilde{\Omega}^{n_0+1}$. Therefore, we can define $\tilde{\Omega}^{n_0+1}$ as the polygon with vertices $T(E) + (0, \eta(1-\epsilon/2))$, $T(E)$, $T(D)$, $T^{n_0+1}(u,v) - (0,\eta) + v_{n_0+1}$ and $T^{n_0+1}(u,v) + (0,\eta) + v_{n_0+1}$.

We have obtained a convex pentagon $\tilde{\Omega}^{n_0+1}$ on which $\tilde{P}^{n_0+1}((u,v), \cdot)$ is lower–bounded by a measure with density lower–bounded by $\left(\frac{\eta\epsilon}{2}\right)^{n_0-1} \mu_0^{n_0+1}$. Because T preserves lengths on horizontal cross-sections, (4.34) implies that the length of $T(D)T(E)$ is equal to that of ED, which is longer by $c_0 = \eta h/4$ than the intersection at n_0.

Step 4 (Induction for $n > n_0 + 1$). Assume that $\tilde{\Omega}^{n_0+1} \cap (\mathbb{R} \times \{v_{\max}\}) = \emptyset$. As a consequence of calculations for Step 2, $\tilde{\Omega}^n_{u,v}$ is growing upwards. Indeed, the calculations rely on Assumption (4.9) and the fact that v_n has a horizontal component whose length we control. Therefore, they adapt to $\tilde{\Omega}^n_{u,v}$, with v_n being the vector from $T(D)$ to $T^{n_0+1}(u,v) - (0,\eta) + v_{n_0+1}$.

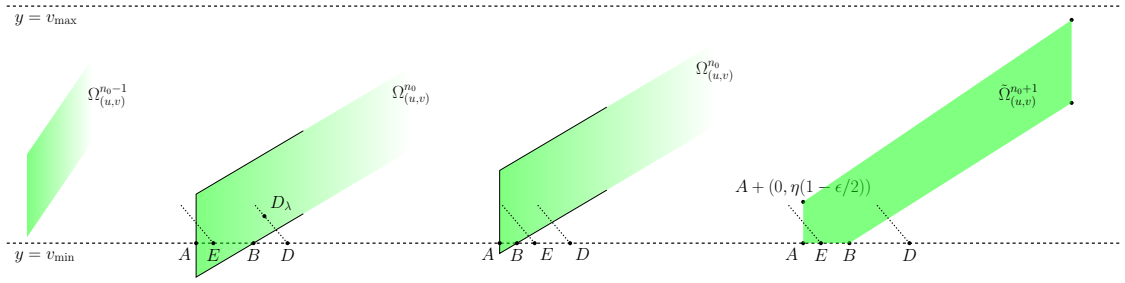


Figure 4.E.3: Illustration of 3 and the proof of (4.34). The leftmost polygon represents Ω^{n_0-1}, at the iteration before the first non-trivial intersection occurs. The two middle parallelograms illustrate the two cases, $x_E \leq x_B$ and $x_B \leq x_E$ respectively, from the proof of (4.34). On the right, the bottom part of the polygon $\tilde{\Omega}^{n_0+1}$ as constructed in 3. The dashed lines represent the integration segments, whose length is measured by L.

In addition, (4.34) still holds. Indeed, redefine A, B, D and E, except with n_0 replaced by n_0+1 in the expression of L. We notice that A, B coincide with $T(E)$ and $T(D)$ from the previous iteration. Because AB is now of length at least $c_0 = h\eta/4$, the proof is easier as we fall in the first case from Section 4.E.2. We define $\tilde{\Omega}^{n_0+2}$ as in Step 3.

We can now iterate this procedure, obtaining a lower–bound of $f^{\star n}_{u,v}$ by a uniform constant, on a convex and polygonal domain $\tilde{\Omega}^n$. Crucially, both the height of $\tilde{\Omega}^n$ and the length of its intersection with $\mathbb{R} \times \{v_{\min}\}$ grow, by uniformly lower–bounded amounts.

Step 5 (Intersection with both boundaries). For some $n_1 \in \mathbb{N}$, the intersection $\tilde{\Omega}^{n_1}_{u,v} \cap (\mathbb{R} \cap \{v_{\max}\})$ is not trivial. By a procedure analogue to that in Step 3, we can define $\tilde{\Omega}^{n_1+1}$, which non–trivially intersects both boundaries. Using the procedure from Step 4, it is clear that the intersection will not only remain non–trivial with n, but also increase.

Step 6 (Cross-sections with length at least 1). By definition, $\tilde{\Omega}^n$ is delimited by a convex, polygonal domain. The length of any horizontal cross-section of $\tilde{\Omega}^n$ is lower–bounded by the minimum of the lengths of the intersections with the lower and upper boundaries[1]. Recall that by Step 4, these two are increasing, and this by at least $h\eta/4$ at each iteration. Hence, for some $n = n_{u,v}$, all horizontal sections of $\tilde{\Omega}^{n_{u,v}}_{u,v}$ are of length at least 1.

By construction of $\tilde{\Omega}^n_{(u,v)}$, we have obtained a region such that for any $n \geq n_{u,v},$,

1. $\tilde{P}^n((u,v),\cdot)$ is lower–bounded by $\left(\frac{\eta\epsilon}{2}\right)^{n-2} \mu_0^n \mu$ on $\tilde{\Omega}^n_{u,v}$, (μ being the Lebesgue measure)
2. $\{\tilde{\Omega}^n_{(u,v)} + (k,0)\}_{k\in\mathbb{Z}}$ is a cover of R.

Step 7 (Uniform lower–bound). We now show that we can choose a uniform $N \in \mathbb{N}$, such that $n_{u,v} \leq N$ for all $(u,v) \in R$. Fix $(u,v) \in R$ and let $\bar{\Omega}^2_{u,v}$ be defined as in (4.27), except with $\frac{\eta}{2}$ instead of η. We can then perform Steps 1 to 6, so we obtain a domain $\bar{\Omega}^{\bar{n}_{u,v}}_{u,v}$ with cross-sections of length at least 1, for some $\bar{n}_{u,v} \geq n_{u,v}$.

Notice that the shrinked parallelogram at $n = 2$ is contained in parallelograms for different initial conditions. Specifically, we have $\bar{\Omega}^2_{u,v} \subset \Omega^2_{x,y}$ for $(x,y) \in C_{u,v}$, where $C_{u,v} = T^{-2}(\bar{\Omega}^2_{u,v})$. In particular, $n_{x,y} \leq \bar{n}_{u,v}$, for all $(x,y) \in C_{u,v}$. Since $(\mathrm{int}(C_{u,v}))_{(u,v)\in[0,1]\times I}$ is an open cover of $[0,1]\times I$, by compacity, we can find a finite cover $\{C_{u_k,v_k}\}_{k=1}^K$. Clearly, $N = \max_{1\leq k\leq K} \bar{n}_{u_k,v_k} < \infty$ gives a uniform bound on $(n_{u,v})_{(u,v)\in[0,1]\times I}$. The bound is also valid on $R \times I$, because the whole construction is invariant with respect to horizontal translations.

Finally, for $(u,v) \in R$, we have that $\tilde{P}^N((u,v),\cdot)$ is lower–bounded by $\left(\frac{\eta\epsilon}{2}\right)^{N-2} \mu_0^N \mu$, on $\Omega_{u,v}$ and $\{\Omega_{(u,v)} + (k,0)\}_{k\in\mathbb{Z}}$ is a cover of R, where $\Omega_{u,v} \coloneqq \tilde{\Omega}^N_{u,v}$.

Step 8 (Conclusion). We can now go back to $(\mathrm{frac}(\gamma_n), V_n)$. By lower–bounding $\tilde{P}^N$ with a uniform measure, we can use the same arguments as in Section 4.E.1 to conclude. For $A \in \mathcal{B}([0,1]\times I)$, we

[1] To see this, consider the parallelogram on the 4 vertices of $\tilde{\Omega}^n$ which belong to the boundary. That parallelogram is included in $\tilde{\Omega}^n$ by convexity, so the lengths of the horizontal sections is between the lengths of both bases.

have

$$\begin{aligned}
\mathrm{frac}_\star \tilde{P}^N((u,v),A) &= \tilde{P}^N((u,v),\mathrm{frac}^{-1}(A)) \\
&\geq \tilde{P}^N((u,v),\mathrm{frac}^{-1}(A)\cap\Omega_{(u,v)}) \\
&\geq C\mu(\mathrm{frac}^{-1}(A)\cap\Omega_{(u,v)}) && \text{(minorating on } \Omega_{(u,v)}) \\
&= C\mu(\cup_{k\in Z}A+(k,0)\cap\Omega_{(u,v)}) && (\{A+(k,0)\}_k \text{ disjoint}) \\
&= C\sum_{k\in Z}\mu(A+(k,0)\cap\Omega_{(u,v)}) \\
&= C\sum_{k\in Z}\mu(A\cap(\Omega_{(u,v)}-(k,0))) && (\mu \text{ translation–invariant}) \\
&\geq C\mu(\cup_{k\in Z}A\cap(\Omega_{(u,v)}-(k,0))) \\
&= C\mu(A) && (\{\Omega_{(u,v)}+(k,0)\}_{k\in\mathbb{Z}} \text{ is a cover of } R),
\end{aligned}$$

where $C = C_{\eta,\epsilon,\mu_0,N} \coloneqq \left(\frac{\eta\epsilon}{2}\right)^{N-2}\mu_0^N$. The lower–bound is uniform in (u,v) and also shows that the measure is non-trivial. We conclude the proof of Proposition 4.25 by applying Theorem 2.17.

Proof of Lemma 4.28

We recall that for some $l,k\in[-1,1]$,

$$(z_1,z_2) = T^{n+1}(u,v) + l\,(0,\eta) + kv_n,$$

where v_n is given in (4.30), so

$$P_\lambda \coloneqq T^{-1}(z_1, z_2+\lambda\eta) = T^n(u,v) + \eta(l+\lambda)(-h,1) + k(1-\epsilon)\left((0,\eta)+v_n\right). \tag{4.35}$$

For a parallelogram Ω generated by vectors x, y and centered around the origin, we have

$$P\in\Omega \iff \left\{\begin{array}{ccccc} x^Ty^\perp & \leq & P^Ty^\perp & \leq & -x^Ty^\perp \\ -y^Tx^\perp & \leq & P^Tx^\perp & \leq & y^Tx^\perp, \end{array}\right. \tag{4.36}$$

where $(x_1,x_2)^\perp = (x_2,-x_1)$. Combining (4.35) with (4.36), we have that $P_\lambda\in\Omega^n_{(u,v)}$ if and only if

$$\begin{aligned}
&\left\{\begin{array}{c} \eta(0,1)\cdot v_n^\perp \leq \eta(l+\lambda)(-h,1)\cdot v_n^\perp + k(1-\epsilon)(\eta\,(0,1)+v_n)\cdot v_n^\perp \leq -\eta(0,1)\cdot v_n^\perp \\ -\eta v_n\cdot(0,1)^\perp \leq \eta^2(l+\lambda)(-h,1)\cdot(0,1)^\perp + k(1-\epsilon)(\eta^2\,(0,1)+\eta v_n)\cdot(0,1)^\perp \leq \eta v_n\cdot(0,1)^\perp \end{array}\right. \\
\iff &\left\{\begin{array}{ccccc} (1,0)\cdot v_n(-1+k(1-\epsilon)) & \leq & -(l+\lambda)(1,h)\cdot v_n & \leq & (1,0)\cdot v_n(1+k(1-\epsilon)) \\ v_n\cdot(1,0)(-1-k(1-\epsilon)) & \leq & (-h\eta)(l+\lambda) & \leq & v_n\cdot(1,0)(1-k(1-\epsilon)). \end{array}\right.
\end{aligned}$$

As $(1,h)\cdot v_n > 0$ and denoting

$$a_n \coloneqq \frac{1}{h\eta}(1,0)\cdot v_n, \qquad b_n \coloneqq 1-\frac{(0,h)\cdot v_n}{(1,h)\cdot v_n},$$

we have

$$P_\lambda\in\Omega^n \iff \left\{\begin{array}{ccccc} b_n(-1-k(1-\epsilon))-l & \leq & \lambda & \leq & b_n(1-k(1-\epsilon))-l \\ a_n(-1+k(1-\epsilon))-l & \leq & \lambda & \leq & a_n(1+k(1-\epsilon))-l. \end{array}\right.$$

Finally, taking into account that $\lambda\in[-1,1]$, we obtain that

$$\begin{aligned}
\lambda\in[&\max(-1, b_n(-1-k(1-\epsilon))-l, a_n(-1+k(1-\epsilon))-l), \\
&\min(1, b_n(1-k(1-\epsilon))-l, a_n(1+k(1-\epsilon))-l)],
\end{aligned}$$

which is of length

$$\begin{aligned}
&\min(1, b_n(1-k(1-\epsilon))-l, a_n(1+k(1-\epsilon))-l) + \\
&+\min(1, b_n(1+k(1-\epsilon))+l, a_n(1-k(1-\epsilon))+l) = \\
&=\min(2\min(1,a_n,b_n), (a_n+b_n)(1-k(1-\epsilon)), 1+b_n(1+k(1-\epsilon))+l, \\
&\qquad 1+a_n(1-k(1-\epsilon))+l, 1+b_n(1-k(1-\epsilon))-l, 1+a_n(1+k(1-\epsilon)-l)
\end{aligned}$$

We claim that for $n \geq 2$,

$$a_n \geq 1, \qquad b_n \geq \frac{1}{2}. \tag{4.37}$$

Combining (4.37) with $l, k \in [-1, 1]$, $0 < \epsilon \leq \frac{1}{2}$, we conclude that the length of (4.33) is at least $\frac{\epsilon}{2}$.

It remains to show (4.37). For a_n, we proceed by induction. Using $v_2 = T(0, \eta)$ for $n = 2$ and $v_3 = (1-\epsilon)\eta(3h, 2)$, we verify that a_2, $a_3 \geq 1$. Notice that $(T(x, y)) \cdot (1, 0) = ((x, y) + (yh, 0)) \cdot (1, 0) = (x, y) \cdot (1, h)$. Then,

$$v_{n+1} \cdot (1, 0) = (1-\epsilon)(T(0, \eta) + T(v_n)) \cdot (1, 0) = (1-\epsilon)\left[h\eta + v_n \cdot (1, h)\right].$$

Using the induction hypothesis, $v_n \cdot (1, 0) \geq h\eta$ combined with $v_n \cdot (0, h) \geq 0$,

$$v_{n+1} \cdot (1, 0) \geq 2h\eta(1-\epsilon) \geq h\eta,$$

since $\epsilon \leq \frac{1}{2}$.

For b_n, we can calculate directly $b_2 = \frac{h\eta}{h\eta + h\eta} = \frac{1}{2}$. For $n \geq 3$, we can express v_n using (4.30), so that

$$\begin{aligned}
\frac{(1, 0) \cdot v_n}{(1, h) \cdot v_n} =& \frac{(1, 0) \cdot (T(0, \eta) + T(v_{n-1}))}{(1, h) \cdot (T(0, \eta) + T(v_{n-1}))} \\
=& \frac{h\eta + (1, 0) \cdot T(v_{n-1})}{2h\eta + (1, h) \cdot T(v_{n-1})} \\
=& \frac{1}{2} + \frac{1}{2} \frac{((1, 0) - (0, h)) \cdot T(v_{n-1})}{2h\eta + (1, h) \cdot T(v_{n-1})} \\
\geq& \frac{1}{2},
\end{aligned}$$

where the last inequality follows from

$$((1, 0) - (0, h)) \cdot T(v_{n-1}) = (1, 0) \cdot v_{n-1} + (0, h) \cdot v_{n-1} - (0, h) \cdot v_{n-1} \geq 0.$$

□

Proof of (4.34)

Notice first that $L(A) = 0$ and $L(P) = 0$ for any $P \in R \times \{v_{\min}\}$ to the left of A, so that $0 \leq x_B, x_D, x_E$. Second, consider $P = A + \left(x_B + (1 - \epsilon/2)\eta h \frac{1}{1-b}, 0\right)$. As $L(P) = \frac{\epsilon}{2}$, we have that $\{P \mid L(P) \geq \frac{\epsilon}{2}\} \neq \emptyset$, so D and E exist. In addition, we know that $x_D \geq x_B + (1 - \epsilon/2)\eta h \frac{1}{1-b}$. In particular, $b \leq 1$ implies that $x_B < x_D$.

Since $A \in \Omega^{n_0}_{u,v}$, we can write $A = T^{n_0}(u, v) + l_A \eta(0, \eta) - v_{n_0}$. By definition, n_0 is the first time such that $\Omega^n_0 \cap R^c$ has non-trivial measure, so, using the relation between $\Omega^{n_0 - 1}$ and Ω^{n_0}, we can conclude that $l_A \leq 0 \leq 1 - \epsilon/2$.

We distinguish two cases, depending which one of $\eta\epsilon h/2$ or x_B is greater. First, if $\eta h\epsilon/2 \leq x_B$, then $x_E = \eta h\epsilon/2$. Indeed, the triangle formed by A, E and $A + (0, \eta\epsilon/2)$ is in Ω^{n_0}, so $L(A + (\eta h\epsilon/2, 0)) \geq \epsilon/2$. Therefore,

$$\begin{aligned}
x_D - x_E \geq& x_B + (1 - \tfrac{\epsilon}{2})\eta h \tfrac{1}{1-b} - \eta h \tfrac{\epsilon}{2} \\
\geq& x_B + \eta h(2(1 - \tfrac{\epsilon}{2}) - \tfrac{\epsilon}{2}) \\
\geq& x_B + \tfrac{5}{4}\eta h,
\end{aligned}$$

where in the last two inequalities, we have use that $\frac{1}{2} \leq b$ and $\epsilon \leq \frac{1}{2}$.

Next, if $x_B < \eta h\epsilon/2$, then $\eta h\epsilon/2 \leq x_E$. So, for $x \leq \eta h$,

$$L(A + (x, 0)) = \frac{1}{\eta h}\left(x - \tfrac{(0,h)\cdot v_n}{(1,h)\cdot v_n}\ (x - x_B)_+\right) = \frac{1}{\eta h}(xb - x_B(1-b)). \tag{4.38}$$

Notice that $L\left(A+(\eta h,0)\right) \geq \frac{\epsilon}{2}$ for $\epsilon \leq \frac{2}{5}$ small enough,

$$\begin{aligned} L\left(A+(\eta h,0)\right) - \frac{\epsilon}{2} &= \frac{1}{\eta h}(\eta h b - x_B(1-b)) - \frac{\epsilon}{2} \\ &= b - (1-b)\frac{x_B}{\eta h} - \frac{\epsilon}{2} \\ &\geq b - (1-b)\frac{\epsilon}{2} - \frac{\epsilon}{2} \\ &= b\left(1+\frac{\epsilon}{2}\right) - \frac{3}{2}\epsilon \\ &\geq \frac{1}{2}\left(1-\frac{5}{2}\epsilon\right), \end{aligned}$$

so $x_E \leq \eta h$. Using (4.38), we find that $x_E = \frac{1}{b}(\eta h \epsilon/2 + x_B(1-b))$. Finally,

$$\begin{aligned} x_D - x_E - x_B &= x_B\left(1-\tfrac{1-b}{b}\right) + \eta h \tfrac{1-\epsilon/2}{1-b} - \tfrac{1}{b}\eta h\epsilon/2 - \eta h \tfrac{\epsilon}{2} \\ &= x_B(2-\tfrac{1}{b}) + \eta h\left(\tfrac{1}{1-b} + \tfrac{\epsilon}{2}(\tfrac{1}{b} - \tfrac{1}{1-b})\right) - \eta h \tfrac{\epsilon}{2} \\ &= x_B(2-\tfrac{1}{b}) + \tfrac{\eta h}{1-b}\left(1 - \tfrac{\epsilon}{b}(b-\tfrac{1}{2})\right)) - \eta h \tfrac{\epsilon}{2}. \end{aligned}$$

Since $\frac{1}{2} \leq b \leq 1$, we have $\frac{b-1/2}{b} \leq 1$ and $\frac{1}{1-b} \geq 2$, so that

$$x_D - x_E - x_B \geq 2(1-\epsilon)\eta h - \eta h\epsilon/2 \geq \eta h(1-\tfrac{3}{2}\epsilon).$$

Combining the two cases with $\epsilon < \frac{1}{2}$, we conclude that

$$x_D - x_E \geq x_B + \eta h \min\left(\left(1-\tfrac{\epsilon}{2}\right), \tfrac{5}{4}\right) \geq x_B + \tfrac{1}{4}\eta h.$$

□

4.F Mixing-preserving operations: mixing coefficients of $(X_n)_{n\in\mathbb{N}}$

We show that the mixing coefficients of X can be upper-bounded by those of $\mathrm{frac}(\gamma)$ and W.

Proposition 4.29. *Let X_n be as in(4.10). For any $k \in \mathbb{N}$,*

$$\beta_X(k+M-1) \leq \beta_S(k) \leq \beta_{\mathrm{frac}(\gamma)}(k) + \beta_W(k).$$

Since ϕ is continuous, Proposition 2.18 implies that $\beta_{\phi(\gamma)}(k) \leq \beta_{\mathrm{frac}(\gamma)}(k)$. Applying Lemma 4.30 to $(\phi(\gamma_n))_{n\in\mathbb{N}}$ and $(W_n)_{n\in\mathbb{N}}$, we obtain $\beta_S(k) \leq \beta_{\mathrm{frac}(\gamma)}(k) + \beta_W(k)$. With Lemma 4.31, we bound $\beta_X(k) \leq \beta_S(k-M+1)$, for $k \geq M$. The proofs of the two lemmata essentially consist in manipulating the definitions.

Lemma 4.30. *For two random variables $U : (\Omega_U, \sigma^U) \to \mathbb{R}$, $V : (\Omega_V, \sigma^V) \to \mathbb{R}$, we have*

$$\beta_{U+V}(k) \leq \beta_U(k) + \beta_V(k),$$

if $(\beta_U(k))_{k\in\mathbb{N}}$, $(\beta_V(k))_{k\in\mathbb{N}} \in \ell^1$. Moreover, the same holds true if U and V are defined on the same probability space, but are independent.

Proof. Define $Z \coloneqq U + V$. Then, Z is (Ω_Z, σ^Z)-measurable, where $\Omega_Z = \Omega_U \times \Omega_V$ and $\sigma^Z = \sigma^U \otimes \sigma^V$. As σ^Z is generated by products of elements from σ^U and σ^V, we only need to consider (countable) partitions $\mathcal{A}_U, \mathcal{B}_\mathcal{U}$ and $\mathcal{A}_V, \mathcal{B}_V$ of $\sigma^U_{-\infty,0}, \sigma^U_{k,\infty}$ and $\sigma^V_{-\infty,0}, \sigma^V_{k,\infty}$ respectively. For any $A_U \in \mathcal{A}_U, A_V \in \mathcal{A}_V$ and $B_U \in \mathcal{B}_U, B_V \in \mathcal{B}_V$, by definition of the product probability measure,

$$\begin{aligned} P((A_U\times A_V)\cap(B_U\times B_V)) - P(A_U\times A_V)P(B_U\times B_V) &= \\ = (P_U(A_U\cap B_U) - P_U(A_U)P_U(B_U))P_V(A_V\cap B_V) & \\ + P_U(A_U)P_U(B_U)(P_V(A_V\cap B_V) - P_V(A_V)P_V(B_V)). & \end{aligned}$$

Since β_U is is summable, $\sum_{A_U,B_U} |P_U(A_U \cap B_U) - P_U(A_U)P_U(B_U)| < \infty$ (idem for V), so we can regroup terms and

$$\begin{array}{rll}
\sum_{\substack{A_U\in\mathcal{A}_U, A_V\in\mathcal{A}_V,\\ B_U\in\mathcal{B}_U, B_V\in\mathcal{B}_V}} & P((A_U\times A_V)\cap(B_U\times B_V)) - P(A_U\times A_V)P(B_U\times B_V) & \\
= & \sum_{A_U\in\mathcal{A}_U,\, B_U\in\mathcal{B}_U}(P_U(A_U\cap B_U) - P_U(A_U)P_U(B_U)) & \\
& \times \sum_{A_V\in\mathcal{A}_V,\, B_V\in\mathcal{B}_V} P_V(A_V\cap B_V) & (=1)\\
& + \sum_{A_U\in\mathcal{A}_U,\, B_U\in\mathcal{B}_U} P_U(A_U)P_U(B_U) & (=1)\\
& \times \sum_{A_V\in\mathcal{A}_V,\, B_V\in\mathcal{B}_V}(P_V(A_V\cap B_V) - P_V(A_V)P_V(B_V)) & \\
\leq & \beta_U(k) + \beta_V(k). &
\end{array}$$

We conclude by taking the sup over partitions of Ω_Z. □

Lemma 4.31. *Consider $S = (S_i)_{i\in\mathbb{N}}$ with coefficients $\beta_S(k)$ and define $X_n = (S_n, \dots, S_{n+M-1})$. Then,*

$$\beta_X(k+M-1) \leq \beta_S(k).$$

Proof. First, note that the σ-algebra generated by a vector coincides with the σ-algebra generated by its components

$$\begin{aligned}
\sigma(X_{n_1}, \dots X_{n_2}) &= \sigma((S_{n_1}, \dots, S_{n_1+M-1}), \dots, (S_{n_2}, \dots, S_{n_2+M-1}))\\
&= \sigma(S_{n_1}, \dots, S_{n_2+M-1})\\
&= \sigma^S_{n_1,n_2+M-1}.
\end{aligned}$$

Then, any partition $\mathcal{A} \subset \sigma^X_{n_1,n_2}$ is also in $\sigma^S_{n_1,n_2+M-1}$. Since β_X is defined as a sup over such partitions, $\beta_X(k+M-1) \leq \beta_S(k)$. For $k \leq M$, we can take $\mathcal{A} = \mathcal{B} \subset \sigma(S_k)$. Since S_k is a continuous random variable, $\beta_X(k) = 1$. □

4.G Lipschitz constant for k^{pi} and $k^{pi,t}$

First, $(x,y) \mapsto \exp(-(x^2+y^2))$ is $\frac{2\sqrt{2}}{e}-$Lipschitz with respect to the Euclidean norm, so $\frac{4}{e}-$Lipschitz for the Minkowski norm. Consider the kernel

$$k^{pi,t}(b,d)(x,y) = \tfrac{1}{2\pi\sigma^2}\left(2 - \tfrac{\|(b,d)-(x,y)\|_\infty}{\sigma}\right)_+^r \exp\left(-\tfrac{(b-x)^2+(d-y)^2}{2\sigma^2}\right).$$

Then, for $r > 1$, the first term is $\frac{2^r r}{\sigma}$-Lipschitz,

$$\begin{aligned}
&\left|\left(2 - \tfrac{\|(b,d)-(x,y)\|_\infty}{\sigma}\right)_+^r - \left(2 - \tfrac{\|(b',d')-(x,y)\|_\infty}{\sigma}\right)_+^r\right|\\
=&\left|\int_0^1 \tfrac{d}{dt}\left(2 - \tfrac{\|(b,d)+(b'-b,d'-d)t-(x,y)\|_\infty}{\sigma}\right)_+^r dt\right|\\
\leq&\int_0^1 \Bigg|r\left(2 - \tfrac{|b+(b'-b)t-x|}{\sigma}\right)_+^{r-1}(-1)^{b-x>b'-bt}\tfrac{(b'-b)}{\sigma}1_{|b+(b'-b)t-x|\geq|d+(d-d')t-y|}\\
&+ r\left(2 - \tfrac{|d+(d'-d)t-y|}{\sigma}\right)_+^{r-1}(-1)^{d-y>d'-dt}\tfrac{(d'-d)}{\sigma}1_{|b+(b'-b)t-x|\leq|d+(d-d')t-y|}\Bigg| dt.\\
\leq&\int_0^1 \tfrac{r}{\sigma}\left(\left(2 - \tfrac{|b+(b'-b)t-x|}{\sigma}\right)_+^{r-1}|b-b'| + r\left(2 - \tfrac{|d+(d'-d)t-y|}{\sigma}\right)_+^{r-1}|d-d'|\right) dt\\
\leq&\tfrac{r}{\sigma}\left((2 - \tfrac{\min(|b-x|,|b'-x|)}{\sigma})_+^{r-1}|b-b'| + r(2 - \tfrac{\min(|d-y|,|d'-y|)}{\sigma})_+^{r-1}|d-d'|\right)\\
\leq&\tfrac{2r}{\sigma}(2 - \tfrac{\min(\|(b,d)-(x,y)\|_\infty,\|(b',d')-(x,y)\|_\infty)}{\sigma})_+^{r-1}\|(b,d)-(b',d')\|_\infty\\
\leq&\tfrac{2^r r}{\sigma}\|(b,d)-(b',d')\|_\infty.
\end{aligned}$$

Then, we obtain

$$
\begin{aligned}
&|k^{pi,t}(b,d)(x,y) - k^{pi,t}(b',d')(x,y)| \\
\leq &\tfrac{1}{2\pi\sigma^2}\left|\left(2 - \tfrac{\|(b,d)-(x,y)\|_\infty}{\sigma}\right)_+^r - \left(2 - \tfrac{\|(b',d')-(x,y)\|_\infty}{\sigma}\right)_+^r\right| \exp\left(-\tfrac{(b-x)^2+(d-y)^2}{2\sigma^2}\right) \\
&+ \tfrac{1}{2\pi\sigma^2}\left(2 - \tfrac{\|(b',d')-(x,y)\|_\infty}{\sigma}\right)_+^r \left|\exp\left(-\tfrac{(b-x)^2+(d-y)^2}{2\sigma^2}\right) - \exp\left(-\tfrac{(b'-x)^2+(d'-y)^2}{2\sigma^2}\right)\right| \\
\leq &\tfrac{1}{2\pi\sigma^2}\tfrac{2^r r}{\sigma}\|(b,d)-(b',d')\|_\infty + \tfrac{1}{2\pi\sigma^2}2^r\tfrac{4}{e}\left\|\left(\tfrac{b-x}{\sigma},\tfrac{d-y}{\sigma}\right) - \left(\tfrac{b'-x}{\sigma},\tfrac{d'-y}{\sigma}\right)\right\|_\infty \\
\leq &\tfrac{2^{r-1}}{\pi\sigma^3}(r+2)\|(b,d)-(b'd')\|_\infty.
\end{aligned}
$$

4.H Moments of the Hölder constant of a stochastic process

In Section 2.1, we gave sufficient conditions for paths of a stochastic process to be Hölder-continuous. In particular, Proposition 2.6 shows that the Hölder constant is a random variable. We now give an upper-bound of its moments.

Theorem 4.32 (Shevchenko (2017)). *Let $r_2 \in \mathbb{N}$ be such that $K_{r_2,\alpha r_2} < \infty$ and $1-\alpha > \frac{1}{r_2}$, $r_2 \geq 2$,*

$$\mathbb{E}[\Lambda_W] \leq 16\ \tfrac{\alpha+1}{\alpha} T K_{r_2,r_2\alpha+1}^{1/r_2}.$$

In addition,

$$\mathbb{E}[\Lambda_W^k] \leq \begin{cases} \left(2^{3+2/r_2}\ \tfrac{\alpha+2/r_2}{\alpha}\right)^k K_{r_2,r_2\alpha+1}^{k/r_2}, & \text{for } 0 < k \leq r_2, \\ \left(2^{3+2/r_2}\ \tfrac{\alpha+2/r_2}{\alpha}\right)^k K_{k,k(\alpha+2/r_2)-1}, & \text{for } k > r_2. \end{cases}$$

Lemma 4.33 (Garsia–Rodemich–Rumsey Inequality (Hu and Le, 2013, Lemma 1.1)). *Let $G : \mathbb{R}_+ \to \mathbb{R}_+$ be a non–decreasing function with $\lim_{x\to\infty} G(x) = \infty$ and $\delta : [0,T] \to [0,T]$ continuous and non–decreasing with $\delta(0) = 0$. Let G^{-1} and δ^{-1} be lower–inverses. Let $f : [0,T] \to \mathbb{R}$ be a continuous functions such that*

$$\int_0^T \int_0^T G\left(\frac{|f(x)-f(y)|}{\delta(x-y)}\right) dxdy \leq B < \infty.$$

Then, for any $s,t \in [0,T]$,

$$|f(s)-f(t)| \leq 8\int_0^{|s-t|} G^{-1}(4B/u^2)d\delta(u).$$

Proof of Theorem 4.32. Consider a path $W(\omega)$ of the stochastic process and set $B(\omega) \coloneqq \int_0^T \int_0^T G\left(\frac{|W_t(\omega)W_s(\omega)|}{\delta(t-s)}\right) dtds$, where $G(u) = u^{r_2}$ and $\delta(u) = u^{\alpha+2/r_2}$. Then, $G^{-1}(u) = u^{1/r_2}$ and $\frac{d}{du}\delta = (\alpha+2/r_2)u^{\alpha+2/r_2-1}$. Applying Lemma 4.33,

$$
\begin{aligned}
|W_t(\omega) - W_s(\omega)| &\leq 8\int_0^{|s-t|} G^{-1}(4B(\omega)/u^2)d\delta(u) \\
&\leq 8\int_0^{|t-s|} \left(\frac{4B(\omega)}{u^2}\right)^{1/r_2} (\alpha+2/p)u^{\alpha+2/r_2-1}du \\
&\leq 8(4B(\omega))^{1/r_2}(\alpha+2/r_2)\int_0^{|t-s|} u^{\alpha-1}du \\
&= 8(4B(\omega))^{1/r_2}\ \tfrac{\alpha+2/r_2}{\alpha}|t-s|^\alpha.
\end{aligned}
$$

As this is valid for any s, $t \in [0,T]$, $\Lambda_W(\omega) \leq 8(4B(\omega))^{1/r_2}\ \frac{\alpha+2/r_2}{\alpha}$. By Jensens' inequality,

$$\mathbb{E}[\Lambda_W] \leq 2^{3+2/r_2}\ \tfrac{\alpha+2/r_2}{\alpha}\mathbb{E}[B(\omega)^{1/r_2}] \leq 2^{3+2/r_2}\ \tfrac{\alpha+2/r_2}{\alpha}\mathbb{E}[B(\omega)]^{1/r_2}. \tag{4.39}$$

By linearity of expectation,

$$\begin{aligned}\mathbb{E}\left[\int_0^T\int_0^T G\left(\frac{|W_t(\omega)W_s(\omega)|}{\delta(t-s)}\right)dtds\right] &= \int_0^T\int_0^T \frac{\mathbb{E}[|W_t(\omega)W_s(\omega)|^{r_2}]}{\delta(t-s)^{r_2}}dtds\\ &= \int_0^T\int_0^T \frac{\mathbb{E}[|W_t(\omega)W_s(\omega)|^{r_2}]}{|t-s|^{p\alpha+2}}dtds\\ &\leq \int_0^T\int_0^T K_{p,p\alpha+1}dtds\\ &= T^2K_{r_2,r_2\alpha+1}.\end{aligned}$$

Finally, $\mathbb{E}[\Lambda_W] \leq 2^{3+2/r_2}\ \frac{\alpha+2/r_2}{\alpha}T^{2/r_2}K^{1/r_2}_{r_2,r_2\alpha+1}$, as long as $r_2\alpha+1 \leq r_2$ and we can simplify the constants if $r_2 > 2$. Consider now the higher moments. If $k \leq r_2$, we can still apply Jensens' inequality in (4.39):

$$\begin{aligned}\mathbb{E}[\Lambda_W^k] &\leq \left(2^{3+2/r_2}\ \frac{\alpha+2/r_2}{\alpha}\right)^k \mathbb{E}[B(\omega)^{k/r_2}]\\ &\leq \left(2^{3+2/r_2}\ \frac{\alpha+2/r_2}{\alpha}\right)^k \mathbb{E}[B(\omega)]^{k/r_2}\\ &\leq \left(2^{3+2/r_2}\ \frac{\alpha+2/r_2}{\alpha}\right)^k K^{k/r_2}_{r_2,r_2\alpha+1}.\end{aligned}$$

However, if $k \geq r_2$,

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^T\int_0^T G\left(\frac{|W_t(\omega)W_s(\omega)|}{\delta(t-s)}\right)dtds\right)^{k/r_2}\right] &= \int_0^T\int_0^T \frac{\mathbb{E}[|W_t(\omega)W_s(\omega)|^{k}]}{\delta(t-s)^{k}}dtds\\ &= \int_0^T\int_0^T \frac{\mathbb{E}[|W_t(\omega)W_s(\omega)|^{k}]}{|t-s|^{k\alpha+2k/r_2}}dtds\\ &\leq \int_0^T\int_0^T K_{k,k(\alpha+2/r_2)-1}dtds\\ &= T^2K_{k,k(\alpha+2/r_2)-1}.\end{aligned}$$

□

Chapter 5

Segmentation of periodic-like signals

Zero-crossings and local extrema appear repeatedly in periodic signals with phase variation. The number of occurrences of both of these elements can be read in the persistence diagram, which is thus a more general feature that also shares their additive properties, as established in Chapter 3. Both zero-crossings and local extrema are used for Problem 1.1: segmenting a reparametrized periodic signal into its constituent periods. In this chapter, we propose to use the persistence diagram to count the number of periods and guide that segmentation.

Let ϕ be a one–periodic, continuous function, $\gamma : [0,1] \to [0,N]$ an increasing bijection for some $N \in \mathbb{N}^*$ and consider

$$\begin{aligned} S : [0,1] &\to \mathbb{R} \\ t &\mapsto (\phi \circ \gamma)(t) + W_t, \end{aligned} \tag{5.1}$$

where $(W_t)_{t\in[0,T]}$ is a stochastic process. We consider the setting where ϕ, γ and N are unknown. Figure 1 illustrates the model with an example of ϕ featuring six local extrema per period, a non–linear γ and $W_t = 0$.

Problem 5.1. Given S, estimate N.

Problem 5.2 (Odometry). Given S and N, find $(t_i)_{i=1}^N$, such that

$$\gamma(t_i) - \gamma(t_{i-1}) = 1, \qquad \forall\ i = 2, \dots, N.$$

Such a sequence will be called an odometric sequence.

Problem 5.1 consists of estimating how many periods of ϕ are present in an observed signal S. Problem 5.2 is a segmentation task, where the number of segments is given. This segmentation, which we call an *odometric sequence*, thus contains finer information: if we know that $(t_i)_{i=1}^N$ is an odometric sequence for S, then $S|_{[t_1,t_N]}$ contains $N-1$ periods. We therefore include the number of periods N as part of the input data to Problem 5.2.

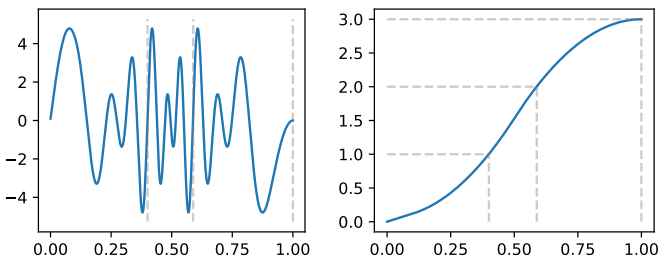


Figure 1: On the left, an example of a noise–free observed function $\phi \circ \gamma$, where a one–periodic ϕ has been composed with γ depicted on the right. On both graphs, $\gamma^{-1}(\mathbb{N})$ are marked by the gray dashed lines.

This work is motivated by the magnetic positioning problem described in Chapter 1. A solution to Problem 5.2, odometric sequence $(t_n)_{n=1}^N$, gives us an approximation of γ:

$$\hat{\gamma}(t) = \sum_{n=1}^{N} 1_{t_n \leq t}. \tag{5.2}$$

We will explore numerical applications of odometry in Chapter 7.

Related literature

Inverse problems similar to Problems 5.1 and 5.2 have been studied in different contexts. In signal processing, the instantaneous phase estimation (Boashash, 2015, Chapter 10) is concerned with recovering $t \mapsto \gamma(t)$ when ϕ is known to belong to a family of functions of simple form, for example $\phi(x) = a\sin(2\pi(x - \phi))$. The work has led to two types of methods. Spectral methods consist of using a frequency representation of S to recover the phase: the exact solution is given by $\gamma(t) = \arctan(H(S)(t)/S(t))$, where H is the Hilbert transform (Boashash et al., 1990). For more complex signals, one can analyze time frequency representations of the signal, using image processing techniques like peak detection and component linking (Rankine et al., 2007). This procedure is even extended to sums of independently-parametrized functions (Khan and Boashash, 2016, Hussain and Boashash, 2002). These methods are aimed at estimating γ' and it is shown possible under hypothesis of separation of different components in the frequency spectrum. Another type of method is concerned with geometric properties of the signal. The simplest example, called the *zero crossings* method relies on counting the number of times that the graph of S crosses zero (Boashash et al., 1990). This principle is generalized to other forms via intrinsic mode functions (Huang et al., 1998). With the empirical mode decomposition algorithm (Huang et al., 1998), ϕ can be decomposed into more elementary functions, connecting local maxima and minima using cubic splines. This algorithm decomposes the signal into multiple, independent, simple components. On one hand, the setting considered in our work is easier than that for the spectral methods, because (5.1) implies two properties of the signal. First, it has constant amplitude. Second, if ϕ is decomposed into multiple components, those components share a reparametrisation. On the other hand, nothing other than that is known about ϕ and the observations are corrupted by additive noise W, which adds spurious local extrema.

The time-delay embedding technique is known to capture the recurrent structure in dynamical systems (Perea, 2019) and has been applied in various contexts, for example Fernández and Mateos (2022), Gidea et al. (2020). As shown in (Gakhar and Perea, 2019), the geometric and topological features depend not only on the periodic function, but also on the frequencies in the signal (see Figure 1.4 for an illustration). The Hilbert transform embedding method proposed in Kennedy et al. (2018) is a different embedding method inspired by the analytic signal transformation found in the signal processing literature. It is shown to maximize the topological signal of a reparametrized, sinusoidal function. In our context, the time delay embedding technique is not adapted, because the frequencies of the signal vary in a range we do not control. The signature of a path (Hambly and Lyons, 2010) is a bounded sequence, which is invariant to any reparametrisation. However, we are not aware of any relation between the signature of a single period and that of multiple periods of a function.

An alternative line of work at the intersection of signal processing and topological methods consists of studying a one dimensional signal directly. Similarly to the zero crossings or empirical mode decomposition from signal processing, it consists of looking at the shape of the signal. More precisely, it is based on the homology of the sublevel sets of the signal. This approach led to a denoising method (Plonka and Zheng, 2016) and also to confidence intervals on a topological descriptor (Myers et al., 2020). In Khasawneh and Munch (2018), the authors propose a period counting method for binary functions ϕ. Similar to the zero crossings, the method counts the number of changes in a binary signal, but discards those which last for only a short period of time. As shown experimentally by the authors, this makes it more robust than the spectral methods. That method solves both Problems 5.1 and 5.2, but in the simple case where ϕ is a binary signal.

As an example, Figure 2 shows three functions ϕ. The two on the left come from the model for the zero crossings method and the topological method from Khasawneh and Munch (2018) respectively. Neither of the two methods was designed for continuous signals with many local extrema per period, as presented on the right, which we attempt to study.

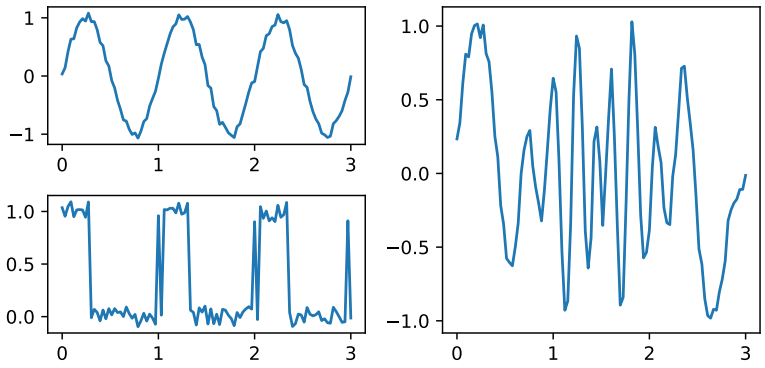


Figure 2: On the left, the graphs of signals from models which the zero crossings and the topological method (Khasawneh and Munch, 2018) respectively, were designed for. On top, a sinusoidal function with one pair of extrema per period. On the bottom, a noisy binary signal. On the right, a signal following the model (5.1) studied in this work.

Contributions

In this work, we propose to use topological methods for Problems 5.1 and 5.2 on γ. We work under genericity assumptions on ϕ and an upper bounded noise W. In Section 5.2, we define an estimator of $\gamma(1)$ based on the insight from Proposition 3.1. We show this estimator solves Problem 5.1 under genericity assumptions on ϕ and on the amplitude of the noise. We propose a robust alternative suitable in practice in Section 5.4.1 and test its performance in Section 5.4.2. Problem 5.2 is tackled in Section 5.3, where we introduce a method to calculate an odometric sequence from the data contained in the persistence diagram. We test the estimator for Problem 5.1 on synthetic signals. We study the applicability of the odometric sequence in practice in Chapter 7.

5.1 Estimation of N in the noiseless setting

Let $\phi : \mathbb{R} \to \mathbb{R}$ be one–periodic and $\gamma : [0, 1] \to [0, N]$ an increasing bijection for $N \in \mathbb{N}^*$, as above. In this section, we consider the noiseless setting $S = \phi \circ \gamma$. Points in the persistence diagram of sub level sets of a periodic function have multiplicity, as described precisely in Proposition 3.1.

Recall from Section 3.1 that $\bar{\phi} : \mathbb{S}^1 \to \mathbb{R}$ is the periodic function defined on the circle. Filtering the circle instead of the interval circumvents problems with boundary conditions, as shown in Proposition 3.1. The persistence diagrams that we consider have points with multiplicity greater than one and so we will identify them with the persistence measures (2.15). We denote by $\mathrm{m}_\phi(p)$ the multiplicity of a point p in the diagram of ϕ, also expressed using the persistence measure μ_ϕ with (2.16). In this case, we have

$$\mu_\phi = \sum_{p \in \mathrm{supp}(\mu_\phi)} \mathrm{m}_\phi(p)\delta_p.$$

For a set $A \subset \mathbb{N}$, the greatest common divisor $\gcd(A)$ is the largest $k \in \mathbb{N}$ such that for all $a \in A$, k divides a. For $A \subset \mathbb{N}^*$ non empty, $1 \leq \gcd(A) < \infty$. We adopt the convention that $\gcd(\emptyset) = 1$. We introduce the following estimator of N

$$N(\phi) = \gcd\{\mathrm{m}_\phi(p) \mid p \in \mathrm{supp}(\mu_\phi)\}.$$

The proposed estimator satisfies a homogeneity property, in the sense detailed by Proposition 5.1.

Proposition 5.1. *Suppose that $\gamma(1) = N$. Then,*

$$N(\phi \circ \gamma) = N \cdot N(\phi|_{[0,1]}).$$

Proof. Consider ϕ_N as defined in Proposition 3.1. The sublevel sets

$$(\phi \circ \gamma)^{-1}(]-\infty, a]) = \{t \mid \phi(\gamma(t)) \leq a\} = \gamma^{-1}(\{x \mid \phi_N(x) \leq a\})$$

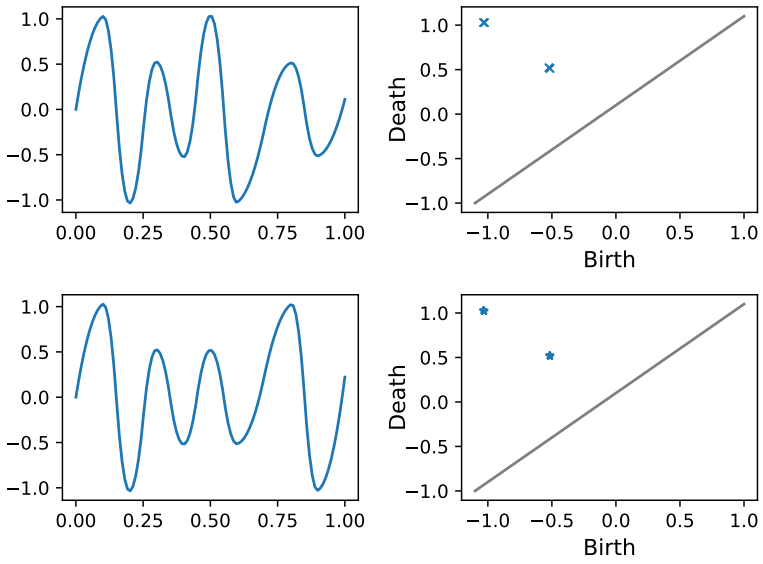


Figure 3: Graphs of two degenerate one–periodic functions and their persistence diagrams.

are homeomorphic to $\phi_N^{-1}(]-\infty, a])$. Hence, the persistence modules are isomorphic by Proposition 2.27a, so $\mu_{\phi_N} = \mu_{\phi\circ\gamma}$ and $\operatorname{supp}(\mu_{\phi_N}) = \operatorname{supp}(\mu_{\phi_1})$. By Proposition 3.1, we obtain $\mu_{\phi\circ\gamma} = \mu_{\phi_N} = N\mu_{\phi_1} = N\mu_{\phi|_{[0,1]}}$, so that

$$\begin{aligned} N(\phi\circ\gamma) &= \gcd\{\mathrm{m}_{\phi\circ\gamma}(p) \mid p \in \operatorname{supp}(\mu_{\phi_N})\} \\ &= \gcd\{N\mathrm{m}_{\phi_1}(p) \mid p \in \operatorname{supp}(\mu_{\phi_1})\} \\ &= N \cdot N(\phi|_{[0,1]}). \end{aligned}$$

□

The estimator $N(\phi)$ is correct only up to a multiplicative constant $N(\phi|_{[0,1]})$, so we need to introduce the notion of non–degenerate functions.

Definition 5.2. We call a 1-periodic function $\phi : \mathbb{R} \to \mathbb{R}$ *non–degenerate* if

$$N(\phi|_{[0,1]}) = 1. \tag{5.3}$$

A function $\phi : \mathbb{R} \to \mathbb{R}$ which does not satisfy (5.3) will be called *degenerate*.

The non-degeneracy of a function ϕ is a condition on the set of pairs of local extreme values and Proposition 5.3 provides a justification for restricting our considerations to non-degenerate functions. If ϕ is non-degenerate, the statement is vacuous. However, if ϕ is a degenerate function, then there exists a function ϕ_2, which has the same persistence measure as ϕ, but for which Problem 5.1 is not identifiable using only the persistence measure.

Proposition 5.3. *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a one–periodic function. There exists a $1/N(\phi|_{[0,1]})$–periodic function $\phi_2 : \mathbb{R} \to \mathbb{R}$, such that*

$$\mu_{\phi|_{[0,1]}} = N(\phi|_{[0,1]})\mu_{\phi_2|_{[0,1]}}.$$

We discuss the identifiability of the problem and the degeneracy in Figure 3, which shows the graphs and persistence diagrams of two degenerate functions. For ϕ in the top row, the problem of inference on N is not identifiable, since $\phi|_{[0,1/2]}$ could be re–parametrized to have the same graph as $\phi|_{[1/2,1]}$. In particular, the function ϕ is degenerate. The function ϕ_2 is also degenerate, because it induces the same persistence measure as ϕ. However, its extremal values occur in a different order, so the same re-parametrisation trick cannot be applied and the problem is identifiable. This situation exemplifies a limitation of our method and more generally, of inference based only on persistence measures. In Section 5.5, we comment on the implications. Non–degeneracy (5.3)

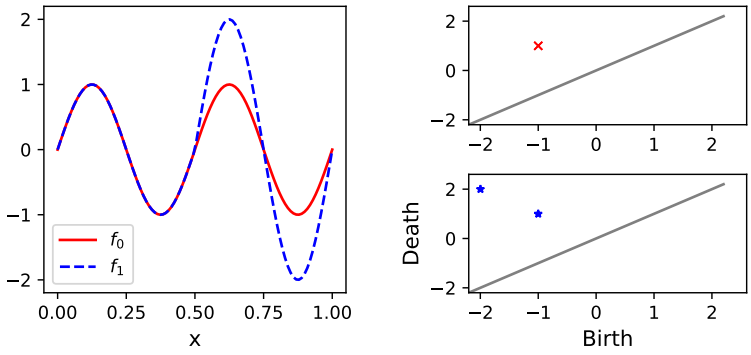


Figure 4: On the left, graphs of $f_0|_{[0,1]}$ and $f_1|_{[0,1]}$. On the right, their persistence diagrams.

restricts the class of considered functions, but it does not quantify the difficulty of the problem of estimation of N. Figure 4 illustrates two functions from a family of functions $(\phi_r)_{r\geq 0}$

$$\phi_r(x) = \begin{cases} \sin(4\pi x), & x \in [0, 1/2] \\ (1+r)\sin(4\pi x), & x \in]1/2, 1] \end{cases},$$

and extended by 1-periodicity to $\mathbb{R}$. Here, ϕ_0 is $\frac{1}{2}$-periodic and degenerate, while ϕ_r, $r > 0$ is not. Indeed, the persistence measure μ_{ϕ_r} is supported on the points $(-1-r, 1+r)$ and $(-1, 1)$, each with multiplicity 1. As $r \to 0$, $\phi_r \xrightarrow{\|\cdot\|_\infty} \phi_0$ and, while $\mathrm{d_b}(D(\phi_r), D(\phi_0)) \to 0$, we have $N(\phi_r|_{[0,1]}) = 1$ for all $r > 0$. To quantify degeneracy, we introduce δ which measures the separation of points in a diagram μ, from each other and from the diagonal

$$\delta = \min(\{d(p,q) \mid p, q \in \mathrm{supp}(\mu),\ p \neq q\} \cup \{d(p,\Delta) \mid p \in \mathrm{supp}(\mu)\}). \tag{5.4}$$

The next proposition lower–bounds how far a non–degenerate function is from a degenerate function, in terms of its separation δ.

Proposition 5.4. *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a non–degenerate function and $\delta > 0$ be the separation of μ_ϕ. For any degenerate $\phi_2 : \mathbb{R} \to \mathbb{R}$ with $N(\phi_2|_{[0,1]}) = n \geq 2$,*

$$\mathrm{d_b}(\mu_{\phi|_{[0,1]}}, \mu_{\phi_2|_{[0,1]}}) \geq \frac{\delta}{2}.$$

Proof. For this proof, we consider the diagram as a multi–set. In the argument below, we need to make a distinction between a point p in a diagram, and its geometric realization in $\mathbb{R}^2$, which we denote by $\overline{p}$, as there can be multiple points in $D(\phi)$ which have the same geometric realizations. Let $\Gamma : (D(\phi|_{[0,1]}) \cup \Delta) \to (D(\phi_2|_{[0,1]}) \cup \Delta)$ be a matching between the persistence diagrams (Definition 2.36).

As $N(\phi_2|_{[0,1]}) = n$ does not divide $1 = N(\phi|_{[0,1]})$, there is $p_0 \in \mathrm{supp}(D(\phi|_{[0,1]}))$ such that

$$n \text{ does not divide } \mathrm{m}_{\phi|_{[0,1]}}(\overline{p_0}). \tag{5.5}$$

Let us denote by $P = \{p \in D(\phi) \mid \overline{p} = \overline{p_0}\}$ the set of points from the diagram, which have the same geometric realization as p_0 and $Q = \Gamma(P) = \{\Gamma(p) \mid p \in P\} \subset \mathbb{R}^2$ the image of these points by the matching.

Consider $Q' = \{q' \in D(\phi_2|_{[0,1]}) \mid \exists q \in Q,\ \overline{q} = \overline{q'}\}$ the subset of $D(\phi_2|_{[0,1]})$ with geometric realizations in Q, and the set $P' = \Gamma^{-1}(Q') \setminus P$. Let us show that $P' \neq \emptyset$. Since $\Gamma(P) = Q \subset Q'$, if $P' = \emptyset$, then $P = \Gamma^{-1}(Q')$. The bijectivity of the matching Γ implies that $|P| = |Q'|$. However, n divides $|Q'| = P = \langle p_0 \rangle_{\phi|_{[0,1]}}$, so we obtain a contradiction with (5.5).

Consider $p_1 \in P'$. As $\Gamma(p_1) \in Q'$, there is $q \in Q$ such that $\overline{\Gamma(p_1)} = q$. Since $Q = \overline{\Gamma(P)}$,

$\overline{\Gamma(p_2)} = q$, for some $p_2 \in P$. Since $\overline{p_1} \neq \overline{p_2}$, $\|p_1 - p_2\|_\infty \geq \delta$, so that

$$\begin{aligned}\max_{p \in D(\phi|_{[0,1]}) \cup \Delta} \|\overline{p} - \overline{\Gamma(p)}\|_\infty &\geq \max(\|\overline{p_1} - \overline{\Gamma(p_1)}\|_\infty, \|\overline{p_2} - \overline{\Gamma(p_2)}\|_\infty) \\ &\geq \frac{1}{2}(\|\overline{p_1} - \overline{q}\|_\infty + \|\overline{p_2} - \overline{q}\|_\infty) \\ &\geq \frac{\delta}{2}.\end{aligned}$$

Since the bottleneck distance $\mathrm{d_b}(D(\phi|_{[0,1]}), D(\phi_2|_{[0,1]}))$ is an infimum over matchings Γ, it is also lower–bounded by $\delta/2$. □

For instance, taking the example from Figure 4, $\mu_{\phi_r|_{[0,1]}}$ is supported on $\{(-1-r, 1+r), (-1, 1)\}$, so that $\delta = \min(r, \frac{1}{2})$. However, the closest persistence measure of a degenerate function is $\mu = 2 \cdot 1_{(-1-r/2, 1+r/2)}$, which, by Proposition 5.3 is realizable as the measure of a function $\phi_2 : [0, 1] \to \mathbb{R}$.

In Section 5, we mentioned the zero–crossings method in the context of estimating N. The number of times that the graph of a periodic function ϕ crosses $y = 0$ can be read from the diagram. For a periodic function, a zero-crossing occurs between two local extrema, both of which have values different from zero. By periodicity, such a pair of extremal values generates exactly two zero-crossings. Hence, the number of zero–crossings is exactly $2\mu_\phi(]-\infty, 0[\times]0, \infty[)$. Using zero-crossing based estimators leads to even more pronounced identifiability issues. In addition, the number of zero-crossings is not stable with respect to perturbations of the signal.

5.2 Estimation of N in the presence of noise

Suppose now that we observe $S = \phi \circ \gamma + W$, where $\|W\| < \epsilon$ is a continuous function $W : [0, 1] \to \mathbb{R}$. In the previous section, since ϕ had a finite number of critical points, the persistence module actually had an interval decompositon. Here, we need to rely on the rectangle measures and the measure diagram, which exists, as argued in Section 2.5.3. In particular, since the module is q-tame, for any $\tau > 0$, the number of points more persistent than 2τ is finite, $\mu_S(\mathbb{R}^2 \setminus \Delta_\tau)$, where $\Delta_\tau = \{(b, d) \mid \mathrm{pers}(b, d) \leq 2\tau\}$ is the *τ-thickening of the diagonal*.

In the case when W is a stationary random process and each W_t is a continuous random variable, we cannot expect the extrema of S to all have the same value. Counting the multiplicity of points in $\phi|_{[0,1]}$ introduced in Section 5.1 needs to be adapted. We replace the notion of multiplicity by that of measures of neighborhoods. For a parameter $\tau > 0$, and $p = (x, y)$,

$$\mathrm{m}_{S,\tau}(p) = \mu_S([x - \tau, x + \tau] \times [y - \tau, y + \tau])$$

is the number of points from μ_S in $B(p, \tau)$, the $\|\cdot\|_\infty$-ball of radius τ centered at p. We generalize the definition of $N(S)$ to

$$\hat{N}(S, \tau) = \gcd\{\mathrm{m}_{S,\tau}(p) \mid p \in \mathrm{supp}(\mu_S) \setminus \Delta_\tau\}. \tag{5.6}$$

Notice that in this second version, the gcd is computed only on points further than τ from the diagonal (in the Chebyshev distance). In the noise–free setting $W = 0$, we have $\hat{N}(\phi \circ \gamma, 0) = N(\phi \circ \gamma)$. In general, τ needs to be sufficiently large compared to the noise ϵ, to account for the local displacement of points. It also needs to be bounded in terms of τ, not to confound the global structure of the diagram. We make those conditions explicit in Proposition 5.5.

Proposition 5.5. *Suppose that ϕ is non–degenerate and that $\epsilon < \delta/6$, where δ is the separation constant introduced above. Then, for $\tau > 0$ satisfying $2\epsilon < \tau < \delta/3$, we have that*

$$N(\phi \circ \gamma) = \hat{N}(S, \tau).$$

The lemma below formalizes the intuition that if points in the diagram are well-separated, then the mass of ϵ-balls does not change.

Lemma 5.6. *Suppose that $\tau > 0$ satisfies $2\epsilon < \tau < \delta/3$. Then, for any $p \in \mathrm{supp}(\mu_{\phi\circ\gamma})$ and $q \in B(p, \epsilon)$,*

$$\mu_S(B(q, \tau)) = \mu_{\phi\circ\gamma}(B(p, \tau)).$$

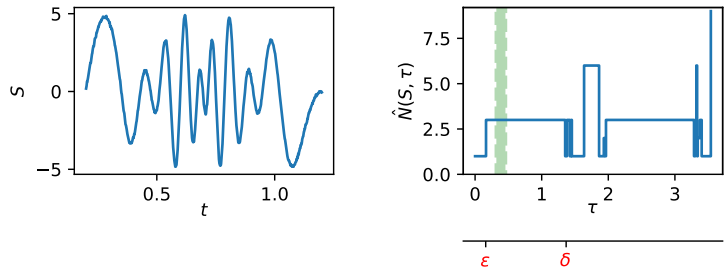


Figure 5: On the left, the graph of a sample from $S = \phi \circ \gamma + W$, with $\gamma(1) = 3$, $\delta_\phi = 1.37$ and $\|W\|_\infty \leq \epsilon = 0.16$. On the right, the graph of the piecewise-linear function $\tau \mapsto \hat{N}(S, \tau)$. Marked in green is the guarantee from Proposition 5.5.

Proof (Lemma 5.6). For any $0 < r < \delta$, the separation of points in the diagram of ϕ ensures that

$$\mu_{\phi\circ\gamma}(B(p, r)) = \mathrm{m}_{\phi\circ\gamma}(p). \tag{5.7}$$

By stability of the persistence measure of $\phi \circ \gamma$, $\mathrm{m}_{\phi\circ\gamma}(p) \leq \mu_S(B(p, \tau))$. As $d(p, q) < \epsilon$ implies $B(p, \epsilon) \subset B(q, 2\epsilon)$, we get $\mu_S(B(p, \epsilon)) \leq \mu_S(B(q, 2\epsilon)) \leq \mu_S(B(q, \tau))$.

For the inequality in the other direction, we use stability, $\mu_S(B(q, \tau)) \leq \mu_{\phi\circ\gamma}(B(q, \tau + \epsilon)) \leq \mu_{\phi\circ\gamma}(B(p, \tau + 2\epsilon))$. Since $\tau + 2\epsilon \leq 2\tau \leq \delta$, we can apply (5.7) for $r = \tau + 2\epsilon$ and $r = \tau$ $\mu_{\phi\circ\gamma}(B(p, \tau + 2\epsilon)) = \mathrm{m}_{\phi\circ\gamma}(p) = \mu_{\phi\circ\gamma}(B(p, \tau))$. □

Proof (Proposition 5.5). Consider $q \in \mathrm{supp}(\mu_S) \backslash \Delta_\tau$. By stability of the persistence diagram, there is $p \in \mathrm{supp}(\mu_{\phi\circ\gamma})$, such that $q \in B(p, \epsilon)$. Lemma 5.6 states that $\mu_S(B(q, \tau)) = \mu_{\phi\circ\gamma}(B(p, \tau))$, so that $N(\phi\circ\gamma)$ divides $\hat{N}(S, \tau)$. For any $p \in \mathrm{supp}(\mu_{\phi\circ\gamma})$, by Lemma 5.6, $\mu_{\phi\circ\gamma}(B(p, \tau)) = \mu_S(B(p, \tau))$, so that $\hat{N}(S, \tau)$ divides $N(\phi \circ \gamma)$. Thanks to the non–degeneracy of ϕ and by Proposition 3.1, we conclude that $\hat{N}(S, \tau) = N(\phi \circ \gamma) = N$. □

Figure 5 shows an example of S, where ϕ, δ and $\|W\|_\infty$ are known. The graph on the right shows that the proposed estimator $\hat{N}(S, \tau)$ is correct ($= 3$) for $\tau \in [0.16, 1.35]$, which almost covers $[\epsilon, \delta]$. The guarantee given by Proposition 5.5, marked in green, is much more pessimistic.

We propose to apply Proposition 5.5 in a random setting, with $(W_t)_{t\in[0,1]}$ being a Gaussian process. One case of practical interest in signal processing is that of white noise. However, we cannot treat it in the continuous setting. We refer the reader to Appendix 5.B, where we first discretize the signal and then devise guarantees. Instead, we propose to examine here the case of a regular Gaussian process. While the noise is now no longer bounded, we can calculate the probability that it remains bounded and that is the basis of our guarantee. We consider regular processes - that is, processes, which have a differentiable covariance function (Azäis and Wschebor, 2009, section 4.3).

Proposition 5.7. *Consider $S = \phi \circ \gamma + W$, with $(W_t)_{t\in[0,1]}$ a Gaussian process with covariance function of the form $Cov(s, t) = \Gamma(|t - s|)$, where*

$$\Gamma(t) = \sigma^2 \exp\left(-\frac{t^2}{2l^2}\right), \qquad \text{for some } \sigma, l > 0. \tag{5.8}$$

Then, for any $\tau \leq \delta/3$,

$$P(\hat{N}(S, \tau) = N) \geq 1 - \left(\frac{1}{l^2\pi} \exp\left(\tfrac{-\kappa^2}{2}\right) + 2\phi(-\kappa)\right), \tag{5.9}$$

where $\kappa = \frac{\tau}{2\sigma}$ and ϕ is the cumulative distribution function of a normal random variable.

Proof. Proposition 5.5 states that $\hat{N}(S, \tau) = N$ whenever S is $\frac{\tau}{2}$–close to $\phi \circ \gamma$, what translates to the following inclusion between the events

$$\{\|\phi \circ \gamma - S\|_\infty \leq \tfrac{\tau}{2}\} \subseteq \{\hat{N}(S, \tau) = N\}.$$

Define $Z = \frac{W}{\sigma}$, where $\sigma = \sqrt{\Gamma(0)}$. Then,

$$P(\|\phi \circ \gamma - S\|_\infty \leq \tfrac{\tau}{2}) = P\left(\sup_{t\in[0,1]} |W_t| \leq \tfrac{\tau}{2}\right)$$
$$= P\left(\sup_{t\in[0,1]} |Z_t| \leq \tfrac{\tau}{2\sigma}\right).$$

Set $u = \frac{\tau}{2\sigma}$. Since $\sup_{t\in[0,1]} |Z_t| = \max(\sup_{t\in[0,1]} Z_t,\ \sup_{t\in[0,1]} -Z_t)$, the events $\{\sup_{t\in[0,1]} Z_t > u\}$ and $\{\sup_{t\in[0,1]} -Z_t > u\}$ are a cover of $\sup_{t\in[0,1]} |Z_t| > u$.

$$P\left(\sup_{t\in[0,1]} |Z_t| \leq \tfrac{\tau}{2\sigma}\right) = 1 - P(|\sup_{t\in[0,1]} Z_t| > u),$$
$$\geq 1 - (P(\sup_{t\in[0,1]} Z_t > u) + P(\sup_{t\in[0,1]} (-Z_t) > u))$$
$$\geq 1 - 2P(\sup_{t\in[0,1]} Z_t > u),$$

where the ultimate inequality follows from the fact that $Z(t)$ and $-Z(t)$ have equal distributions. Let $r(s_1, s_2) = \mathbb{E}[Z_{s_1} Z_{s_2}]$ and $r_{1,1}(s,t) = \frac{\partial^2}{\partial s_1 \partial s_2} r(s,t)$. By Proposition 2.4,

$$P\left(\sup_{t\in[0,1]} Z_t > u\right) \leq \frac{\exp(-u^2/2)}{2\pi} \int_0^1 \sqrt{r_{1,1}(t,t)}\mathrm{d}t + 1 - \phi(u). \tag{5.10}$$

Here, $r(s_1, s_2) = \Gamma(s_1 - s_2) = \exp\left(-\frac{(s_1-s_2)^2}{2l^2}\right)$, so computing the derivative of the covariance function

$$r_{1,1}(s,t) = \frac{\partial^2}{\partial s_1 \partial s_2} r(s,t) = \frac{s-t+1}{l^2} \exp\left(-\frac{(s-t)^2}{2l^2}\right).$$

For $s = t$,

$$r_{1,1}(t,t) = \frac{1}{l^2}.$$

Computing the right–hand side of (5.10) and putting it all together, we obtain

$$P(\hat{N}(S,\tau) = N) \geq 1 - \left(\frac{1}{l^2\pi} \exp\left(\tfrac{-\kappa^2}{2}\right) + 2\phi\left(-\kappa\right)\right).$$

□

Let us analyze the bound (5.9). The parameter l in (5.8) quantifies the horizon of dependence of the stochastic process. The bound is increasing in l, implying that a long dependence (large l) yields a simpler structure for the method. In the limit $l \to \infty$, the expression on the right, $P(\hat{N}(S,\tau) = N) \geq 1 - 2\phi\left(-\kappa\right)$, is the same as the probability that $\|(\phi \circ \gamma)(0) - S(0)\|_\infty < \tau/2$. On the other hand, when the interactions become short-term (l small), the bound becomes trivial: there is a constant l_0 depending on κ, such that for $l \in]0, l_0]$, the bound is 0. The other governing parameter is κ - the ratio between the chosen scale τ and the standard deviation σ of the process. So, since $\tau \leq \frac{\delta}{3}$ has to be smaller than a fraction of the separation of f, $\kappa \leq \frac{\delta}{6\sigma}$ is also bounded. Regardless of how good of a method to choose τ we have, we obtain the best lower–bound of the probability of correctness only in terms of the quantities characteristic of the signal and the stochastic process. Experimental observations regarding l are discussed in Section 5.4.2.

5.3 Inference of an odometric sequence

In this section, we propose an odometric sequence for Problem 5.2, which we take as a subset of the local minima of the signal, chosen using the estimator of N proposed in Section 5.2. We first examine the noiseless case $\phi \circ \gamma$, with $\gamma : [0,1] \to [0,N]$. Let $\mathcal{C}$ be the set of local minima of $\phi \circ \gamma$. If K is the number of local minima in a single period of ϕ, then $\mathrm{card}(\mathcal{C}) = NK$. By picking every K-th minimum in $\mathcal{C}$, we obtain an odometric sequence. More precisely, we lexicographically index

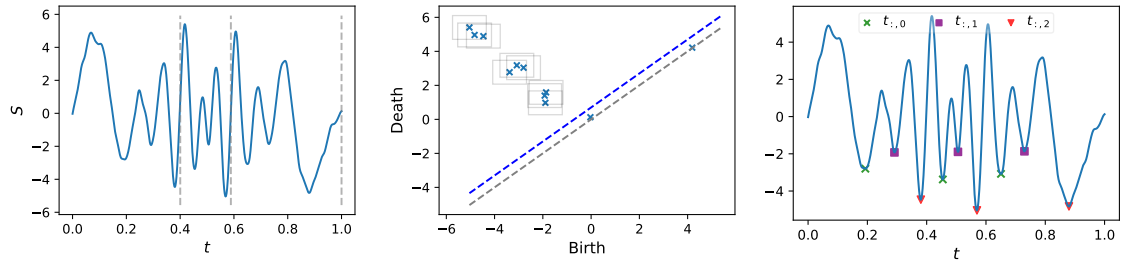


Figure 6: Example of an odometric sequence. On the left, an observed signal, with $N = 3$. In the center figure, its persistence diagram, with $\|\cdot\|_\infty$-balls of radius 0.7, marking the sets whose measure is evaluated. On the right, the observed signal with the odometric sequences stemming from persistent local minima. For each $k = 0, 1, 2$, $(\hat{t}_{n,k})_{n=1}^3$ is an odometric sequence.

the ordered minima $\mathcal{C} = (t_{n,k})_{\substack{n=1,\dots,N \\ k=1,\dots,K}}$ and then, for any $k = 1, \dots, K$, and $n = 2, \dots, N$, we have $\gamma(t_{n,k}) = \gamma(t_{n-1,k}) + 1$.

Notice that since N is part of the data in Problem 5.2, we do not even need to assume that ϕ is non–degenerate. If we wanted to use the same technique with N unknown, we would need to estimate it and the non–degeneracy of f is necessary for Proposition 5.1.

Let us consider the case of a signal corrupted by noise $S = \phi \circ \gamma + W$. While the location of the local minima of S is slightly different than those of $\phi \circ \gamma$, we can expect to be able to find a set of 'persistent' minima, in one-to-one correspondence with those of $\mathcal{C}$. Thus, the odometric property will not be exactly satisfied. An example is shown in Figure 6. We observe a signal, where $\gamma(1) = 3$. Note that two points lie between the diagonal and it's offset, marked by the dashed lines. These are not used in the estimation of N and the corresponding local minima are ignored when selecting the odometric sequences, shown on the right. Proposition 5.8 makes precise the guarantees on the existence of the sequence. It also states that it satisfies the odometric property, up to a constant which depends on ϕ, τ and ϵ.

Proposition 5.8. *Let $\tau > 0$ and $\hat{\mathcal{C}}_\tau$ be the set of local minima of S, corresponding to points in the diagram with persistence exceeding τ. If $\tau \in]2\epsilon, \delta/3[$, then*

$$\operatorname{card}(\hat{\mathcal{C}}_\tau) = NK.$$

In addition, if we order the minima $\hat{\mathcal{C}}_\tau = \{\hat{t}_{1,1}, \dots \hat{t}_{1,K}, \hat{t}_{2,1}, \dots, \hat{t}_{N,K}\}$, then

$$|\gamma(\hat{t}_{n,k}) - \gamma(\hat{t}_{n-1,k}) - 1| \leq 2R(\tau + 2\epsilon),$$

where

$$R(\nu) = \sup_{x \in \gamma(\mathcal{C})} \inf_{r>0} \{r \mid \phi(x+r) - \phi(x) > \nu,\ \phi(x-r) - \phi(x) > \nu\}$$

is a constant depending on ϕ and not depending on γ and N.

Before we proceed to the proof, we recall a result from the theory of persistence: between any two consecutive local minima of ϕ, there is a local maximum, larger than those minima by at least δ, of which we provide a short proof.

Lemma 5.9. *Let $x_1 < x_2$ be two, consecutive local minima of ϕ and suppose that the separation constant* (5.4) *for μ_ϕ is $\delta > 0$. Then,*

$$[\min(\phi(x_1), \phi(x_2)), \max(\phi(x_1), \phi(x_2)) + 2\delta] \subset \phi([x_1, x_2]).$$

Proof of Lemma 5.9. The lemma is in fact a direct consequence of Chazal et al. (2016, Lemma 3.38). Consider the 0-simplices $\sigma_1 \equiv x_1$ and $\sigma_2 \equiv x_2$, let $r = \min(\phi(x_1), \phi(x_2))$, $s = \max(\phi(x_1), \phi(x_2))$ and $t \coloneqq \max \phi|_{[x_1,x_2]}$. Because x_1 and x_2 are isolated local minima, the homology class of $\sigma_1 - \sigma_2$ is non trivial in $H_0(\phi^{-1}(]-\infty, s]) \cap [x_1, x_2])$. However, $\iota_s^t(\sigma_1 - \sigma_2)$ is trivial in $H_0(\phi^{-1}(]-\infty, t]) \cap [x_1, x_2])$. The morphism ι_s^t is therefore not injective and the rectangle $]-\infty, s] \times [t, \infty[$ in the diagram of $\phi|_{[x_1,x_2]}$ is not empty by Chazal et al. (2016, Lemma 3.38 (2)). On the other hand, by

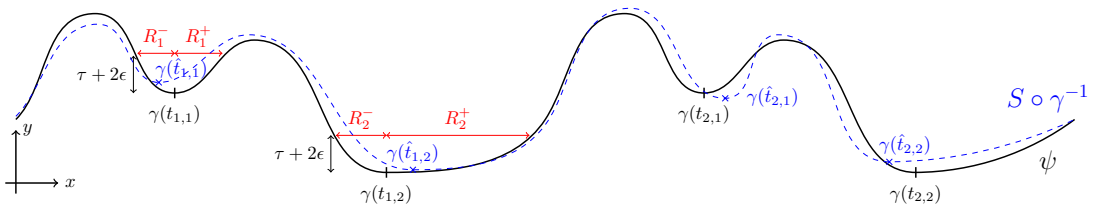


Figure 7: The figure shows an example of the graphs of ϕ and $S \circ \gamma^{-1}$, with local minima of both functions and R_k^-, R_k^+ from (5.11) illustrated.

(1) from that same Lemma, there is no point in $[r, s] \times [s, \infty]$, because ι_r^s is surjective. Therefore, there must be a point in the triangle defined by the vertices (s, s), (s, t), (t, t) and which is not on the diagonal Δ. Finally, the separation constant δ lower-bounds the distance of any point in the diagram to the diagonal, so in particular, $\delta < (t - s)/2$ and therefore $t \geq s + 2\delta$. □

Proof of Proposition 5.8. When $2\epsilon < \tau < \delta/3$,

$$\begin{aligned} \operatorname{card}(\hat{\mathcal{C}}_\tau) &= \mu_S(\mathbb{R}^2 \setminus \Delta_\tau) \\ &= \mu_{\phi\circ\gamma}(\mathbb{R}^2 \setminus \Delta) \\ &= \operatorname{card}(\mathcal{C}) \\ &= NK, \end{aligned}$$

where the second inequality follows from Lemma 5.6. This yields the first statement.
For the second part, let us start by defining $x_k = \gamma(t_{1,k}) \in [0, 1]$. By definition, $\gamma(t_{n,k}) - \gamma(t_{n-1,k}) = x_k + (n-1) - (x_k + (n-2)) = 1$, so

$$\begin{aligned} |\gamma(\hat{t}_{n,k}) - \gamma(\hat{t}_{n-1,k}) - 1| &= |\gamma(\hat{t}_{n,k}) - \gamma(\hat{t}_{n-1,k}) - (\gamma(t_{n,k}) - \gamma(t_{n-1,k}))| \\ &\leq |\gamma(\hat{t}_{n,k}) - \gamma(t_{n,k})| + |\gamma(\hat{t}_{n-1,k}) - \gamma(t_{n-1,k}))|. \end{aligned}$$

Let us introduce $R_k = \max(R_k^-, R_k^+)$, with

$$\begin{aligned} R_k^- &= \sup_r \{r \mid \forall y \in [x_k - r, x_k],\ \phi(y) \leq \phi(x_k) + \tau + 2\epsilon\}, \\ R_k^+ &= \sup_r \{r \mid \forall y \in [x_k, x_k + r],\ \phi(y) \leq \phi(x_k) + \tau + 2\epsilon\}, \end{aligned} \tag{5.11}$$

which upper–bounds how far one needs to look in either direction from x_k, so ϕ leaves the interval $[\phi(x_k), \phi(x_k) + \delta]$. We now need to show that, for all n, k,

$$|\gamma(\hat{t}_{n,k}) - \gamma(t_{n,k})| \leq \max\{R_k^+, R_k^-\}.$$

More specifically, we show that for any n, k, $\operatorname{card}(\hat{\mathcal{C}}_\tau \cap I_{n,k}) = 1$. Since there are exactly $\operatorname{card}(\hat{\mathcal{C}}_\tau) = NK$ intervals $I_{n,k}$, $\hat{t}_{n,k} \in I_{n,k}$. We illustrate this in Figure 7 and conclude by setting $R(\tau + 2\epsilon) = \max_k\{R_k^-, R_k^+\}$.
We now show $\operatorname{card}(\hat{\mathcal{C}}_\tau \cap I_{n,k}) = 1$. Observe that since S is continuous, it has a minimum a^* over $I_{n,k}$. This minimum is not achieved at the extremities of $I_{n,k}$, since

$$\begin{aligned} S(\gamma^{-1}(\gamma(t_{n,k}) + R_k^+)) - S(t_{n,k}) &\geq \tau + 2\epsilon + (W_{\gamma^{-1}(\gamma(t_{n,k})+R_k^+)} + W_{t_{n,k}}) > \tau. \\ S(\gamma^{-1}(\gamma(t_{n,k}) - R_k^-)) - S(t_{n,k}) &\geq \tau + 2\epsilon + (W_{\gamma^{-1}(\gamma(t_{n,k})-R_k^-)} + W_{t_{n,k}}) > \tau. \end{aligned}$$

The inequalities above also show that the a^* has a persistence greater than τ, so $a^* \in \hat{\mathcal{C}}_\tau$.
Suppose now that two elements $a^* < b^* \in \hat{\mathcal{C}}_\tau$ are contained in the same interval $I_{n,k}$. Lemma 5.9 states that there is a local maximum $c^* \in]a^*, b^*[\subset I_{n,k}$. If $t_{n,k} < c^*$, $S(c^*) - S(a^*) > \delta$ contradicts the minimality of R_k^+ and therefore $a^* = b^*$. If $t_{n,k} > c$, we obtain a contradiction of the minimality of R_k^- since $S(c^*) - S(b^*) > \delta$. □

Proposition 5.8 is similar in spirit to Chazal et al. (2017, Lemma 17). It is interesting to note that the guarantee is not cumulative: more specifically, for any $0 \leq n, m \leq N$, we have

$$|\gamma(\hat{t}_{n,k}) - \gamma(\hat{t}_{m,k}) - (n-m)| \leq 2R(\tau + 2\epsilon).$$

The relative error is therefore small for $m - n$ large.

As we would expect, the more strictly convex ϕ is around prominent local minima, and the smaller τ and ϵ, the smaller R_k is. On the other hand, convexity of ϕ implies that the sampling error of ϕ will be greater.

5.4 Numerical experiments

We introduce more robust versions of $\hat{N}$ and we quantify their performance on synthetic signals.

5.4.1 Practical adaptations of $\hat{N}$

We introduce $\hat{N}_c$, obtained by replacing in (5.6) the number of points in neighborhoods by the sizes of some partitions, and $\hat{N}_c^T$, which does not require a selection of τ. Specifically, let $\mathcal{A}_\tau$ be a partition of $\mathrm{supp}(\mu_S) \cup \Delta$ obtained via single–linkage hierarchical clustering with a scale-parameter τ. We define

$$\hat{N}_c(S, \tau) = \gcd\{\mu_S(A) \mid A \in \mathcal{A}_\tau, A \cap \Delta = \emptyset\}. \tag{5.12}$$

Let us now consider the map $h_S : \tau \mapsto \hat{N}_c(S, \tau)$, for $\tau \in \mathbb{R}_+^*$, with the convention that $\gcd(\emptyset) = 1$. The map h_S is piece–wise constant and equal to 1 on $](\max(S) - \min(S)), \infty[$, since the persistence of all points is bounded by $(\max(S) - \min(S))/2$. Let $\mathcal{I}_n$ be the collection of maximal intervals, where $h_S(\tau) = n$. Finally, we set $\hat{N}_c^T$ to the value which is realized over the longest interval

$$\hat{N}_c^T(S) = h_S(\tau), \qquad \text{for } \tau \in \operatorname*{argmax}_{I \in \mathcal{I}_n} \mathrm{length}(I). \tag{5.13}$$

Let us motivate the introduction of $\hat{N}_c$. Experiments show that points in the diagrams μ_S form non–overlapping clusters. The diameters of those clusters can be greater than the distance which separates them (when the condition $6\epsilon < \delta$ is not satisfied). In such a situation, the estimator (5.6) will be incorrect, even if, to the human eye, the clusters are easily identifiable. In the proof of Lemma 5.6, we notice that if the upper bound of noise ϵ is smaller than the separation $\delta/3$, the points in the persistence diagram are partitioned into neighborhoods of size τ. In particular, if two points in the diagram share a neighbor at distance no greater than τ, their τ-neighborhoods are equal due to the separation property satisfied by ϕ. This is exactly the property satisfied by a partition produced by single–linkage clustering. Namely, that algorithm produces the coarsest partition $\mathcal{A}_\tau$ of $\mathbb{R}^2$ such that: for any $p_1, p_2 \in \mathrm{supp}(\mu_S)$, if there exists $q \in \mathrm{supp}(\mu_S)$ such that $d(p_1, q) < \tau$ and $d(q, p_2) < \tau$, then there exists $A \in \mathcal{A}$ such that $p_1, p_2 \in A$. This motivates replacing $(B(p, \tau))_{p \in \mathrm{supp}(\mu_S \setminus \Delta_\tau)}$ with a partition $\mathcal{A}_\tau$ from a single-linkage algorithm, leading to an algorithm more robust to noise and to the choice of τ.

The second estimator $\hat{N}_c^T$ allows us to avoid the selection of τ: it takes the value for which h_S is constant for the longest interval. Suppose there exists $\tau_0 > 0$ such that $h_S(\tau_0) = N$ and that no point in A_τ is isolated. Then N divides $h_S(\tau)$, for all $\tau \in]\tau_0, \tau_{\max}[$, where $\tau_{\max}$ is the smallest value of τ such that A_τ contains a single cluster, intersecting the diagonal. By a similar argument as in Proposition 5.5, $h_S(\tau) = N$ for $\tau \in]2\epsilon, \delta/3[$. So, there exists $I \in \mathcal{I}_N$ such that $]2\epsilon, \delta/3[\subset I$. Usually, this inclusion will be strict, because the clustering scheme is less sensitive to ϵ.

The automatic choice of τ comes with some limitations. Since we set the domain of h_S in a data–driven way to $[0, (\max S - \min S)/2]$, we encounter problems when δ is small compared to the amplitude of the signal. It is particularly visible when the signal ϕ is close to a degenerate–function, even in the noiseless setting. For ϕ_r from Figure 4, we have $\delta = \min(\|(-1-r, 1+r) - (-1, 1)\|_\infty, \|(-1-r, 1+r)\|_\infty) = \min(r, 1-r)$, so

$$h_{\phi_r|_{[0,1]}}(\tau) = \begin{cases} 1 & \text{for } \tau < r, \\ 2 & \text{for } r \leq \tau < 1 - r, \\ 1 & \text{for } 1 - r \leq \tau \leq 1, \end{cases} \qquad \text{if } 0 \leq r < \tfrac{1}{2},$$

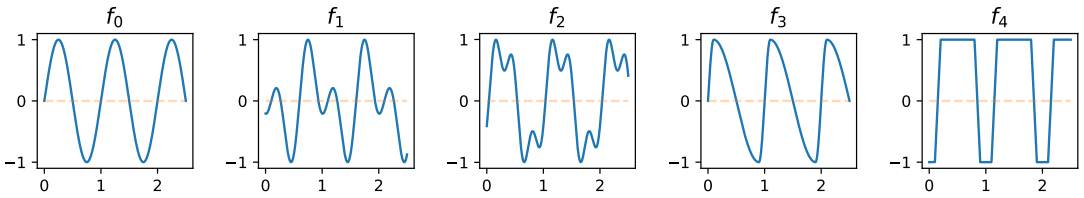


Figure 8: The graphs of five one–periodic functions over $[0, 2.5]$.

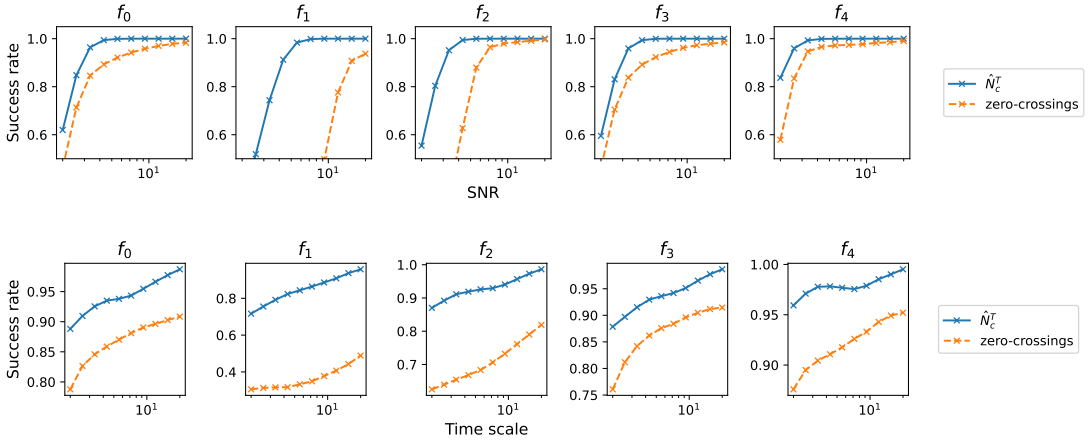


Figure 9: The success rates for estimating N, while varying the noise level σ (top row) and the time–scale l (bottom row) of the Gaussian process W with covariance (5.8).

and $h_{\phi_r|_{[0,1]}}(\tau) \equiv 1$ for $\frac{1}{2} \leq r \leq 1$. Therefore, $\hat{N}_c^T(\phi_r|_{[0,1]}) = 1$ if and only if $r > 1/3$. When $r \leq 1/3$ is too small, the function looks like a degenerate function for a range of scale τ too large. Such behavior is to be expected from any method of automating the choice of τ: estimating N amounts to looking for symmetries in the diagram of S, so if the true symmetry is present only for a narrow range of scales τ, it will be harder to capture that scale correctly and automatically.

Ultimately, we are often interested in computing the odometric sequence for which a value of τ is needed. We take τ to be the middle of the longest interval on which h_S is constant, so that $\hat{N}_c(S, \tau) = \hat{N}_c^T(S)$.

5.4.2 Estimating N on synthetic data

We choose five template functions $\phi_0, \ldots, \phi_4$ and we picture their periods in Figure 8. The functions ϕ_0, ϕ_3 and ϕ_4 all have a single pair of extrema per period, while ϕ_1 and ϕ_2 have two and three pairs per period respectively. All the functions are normalized so that their range is $[-1, 1]$, but their separation parameters δ differ: $\delta_0 = 1 = \delta_3 = \delta_4$, $\delta_1 = 0.21$ and $\delta_2 = 0.13$. We generate 100 reparametrisations γ with the following procedure: sample N_k uniformly from $\{5, \ldots, 50\}$ and $(t_l^k)_{l=1}^{N_k}$ uniformly in $[0, 1]$, then linearly interpolate $(t_l^k, l)_{l=1}^{N_k}$ to obtain γ_k. Ordered ascendingly, $(t_l^k)_{l=1}^{N_k}$ define the starts of the periods. This yields a collection of 500 signals, that we perturb with noise $(W_t)_{t\in[0,1]}$, a Gaussian process with covariance as in (5.8), with parameters $\sigma \in [0.0001, 6]$ and $\gamma \in [0.01, 0.4]$.

We compare two estimators. We define an oracle, based on the zero crossings method, by specifying the number of zero crossings in a period of ϕ_k - information that is inaccessible in practice. We benchmark $\hat{N}_c^T$ against this zero crossings oracle, with the fraction of samples for which N was estimated correctly as a metric. We present the results for each template function f_k in Figure 9. Globally, the performance of both estimators is increasing as the SNR increases (σ decreases). The estimator $\hat{N}_c^T$ outperforms the oracle in all scenarios. The sensitivity of the zero–crossings method to local extrema close to zero is confirmed by the poor performance on ϕ_1: a noise of small amplitude can create additional zero–crossings. Even though $\hat{N}_c^T$ also shows a degraded performance for this periodic function, it outperforms the benchmark and that by the

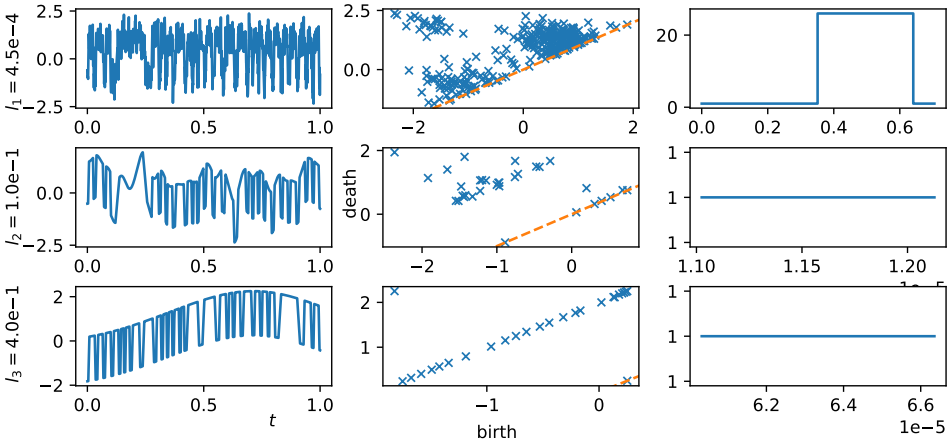


Figure 10: The left column represents realizations of $S = (\phi_4 \circ \gamma)(t) + W_t$, for W_t sampled with different time–scales l_1, l_2, l_3. From left to right, each row contains the graph of S, the diagram μ_S and h_S.

biggest margin. The difference in performance between the two methods is also most noticeable on ϕ_1. At high noise levels, the pairs of local and global extrema get identified as the same, as it would be the case for a degenerate function.

Since $\|W\|$ is decreasing in the time–scale l, we expect the success rate to be monotonously increasing in l. However, it decreases and reaches a minimum at around $l = 0.1$. We believe that the initial increase is indeed due to the decrease in $\|W\|_\infty$. In this range of l, the noise is high–frequency and of small amplitude, leading to many points in the diagram, as shown in the top row in Figure 10. The difficulty in estimating N comes essentially from the presence of many points close to the diagonal. However, the points corresponding to minima of ϕ are generally separable from those spurious points. When the time scale increases to around 0.05 - 0.1, the noise has lower frequencies, almost in the range of γ'. Looking at the graph on the left, the noise shifts some periods vertically, creating isolated points in the persistence diagram. For a wide range of scales, those points are isolated instead of being clustered with the rest. When l increases even further, the induced oscillations are generally of smaller amplitude and resemble a slow drift of the period. In the diagram, such a drift manifests itself as a densely–sampled line, which can be identified as a single cluster.

5.5 Conclusion and perspectives

We formulated two inverse problems: estimation of N and finding odometric sequences. We use the homogeneity property to propose an estimator of N, which generalizes that of zero-crossings. We prove that the estimation is correct for a wide class of functions, even when the signal is obfuscated with additive noise. We propose a more robust variant $\hat{N}_c$ and a parameter–free version $\hat{N}_c^T$. We show that for certain types of functions that we describe, this estimator perform similarly to a standard approach. Using the estimation of N, we propose a method of obtaining odometric sequences, which are stable with respect to perturbations of the input signal.

From a theoretical perspective, using the persistence diagrams as the sole descriptor of a signal limits identifiability, as explained via Figure 4. A richer descriptor which does not suffer from similar problems is the merge tree, see for example Pegoraro and Secchi (2021). However, we are not aware of a way to exploit trees for the problem of estimating N. A possible avenue suggested to us would be to use merge trees to validate the segmentation: compare merge trees associated to each period. If the edit distance between the trees is greater than the separation, it indicates that the periodic structure has not been well-identified.

In practical scenarios, we never observe an exact number of revolutions, so the assumption that $N \in \mathbb{N}$ is a significant limitation of our work. It is not a noticeable problem in the case of the numerical experiments, because that constraint is satisfied for the synthetic examples, and the

magnetic signals encountered in the application motivating this thesis often have a single pair of local extrema per period.

Another limitation is the use of the greatest common divisor in (5.6). It is not robust to any perturbation: a change in a single value in the set leads to the gcd being 1. For this reason, the whole procedure is very sensitive to outliers. Algorithms which solve the approximate common divisor (ACD) problem, like the Simultaneous Diophantine approximation (SDA) (Galbraith et al., 2016), could be used instead of the gcd. They are shown to be stable with respect to perturbations of the individual elements, and, preliminary tests on a handful of examples show that the estimation is much more robust. However, the algorithms require additional information: for example, before applying SDA in our case, we would need to give an integer lower–bound for $\log_2(N)$. While the bound should be sharp in theory, our preliminary experiments on a handful of examples have shown promise. However, a drawback is the increased computational complexity in the number of points. A more serious problem is how to adapt the segmentation method. In the definition of the odometric sequence, we heavily rely on the fact that there are exactly NK persistent points in the persistence diagram, while with these algorithms, it is not necessarily the case.

Finally, the presented method is designed for univariate data. While it can be applied to a single covariate from a multi–variate signal, a well–designed approach for several variables could make the estimation more robust. The restriction to univariate data is related to how we construct the persistence diagram and it is not immediately clear how to generalize for multi–variate data while preserving properties like those shown in Chapter 3. We explore this aspect in Chapter 6.

Appendix

5.A Proof of Proposition 5.3

Before the proof Proposition 5.3, we show that certain multi-sets can be realized as the persistence diagrams of sub level sets of a continuous function. Lemma 5.10 is similar to Lesnick (2015, Proposition 5.8), but we include the proof for two reasons: first, to be self–contained and, second, to provide a simple construction specific to our case, where we characterize the period of ϕ.

Lemma 5.10. *Let D be a non–empty persistence diagram with a unique, most persistent point $p_0 = (b_0, d_0)$. Assume that $D \setminus \Delta$ is finite and $D \subset [b_0, d_0] \times [b_0, d_0]$. Then, there exists a continuous function $\phi : \mathbb{R} \to \mathbb{R}$ with period 1, such that $D = D(\phi|_{[0,1]})$, $\phi(0) = \phi(1)$.*

Proof. We first construct a candidate function ϕ as shown in Figure 5.A.1 and then, prove that it satisfies the desired properties. Let us enumerate the points in $D \setminus \Delta$, in lexicographic order of increasing birth and decreasing death values: $(b_0, d_0), \dots (b_{K-1}, d_{K-1})$, where $K = |D \setminus \Delta|$. Then, consider the function

$$\begin{array}{rccc} \phi_K : & \{0, \dots, 2K\} & \to & \mathbb{R} \\ & k & \mapsto & \begin{cases} d_{k/2}, & \text{if } k \text{ is pair} \\ b_{(k-1)/2}, & \text{otherwise,} \end{cases} \end{array}$$

with the convention $d_K = d_0$, and the linear interpolation $\phi : [0, 1] \to \mathbb{R}$.

First, note that ϕ is continuous and that $\phi(0) = \phi_K(0) = d_0 = d_K = \phi_K(2K) = \phi(1)$. The sequence of extrema is aperiodic if either $K = 0$ or $(b_0, d_0) \neq (b_K, d_K)$. An aperiodic sequence leads to ϕ_K aperiodic and so ϕ can be extended to $\mathbb{R}$ to have period 1. If $(b_0, d_0) = (b_K, d_K)$, modify the subdivision of $[0, 1]$ to be non–uniform, in which case the period of ϕ is also 1.

We argue that $D = D(\phi)$. It can be seen by noticing that every local minimum is paired with the local maximum on its left. Formally, we calculate $D(\phi)$ with Algorithm 1. The local minima and maxima of ϕ are $(x_{b_k})_{k=1}^K$, $(x_{d_k})_{k=1}^K$ respectively. We only need to consider, for every d_k, what point it is the second coordinate of. The connected component I of x_{d_k} in $\phi^{-1}(]-\infty, d_k])$ is partitioned into two non–empty sets $I_- = [0, x_{d_k}] \cap I$ and $I_+ = [x_{d_k}, 1] \cap I$. Denote $M := \max(\phi)$ and suppose that $d_k < M$. Then, since the values of the local minima are increasing by definition

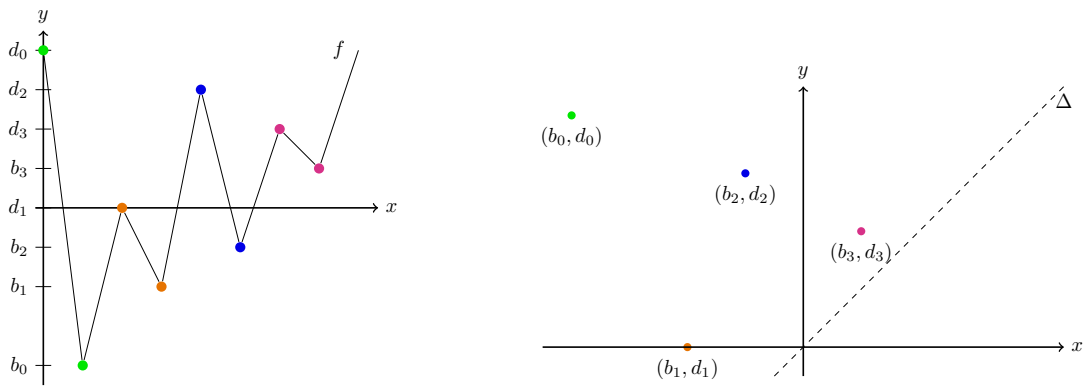


Figure 5.A.1: Illustration of the function ϕ.

of ϕ, $\min \phi|_{I_-} < \min \phi|_{I_+}$ and $x_{b_k} = \operatorname{argmin} \phi|_{I_+}$, so d_k becomes the second coordinate of b_k. Therefore, there are exactly N copies of each point (b_k, d_k), such that $d_k < M$. If $d_k = M$, the second coordinate of a point whose second coordinate has not been set yet, is set to d_k. □

Proof of Proposition 5.3. Let $\mu = \frac{1}{N(\phi|_{[0,1]})}\mu_{\phi|_{[0,1]}}$. Then, μ is a persistence measure: by definition, it is a finite sum of weighted point measures. The weights are integral, because $N(\phi|_{[0,1]})$ divides the multiplicity of every point.

Thanks to Proposition 2.34, we can apply Lemma 5.10to the persistence diagram of $\mu_{\phi|_{[0,1]}}$, which gives us $\phi_2 : \mathbb{R} \to \mathbb{R}$ such that $\mu_{\phi_2|_{[0,1]}} = \mu = \frac{1}{N(\phi|_{[0,1]})}\mu_{\phi|_{[0,1]}}$. □

5.B $\hat{N}(S)$ for a signal with white noise

In Section 5.2, we discuss the estimation of $\gamma(1)$ in the presence of noise. The aim of this section is to derive guarantees in the case where the noise entries are independent. Such noise is called "white noise" and it is often considered in applications. Continuous equivalents of white noise processes exist, but they are difficult to analyze and many objects, like persistence diagrams, are not defined for such processes. Since this consideration is motivated by practical considerations, we choose to derive guarantees on a discrete-time signal, obtained by sampling the continuous signal at a frequency $\omega > 0$. For example, in magnetic odometry motivating this work, $\omega = 125Hz$.

Let $M = \lfloor \omega \rfloor$. We define the sampling operator $L : C^0([0,1]) \to \mathbb{R}^M$ and the piecewise linear interpolation operator $F : \mathbb{R}^{M+1} \to C^0([0, \frac{M}{\omega}])$. Specifically, for $h \in C^0[0,1]$ and $a \in \mathbb{R}^{M+1}$,

$$\begin{aligned} L(h) &\coloneqq (h(\tfrac{m}{\omega}))_{m=0}^{M}, \\ F(a)(t) &\coloneqq a_m(\omega t - (m-1)) + a_{m-1}(m - \omega t), \quad \text{for any } t \in [\tfrac{m-1}{\omega}, \tfrac{m}{\omega}[. \end{aligned}$$

We denote the composition of the two by $T = F \circ L : \mathcal{C}^3([0,1]) \to \mathcal{C}^0([0, \frac{M}{\omega}])$. We examine the approximation obtained by interpolating a uniform sample from $f \circ \gamma$.

Lemma 5.11. *Suppose that ϕ and γ are $\mathcal{C}^3$. Then, for $t \in \left[0, \frac{M}{\omega}\right[$,*

$$|T(\phi \circ \gamma)(t) - (\phi \circ \gamma)(t)| \leq \tfrac{2}{\omega^2}(\|\phi''\|_\infty \|\gamma'\|_\infty^2 + \|\phi'\|_\infty \|\gamma''\|_\infty) + \mathcal{O}(\omega^{-3}).$$

Proof. Let $h = \phi \circ \gamma$. First, we show that $|T(h) - h| = \mathcal{O}\left(\frac{\|h''\|_\infty}{\omega^2}\right)$. For m such that $t \in \left[\frac{m-1}{\omega}, \frac{m}{\omega}\right[$,

$$\begin{aligned} |T(h)(t) - h(t)| &= \left|(h\left(\tfrac{m}{\omega}\right) - h(t))(\omega t - (m-1)) + \left(h\left(\tfrac{m-1}{\omega}\right) - h(t)\right)(m - \omega t)\right| \\ &= |(h'(\tfrac{m}{\omega})(t - \tfrac{m}{\omega}) + \tfrac{1}{2}h''(t_m^*)\left(t - \tfrac{m}{\omega}\right)^2)(\omega t - (m-1)) \\ &\quad + (h'\left(\tfrac{m-1}{\omega}\right)\left(t - \tfrac{m-1}{\omega}\right) + \tfrac{1}{2}h''(t_{m-1}^*)\left(t - \tfrac{m-1}{\omega}\right)^2)(\omega t - (m-1))|, \end{aligned}$$

where $t_{k-1}^* \in [\frac{m-1}{\omega}, t]$, $t_k^* \in [t, \frac{m}{\omega}]$ are given by the Taylor–Lagrange expansion of h. By an expansion of h' around $\frac{m-1}{\omega}$,

$$\begin{aligned} |T(h)(t) - h(t)| &\leq \tfrac{1}{\omega^2}\left(\left|h''\left(\tfrac{m-1}{\omega}\right)\right| + |h''(t_1)| + \tfrac{1}{2\omega}h'''(t_2)\right) \\ &\leq \tfrac{1}{\omega^2}\left(2\|h''\|_\infty + \mathcal{O}\left(\tfrac{1}{\omega}\right)\right). \end{aligned}$$

□

The following proposition is an application of the stability result to $[0, \frac{M}{\omega}]$.

Proposition 5.12. *Let ϕ, γ be $\mathcal{C}^3$ as above. In addition, suppose that ω is large enough so that $[0, \frac{M}{\omega}]$ contains all the local extrema of $\phi \circ \gamma$. Then,*

$$\mathrm{d_b}(D(T(\phi \circ \gamma)), D(\phi \circ \gamma)) \leq \tfrac{2}{\omega^2}(\|\phi''\|_\infty \|\gamma'\|_\infty^2 + \|\phi'\|_\infty \|\gamma''\|_\infty) + \mathcal{O}(\omega^{-3}).$$

Proof. Since all extrema of $\phi \circ \gamma$ are included in $I = [0, \frac{M}{\omega}]$, we have equality between the diagrams $D\left(\phi \circ \gamma|_I\right) = D(\phi \circ \gamma)$. Using the stability of the persistence diagram, on the filtrations induced by $T_\omega(\phi \circ \gamma)$ and $\phi \circ \gamma$ on I, we conclude with the Lemma 5.11. □

Proposition 5.12 relates the diagram of the noise–less signal to the sampled version. Let us now consider a noisy sample from $p \circ \gamma$. More precisely,

$$(S_m)_{m=1}^M = (\phi(\gamma(\tfrac{m-1}{\omega})) + W_m)_{m=1}^M,$$

where W_m are independent, identically distributed centered Gaussian random variables with standard deviation σ.

Proposition 5.13. *Let $\omega > 0$ and $0 < \tau < \delta/3$ be such that $\alpha := \tau/2 - \frac{1}{\omega^2} C_{\phi,\gamma} > 0$, where*

$$\begin{aligned} C_{\phi,\gamma} =& \|\phi''\|_\infty \|\gamma'\|_\infty^2 + \|\phi'\|_\infty \|\gamma''\|_\infty \\ &+ \tfrac{1}{2}(\|\phi'''\|_\infty \|\gamma'\|_\infty^3 + 3\|\phi''\|_\infty \|\gamma'\|_\infty \|\gamma''\|_\infty + \|\phi'\|_\infty \|\gamma'''\|_\infty). \end{aligned}$$

Then,

$$P(\hat{N}(F((S_m)_{m=1}^M)) = N, \tau) \geq (1 - \phi(\tfrac{\alpha}{\sigma}))^\omega.$$

Proof. We show that the linear approximation of the exact signal is good enough, so that the entries can be perturbed with noise. Indeed, $\|F((S_m)_{m=1}^M) - \phi \circ \gamma\|_\infty \leq \|(W_m)\|_\infty + \|T(\phi \circ \gamma) - f \circ \gamma\|_\infty$. By Lemma 5.11, $\|F((S_m)_{m=1}^M) - \phi \circ \gamma\|_\infty \leq \|(W_m)\|_\infty + \frac{1}{\omega^2} C_{\phi,\gamma}$. Since W_m are independent, centered normal variables with variance σ^2, $P(W_m \leq \alpha) = (1 - \phi(\frac{\alpha}{\sigma}))^\omega$. For $\alpha = \tau/2 - \frac{1}{\omega^2} C_{\phi,\gamma}$, $\|F((S_m)_{m=1}^M) - \phi \circ \gamma\|_\infty \leq \alpha$ with probability $(1 - \phi(\frac{\alpha}{\sigma}))^\omega$. By Proposition 5.5, $P(\hat{N}(F((S_m)_{m=1}^M)) = N, \tau) < (1 - \phi(\frac{\alpha}{\sigma}))^\omega$. □

The bound in Proposition 5.13 is decreasing in ω, as soon as the approximation is good enough: $\alpha > 0$. There is a trade-off between the probability of $\hat{N}$ being correct and the precision of the odometric sequence, which increases with w.

Chapter 6

From periodic functions to curves

In the previous chapters, we have considered two problems on uni-dimensional signals, samples from a reparametrized periodic function. In the car positioning application that motivates this thesis, there are several magnetometers, each providing a three-dimensional signal. In this chapter, we discuss strategies to apply the estimators of the number of periods developed in Chapter 5 to the three-dimensional data from a single magnetometer,

$$\begin{array}{rccl} \mathbf{S}: & [0,T] & \to & \mathbb{R}^3 \\ & t & \mapsto & \boldsymbol{\phi}(\gamma(t)) + \mathbf{W}(t), \end{array} \tag{6.1}$$

where $\boldsymbol{\phi} : \mathbb{S}^1 \to \mathbb{R}^3$ and $\mathbf{W} : [0,T] \to \mathbb{R}^3$ are continous and the latter represents noise. Some assumptions and methods developed in this chapter are specific to the application. We also emphasize that this chapter is more exploratory in nature and guided by experimental results. It is certainly less complete than Chapters 3-5.

Our point of view is to reduce the problem to the one-dimensional Problem 5.1. To do so, we use linear projections $\mathbb{R}^3 \to \mathbb{R}$ which are a simple, yet sufficient family of functions. Indeed, historically, the procedures from Le Goff et al. (2012) and Zabulon et al. (2019) use the intensity of the magnetic field measured only along a single axis, which is chosen to maximize the signal amplitude. A second reason is specific to the structure of the magnetic signal $\mathbf{S}$: there exists $v^\star \in \mathbb{S}^2$, such that the projection along $v^\star$ annihilates a major component of the noise $\mathbf{W}$. Our driving question is deciding between two concurring approaches: maximize the signal or minimize the noise, explored in Sections 6.1 and 6.2 respectively. We propose several computable choices in Section 6.3, and we evaluate their performance in a simple numerical experiment in Section 6.4.

We propose strategies to choose a single projection or a measure, from which to sample projections, with the goal of maximizing the performance of an estimator, $\hat{N} : C([0,T],\mathbb{R}) \to \mathbb{N}$. To a vector $v \in \mathbb{S}^2$, which we will also call *direction*, we can associate a projection $\mathbb{R}^3 \to \mathbb{R}$ and we denote by $\mathbf{S}_v : [0,T] \to \mathbb{R}$ the signal $\mathbf{S}$ projected along v. For a probability measure ν on $\mathbb{S}^2$, we define the estimator

$$I(\mathbf{S},\nu) := \max \mathcal{L}(\hat{N}(\mathbf{S}_v)), \qquad v \sim \nu. \tag{6.2}$$

Note that $\mathbf{S}$ is considered fixed and the mode is taken with respect to the distribution on $\mathbb{N}$ induced by $v \mapsto \hat{N}(\mathbf{S}_v)$. Recall that the mode of an $\mathbb{N}$-valued random variable X distributed according to $\mathbb{P}$ is $\mathrm{mode}(X) := \mathrm{argmax}_x \mathbb{P}(X = x)$ and the maximum extends (6.2) to cases where the mode is not unique. Our goal in this chapter is to propose a data-driven choice of a probability measure $\nu = \nu(\mathbf{S})$ on $\mathbb{S}^2$ which minimizes the risk $\mathbb{P}_{\mathbf{S}}(|I(\mathbf{S},\nu(\mathbf{S})) - (\gamma(T) - \gamma(0))| < 1)$, where $\mathbf{S}$ is a random signal from (6.1).

We will consider two particular types of distributions ν. First, ν can be a discrete measure, sometimes supported on a singleton: to $v \in \mathbb{S}^2$, we can associate $\nu = \delta_v$. Second, distributions with a density with respect to the uniform measure μ on $\mathbb{S}^2$, with $\mu(\mathbb{S}^2) = 4\pi$. For a non-negative and measurable function[1] $f : \mathbb{S}^2 \to \mathbb{R}$ with $\int f d\mu > 0$, we denote by $f\mu$ a measure with density $\frac{f}{\int_{\mathbb{S}^2} f d\mu}$ with respect to μ.

[1] In cases that we will consider, f is continuous.

6.1 Maximizing the signal: projecting along directions with high amplitude

In Tazaki et al. (2001), the authors place the magnetic sensor in an orientation which maximizes the amplitude of the signal. Since projections are linear, a heuristic to maximize the signal is to select $\nu = \delta_{v_A}$, where $v_A \in \operatorname{argmax}_{v\in\mathbb{S}^2} A(\boldsymbol{\phi}_v)$ and $A(\boldsymbol{\phi}_v) \coloneqq \max \boldsymbol{\phi}_v - \min \boldsymbol{\phi}_v$. Depending on the curve, the true maximum might not be unique. With noise and the optimization error, the results of such a selection procedure have large variability. Hence, instead of selecting the maximum, we propose to sample $v \in \mathbb{S}^2$ proportionally to $A(\boldsymbol{\phi}_v)$. Specifically, we define $\mu_{\boldsymbol{\phi}} \coloneqq (v \mapsto A(\boldsymbol{\phi}_v))\mu$.

In one-dimensional odometry, estimators $\hat{N}$ and $\hat{N}_c$ require choosing a parameter $\tau > 0$. It is related to the separation constant introduced in (5.4),

$$\delta(\boldsymbol{\phi}_v) = \min(\{d(p,q) \mid p,q \in D(\boldsymbol{\phi}_v),\, p \neq q\} \cup \{d(p,\Delta) \mid p \in D(\boldsymbol{\phi}_v)\}),$$

which quantifies the difficulty of Problem 5.1, as stated by Propositions 5.4 and 5.5. The purpose of this section is to present and show Proposition 6.5, which establishes a heuristic for selecting τ in terms of $C_{\boldsymbol{\phi}}$, a quantity related to the geometry of the curve.

Let us for now fix $v \in \mathbb{S}^2$ and reason on uni-dimensional signals, $\boldsymbol{\phi}_v$ and $\mathbf{S}_v$. Recall the estimator $\hat{N}_c(\mathbf{S}_v, \tau)$ first introduced in (5.12).

Definition 6.1 ($\hat{N}_c$ from (5.12))**.** Let $D(\mathbf{S}_v)$ be the persistence diagram of $\mathbf{S}_v$ with periodic boundary conditions, as defined in Section 3.1. We denote by $\mathcal{C}^\tau$ the clusters obtained by single linkage at scale $\tau > 0$ on the persistence diagram $D(\mathbf{S}_v) \sqcup \Delta$. Then, we define

$$\hat{N}_c(\mathbf{S}_v, \tau) = \gcd\{\operatorname{card}(D(\mathbf{S}_v) \cap C) \mid C \in \mathcal{C}^\tau, C \cap \Delta = \emptyset\}.$$

Consider $\delta(\boldsymbol{\phi}_v)^\tau = \min_{C,C'\in\mathcal{C}^\tau}\{d(p,q) \mid p \in C, q \in C'\}$, the minimum separation between any two clusters C, $C' \in \mathcal{C}^\tau$. We introduce two multi-scale versions of (5.4),

$$\delta(\boldsymbol{\phi}_v)^\star = \max_\tau \delta(\boldsymbol{\phi}_v)^\tau, \qquad \delta(\boldsymbol{\phi}_v)^\dagger = \max_{\tau \mid \hat{N}(\boldsymbol{\phi}_v,\tau)=1} \delta(\boldsymbol{\phi}_v)^\tau.$$

The separation constant $\delta(\boldsymbol{\phi}_v)^\star$ is the maximal separation of clusters, achieved for clustering at a certain scale. The second constant is similar, although it takes into account the symmetries in the function. Figure 1 illustrates the clustering for different values of τ and the separation constants on an example. Let $f : \mathbb{S}^1 \to \mathbb{R}$ be a continuous function.

Proposition 6.2. *If f is non–degenerate in the sense of Definition 5.2, then*

$$\delta(f) \leq \delta(f)^\dagger \leq \delta(f)^\star.$$

Proof. For any $\tau > 0$, the minimum in $\delta(f)^\tau$ is taken only on a subset of pairs of points, so $\delta(f) \leq \delta(f)^\tau$. Suppose that $\delta(f) > 0$. Then, for $\tau = \frac{\delta(f)}{2}$, the clusters in $\mathcal{C}^\tau$ are of cardinality 1 and therefore, $\hat{N}(f,\tau) = 1$ so $\delta(f)^\tau \leq \delta(f)^\dagger$. The inequality $\delta(f)^\dagger \leq \delta(f)^\star$ follows from the definitions. □

Figure 1 shows that the separation constants differ in situations where a single period of f is close in the sup norm to several periods of another function, $\tilde{f}$. The particularity of $\delta(f)^\star$ is that it upper-bounds nearest-neighbor distances.

Proposition 6.3. *Let $x_0 \coloneqq (\min f, \max f)$. If $x_0 \notin \Delta$, then*

$$0 < \min_{x_1 \in D(f)\sqcup\Delta\setminus\{x_0\}} d(x_0, x_1) \leq \delta(f)^\star \leq \tfrac{A(f)}{2}.$$

Proof. Let $\tau > 0$ and consider the associated single-linkage partition $\mathcal{C}^\tau$ of $D(f)$. For any two distinct clusters $C, C' \in \mathcal{C}^\tau$, we have $d(C,C') \geq \tau$ by definition of the clustering, so $\delta(f)^\tau \geq \tau$. Since $A(f) > 0$, we have $x_0 \notin \Delta$. Because the persistence module associated to f is q-tame, x_0 is isolated: there is $\tau > 0$, such that $d(x_0, x_1) > \tau$ for any $x_1 \in D(f)$, $q \neq p$. Therefore, $\mathcal{C}^\tau$ contains at least two distinct clusters, namely $C_{x_0} = \{x_0\}$ and C_Δ. The upper-bound is straightforward: for any point $x \in D(f)$, $d(x,\Delta) \leq \frac{A(f)}{2}$, so that $\delta(f)^\tau \leq \frac{A(f)}{2}$ for any $\tau \geq 0$. □

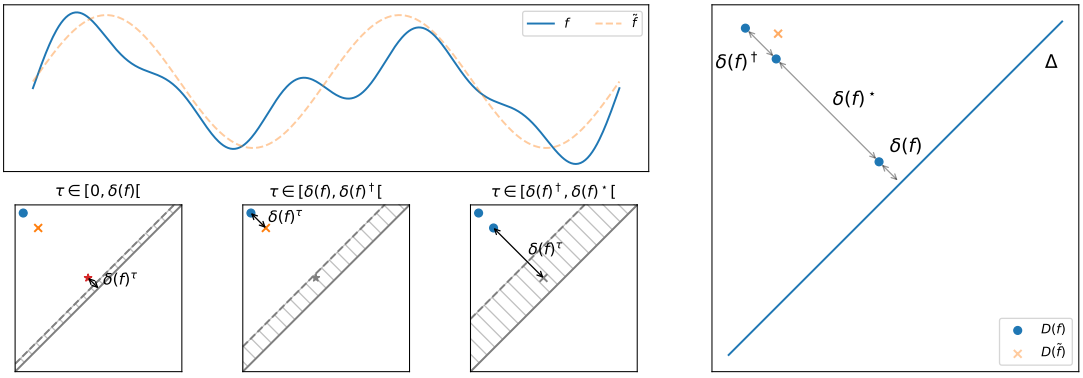


Figure 1: On the top left, the graphs of a one-periodic function f and a 0.5-periodic function $\tilde{f}$. Below, $D(f)$ clustered at different scales τ, with the constants $\delta(f)^\tau$ marked by the arrow. On the right, the persistence diagrams $D(f)$ and $D(\tilde{f})$, with the separation constants $\delta, \delta^\dagger$ and $\delta^\star$ marked for the former.

Let us now go back to the three-dimensional curve $\boldsymbol{\phi}$. We will establish a link between $(\delta(\boldsymbol{\phi}_v)^\star)_{v\in\mathbb{S}^2}$, the length of $\boldsymbol{\phi}$, denoted by $L(\boldsymbol{\phi})$, and $\bar{A}(\boldsymbol{\phi}) := \frac{1}{4\pi}\int_{\mathbb{S}^2} A(\boldsymbol{\phi}_v)d\mu$. We introduce a regularity parameter of $\boldsymbol{\phi}$,

$$C_{\boldsymbol{\phi}} := L(\boldsymbol{\phi})/2\bar{A}(\boldsymbol{\phi}) - 1.$$

Informally, $C_{\boldsymbol{\phi}}$ quantifies how "twisted" $\boldsymbol{\phi}$ is. Later, in (6.8), we will see a different interpretation of $C_{\boldsymbol{\phi}}$ which shows that it is at least 1.

Example 6.4. Let $\boldsymbol{\phi}$ be a circle of radius 1 contained in the x-y-plane. The length is $L(\boldsymbol{\phi}) = 2\pi$ and the amplitude along v is $A(\boldsymbol{\phi}_v) = 2\sin\theta_1$, for $v = (\sin\theta_1\cos\theta_2, \sin\theta_1\sin\theta_2, \cos\theta_1)$ which makes an angle of $\frac{\pi}{2} - \theta_1$ with the plane that contains $\boldsymbol{\phi}$. Then,

$$\bar{A}(\boldsymbol{\phi}) = \frac{1}{4\pi}\int_0^\pi \int_0^{2\pi} 2\sin\theta_1 \sin\theta_1 d\theta_2 d\theta_1 = \frac{2\pi}{4\pi}\int_0^\pi 2\sin^2(\theta_1)d\theta_1 = \frac{\pi}{2},$$

so that $C_{\boldsymbol{\phi}} = \frac{2\pi}{2\pi/2} - 1 = 1$.

We are now ready to state the main result of this section, which relates τ, $\mu_{\boldsymbol{\phi}}$ and $C_{\boldsymbol{\phi}}$. Recall that $\mu_{\boldsymbol{\phi}}$ is the probability measure on $\mathbb{S}^2$, with density with respect to μ proportional to $v \mapsto A(\boldsymbol{\phi}_v)$.

Proposition 6.5. *For any* $\xi \in [0, \frac{1}{2}[$, *let* $\tau_v \leq \min\left(\frac{1-2\xi}{2C_{\boldsymbol{\phi}}-(1+2\xi)}, 1\right)\frac{A(\boldsymbol{\phi}_v)}{6}$. *Then*

$$\mu_{\boldsymbol{\phi}}\left(\tau_v \leq \frac{\delta(\boldsymbol{\phi}_v)^\star}{3}\right) \geq \frac{1}{2} + \xi. \tag{6.3}$$

For $\boldsymbol{\phi}$ as in Example 6.4, for any ξ, $\frac{1-2\xi}{2C_{\boldsymbol{\phi}}-(1+2\xi)} = 1$, so for any $\tau_v \leq \frac{A(\boldsymbol{\phi}_v)}{6}$, we obtain that $\mu_{\boldsymbol{\phi}}\left(\tau_v \leq \frac{\delta(\boldsymbol{\phi}_v)^\star}{3}\right) = 1$. It confirms that for the circle, only the amplitude along v is needed to set τ_v.

An immediate corollary is that if we choose $\tau_v = \frac{A(\boldsymbol{\phi}_v)}{6(2C_{\boldsymbol{\phi}}-1)}$, then $\tau_v \leq \frac{\delta_v^\star}{3}$ for at least half of the directions. If we decrease the scale further, the lower–bound in (6.3) increases at the expense of robustness to noise. The implication for estimation of N is that we can set the scale τ_v to a constant fraction $\rho \in]0,1[$ of the amplitude $A_{\boldsymbol{\phi}_v}$. Unfortunately, $\delta(f)^\star$ is not the right separation constant, so no estimation guarantee can be provided. In addition, τ_v is chosen in terms of $C_{\boldsymbol{\phi}}$ which is unknown in practice.

Remark 6.6. In many practical cases of interest, the curve lacks large-scale self-symmetry. Then, $\delta(\boldsymbol{\phi}_v)^\star$ should be close to $\delta(\boldsymbol{\phi}_v)^\dagger$, which is related to the correctness of $\hat{N}_c$. We will see that the regularity constant is also not too large, typically, $C_{\boldsymbol{\phi}} \in [1., 1.5]$.

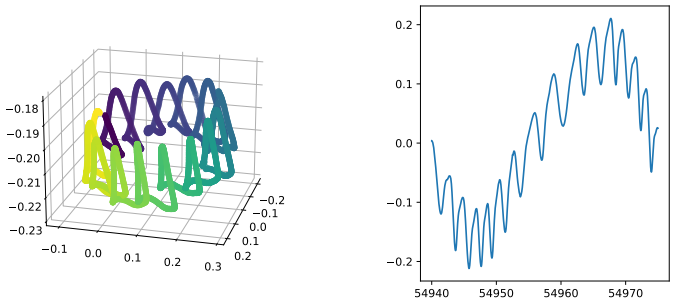


Figure 2: Magnetic signal recorded during a full, right-hand side turn. On the right, the 3-dimensional curve, with color denoting the time. On the right, the projection along the x-axis.

6.2 Minimizing noise via local to global stability: directions normal to heading

The algorithms described in Chapter 5 are tailored to reparametrised periodic functions and they are sensitive to noise in the supremum norm. We present two modelling assumptions which will help us decompose noise and reduce one of its components. For more details about the nature and the structure of magnetic measurements in a moving car, see Sections 7.1.1 and 7.1.2.

6.2.1 Noise decomposition

First, what we considered noise in (6.1), $\mathbf{W} : [0, T] \to \mathbb{R}^3$, can be decomposed into a heading component and the fingerprint of external phenomena. Specifically, by (7.1) and (7.2), we have

$$\mathbf{W}(t) = \boldsymbol{\phi}_{\boldsymbol{E}}(\gamma_h(t)) + \mathbf{W}_0(t) + \mathbf{B}_{HI}, \tag{6.4}$$

where $\boldsymbol{\phi}_{\boldsymbol{E}} : \mathbb{S}^1 \to \mathbb{R}^3$ is an ellipse, $\gamma_h : [0, T] \to \mathbb{S}^1$ represents the heading of the vehicle, $\mathbf{W}_0$ is residual noise and $\mathbf{B}_{HI} \in \mathbb{R}^3$ a constant. As the heading of the car changes, its orientation with respect to the Earths' magnetic field changes. Combined with the distortions of magnetic intensity in certain directions, we observe an ellipse, $\boldsymbol{\phi}_{\boldsymbol{E}}$. Residual noise coming from passing vehicles or nearby electric infrastructure has in general amplitude smaller than the diameter of $\boldsymbol{\phi}_{\boldsymbol{E}}$. Figure 1 shows an example of real data, where, the trajectory is almost constant on the left, and the car has a circular trajectory on the right.

Let us denote by v^1, v^2 unit vectors corresponding to principal axes of the ellipse $\boldsymbol{\phi}_{\boldsymbol{E}}$ and by $v^\star$ a unit vector normal to the direction of the plane containing $\boldsymbol{\phi}_{\boldsymbol{E}}$. That last is orthogonal to the changes in the heading: specifically, $t \mapsto \boldsymbol{\phi}_{\boldsymbol{E}}(\gamma_h(t))_{v^\star}$ is constant. Setting $\nu = \delta_{v^\star}$ removes a major component of the noise. Unfortunately, estimating $v^\star$ is not straightforward. In fact, it is the most difficult in the seemingly simplest scenario: when γ_h is constant or varies very little. On the other hand, in that case, $\mathbf{S}(t) = \boldsymbol{\phi}(\gamma(t)) + \mathbf{W}_0(t) + \text{const}$, so for any direction $v \in \mathbb{S}^2$, $\mathbf{S}_v$ looks like a reparametrized periodic function.

6.2.2 Minimal turning radius

A second insight from the model is that there exists $C > 1$, such that

$$C|\gamma_h(s) - \gamma_h(t)| < |\gamma(s) - \gamma(t)|. \tag{6.5}$$

The constant C is proportional to the minimal turning radius. It implies that whenever the heading changes, γ has changed by a greater amount. Another way of saying this is that the heading changes $t \mapsto \boldsymbol{\phi}_{\boldsymbol{E}}(\gamma_h(t))$ are in lower frequencies than $t \mapsto \boldsymbol{\phi}(\gamma(t))$. Figure 2 shows an example. In realistic scenarios, the heading variations are generally subtle and the larger ones are localized in time.

6.2.3 Identifying directions normal to heading changes

From Section 6.2.1, we conclude that it is important to choose directions $v \in \mathbb{S}^2$, which are orthogonal to the heading changes. Section 6.2.2 shows that heading changes and angular rotations

of the wheels occur at different time scales. It gives a heuristic to choose a direction $v \in \mathbb{S}^2$, by considering the signal at shorter and longer time scales. Indeed, if the signal at both time scales "look similar", it means that either the heading has not changed, or that v is orthogonal to those changes. Therefore, we propose to choose v for which $\mathbf{S}_v$ "looks most similar" at shorter and longer time-scales. We quantify that similarity through some descriptors of the signal.

As an example, consider Figure 5 where five shorter series are pictured by colored windows and the longer series is the whole signal. The descriptors, which we will define shortly are shown on in the bottom row: the lower-frequency signals shift the signal or change its amplitude slightly, making the descriptors less aligned.

A suitable descriptor will satisfy a property like Proposition 4.5. Examples of such descriptors are those from Chapter 4, but also simpler ones, like the amplitude of the signal. In fact, Proposition 4.5 is a consequence of convergence to a limit descriptor as the number of periods increases, and of stability with respect to additive noise, so any descriptor with similar properties will be suitable.

We now describe the construction, which is summarized in the pseudo-code in Algorithm 2. In what follows, the data is a time series $(\mathbf{S}_n)_{n=1}^N$, where $\mathbf{S}_n = \mathbf{S}(t_n)$ is a sample from the continuous model (6.1). Consider $F : \bigsqcup_{n\in\mathbb{N}} \mathbb{R}^n \to (\mathcal{H}, d)$, a descriptor of $(\mathbf{S}_n)_{n=1}^N$, where $(\mathcal{H}, d)$ is a metric space. We choose $N_1 < N \in \mathbb{N}$ and we compare the descriptor of the whole signal (of length N) with that computed on a window of length N_1. If F satisfies an analogue of Proposition 4.5, we expect $d(F((\langle \mathbf{S}_n, v\rangle)_{n=1}^{N_1}), F((\langle \mathbf{S}_n, v\rangle)_{n=1}^{N}))$ to be small exactly when $v \in \mathbb{S}^2$ is close to orthogonal to the heading-induced shifts. We generate a uniform grid on $\mathbb{S}^2$ and we choose v among the elements of the grid which minimize the difference between the two descriptors.

We will consider two kinds of descriptors. The topological descriptor is the persistence silhouette evaluated on a fixed grid,

$$F_{\bar\rho}((\mathbf{S}_1, \dots, \mathbf{S}_{N_1})_v) = \bar\rho_{k^s,0,2}(D((\mathbf{S}_1, \dots, \mathbf{S}_{N_1})_v)),$$

where $\bar\rho$ has been introduced in Chapter 4 and D denotes the persistence diagram of the sub level sets on an interval. In this case, $(\mathcal{H}, d) = (C(\mathbb{R}, \mathbb{R}), \|\cdot\|_\infty)$. The topological descriptor is sensitive to a wide range of changes in the underlying periodic signal. Since the noise is additive (6.4), a change in the heading shifts $\boldsymbol{\phi}$, and is already captured by the extremal values of the observation. Therefore, we also consider simpler descriptors, the difference and the mean of global extremal values,

$$\begin{aligned} F_A((\mathbf{S}_1, \dots, \mathbf{S}_{N_1})_v) =& A(\mathbf{S}_v) = \max_n (\mathbf{S}_n)_v - \min_n (\mathbf{S}_n)_v, \\ F_L((\mathbf{S}_1, \dots, \mathbf{S}_{N_1})_v) =& \frac{\max_n (\mathbf{S}_n)_v + \min_n (\mathbf{S}_n)_v}{2}, \end{aligned}$$

for which $\mathcal{H} = \mathbb{R}$. We also combine the two in $F_{AL} = (F_A, F_L)$, for which $\mathcal{H} = \mathbb{R}^2$ with the Euclidean norm. These descriptors share the invariance properties of topological descriptors (analogue to Propositions 2.35 and Theorem 4.2), even though they are less expressive. In addition, they do not depend on any parameters and are also faster to compute than the topological counterpart.

Remark 6.7. The limit of $F_{\bar\rho}$ as $p \to \infty$ is closely related to F_{AL}. For a continuous function f, we have $\bar\rho_{k^s,0,p}(D(f)) \xrightarrow{\|\cdot\|_\infty} \Lambda_{\min f, \max f}$ as $p \to \infty$. For two functions f and g,

$$\begin{aligned} \|\Lambda_{\min f,\max f} - \Lambda_{\min g,\max g}\|_\infty &= |A(f) - A(g)| + 2|\tfrac{\min f + \max f}{2} - \tfrac{\min g + \max g}{2}| \\ &= |F_A(f) - F_A(g)| + 2|F_L(f) - F_L(g)|. \end{aligned}$$

Before moving to estimation of N using $\hat v$, we illustrate the computation of $\hat v$ for different descriptors on an example. We randomly generate $\boldsymbol{\phi}$ and we show its image and the projections in Figure 3. We generate a random, 30-second reparametrisation γ with $R = 40$ full revolutions and random heading, γ_h a path from a Gaussian process. We let $\mathbf{S}$ as in (6.1) and (6.4), with $\boldsymbol{\phi}_{\boldsymbol{E}}$ randomly oriented, with axes of diameters 6 and 0.2 respectively. Q a random rotation matrix, $a = 3.$, $b = 0.1$ and $\mathbf{W}_0 = 0$ and Figure 4 shows the considered sample realization. Figure 5 shows the windows and descriptors thereof from Algorithm 2, for $F = \bar\rho$. The left and right columns correspond to projections along two different directions.

The left column in Figure 6 shows a representation of $v \mapsto d(\sup_{n=1,\dots,N_w}(d(F((\mathbf{X}_n)_v), F(\mathbf{S}_v))))$, for $F_{\bar\rho}$, F_A and F_{AL}. The shapes of the loss functions are roughly similar for the 3 descriptors:

Algorithm 2: Choosing a direction normal to heading changes

Data: An observation $(\mathbf{S}_n)_{n=1}^N$
Input: A descriptor F, integers N_1, N_w, M

```
/* Generate windows                                                    */
```
for $n \in \{1, \dots, N_w\}$ **do**
  Draw $k \sim \mathcal{U}(1, \dots, N - N_1 + 1)$;
  Set $\mathbf{X}_n = (\mathbf{S}_k, \dots, \mathbf{S}_{k+N_1-1})$;
end
```
/* Create a grid on the sphere                                         */
```
Generate $v_1, \dots, v_M$, a uniform grid on the sphere;
```
/* Compare the descriptors along different directions                  */
```
for $m \in \{1, \dots, M\}$ **do**
  $d_m = \sup_{n=1,\dots,N_w}(d(F((\mathbf{X}_n)_{v_m}), F((\mathbf{S})_{v_m})))$;
end
Set $m = \operatorname{argmin}_m d_m$ and $\hat{v} = v_m$;
return $\hat{v}$

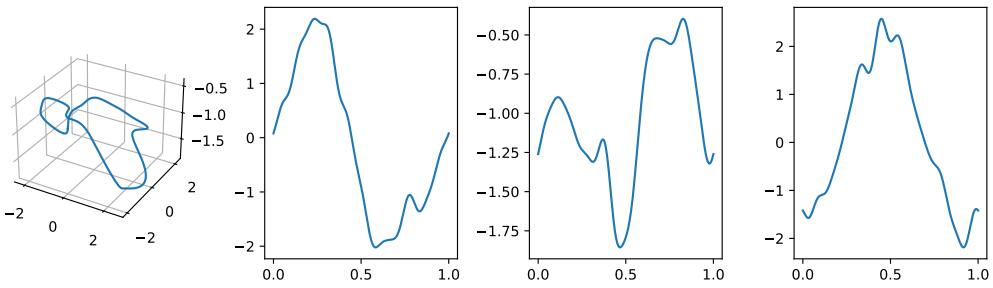


Figure 3: In the top left corner, a randomly-generated closed curve. The three other plots represent one-dimensional projections of $\boldsymbol{\phi}$, along the x-, y- and z-axis respectively.

there is a pronounced global maximum and a one-dimensional space where that distance is small. However, the global minimum for $F_{\tilde{\rho}}$ is far from those of F_A and F_{AL}, which happen to estimate $v^\star$ very well. In any case, all three minimizers of the respective criteria reduce the heading noise, as is shown on the projected signals in the middle column of Figure 6.

6.3 Estimation of $\hat{N}$ with direction selection

Recall that the problem consists of adapting an estimator of $\hat{N}$ to three-dimensional data. For ν a probability measure on $\mathbb{S}^2$ and an estimator $\hat{N}$, we will consider an estimator of the form (6.2). We now define several choices for ν which fall into 2 categories: maximizing the signal and minimizing the variability, based on results and heuristics from Sections 6.1 and 6.2 respectively. We make sure that ν can be computed from $\mathbf{S}$ and easily approximated.

Maximizing the signal

The function $v \mapsto A(\boldsymbol{\phi}_v)$ is not known, so the estimator $I(\mathbf{S}, \mu_{\boldsymbol{\phi}})$ motivated by Proposition 6.5 is not computable in practice. Using the most naive approach, we replace $A(\boldsymbol{\phi}_v)$ by $A(\mathbf{S}_v)$ and we define

$$\mu_{\mathbf{S}} \coloneqq (v \mapsto A(\mathbf{S}_v))\mu.$$

Unfortunately, the heading and the noise introduce a bias towards directions orthogonal to $v^\star$. In an attempt to reduce it, we consider local amplitudes instead,

$$\mu_{\mathbf{S},d} \coloneqq (v \mapsto \min_{n=1,\dots N_w} A((\mathbf{X}_n)_v))\mu,$$

where $(\mathbf{X}_n)_{n=1}^{N_w}$ is as in Algorithm 2.

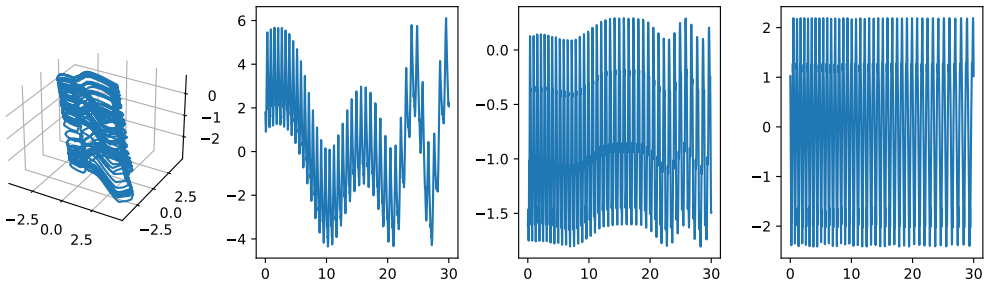


Figure 4: In the top left corner, the observation $\mathbf{S}$. The three other plots represent one-dimensional projections, $t \mapsto \mathbf{S}_v(t)$, for v^1, v^2 the principal directions of $\boldsymbol{\phi_E}$ and v^* the normal.

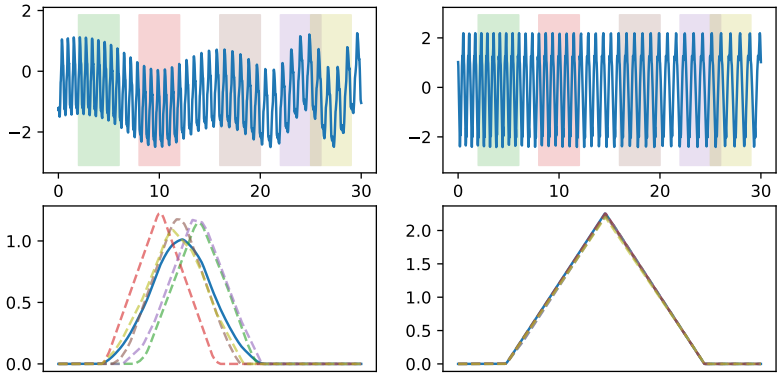


Figure 5: On the top row, the observation $\mathbf{S}$ projected along two different directions v_1, $v_2 \in \mathbb{S}^2$, along with 5 color-marked regions. The direction $v_1 \in \mathrm{Span}(v^1, v^2)$ is contained in the plane of the ellipse, while $v_2 = v^*$. The coloured time intervals delimit the windows $\mathbf{X}_1, \dots \mathbf{X}_5$ from Algorithm 2. The graphs in the bottom rows show the descriptors $F_{\bar{\rho}}((\mathbf{X}_n)_v)$ for $n = 1, \dots, 5$ and $F_{\bar{\rho}}(\mathbf{S}_v)$, in dashed and solid lines respectively.

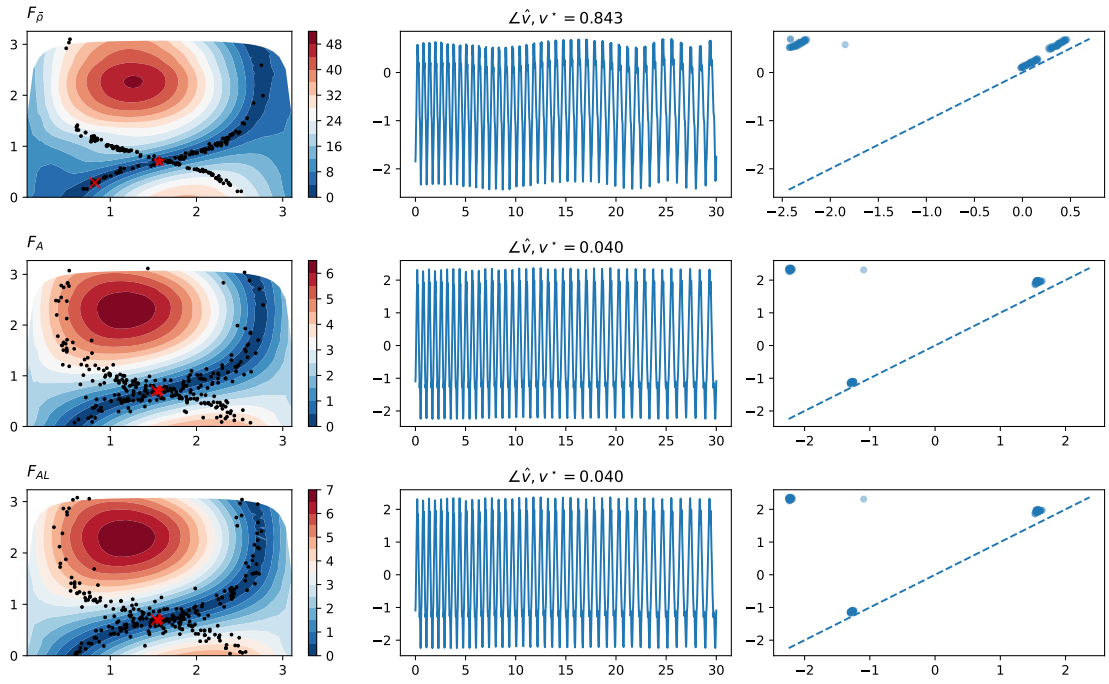


Figure 6: Each row corresponds to a different descriptor, $F_{\bar{\rho}}$, F_A and F_{AL} respectively. On the left, a visualization of $v \mapsto \sup_{n=1,\dots,N_w}(d(F((\mathbf{X}_n^1)_v), F(\mathbf{S}_v)))$, approximated in Algorithm 2, where $v \in \mathbb{S}^2$ is parametrized with two angles. Marked in red by a cross and star respectively are $\hat{v}$ and $v^\star$. The black points are a sample of $K = 300$ directions from (6.7). The central column shows $\mathbf{S}_{\hat{v}}$. On the right, the corresponding persistence diagram.

Minimizing the heading noise

For a descriptor F, Algorithm 2 outputs a single direction, $\hat{v}$, which is a good candidate to minimize the amount of noise in $\mathbf{S}_v$. We associate to it the measure

$$\mu_F = \delta_{\hat{v}}. \tag{6.6}$$

We will consider the descriptors F_A, F_L, F_{AL} as well as $F_{\bar{\rho}}$.

We also propose a hybrid approach. While the minimizers $\hat{v}$ achieved in the example from Figure 6 seem sensible, it is only an approximation. Not only is the cost function approximated, but the obtained minimum is also conditioned on the choice of windows $\mathbf{X}_1, \dots, \mathbf{X}_{N_1}$. Therefore, we propose a less degenerate measure: we put mass on directions which also exhibit stability across time-scales. For any descriptor F, we define $\mu_{F,A}$ as proportional to

$$\mu_{F,A} = (v \mapsto \exp(-c_1 \| d(F(\mathbf{S}_v) - F(\mathbf{X}_v))/F_A(\mathbf{X}_v) \|_{\ell^p}))\mu, \tag{6.7}$$

for some c_1, $p > 1$. In the experiments in the next section, we set $p = 10 = c_1$ and $F = F_{AL}$. We will denote the corresponding measure by $\mu_{AL,A}$.

Sampling from μ_F only requires executing Algorithm 2. The other measures are of the form $\nu = f\mu$, defined by the weight function f, in which case we proceed in two stages. We sample $V_1^0, \dots, V_M^0 \sim \mu$ on $\mathbb{S}^2$ and we approximate ν by the empirical measure

$$\hat{\nu} = \frac{1}{\sum_{m=1}^M f(V_m^0)} \sum_{m=1}^M f(V_m^0)\delta_{V_m^0}.$$

We justify such a procedure by continuity of f. Algorithm 3 summarizes this in pseudo-code.

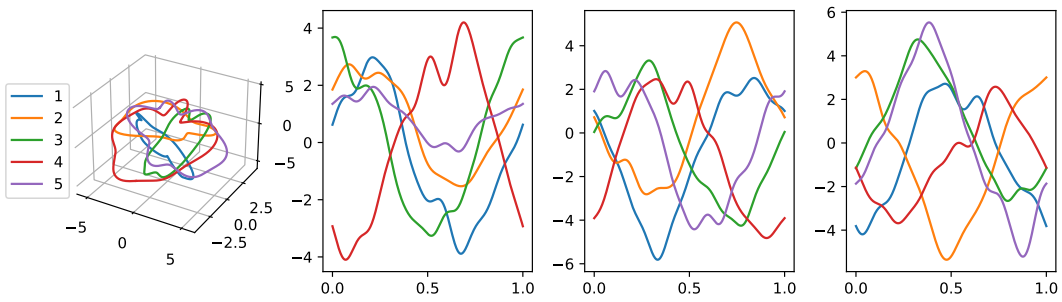


Figure 7: Illustration of 5, randomly generated curves, $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_5$ and their projections along the x-, y- and z-axis respectively.

Algorithm 3: Estimation along random directions $I(\mathbf{S}, \hat{\nu})$

Data: An observation $\mathbf{S}$
Input: A weight function f, a parameter $\rho > 0$ and integers K, M

`/* Draw from` $\mu_{\boldsymbol{\phi}}$ `*/`
Draw $V_1^0, \ldots, V_M^0 \sim \mu$ uniformly on $\mathbb{S}^2$;
Sample $m_1, \ldots, m_K \in \{1, \ldots, M\}$, where m has mass $\frac{f(V_m^0)}{\sum_{m=1}^M f(V_m^0)}$;
`/* Estimation along each direction */`
Initialize $Ns = [\,]$;
for $k \in \{1, \ldots, K\}$ **do**
  Set $y = \langle \mathbf{S}, V_{m_k}^0 \rangle$;
  Set $A = \max y - \min y$;
  Set $\hat{N} = \hat{N}_c(y, \rho A/2)$;
  if $\hat{N} \neq 1$ **then**
    Append $\hat{N}$ to Ns;
  end
end
return *mode*(Ns)

6.4 Numerical example: direction selection

We compare the two proposed algorithms on a synthetic dataset. Specifically, we consider the three proposed estimators, $\hat{N}$, $\hat{N}_c$ and $\hat{N}_c^T$, used in combination with the direction distributions from Section 6.3.

We now describe the dataset. We generate 5 curves $\mathbb{S}^1 \to \mathbb{R}^3$, shown in Figure 7, and 50 pairs (γ, γ_h) of random trajectories, with the number of periods in γ, $R \in [38, 41]$, not necessarily an integer. The heading $\gamma_h : [0, T] \to \mathbb{R}$ is a 0-mean Gaussian process with square exponential covariance with standard deviation 0.5 and time-scale 15. To sample $\boldsymbol{\phi}_{\boldsymbol{E}}$, we fix $a = 10$ and $b = 3$ and we generate a random rotation matrix, which fixes determines the ellipse. For each curve $\boldsymbol{\phi}$ and pair (γ, γ_h), we generate a sample according to (6.1), where $\mathbf{W}$ is sampled from a 0-mean Gaussian process with square exponential covariance with standard deviation 0.3 and time-scale $\tau = 0.5$.

Remark 6.8. Note that while we set the timescale in γ_h to a relatively large value, we do not enforce (6.5) explicitly.

As in Chapter 5, we use accuracy (proportion of samples for which N was estimated correctly) as the metric. Here, the number of periods R is not an integer, so we say that the estimation is correct if $|\hat{N} - R| < 1$. We also evaluate the choice of direction using two metrics that measure the (mean) heading variation and the (mean) angle between the chosen (or sampled) direction and $v^\star$. Specifically, since the data is synthetic, we know $\boldsymbol{\phi}_{\boldsymbol{E}}$ and γ_h, so we measure the heading variation with the amplitude $\nu A_h := \int_{\mathbb{S}^2} A(\boldsymbol{\phi}_{\boldsymbol{E}v} \circ \gamma_h) d\nu$. The angle between $v^\star$ and v is $\nu\angle := \int_{\mathbb{S}^2} \arccos((v^\star)^T v) d\nu$. For both quantities, the lower the better.

ν	$\hat{N}_c^T$	$\hat{N}$	$\hat{N}_c$
μ	0.856	0.756	0.796
$\hat{\mu}_{\mathbf{S}}$	0.812	0.740	0.764
$\hat{\mu}_{\mathbf{S},d}$	0.860	0.780	0.808
$\hat{\mu}_A$	0.984	0.808	0.816
$\hat{\mu}_{AL}$	0.980	0.812	0.824
$\hat{\mu}_L$	0.988	0.812	0.824
$\hat{\mu}_{\bar{\rho}}$	0.876	0.704	0.724
$\hat{\mu}_{AL,A}$	**0.992**	**0.964**	**0.988**

Table 1: Accuracy for estimation of N for different direction selection methods and N estimators.

ν	A_h	$\angle$
μ	9.000	0.999
$\hat{\mu}_{\mathbf{S}}$	10.004	1.053
$\hat{\mu}_{\mathbf{S},d}$	8.985	1.006
$\hat{\mu}_{AL}$	**0.527**	0.132
$\hat{\mu}_A$	0.536	0.127
$\hat{\mu}_L$	0.536	**0.121**
$\hat{\mu}_{\bar{\rho}}$	0.978	0.265
$\hat{\mu}_{AL,A}$	1.760	0.378

Table 2: Metrics for estimation of the direction.

For Algorithm 2, we set the window size N_1 such that it corresponds to 4 seconds and we sample $N_w = 20$ windows. We use $M = 40$ points for the optimization in $\mathbb{S}^2$. For the random directions, as described in Algorithm 3, we approximate the density with $M = 10000$ directions, from which we sample $K = 100$ elements. For estimation of N, we vary ρ depending on the estimator, with $\rho_{\hat{N}} = \frac{1}{3}$ and $\rho_{\hat{N}_c} = \frac{1}{6}$. The estimator $\hat{N}_c^T$ does not require a choice of the scale.

We discuss the experimental results for estimation of N, shown in Table 1. The accuracy of any estimator combined with $\hat{\mu}_{\mathbf{S}}$ is worse than uniform sampling μ. This is due to the bias in the estimation of the amplitude, as corroborated by $\hat{\mu}_{\mathbf{S}} A_h \geq \max(\mu A_h, \hat{\mu}_{\mathbf{S},d}) A_h$ in Table 2 and the slightly better performance obtained when using $\hat{\mu}_{\mathbf{S},d}$. That last does not seem to work particularly well and has performance between that of uniform sampling μ and that of best 'single-direction' methods. Consider now estimators computed on a single direction. The one based on the topological descriptor, $\hat{\mu}_{\bar{\rho}}$, has accuracy comparable to the uniform sampling μ. The metrics for the direction selection $\hat{\mu}_{\bar{\rho}} A_h$ and $\hat{\mu}_{\bar{\rho}} \angle$ are worse (higher) almost by a factor of 2 than for other criteria. The other 'single direction' methods $\hat{\mu}_A$, $\hat{\mu}_L, \hat{\mu}_{AL}$ have comparable metrics, although $\hat{\mu}_L$ seems to be slightly better, both on accuracy for all estimators, but also for the direction selection metrics. The direction selection method which leads to the best accuracy is by far $\hat{\mu}_{AL,A}$. It is the only method for which the accuracy figures for all three estimators $\hat{N}_c^T$, $\hat{N}$ and $\hat{N}_c$ are so close: for other methods, the accuracy of $\hat{N}_c^T$ exceeds that of $\hat{N}_c$, which is itself greater than $\hat{N}$. It is not the method which shows best metrics for estimation of the direction: sampled directions are not all close to the minimum, as illustrated in the lower-left graph in Figure 6. We explain the surprisingly good accuracy in estimation of N by the fact that while the chosen direction might be close to $v^\star$, that last might not be the best direction. Since R is not an integer, there are spurious points in the persistence diagram due to boundary conditions, sometimes most pronounced along a direction close to $v^\star$. Using several directions and the mode to aggregate the resulting estimations increases robustness to the boundary conditions.

6.5 Perspectives

We have considered the problem of applying estimators for Problem 5.1 from Chapter 5 to three-dimensional curves. Our approach relies on finding suitable directions along which to project the signal and apply one-dimensional estimators.

Based on our experiments, it is clear that choosing projections which minimize the noise is more

important than maximizing the amplitude. While the estimator (6.2) with $\mu_{AL,L}$ and any of $\hat{N}$, $\hat{N}_c$ or $\hat{N}_c^T$ performs the best, Algorithm 3 is computationnally expensive, as it requires calculating K persistence diagrams. Therefore, in case of computational issues, we also recommend to use any of μ_A, μ_L, μ_{AL} with $\hat{N}_c^T$.

We have not considered the odometric problem, Problem 5.2, explicitly. For the single-directions methods, once we have $\hat{v}$ from Algorithm 2, we apply the odometric method from Chapter 5 to $\mathbf{S}_{\hat{v}}$: estimate $\gamma(T) - \gamma(0)$, choose one of the odometric sequences and construct $\hat{\gamma}$ the odometric estimator of γ as in (5.2). When sampling, a straightforward adaptation of Algorithm 3 is to construct $\hat{\gamma}_v$ for all $v = v_m$ such that $\hat{N}(\mathbf{S}_{v_m}) = I(\mathbf{S}, \nu)$ and average those sequences.

We have shown a simple connection between $C_{\boldsymbol{\phi}}$ and $\delta^\star$, which we interpret as a relation between the geometry of $C_{\boldsymbol{\phi}}$ and the difficulty of the estimation of N. Formalized in Proposition 6.5, it gives a heuristic for setting the scale parameter for the estimators from Chapter 5.

We believe that further research could exploit the Euler Characteristic transform (ECT) (Curry et al., 2022), known to fully characterize certain classes of sets, to further such connections. For a class of curves, the stability of the ECT and the curvature have been linked in Marsh and Beers (2023). That last being also related to $C_{\boldsymbol{\phi}}$, the structure of the curve revealed in the ECT might be helpful to fully characterize the difficulty of the problem, not unlike Proposition 5.4 for the one-dimensional case.

Appendix

6.A Proof of Proposition 6.5

The proof relies on two elements. One, Proposition 6.10, gives a the lower–bound of the separation constant in terms of persistence and of the amplitude. The other, Proposition 6.9, makes a link between the integral of $\text{pers}(\boldsymbol{\phi}_v)$ with respect to μ and the length of the curve. For $\rho \in]0,1]$, we now define $\Omega_\rho \coloneqq \left\{ v \in \mathbb{S}^2 \mid \delta(\boldsymbol{\phi}_v)^* \geq \rho \frac{A(\boldsymbol{\phi}_v)}{2} \right\}$. Both $v \mapsto \delta(\boldsymbol{\phi}_v)^*$ and $v \mapsto A(\boldsymbol{\phi}_v)$ are continuous, so Ω_ρ is measurable.

Proposition 6.9. *For any $\rho \in [0,1[$, we have*

$$\mu_{\boldsymbol{\phi}}(\Omega_\rho) \geq \frac{1 - \rho C_{\boldsymbol{\phi}}}{1-\rho}.$$

Proof. We apply Lemma 6.10 pointwise for $v \in \mathbb{S}^2$,

$$\begin{aligned}\int_{\mathbb{S}^2} \text{pers}_1(\boldsymbol{\phi}_v) d\mathbb{S}^2 &\geq \int_{\mathbb{S}^2} \frac{A(\boldsymbol{\phi}_v)}{2} \left(\frac{A(\boldsymbol{\phi}_v)}{2\delta(\boldsymbol{\phi}_v)^*} + 1 \right) d\mathbb{S}^2 \\ &= \int_{\mathbb{S}^2 \setminus \Omega_\rho} \frac{A(\boldsymbol{\phi}_v)}{2} \left(\frac{A(\boldsymbol{\phi}_v)}{2\delta(\boldsymbol{\phi}_v)^*} + 1 \right) d\mathbb{S}^2 + \int_{\Omega_\rho} \frac{A(\boldsymbol{\phi}_v)}{2} \left(\frac{A(\boldsymbol{\phi}_v)}{2\delta(\boldsymbol{\phi}_v)^*} + 1 \right) d\mathbb{S}^2.\end{aligned}$$

We have $\frac{A(\boldsymbol{\phi}_v)}{2\delta(\boldsymbol{\phi}_v)^*} \geq \frac{1}{\rho}$ for $v \in \Omega_\rho^C$ and $\frac{A(\boldsymbol{\phi}_v)}{2\delta(\boldsymbol{\phi}_v)^*} \geq 1$ for any $v \in \mathbb{S}^2$, so

$$\begin{aligned}\int_{\mathbb{S}^2} \text{pers}_1(\boldsymbol{\phi}_v) d\mathbb{S}^2 &\geq \int_{\Omega_\rho^c} \frac{A(\boldsymbol{\phi}_v)}{2} \left(\frac{1}{\rho} + 1 \right) d\mathbb{S}^2 + 2 \int_{\Omega_\rho} \frac{A(\boldsymbol{\phi}_v)}{2} d\mathbb{S}^2 \\ &= \int_{\mathbb{S}^2} \frac{A(\boldsymbol{\phi}_v)}{2} \left(\frac{1}{\rho} + 1 \right) d\mathbb{S}^2 - \left(\frac{1}{\rho} + 1 \right) \int_{\Omega_\rho} \frac{A(\boldsymbol{\phi}_v)}{2} d\mathbb{S}^2 + 2 \int_{\Omega_\rho} \frac{A(\boldsymbol{\phi}_v)}{2} d\mathbb{S}^2.\end{aligned}$$

Dividing by $\int_{\mathbb{S}^2} \frac{A(\boldsymbol{\phi}_v)}{2} d\mathbb{S}^2 = 2\pi \bar{A}(\boldsymbol{\phi})$ and re-arranging the terms,

$$\begin{aligned}\left(\frac{1}{\rho} - 1 \right) \frac{1}{2\pi \bar{A}(\boldsymbol{\phi})} \int_{\Omega_\rho} \frac{A(\boldsymbol{\phi}_v)}{2} d\mathbb{S}^2 &\geq \left(1 + \frac{1}{\rho} \right) - \frac{1}{2\pi \bar{A}(\boldsymbol{\phi})} \int_{\mathbb{S}^2} \text{pers}_1(\boldsymbol{\phi}_v) d\mathbb{S}^2. \\ &= \frac{1}{\rho} - \left(\frac{1}{2\pi \bar{A}(\boldsymbol{\phi})} \int_{\mathbb{S}^2} \text{pers}_1(\boldsymbol{\phi}_v) d\mathbb{S}^2 - 1 \right).\end{aligned}$$

Multiplying by $\frac{\rho}{1-\rho}$ both sides, we have

$$\frac{1}{2\pi \bar{A}(\boldsymbol{\phi})} \int_{\Omega_\rho} \frac{A(\boldsymbol{\phi}_v)}{2} d\mathbb{S}^2 \geq \frac{1 - \rho C}{1 - \rho}, \qquad \text{where } C = \frac{1}{2\pi \bar{A}(\boldsymbol{\phi})} \int_{\mathbb{S}^2} \text{pers}_1(\boldsymbol{\phi}_v) d\mathbb{S}^2 - 1.$$

It remain to show that $C = C_{\boldsymbol{\phi}}$. From Cohen-Steiner and Edelsbrunner (2007, Length Inter-

pretation), we recall that

$$\begin{aligned}
L(\boldsymbol{\phi}) &= \frac{1}{\pi}\int_{\mathbb{S}^2}\int_{\mathbb{R}}\sum_{(b,d)\in D_0\cup D_1} 1_{b,d}(z)dzd\mu(v)\\
&= \frac{1}{\pi}\int_{\mathbb{S}^2}\int_{\mathbb{R}}\left(1_{[\min\boldsymbol{\phi}_v,\infty]}(z) - 1_{[\max\boldsymbol{\phi}_v,\infty]}(z) + \sum_{(b,d)\in D_0 \text{ finite}} 1_{b,d}(z)\right)dzd\mu(v)\\
&= \frac{1}{\pi}\int_{\mathbb{S}^2}\max\boldsymbol{\phi}_v - \min\boldsymbol{\phi}_v + \mathrm{pers}_{1,\text{finite}}(\boldsymbol{\phi}_v)d\mu(v)\\
&= \frac{1}{\pi}\int_{\mathbb{S}^2}\mathrm{pers}_1(\boldsymbol{\phi}_v)d\mu(v).
\end{aligned}$$

□

The proof give another interpretation to the constant $C_{\boldsymbol{\phi}}$, as the $\mu_{\boldsymbol{\phi}}$-mean "excess persistence", compared to the one resulting only from the amplitude,

$$\begin{aligned}
C_{\boldsymbol{\phi}} &= \frac{2\pi L(\boldsymbol{\phi})}{4\pi\bar{A}(\boldsymbol{\phi})} - 1\\
&= 2\pi L(\boldsymbol{\phi})/\left(\int_{\mathbb{S}^2} A(\boldsymbol{\phi}_v)d\mu\right) - 1\\
&= \int_{\mathbb{S}^2}\left(2\frac{\mathrm{pers}(\boldsymbol{\phi}_v)}{A(\boldsymbol{\phi}_v)} - 1\right)A(\boldsymbol{\phi}_v)d\mu \,/\left(\int_{\mathbb{S}^2} A(\boldsymbol{\phi}_v)d\mu\right)\\
&= \mathbb{E}_{\mu_{\boldsymbol{\phi}}}\left[2\frac{\mathrm{pers}(\boldsymbol{\phi}_v)}{A(\boldsymbol{\phi}_v)} - 1\right].
\end{aligned} \tag{6.8}$$

The remaining ingredient for Proposition 6.5 is not specific to the case of curves and we present it for functions on an interval.

Lemma 6.10. *For any continuous function $f : X \to \mathbb{R}$ on X a circle or a compact interval,*

$$\mathrm{pers}_1(f) \geq \frac{A(f)}{2}\left(\frac{A(f)}{2\delta(f)^*} + 1\right).$$

Proof. Let $K_0 \coloneqq \left\lfloor \frac{A(f)}{2\delta(f)^*} \right\rfloor$ and let $\delta_0 \coloneqq A(f) - 2K_0\delta(f)^*$. We let $K = \mathrm{card}(D(f)\cap\{(b,d) \mid d-b \geq \delta_0\})$ and we order $x_0,\dots,x_K \in D(f)\cap\{(b,d) \mid d-b\geq\delta_0\}$ by decreasing persistence. Then, since $\delta(f)^*$ designates the maximal separation between points,

$$\mathrm{pers}(x_k) \geq \mathrm{pers}((\min f, \max f) - k\delta(f)^*(1,-1)) = \frac{A(f)}{2} - 2k\delta(f)^* = \delta_0 + 2(K_0-k)\delta(f)^*.$$

So,

$$\begin{aligned}
\mathrm{pers}_1(f) &\geq \sum_{k=0}^{K_0}\mathrm{pers}(x_k)\\
&\geq \sum_{k=0}^{K_0}(\delta_0 + 2(K_0-k)\delta(f)^*)\\
&= (K_0+1)\delta_0 + 2\delta(f)^*\sum_{k=0}^{K_0} k\\
&= (K_0+1)(\delta_0 + K_0\delta(f)^*)\\
&= \left(\frac{A(f)}{2\delta(f)^*} + 1 - \frac{\delta_0}{2\delta(f)^*}\right)\left(\delta_0 + \frac{A(f)}{2} - \frac{\delta_0}{2}\right)\\
&= \left(\frac{A(f)}{2\delta(f)^*} + 1\right)\frac{A(f)}{2} + \frac{\delta_0}{2}\left(1 - \frac{\delta_0}{2\delta(f)^*}\right).
\end{aligned}$$

By definition, $0 \leq \delta_0 \leq 2\delta(f)^*$, so

$$\mathrm{pers}_1(f) \geq \left(\frac{A(f)}{2\delta(f)^*} + 1\right) \frac{A(f)}{2}.$$

□

Lemma 6.10 is tight for functions f for which the points in the diagram lie evenly spaced on the segment with endpoints $(\min f, \max f)$ and $((\max f + \min f)/2, (\max f + \min f)/2)$. A trivial example is $f = \sin(2\pi\cdot)$, for which $\delta(f) = 1 = \delta(f)^\star$. On the other hand, $f = \sin(4\pi\cdot)$ is degenerate, so the constants $\delta(f)$ and $\delta^\dagger(f)$ are not defined, while $\delta(f)^\star = 1$, and, naturally the bound in Lemma 6.10 is not tight. We conclude this section with a short proof of the main result.

Proof of Proposition 6.5. Let $\epsilon > 0$ and set $\rho = \min\left(\frac{1-2\xi}{2C_{\boldsymbol{\phi}}-(1+2\xi)}, 1-\epsilon\right)$. As $\rho < 1$, by Proposition 6.9,

$$\mu_{\boldsymbol{\phi}}(\tau_v \leq \delta(\boldsymbol{\phi}_v)^\star/3) = \mu_{\boldsymbol{\phi}}\left(\rho\frac{A(\boldsymbol{\phi}_v)}{2} \leq \delta(\boldsymbol{\phi}_v)^\star\right) \geq \frac{1-\rho C_{\boldsymbol{\phi}}}{1-\rho}.$$

Because $\rho \leq \frac{1-2\xi}{2C_{\boldsymbol{\phi}}-(1+2\xi)}$, we have $\frac{1-\rho C_{\boldsymbol{\phi}}}{1-\rho} \geq \frac{1}{2} + \xi$. □

Chapter 7

Application to vehicle positioning

In this chapter, we return to the magnetic positioning problem that motivates this thesis. Our experiments show that topological odometry is more accurate than the zero-crossings method, which we use as a benchmark. First, we introduce the industrial context. Then, we present the model for the magnetic field which compliments the discussions from Chapters 4 to 6. We formally state the problem, concisely re-introduce the methods and metrics in Section 7.2. In Section 7.3, we validate our approach on 'simple' scenarios and then apply our methods on more challenging examples in Section 7.4.

7.1 Industrial context: positioning systems and model for the magnetic signal

Car positioning systems provide an estimate of the vehicles' location in a certain frame of reference. The Global Navigation Satellite System (GNSS) outputs longitude and latitude coordinates. However, the satellite signal is not available in certain scenarios, for example, under ground, in buildings, and it is also easy to jam. More advanced and reliable systems are therefore comprised of multiple estimators, each with their own weaknesses, all aggregated for example via filtering (Haykin, 2001). If the sources of data are independent from each other, those can complement each other when one is not available. We now give a few examples of such estimators.

In Cruz and Aguiar (2020), it has been proposed to measure the magnetic signal in an area and construct a "magnetic map". Urban infrastructure, especially bridges, tunnels and high-voltage cables, produce disturbances of the magnetic field. Driving in that area, it is possible to locate the car by comparing the observed signal with the pre-recorded reference magnetic map. However, it requires the car to remain in the pre-defined area.

The GPS and the magnetic map are examples of positioning systems, where the reference is external and fixed. *Dead reckoning* is a different navigation principle, where the displacement from an initial location is estimated. This amounts to positioning the vehicle with respect to the inertial frame, fixed by the initial condition. For a car moving on a plane, it is enough to know the heading $t \mapsto \theta_h(t)$ and the forward motion $t \mapsto \theta(t)$ to reconstruct the displacement vector $t \mapsto (x(t), y(t))$ from that initial point.

A common example of dead reckoning is an Inertial Navigation System (INS). It relies on data from an Inertial Measurement Unit (IMU), equipped with accelerometers and gyroscopes - motion and rotation sensors respectively. Combining acceleration and rotation data, one can determine the position of the IMUs' frame in the initial, inertial frame of reference (Stovall, 1997). A major problem in INS is the bias in the motion sensors of the IMU: any bias results in a linear drift in the velocity estimation and a quadratic error in position estimation. This problem is mitigated in the most high-precision IMUs, used in military and industrial applications, but such IMUs are too bulky and expensive for consumer applications.

INS are for scenarios where the drift can be controlled. For example, in positioning of pedestrians with a foot-mounted IMU, the zero-velocity update technique consists in the velocity to zero at a specific moment in the gait, which can be detected (Ma et al., 2018). Thus, exploiting the structure of the human gait (Qigao et al., 2018) allows to regularly reduce the drift. While the same cannot be done for vehicles, an independent estimator of speed available at regularly-spaced

time intervals could play a similar role.

A source of data which could be used to construct such an estimator is the odometer, which records the angular position of the cars' wheels in time. The signal is processed and the computed speed displayed to the driver, but rarely directly accessible: at best, only a speed estimate can be read from the cars' computer. In any case, it requires retro fitting, and in particular, it cannot be assumed available in location tracking without consent.

As an alternative, Bristeau (2012) and Zabulon et al. (2019) propose estimators of the speed based on magnetic signal. As we will see in the next section, the magnetic signal recorded in a moving car contains a signature of the angular position of one of its wheels. In Tazaki et al. (2001), an odometer based on magnetic signals is proposed.

7.1.1 Model for the magnetic signal

Historically, the Earths' magnetic field has been used for heading estimation: in an unperturbed environment, the direction of the intensity vector points (roughly) towards the North (exact references are available (Alken et al., 2021)). The magnetic field measured by a *magnetometer* inside a moving car is the Earths' magnetic field $\mathbf{B}_E \in \mathbb{R}^3$ affected by three sources of perturbation, making heading estimation challenging, but magnetic odometry possible.

The first source of perturbation is the rotation of the wheels of a car. The magnetisation of a wheel comes from the steel components in the tire (Le Goff et al., 2012). As the wheel rotates, different parts of the tire are closer to the magnetometer, thus, creating a signature

$$\begin{array}{rccc} \mathbf{B}_u: & \mathbb{S}^1 & \to & \mathbb{R}^3 \\ & \theta & \mapsto & \mathbf{B}_u(\theta). \end{array}$$

Figure 1.2 shows an example of $\mathbf{B}_u$, from which we deduce that $\mathbf{B}_u$ might have a complex structure. Since the magnetization is the result of the production process and the history of the steel, it is individual to each tire and unknown, but it can be assumed invariant on the scale of a trajectory. Measured right next to the tire, the norm of the signatures' intensity is of similar magnitude to the Earths' magnetic field $\mathbf{B}_E \in \mathbb{R}^3$, so it is not negligible in general. The magnetic field intensity decreases cubically with the distance to the source, so if we place the magnetometer close to one of the wheels, the magnetic signatures from other wheels are not visible. Apart from the tire, wheels are usually made of steel or aluminium and only the former is magnetic. Since the wheel has no electric charge, no additional magnetic field is generated as a result of the angular velocity. Crucially, $\mathbf{B}_u$ depends on the angular position of the wheel, not its speed.

The second source of perturbation is non-controlled: it includes passing vehicles, nearby infrastructure (like high voltage cables), tunnels, bridges. These components were used in Cruz and Aguiar (2020) to create the magnetic map, but are a nuisance in a dead reckoning approach. It also includes elements in the car which do not depend on its dynamics, like windshield wipers, electric windows. Some of these components are negligible with respect to $\mathbf{B}_u$ and $\mathbf{B}_E$, but we will see examples of perturbations in a tunnel which are not. We will denote this component by $\mathbf{W}_0$.

The third source is the distortion of $\mathbf{B}_E$ by ferromagnetic materials in the car, described in Bristeau (2012). Typically, the frame and the body of the car are made of a variety of materials, including steel, aluminium and sometimes carbon. The ferromagnetic materials can be divided into soft- and hard-iron, each have a different effect on the measured signal. Similarly to the tire, hard irons can be magnetized in the manufacturing process. If the magnetometer is at rest with respect to the car, hard irons act by translation on the magnetic signal, by say $\mathbf{B}_{HI} \in \mathbb{R}^3$. Soft irons amplify the magnetic field intensity in certain directions. The action of both hard and soft iron depends on the location of the magnetometer inside the car, while its effect on $\mathbf{B}_E$ depends on its orientation.

Assume that the magnetometer is positioned and fixed in the car near one of the cars' wheels, and that the Earth's magnetic field is constant. A model for the magnetic signal which includes the above phenomena is given in Bristeau (2012),

$$\mathbf{B}_m = Q_{SI}(Q\mathbf{B}_E + \mathbf{B}_u + \mathbf{B}_{HI} + \mathbf{W}_0),$$

where $Q \in \mathrm{SO}(3)$ is a rotation matrix determined by the orientation of the car in the inertial frame and Q_{SI} is a real, symmetric matrix representing a homothety due to soft iron effects.

The orientation of the car is determined by the heading, roll and pitch angles. In this work, we will simplify this model. We will assume that the roll and pitch of the car are constant, that is,

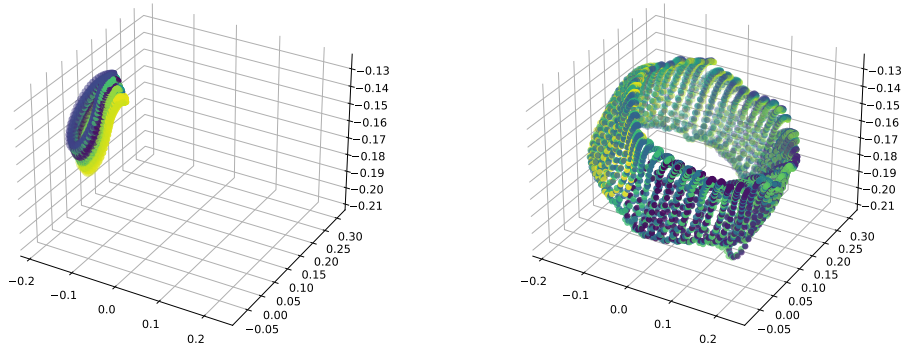


Figure 1: A scatter plot of magnetic measurements $\mathbf{S}$, recorded in a moving car. On the left, the trajectory of the car is almost straight (γ_h almost constant), so the ellipse is not visible. On the right, the car is moving in a circular trajectory and the ellipse is identifiable. The color indicates the time

there is no sideways' and 'forward-backwards' rotation, and we attribute any deviation from this assumption to the noise $\mathbf{W}_0$. This leaves us with two angles, θ, $\theta_h \in \mathbb{S}^1$, representing the heading of the vehicle and the angular position of one of the wheels,

$$\mathbf{B}_m(\theta, \theta_h) = Q_{SI}(Q(\theta_h)\mathbf{B}_E + \mathbf{B}_u(\theta) + \mathbf{B}_{HI} + \mathbf{W}_0). \tag{7.1}$$

The magnetic field intensity is thus a map from the torus $\mathbb{S}^1 \times \mathbb{S}^1$ to $\mathbb{R}^3$, where one of the circles is embedded as an ellipse and the other as $\mathbf{B}_u$.

Consider now the situation where the car is moving. If we denote by $\gamma_h, \gamma : [0, T] \to \mathbb{R}$ the heading of the car and the phase of the cars' wheel respectively, then the data recorded by the magnetometer is[1]

$$\begin{array}{rccl} \mathbf{S}(t): & [0,T] & \to & \mathbb{R}^3 \\ & t & \mapsto & \mathbf{B}_m(\gamma(t), \gamma_h(t)). \end{array} \tag{7.2}$$

An example is shown in Figure 1. For dead reckoning, we are interested in estimating γ, γ_h.

Le Goff et al. (2012), Bristeau (2012) analyse the magnetic signal and the latter remarks that heading variations are problematic for the detection of periodicity in (7.2). One solution to this problem consists of considering the gradient of the magnetic field instead (Tazaki et al., 2001). Specifically, for two magnetometers measuring magnetic field intensities $\mathbf{B}^1_m$, $\mathbf{B}^2_m$, we call the gradient $\mathbf{B}^2_m - \mathbf{B}^1_m$, and we use $(\mathbf{S}^2 - \mathbf{S}^1)(t)$ instead of (7.2). If the magnetometers are close (a few centimeters apart) and the axes of their reference frames are parallel, then $Q^2_{SI} - Q^1_{SI}$ and $\mathbf{B}^2_{HI} - \mathbf{B}^1_{HI}$ will be close to zero. On the other hand, since the magnetic field decreases cubically with the distance to the source, $\mathbf{B}^2_u - \mathbf{B}^1_u$ has a small but non-zero amplitude. Then,

$$\begin{aligned} (\mathbf{B}^2_m - \mathbf{B}^1_m)(\theta, \theta_h) =& (Q^2_{SI} - Q^1_{SI})(Q(\theta_h)\mathbf{B}_E + \mathbf{W}_0) + Q^2_{SI}\mathbf{B}^2_{HI} - Q^1_{SI}\mathbf{B}^1_{HI} \\ &+ Q^2_{SI}\mathbf{B}^2_u(\theta) - Q^1_{SI}\mathbf{B}^1_u(\theta) \\ &= Q^2_{SI}\mathbf{B}^2_u(\theta) - Q^1_{SI}\mathbf{B}^1_u(\theta) + \mathcal{O}(\|Q^2_{SI} - Q^1_{SI}\| + \|\mathbf{B}^1_{HI} - \mathbf{B}^2_{HI}\|) \\ &= Q^1_{SI}(\mathbf{B}^2_u - \mathbf{B}^1_u)(\theta) + \mathcal{O}(\|Q^2_{SI} - Q^1_{SI}\| + \|\mathbf{B}^1_{HI} - \mathbf{B}^2_{HI}\|). \end{aligned}$$

The disadvantage of using the gradient is the potentially more complicated structure of the embedding of $\mathbb{S}^1$ via $(\mathbf{B}^2_u - \mathbf{B}^1_u)$.

7.1.2 Dynamics of a vehicle

While $\mathbf{S}$ is a trajectory on an immersed torus, (γ, γ_h) represent the displacement and heading of the car, so we can make additional assumptions about its structure. First, we will assume that the car is only moving forward, that is, $t \mapsto \gamma(t)$ is non-decreasing. Reversing is a manoeuvre that often occurs at the beginning or the end of a typical trajectory, and can be usually detected (Bristeau, 2012).

[1] We see $\mathbf{B}_m$ as a function defined on $\mathbb{R} \times \mathbb{R}$ and 1-periodic in each of the variables.

When the trajectory is a straight line, $\mathbf{S}$ revolves around $\mathbf{B}_u$. When the vehicle is turning, unless the car is skidding, the variations of γ_h are upper-bounded by those of γ. For a given steering angle α, wheelbase L_W (distance between the front and rear axles), the turning radius is $\frac{W}{\sin(\alpha)}$. If we denote the track (distance between the left and right wheel) by L_T, then the distance travelled by the rear, inside wheel, is lower–bounded by the heading change as

$$\left(\frac{L_W}{\sin(\alpha)} - L_T\right) \frac{1}{r}|\gamma_h(s) - \gamma_h(t)| \leq |\gamma(s) - \gamma(t)|, \qquad \text{if } \gamma \text{ is monotone on } [s,t]. \tag{7.3}$$

In most modern cars, the maximum steering angle is $\alpha = 30°$. For the experimental vehicle that we used, $L_W \approx 2.6m$, $L_T \approx 1.4$ and $r \approx 0.33m$, so that factor is at least around 11.5. For example, during a right-hand-side U-turn, the right rear wheel will do 5 revolutions. Looking at the right plot in Figure 2, it is clear that the wheels induce a frequency higher than that of the heading changes.

The signal is sampled at $\omega = 125Hz$. The speed of a vehicle varies, from 0 to around $50m/s$, so for our experimental vehicle, the fundamental frequency of $\mathbf{B}_u$ ranges from 0 to $23Hz$. It is important to note that the signature itself might contain higher frequencies (as for example shown in Figure 1.2).

7.2 Magnetic odometry: data, methods and metrics

The magnetic measurement unit records measurements from five magnetometers $(\mathbf{S}^i)_{i=1,\dots,5}$ as well as inertial data. The aim is to estimate γ, but we consider a proxy: by constructing an *odometric sequence*. Recall that $(t_n)_{n=1}^N$ is called an odometric sequence for γ if $|\gamma(t_{n-1}) - \gamma(t_n) - 1|$ is small. One can still construct $\hat{\gamma}$ from the odometric sequence (for example as shown in (5.2)), as an estimator of γ.

We consider three methods which, given a one-dimensional time-series $S : [0,T] \to \mathbb{R}$, output such sequences. We treat those sequences as segmentations of $[0,T]$ into periods and we construct estimators of γ only to compute the metrics.

Zero-crossings

We say that t_n is a zero-crossing, if $S(t_{n-1})$ and $S(t_n)$ are of different signs. Let $(t_k)_{k=0}^K$ be the collection of zero-crossings in increasing order. We define $\mathrm{ZC}(S) = (t_{2n})_{n=0}^N$, for $N = \lfloor K/2 \rfloor$, since zero-crossings come in pairs.

Persistent local minima

We denote by $D(S)$ the persistence diagram of sublevel sets of S with periodic boundary conditions, as defined in Section 3.1. For a scale $\tau > 0$, we let $\mathrm{Hom}_0 = (t_n)_{n=0}^N$ be the local minima which are associated to points in $D(S) \setminus \Delta_\tau$, that is, more persistent than 2τ.

Remark 7.1. With respect to our considerations from Chapter 5, there is in general no reason for Hom_0 to be a good odometric sequence. However, we will see that it is when the period consists of a single prominent pair of extrema.

Clustering on persistence diagrams

We define the sequence Hom in terms of the estimator $\hat{N}_c$ of $\gamma(T) - \gamma(0)$, introduced in (5.13). Let $\mathcal{C}^\tau$ be the partition $D(S)$ obtained by single-linkage clustering at scale τ and denote by $\tilde{\mathcal{C}}^\tau = \{C \in \mathcal{C}^\tau \mid C \cap \Delta_\tau = \emptyset\}$ the clusters where all the points have persistence more than τ. Then, let $(t_k)_{k=0}^K$ be the local minima associated to points in $\bigcup_{C \in \tilde{\mathcal{C}}^\tau} C$. Because $\hat{N}_c$ is defined as a greatest common divisor of cardinalities, we have $m \coloneqq K/\hat{N}_c(S,\tau) \in \mathbb{N}$. We define $\mathrm{Hom} = (t_{mn})_{n=0}^{\hat{N}_c(S,\tau)}$ to be every m-th element from the sequence of all persistent local minima.

7.2.1 Selection of τ

The sequences Hom_0 and Hom depend on the parameter τ. We select it for each sample, using one of the two methods. First, recall that in the process of computing $\hat{N}_c^T(S)$ (5.13), we obtain an interval decomposition of $\tau \mapsto \hat{N}_c(S,\tau)$. Let $I = [a,b[$ be the longest interval for which

the multi-scale estimation is realized, that is $\hat{N}_c(S,\tau) = \hat{N}_c^T(S)$ for all $\tau \in I$. Then, we set $\tau_1 \coloneqq \tau_1(S) = a + (b-a)/4$. The second choice that we will use is $\tau_2 \coloneqq \tau_2(S) = A(S)/6$, where $A(S) = \max S - \min S$ is the amplitude. It is motivated by Proposition 6.5. More generally, $\tau \in [\frac{1}{6}, \frac{1}{3}]$ seems a reasonable choice.

7.2.2 Projection direction selection

We will consider four ways to define a one-dimensional signal S from the 15-dimensional $(\mathbf{S}^i)_{i=1,\dots,5}$. At the level of magnetometers, our first approach is to use the magnetic field intensity provided by one sensor, say $\mathbf{S}^1$, as described in Bristeau (2012). A second approach consists in analysing the gradient of the magnetic field instead (Tazaki et al., 2001). The authors use the axis of symmetry of the car, which corresponds to the difference $\mathbf{S}^2 - \mathbf{S}^1$ in our setup[2]. As for the direction $v \in \mathbb{S}^2$, we consider two choices. We project along the x-axis as a baseline, $v_1 = (1,0,0)$. We also consider the direction selection procedure described in Chapter 6, with $\nu = \mu_L = \delta_{\hat{v}}$ as defined in (6.6). Recall that this direction is selected based on three-dimensional data. This yields a total of four combinations, $\mathbf{S}^1_{v_1}$, $\mathbf{S}^1_{\hat{v}(\mathbf{S}^1)}$, $(\mathbf{S}^2 - \mathbf{S}^1)_{v_1}$, $(\mathbf{S}^2 - \mathbf{S}^1)_{\hat{v}(\mathbf{S}^2 - \mathbf{S}^1)}$.

7.2.3 Metrics

We introduce four metrics to evaluate the quality of our estimators. We assume that a reference $\gamma : [0,T] \to \mathbb{R}$ is available, with $\gamma(0) = 0$.

To measure the odometric quality of a sequence $(t_n)_{n=0}^N$, we introduce *False Positives* (FP) and *False Negatives* (FN), which count the number of intervals where the error in locating the endpoints is above $0 < \alpha < 1$

$$\begin{aligned} FP &= \operatorname{card}\{1 \le n \le N \mid \gamma(t_n) - \gamma(t_{n-1}) \le (1-\alpha)\}, \\ FN &= \operatorname{card}\{1 \le n \le N \mid \gamma(t_n) - \gamma(t_{n-1}) \ge (1+\alpha)\}. \end{aligned} \tag{7.4}$$

Unless specified otherwise, we will use $\alpha = 0.1$. We also define the cumulative error inside (E_I) and outside (E_O) of the odometric sequence

$$E_I = |\gamma(t_N) - \gamma(t_0) - N|, \qquad E_O = \gamma(t_0) - \gamma(0) + \gamma(T) - \gamma(t_N). \tag{7.5}$$

The first measures the displacement estimation error inside the odometric sequence, while the latter - the displacement on $[0,t_0]$ and $[t_n,T]$, which is undetected.

The metrics are complimentary. Indeed, while FN and FP measure the quality of the odometric sequence, they fail to capture large errors. For example, a single mislocalized element t_n, or missing 10 periods between t_{n-1} and t_n both contribute equally to FN only by 1. For this purpose, we introduce E_I, which is a metric on $\hat{\gamma}$. Our estimators are based on the principle of odometry and, if integrated in a navigation system, could be used as such. For example, if the odometric sequence was used to correct the drift in an INS, then FN and FP are more relevant. If however, $\hat{\gamma}$ was used as an estimator of travelled distance γ, alongside an INS, then E_I measures its performance better.

7.3 Magnetic odometry on straight lines

We validate our procedures on a collection of straight-line trajectories in a controlled environment. We fix the direction to $v = v_1$ and we evaluate the performance of odometry on the magnetic measures $\mathbf{S}^1_{v_1}$ and their gradients $(\mathbf{S}^2 - \mathbf{S}^1)_{v_1}$.

Data: experimental setup and preprocessing

The data consists of a two-hour recording, from which 20-minutes partitioned into 24 configurations are extracted. In each configuration the vehicle traverses a segment (A or B) in either direction (0 or 1) at a certain speed (10, 30, 50 km/h). The measurements are performed with the magnetometer in one of two positions: it is situated above the rear axle of the vehicle, either in the center or near

[2]The gradient along a different direction, differentiating between other magnetometers, say $\mathbf{S}^3 - \mathbf{S}^1$, produced signals of very similar form and amplitude, and in this setting, result in the same performance figures. Therefore, we restrict our gradient based considerations to $\mathbf{S}^2 - \mathbf{S}^1$.

the rear right wheel of the vehicle (CR or RR). The partitioning of the recording into segments is performed manually, based on the GPS position.

Remark 7.2. The experiments were performed in normal traffic conditions, so the experimental protocol could not be perfectly fulfilled. Specifically, the speed had to be adjusted to traffic conditions and the heading of the car is not exactly constant on each of the segments, due to a vehicle parked alongside on segment B. Finally, segment A contains a speed hump, which introduces pitch variations and the suspension movement causes changes in the distance between the magnetometer and the wheel, thus violating our assumptions from Section 7.1.1 in two ways.

To reduce the influence of these phenomena, we preprocess each series with a high-pass filter: we subtract from the signal its 5-second rolling median. This allows us to remove part of the heading noise without sacrificing the ability to work with low speed scenarios[3].

Results and discussion

Table 1 shows summary statistics of the performance of the different methods. First, we note that the topological methods (Hom and Hom_0) have in general lower FP and FN than ZC. This is a consequence of local extrema reflecting the angular position more precisely than the zero-crossings, which are susceptible to low-frequency shifts. This is clearly visible in Figure 2, where the slope of $\hat{\gamma} - \gamma$ varies more for ZC than for the topological estimators. The few exceptions to FP and FN being worse for topological methods occur for Hom_0 in the CR position. In these particular cases, the scale τ seems to be underestimated, leading to local minima due to noise not being removed. Here, we can see the particularity of Hom. In contrast to Hom_0, points of small persistence are not removed prior to clustering, so a segmentation is produced only when there is a big enough gap in the persistence diagram between the noise and the prominent local extrema. It makes Hom more fragile to noise, as it does not always provide a sequence, as clearly shown by the high values of E_O. When a sequence is provided, it is more accurate: FP, FN and E_I are all lower for Hom than Hom_0.

Using the gradient $(\mathbf{S}^2 - \mathbf{S}^1)_v$ instead of the series $\mathbf{S}^1_v$ boosts the performance of all estimators in all settings. The amplitude of the series is much smaller, but part of the influence of the Earths' magnetic field has been removed. We observe that it is particularly beneficial for Hom, which is sensitive to lower-frequency variations[4]. On the other hand, the differentiated signal contains small local extrema, visible for example in the left graph of Figure 3. While such extrema do not impact the topological methods provided that τ is chosen sufficiently large, they induce spurious elements in the ZC sequence. The presence of these extrema is more frequent in the RR position, leading to a FP higher for ZC in that setting. The fact that FP for ZC is worse in the CR setting is due only to a particularly irregular series, where the periodicity seems to change.

7.4 Magnetic odometry: complex trajectories

We apply our method on recordings from three more challenging scenarios: on a rough road, in a tunnel and on circular trajectories. The last is known as particularly challenging for frequency-decomposition based methods, but we also consider the first two since the methods we study are of a different nature.

In contrast to the previous section, an accurate reference $\hat{\gamma}$ is not generally not available. On the one hand, it is due to losses of GPS measurements, sometimes on periods too long for the reference to be considered reliable. On the other, it is a result of a lack of synchronization of the magnetic measurements with the GPS-based position estimate or readings from an odometer.

7.4.1 Rough road

The dataset consists of measurements performed on a rough, country-side road. Although paved, the 2.5 kilometre long road is dense in patches and holes. It also features a gentle 45 degree turn around the middle. The data consists of three recordings of a total of thirty minutes during which

[3] A 5-second rotation period corresponds to a velocity of less than 2 km/h, so we can still detect velocities above 5 km/h.

[4] We point to the forth row in Figure 3.3 for an illustration of the effect of low frequency noise on the persistence diagram.

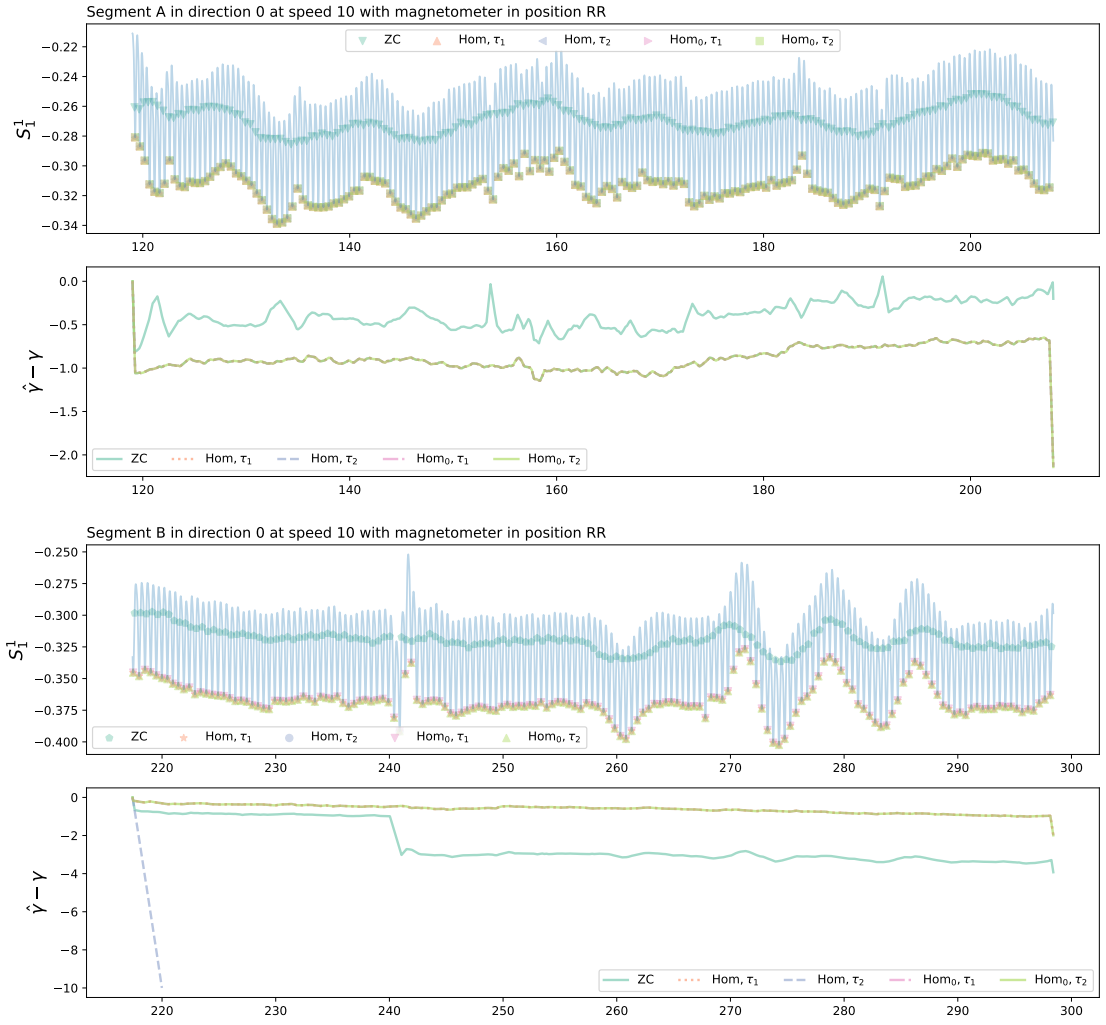


Figure 2: Examples of two recordings from the straight lines dataset with the segmentations provided by the different methods, as well as the resulting estimators $\hat{\gamma}$. The sequences from topological methods all coincide, except Hom with τ_2 in the second signal.

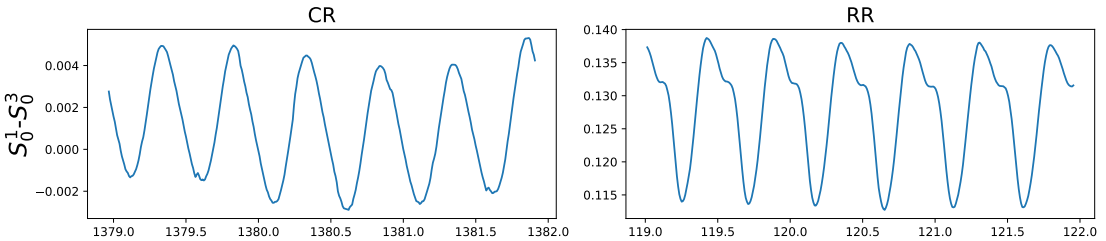


Figure 3: Illustration of the gradient $\mathbf{S}^2 - \mathbf{S}^1$, for the magnetic measurement unit in two positions, "CR" and "RR".

Series	Position	Method		FP	FN	E_O	E_I	nb_detected
$\mathbf{S}^1_{v_1}$	CR	Hom	τ_1	2	1	1254.27	87.07	647
		Hom	τ_2	1	1	1572.72	86.62	333
		Hom_0	τ_1	5	130	24.27	339.06	1629
		Hom_0	τ_2	2041	2	14.27	1888.83	3862
		ZC		299	132	22.81	241.78	1947
	RR	Hom	τ_1	0	1	15.75	3.02	1990
		Hom	τ_2	0	2	749.17	78.60	1181
		Hom_0	τ_1	0	1	15.75	3.02	1990
		Hom_0	τ_2	0	1	15.75	3.02	1990
		ZC		12	15	9.91	6.35	1993
$(\mathbf{S}^2 - \mathbf{S}^1)_{v_1}$	CR	Hom	τ_1	2	2	16.46	4.00	1973
		Hom	τ_2	2	2	356.52	2.85	1633
		Hom_0	τ_1	2	2	16.36	4.00	1973
		Hom_0	τ_2	2	2	16.36	4.00	1973
		ZC		103	25	9.47	48.49	2028
	RR	Hom	τ_1	0	1	16.66	3.01	1989
		Hom	τ_2	0	1	16.66	3.01	1989
		Hom_0	τ_1	0	1	16.66	3.01	1989
		Hom_0	τ_2	0	1	16.66	3.01	1989
		ZC		35	5	5.62	16.51	2014

Table 1: Metrics of different methods for sequences computed on the straight line dataset. For all methods, lower is better.

the road was traversed at different speeds and directions. Such a setting allows us to determine the sensitivity of the methods to the working of the cars suspension and to sensitivity of the heading.

We present the results in Table 2. We exclude the FP and FN metrics, because of a lack of reliable reference. We use E_O and E_I, since the speed of the vehicle was kept approximately constant and we believe the misalignment to be only on the order of a few seconds, which does not severely affect the global metrics.

Among the topological methods, Hom_0 with τ_1 seems to best identify the number of wheel revolutions. The more sensitive method Hom fails due to outliers[5] and some discontinuities in the time series. The ZC method shows poor performance in two scenarios. Aside from the gradient-induced local extrema described previously, spurious zero-crossings are detected when the vehicle is at rest or driving at very low speeds. We note that $v^\star(\mathbf{S}^1)$ seems to point to a direction normal to the heading, as shown on Figure 4. Combining the gradient $\mathbf{S}^2 - \mathbf{S}^1$ with $v^\star$ does not seem better than v_1 geometrically, although it improves the metrics.

7.4.2 Tunnel

As discussed in Section 7.1.1, power-lines, cables and other electric infrastructure generate a magnetic field, and metallic structures further alter it. Tunnels are thus a very challenging environment with many sources of noise. We obtain two, 6-minute recordings of a car driven on a high-speed road and through a tunnel. Unfortunately, no reliable reference is available for those recordings. In the following, we only inspect subsets of the series and the resulting segmentations.

The recurrent character of the signal is clearly identifiable when the vehicle is on the high-speed road. Inside the tunnel, the series is dominated by the noise. All methods fail completely in analysing $\mathbf{S}^1$, along any direction $v \in \mathbb{S}^2$. In the differentiated signal $\mathbf{S}^2 - \mathbf{S}^1$, the recurrent structure is identifiable, although some high-frequency noise persist.

We focus on $S = (\mathbf{S}^2 - \mathbf{S}^1)_{\hat{v}(\mathbf{S}^2 - \mathbf{S}^1)}$, an excerpt of which we show in Figure 5. The odometric sequence obtained on with Hom_0 with τ_1 on the first recording seems generally correct. However, τ_1 chosen on the second recording is too large, leading to an odometric sequence that is too sparse.

[5] Due to their short duration (only 2-4 consecutive values) and very high values, we believe these are reading errors.

Series	v	method		E_O	E_I	nb_detected
$\mathbf{S}^2 - \mathbf{S}^1$	v_1	Hom	τ_1	1318.08	1.23	2702
		Hom_0	τ_1	10.04	2.27	4009
		ZC		0.90	363.58	4384
	$\hat{v}$	Hom	τ_1	12.50	3.81	4005
		Hom_0	τ_1	11.83	1.73	4010
		ZC		0.65	259.54	4279
$\mathbf{S}^1$	v_1	Hom	τ_1	17.04	2.98	4005
		Hom_0	τ_1	17.04	2.98	4005
		ZC		2.83	55.90	4072
	$\hat{v}$	Hom_0	τ_1	2707.67	0.65	1312
		Hom	τ_1	11.71	5.98	4007
		ZC		1.12	110.81	4131

Table 2: Metrics for three methods on the rough road.

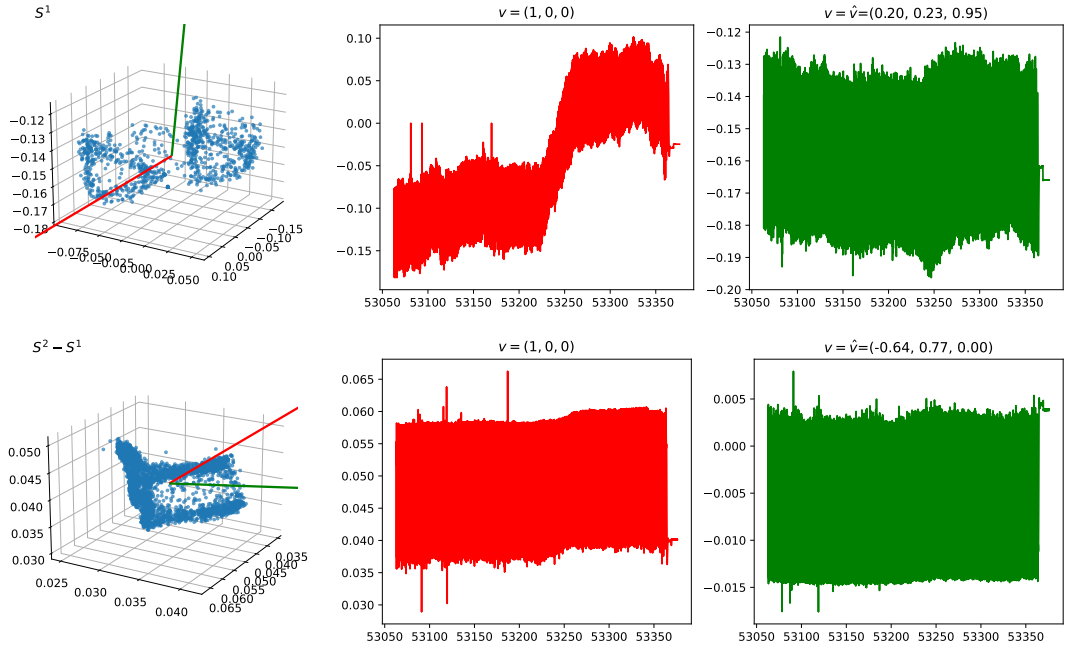


Figure 4: The magnetic signal recorded when traversing the rough road segment. In the top row, on the left, a subsample of $\mathbf{S}^1$, with the two directions, v_1 and $\hat{v}$, along which the signal is projected. In the bottom row, analogue figures for the differentiated series $\mathbf{S}^2 - \mathbf{S}^1$.

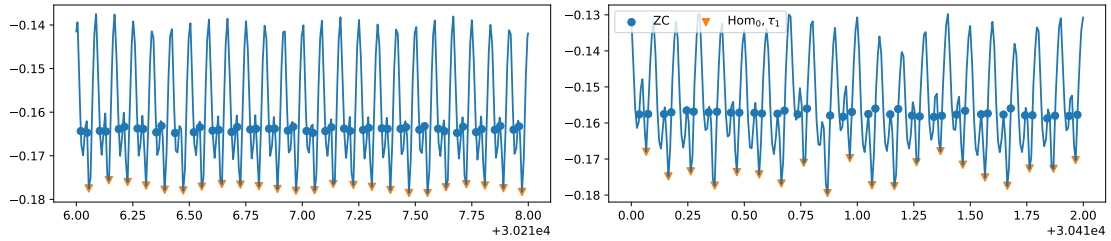


Figure 5: A projection of $\mathbf{S}^2 - \mathbf{S}^1$ along $\hat{v}$ outside and inside the tunnel. The markers show the detected sequences.

Series	v	method		FP	FN	E_O	E_I	nb_detected
$\mathbf{S}^2 - \mathbf{S}^1$	v_1	Hom	τ_1	410	334	13.92	10.71	1922
		Hom$_0$	τ_1	412	335	12.73	9.94	1924
		ZC		561	312	4.96	107.54	2036
	$\hat{v}$	Hom	τ_1	411	331	15.56	10.85	1921
		Hom$_0$	τ_1	413	331	15.56	9.85	1922
		ZC		485	320	4.73	68.73	1992
$\mathbf{S}^1$	v_1	Hom	τ_1	0	34	606.52	1284.90	43
		Hom$_0$	τ_1	316	372	32.23	235.52	1667
		ZC		498	333	7.96	46.92	1965
	$\hat{v}$	Hom	τ_1	334	276	109.75	211.42	1624
		Hom$_0$	τ_1	410	332	14.58	10.29	1921
		ZC		485	331	6.13	54.00	1974

Table 3: Metrics for three methods on the turns dataset.

Exactly the opposite is true of τ_2, where τ_2 is too small on the first recording and suitable for the second. We do not have enough data to be confident about the performance of our approach in tunnels.

7.4.3 Circular trajectories

The "turns" dataset is composed of 9 recordings of a total of over 16 minutes, during which almost 4km are travelled. A total of five trajectories are recorded on round-abouts: two and three recordings on roundabouts of diameters 35m and 70m respectively. The remaining four trajectories are tight turns on a parking lot, executed in both directions. The fundamental frequencies of the heading changes and the rotations of the wheels are separated by factors of around 100 and 10 in the "big roundabout" and "parking lot" settings respectively, the latter being much harder from a spectral point of view.

In order to determine the relevance of differentiating the series and of using the direction $\hat{v}$, we do not apply the high-pass filter as a preprocessing step for the topological methods. We still apply it prior to ZC in order to center the signal.

Table 3 shows the metrics for odometric sequences obtained using the different methods. When we examine E_O and E_I, the best results for Hom and Hom$_0$ are obtained for the differentiated signal, where the choice of direction does not matter. Interestingly, ZC performs better on non-differentiated series, rivaling the topological sequences. Examining the zero-crossings reveals a few spurious detections of the local minima, as shown in Figure 3. Some zero-crossings are also detected at very low speeds of the vehicle. While the sequences "look" correct, the FP and FN metrics do not reflect it. We believe that this is due to the difficulty of constructing a precise and well-synchronized ground-truth for γ in this setting.

Remark 7.3. Constructing a reference for trajectories with tight turns and abrupt heading changes is more delicate. The speed measurements provided by the GPS are the norm of the velocity vector of the antenna, which is usually positioned on the axis of symmetry of the car $y = 0$.

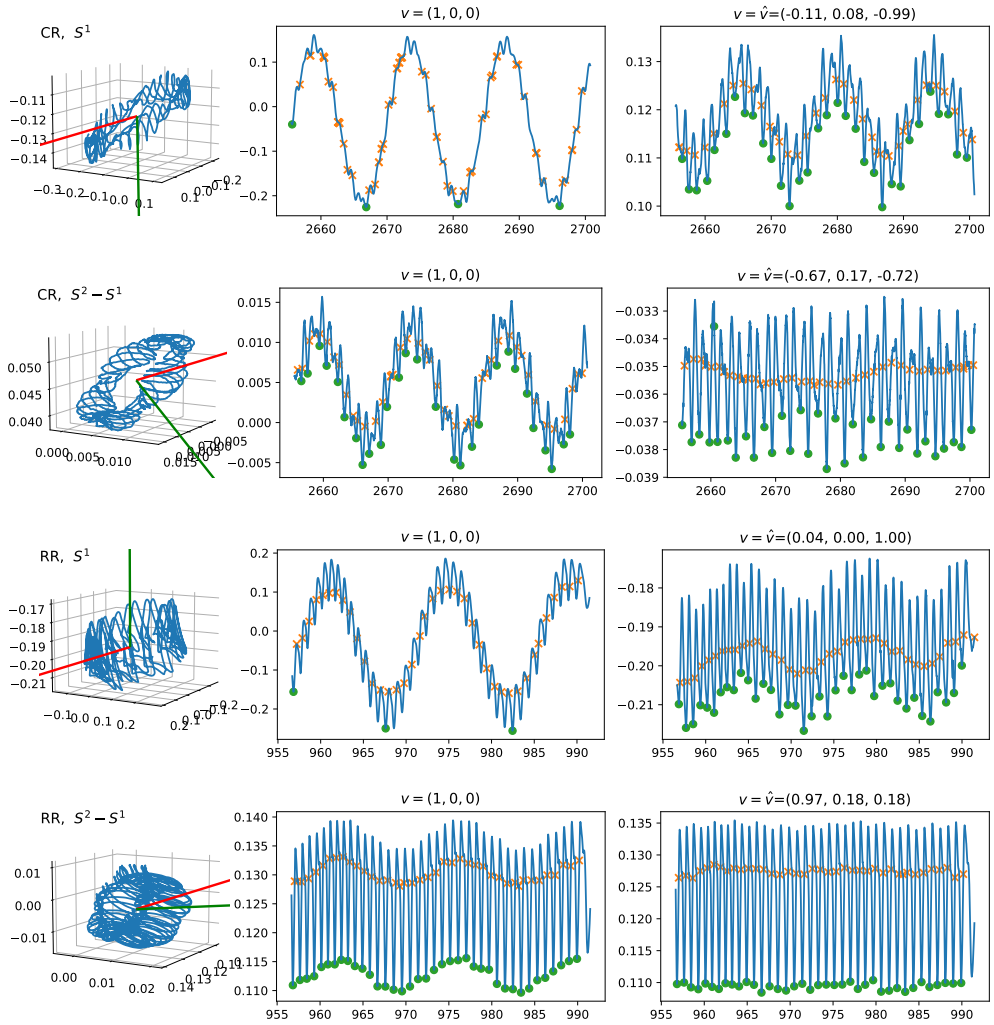


Figure 6: Magnetic signals recorded a left-hand side, circular trajectory. The top two rows show the magnetic field and its gradient, recorded with the magnetometer in position "CR", in the center of the vehicle. The bottom two rows show analogous data, except that the magnetometer was on the right of the vehicle 'RR'.

During a turn, the inside (resp. outside) wheel travels a smaller (resp. greater) distance than the reference constructed by integrating the GPS. For tight turns, this difference becomes non-negligible. Mathematically, creating a suitable ground-truth requires calculating the velocity of the wheel, what can be done using angular velocities measured by the IMU and the relative position of the wheel with respect to the GPS antenna.

We believe it is also interesting to qualitatively examine the impact of the signal-to-noise ratio. Specifically, we have two signals recorded through comparable trajectories of tight turns on a parking lot. The magnetic measurement unit was positioned in the settings "CR" and "RR" respectively. Figure 6 shows four signals, $\mathbf{S}^2 - \mathbf{S}^1$ and $\mathbf{S}^1$ for each of the configurations, and the odometric sequences ZC and Hom_0 with τ_1. In the end, for both "CR" and "RR", the combination of not differentiating and projecting along v_1 results in erroneous sequence from Hom_0. The rightmost configuration illustrates a problem with using Hom_0: the scale τ_1 has been chosen reasonably, but a single period features two prominent pairs of extrema, which are counted as two realizations of the same pair by Hom_0. The ZC method seems to perform reasonably well, what indicates that influence of the heading can be eliminated for this kind of method and can be well-isolated using either the gradient or by choosing a projection appropriately.

7.5 Discussion and perspectives

Our results confirm that magnetic odometry has potential. In straight-line trajectories, rotations of the wheels are well-identified and the resulting sequences provided by Hom and Hom_0 are accurate. For circular trajectories, the methods are successful provided that the influence of the heading noise is reduced, by a low-pass filter, by choosing the projection appropriately or analysing the gradient of the magnetic field instead of the field itself. That last also reduces the high-frequency witnessed in a tunnel, but at the expense of complicating the structure of ϕ.

In the perspective of integration in a positioning system, the topological methods present an interesting trade-off. The odometric sequences produced by Hom are precise (low FN and FP), but are not always available, especially in challenging environments. On the other hand, Hom_0 almost always provides a sequence, but its precision is more questionable: it is sensitive to the choice of the scale τ and does not always capture the right modes.

We find that using the gradient of the magnetic field is preferable to using the magnetic field intensity. We also confirm the observations from Tazaki et al. (2001) that, with the gradient, the choice of the projection is not particularly important. On the other hand, if measurements from a single magnetometer are available, it is crucial to select the direction appropriately and $\hat{v}$ is often a sensible way to reduce the heading noise. The low-pass filtering is a reasonable addition, to which we found little disadvantage.

More data should be acquired to calculate estimate the errors in the odometric sequences produced by the proposed methods. A first step would be to construct a ground-truth from an odometer measuring the angular position of the same wheel as the magnetometer, and well-synchronized with that last. Second, different vehicles and tires should be used. Our tests were conducted using a single vehicle, and over a relatively short time. We should expect that using newer tires or other a different vehicle will reveal other magnetic signatures, what might be more challenging to some methods than others.

We are interested whether magnetic odometric sequences constructed independently on the left and right-hand sides could be used to determine the heading of the vehicle.

Finally, it should be noted that other approaches could be considered, two of which we mention. First, the spectral method from Bristeau et al. (2010) might show better performance on well-selected directions on $(\mathbf{S}^2 - \mathbf{S}^1)_{\hat{v}}$. Second, a geometric approach based on estimating the support of $\mathbf{B}_u$ could be developed. Such an approach would probably be more robust in low-noise scenarios.

Bibliography

Aamari, E., Aaron, C., and Levrard, C. (2023). Minimax Boundary Estimation and Estimation with Boundary. arXiv, arXiv:2108.03135.

Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F., and Ziegelmeier, L. (2017). Persistence Images: A Stable Vector Representation of Persistent Homology. *The Journal of Machine Learning Research*, 18(1):218–252.

Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Infinite Dimensional Analysis. Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg, 3rd ed edition.

Alken, P., Thébault, E., Beggan, C. D., Amit, H., Aubert, J., Baerenzung, J., Bondar, T. N., Brown, W. J., Califf, S., Chambodut, A., Chulliat, A., Cox, G. A., Finlay, C. C., Fournier, A., Gillet, N., Grayver, A., Hammer, M. D., Holschneider, M., Huder, L., Hulot, G., Jager, T., Kloss, C., Korte, M., Kuang, W., Kuvshinov, A., Langlais, B., Léger, J.-M., Lesur, V., Livermore, P. W., Lowes, F. J., Macmillan, S., Magnes, W., Mandea, M., Marsal, S., Matzka, J., Metman, M. C., Minami, T., Morschhauser, A., Mound, J. E., Nair, M., Nakano, S., Olsen, N., Pavón-Carrasco, F. J., Petrov, V. G., Ropp, G., Rother, M., Sabaka, T. J., Sanchez, S., Saturnino, D., Schnepf, N. R., Shen, X., Stolle, C., Tangborn, A., Tøffner-Clausen, L., Toh, H., Torta, J. M., Varner, J., Vervelidou, F., Vigneron, P., Wardinski, I., Wicht, J., Woods, A., Yang, Y., Zeren, Z., and Zhou, B. (2021). International Geomagnetic Reference Field: The thirteenth generation. *Earth, Planets and Space*, 73(1):49.

Azäis, J.-M. and Wschebor, M. (2009). *Level Sets and Extrema of Random Processes and Fields: Azaïs/Level Sets and Extrema of Random Processes and Fields*. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Berry, E., Chen, Y.-C., Cisewski-Kehe, J., and Fasy, B. T. (2020). Functional summaries of persistence diagrams. *Journal of Applied and Computational Topology*, 4(2):211–262.

Biswas, R., Cultrera di Montesano, S., Edelsbrunner, H., and Saghafian, M. (2023). Geometric characterization of the persistence of 1D maps. *Journal of Applied and Computational Topology*.

Bleile, B., Garin, A., Heiss, T., Maggs, K., and Robins, V. (2021). The Persistent Homology of Dual Digital Image Constructions. In *Research in Computational Topology 2*, pages 1–26. Springer International Publishing, Cham.

Boashash, B. (2015). *Time-Frequency Signal Analysis and Processing : A Comprehensive Reference*. Elsevier Science, 2 edition.

Boashash, B., O'Shea, P., and Arnold, M. (1990). Algorithms for instantaneous frequency estimation: A comparative study. In *Advanced Signal Processing Algorithms, Architectures, and Implementations*, volume 1348, pages 126–148. SPIE.

Bobbia, B., Doukhan, P., and Fan, X. (2022). A Review on some weak dependence conditions. HAL, 03325994v2.

Bobrowski, O. and Borman, M. S. (2012). Euler Integration of Gaussian Random Fields and Persistent Homology. *Journal of Topology and Analysis*, 04(01):49–70, 1003.5175.

Bogachev, V. I. (2007). *Measure Theory*. Springer, Berlin ; New York.

Bois, A., Tervil, B., Moreau, A., Vienne-Jumeau, A., Ricard, D., and Oudre, L. (2022). A topological data analysis-based method for gait signals with an application to the study of multiple sclerosis. *PLOS ONE*, 17(5):e0268475.

Bradley, R. C. (2005). Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. *Probability Surveys*, 2:107–144.

Bristeau, P.-J. (2012). *Techniques d'estimation du déplacement d'un véhicule sans GPS et autres exemples de conception de systèmes de navigation MEMS*. PhD thesis, Ecole Nationale Superieure des Mines de Paris.

Bristeau, P.-J., Petit, N., and Praly, L. (2010). Design of a navigation filter by analysis of local observability. In *49th IEEE Conference on Decision and Control (CDC)*, pages 1298–1305, Atlanta, GA, USA. IEEE.

Bubenik, P. (2015). Statistical Topological Data Analysis using Persistence Landscapes. *Journal of Machine Learning Research*, 16(1):77–102.

Bühlmann, P. (1995). The blockwise bootstrap for general empirical processes of stationary sequences. *Stochastic Processes and their Applications*, 58(2):247–265.

Bühlmann, P. (2002). Bootstraps for Time Series. *Statistical Science*, 17(1):52–72.

Bühlmann, P. and Künsch, H. R. (1999). Block length selection in the bootstrap for time series. *Computational Statistics & Data Analysis*, 31(3):295–310.

Carrière, M., Chazal, F., Ike, Y., Lacombe, T., Royer, M., and Umeda, Y. (2020). PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108, pages 2786–2796. PMLR.

Chazal, F., de Silva, V., Glisse, M., and Oudot, S. (2016). *The Structure and Stability of Persistence Modules*. Number 2191-8198 in SpringerBriefs in Mathematics. Springer, Cham.

Chazal, F., Fasy, B., Lecci, F., Michel, B., Rinaldo, A., and Wasserman, L. (2017). Robust Topological Inference: Distance To a Measure and Kernel Distance. *Journal of Machine Learning Research*, 18(1):5845–5884.

Chazal, F., Fasy, B. T., Lecci, F., Rinaldo, A., and Wasserman, L. (2014). Stochastic Convergence of Persistence Landscapes and Silhouettes. In *Annual Symposium on Computational Geometry - SOCG'14*, pages 474–483, Kyoto, Japan. ACM Press.

Chazal, F., Glisse, M., Labruere, C., and Michel, B. (2015). Convergence Rates for Persistence Diagram Estimation in Topological Data Analysis. *Journal of Machine Learning Research*, 16(110):3603–3635.

Chazal, F. and Michel, B. (2021). An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists. *Frontiers in Artificial Intelligence*, 4.

Chung, Y.-M. and Lawson, A. (2022). Persistence Curves: A canonical framework for summarizing persistence diagrams. *Advances in Computational Mathematics*, 48(6):42.

Cohen-Steiner, D. and Edelsbrunner, H. (2007). Inequalities for the Curvature of Curves and Surfaces. *Foundations of Computational Mathematics*, 7(4):391–404.

Cohen-Steiner, D., Edelsbrunner, H., Harer, J., and Mileyko, Y. (2010). Lipschitz Functions Have L p -Stable Persistence. *Foundations of Computational Mathematics*, 10(2):127–139.

Cruz, S. B. and Aguiar, A. (2020). MagLand: Magnetic Landmarks for Road Vehicle Localization. *IEEE Transactions on Vehicular Technology*, 69(4):3654–3667.

Curry, J., Mukherjee, S., and Turner, K. (2022). How Many Directions Determine a Shape and other Sufficiency Results for Two Topological Transforms. *Transactions of the American Mathematical Society, Series B*, 9(32):1006–1043, 1805.09782.

Dedecker, J., Doukhan, P., Lang, G., R. Léon, J., Louhichi, S., and Prieur, C. (2007). *Weak Dependence: With Examples and Applications*. Number 190 in Lecture Notes in Statistics. Springer, New York, 1 edition.

Di Fabio, B. and Landi, C. (2011). A Mayer-Vietoris Formula for Persistent Homology with an Application to Shape Recognition in the Presence of Occlusions. *Foundations of Computational Mathematics*, 11:499–527.

Divol, V. and Chazal, F. (2019). The density of expected persistence diagrams and its kernel based estimation. *Journal of Computational Geometry*, 10(2):127–153.

Divol, V. and Lacombe, T. (2021). Understanding the topology and the geometry of the space of persistence diagrams via optimal partial transport. *Journal of Applied and Computational Topology*, 5(1):1–53.

Divol, V. and Polonik, W. (2019). On the choice of weight functions for linear representations of persistence diagrams. *Journal of Applied and Computational Topology*, (3):249–283.

Dłotko, P., Qiu, W., and Rudkin, S. (2019). Cyclicality, Periodicity and the Topology of Time Series. arXiv, arXiv:1905.12118.

Doukhan, P. (1995). *Mixing*, volume 85 of *Lecture Notes in Statistics*. Springer New York, NY.

Erden, F. and Cetin, A. (2017). Period Estimation of an Almost Periodic Signal Using Persistent Homology With Application to Respiratory Rate Measurement. *IEEE Signal Processing Letters*, 24:958–962.

Fermanian, A. (2021). *Learning Time-Dependent Data with the Signature Transform*. PhD thesis, Sorbonne université, Paris, France.

Fernández, X., Borghini, E., Mindlin, G., and Groisman, P. (2023). Intrinsic Persistent Homology via Density-based Metric Learning. *Journal of Machine Learning Research*, 24(75):1–42.

Fernández, X. and Mateos, D. (2022). Topological biomarkers for real-time detection of epileptic seizures. arXiv, arXiv:2211.02523.

Gakhar, H. and Perea, J. A. (2019). K\"unneth Formulae in Persistent Homology. arXiv, arXiv:1910.05656.

Galbraith, S. D., Gebregiyorgis, S. W., and Murphy, S. (2016). Algorithms for the approximate common divisor problem. *LMS Journal of Computation and Mathematics*, 19(A):58–72.

Garin, A. and Tauzin, G. (2019). A Topological "Reading" Lesson: Classification of MNIST using TDA. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 1551–1556. IEEE.

Gasser, T. and Wang, K. (1997). Alignment of curves by dynamic time warping. *The Annals of Statistics*, 25(3):1251–1276.

Ghil, M. and Sciamarella, D. (2023). Review Article: Dynamical Systems, Algebraic Topology, and the Climate Sciences. EGUsphere, 2023-216.

Gidea, M., Goldsmith, D., Katz, Y., Roldan, P., and Shmalo, Y. (2020). Topological recognition of critical transitions in time series of cryptocurrencies. *Physica A: Statistical Mechanics and its Applications*, 548:123843.

Gidea, M. and Katz, Y. (2018). Topological Data Analysis of Financial Time Series: Landscapes of Crashes. *Physica A: Statistical Mechanics and its Applications*, 491:820–834.

Glisse, M. (2023). Fast persistent homology computation for functions on $\mathbb{R}$. arXiv, arXiv:2301.04745.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C. K., and Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):E215–220.

Hacquard, O. (2023). Statistical learning on measures: An application to persistence diagrams. arXiv, arXiv:2303.08456.

Hambly, B. and Lyons, T. (2010). Uniqueness for the signature of a path of bounded variation and the reduced path group. *Annals of Mathematics*, 171(1):109–167.

Hatcher, A. (2002). *Algebraic Topology*. Cambridge University Press.

Haykin, S. S., editor (2001). *Kalman Filtering and Neural Networks*. Adaptive and Learning Systems for Signal Processing, Communications, and Control. Wiley, New York.

Herbert Edelsbrunner and John Harer (2010). *Computational Topology: An Introduction*. American Mathematical Society.

Herbert Edelsbrunner, Ranita Biswas, Sebastiano Cultrera di Montesano, and Morteza Saghafian (2023). Window to the Persistence of 1D Maps: Geometric Characterization of Critical Point Pairs. *Journal of Applied Computational Topology*, 48(2):47.

Hofer, C. D., Graf, F., Rieck, B., Niethammer, M., and Kwitt, R. (2020). Graph Filtration Learning. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 4314–4323. PMLR.

Hu, Y. and Le, K. (2013). A multiparameter Garsia-Rodemich-Rumsey inequality and some applications. *Stochastic Processes and their Applications*, 123(9):3359–3377.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995.

Hussain, Z. M. and Boashash, B. (2002). Adaptive instantaneous frequency estimation of multicomponent FM signals using quadratic time-frequency distributions. *IEEE Transactions on Signal Processing*, 50(8):1866–1876.

Kaczynski, T., Mischaikow, K. M., and Mrozek, M. (2011). *Computational Homology*. Applied Mathematical Sciences. Springer, New York, NY, 1 edition.

Kahle, M. (2011). Random geometric complexes. *Discrete & Computational Geometry*, 45(3):553–573.

Kennedy, S. M., Roth, J. D., and Scrofani, J. W. (2018). A Novel Method for Topological Embedding of Time-Series Data. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 2350–2354, Rome. IEEE.

Khan, N. A. and Boashash, B. (2016). Multi-component instantaneous frequency estimation using locally adaptive directional time frequency distributions. *International Journal of Adaptive Control and Signal Processing*, 30(3):429–442.

Khasawneh, F. A. and Munch, E. (2016). Chatter detection in turning using persistent homology. *Mechanical Systems and Signal Processing*, 70–71:527–541.

Khasawneh, F. A. and Munch, E. (2018). Topological data analysis for true step detection in periodic piecewise constant signals. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2218):20180027.

Khasawneh, F. A., Munch, E., and Perea, J. A. (2018). Chatter Classification in Turning using Machine Learning and Topological Data Analysis. *IFAC-PapersOnLine*, 51(14):195–200.

Khorram, S., McInnis, M. G., and Provost, E. M. (2019). Trainable Time Warping: Aligning Time-Series in the Continuous-Time Domain. In *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3502–3506.

Kim, K. and Jung, J.-H. (2022). Exact multi-parameter persistent homology of time-series data: One-dimensional reduction of multi-parameter persistence theory. arXiv, arXiv:2211.03337.

Kneip, A. and Gasser, T. (1992). Statistical Tools to Analyze Data Representing a Sample of Curves. *The Annals of Statistics*, 20(3):1266–1305.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics. Springer New York, New York, NY.

Krebs, J. (2021). On limit theorems for persistent Betti numbers from dependent data. *Stochastic Processes and their Applications*, 139:139–174.

Lasalle, E. (2023). Heat diffusion distance processes: A statistically founded method to analyze graph data sets. *Journal of Applied and Computational Topology*.

Le Goff, A., Lacoume, J.-L., Blanpain, R., Dauvé, S., and Serviere, C. (2012). Automobile wheel clearance estimation using magnetism. *Mechanical Systems and Signal Processing*, 26:315–319.

Lesnick, M. (2015). The Theory of the Interleaving Distance on Multidimensional Persistence Modules. *Foundations of Computational Mathematics*, 15(3):613–650.

Lyons, T. J., Caruana, M., and Lévy, T. (2007). *Differential Equations Driven by Rough Paths: École d'Été de Probabilités de Saint-Flour XXXIV - 2004*, volume 1908 of *Lecture Notes in Mathematics*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Ma, M., Song, Q., Gu, Y., Li, Y., and Zhou, Z. (2018). An Adaptive Zero Velocity Detection Algorithm Based on Multi-Sensor Fusion for a Pedestrian Navigation System. *Sensors (Basel, Switzerland)*, 18(10):3261.

Marron, J. S., Ramsay, J. O., Sangalli, L. M., and Srivastava, A. (2015). Functional Data Analysis of Amplitude and Phase Variation. *Statistical Science*, 30(4):468–484.

Marsh, L. and Beers, D. (2023). Stability and Inference of the Euler Characteristic Transform. arXiv, arXiv:2303.13200.

Meyn, S. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Communications and Control Engineering. Springer-Verlag, 1 edition.

Myers, A. D., Khasawneh, F. A., and Fasy, B. T. (2020). Separating Persistent Homology of Noise from Time Series Data Using Topological Signal Processing. arXiv, arXiv:2012.04039.

Panaretos, V. M. and Zemel, Y. (2020). *An Invitation to Statistics in Wasserstein Space*. SpringerBriefs in Probability and Mathematical Statistics. Springer Cham, 1 edition.

Pegoraro, M. and Secchi, P. (2021). Functional Data Representation with Merge Trees. arXiv, arXiv:2108.13147.

Perea, J. A. (2019). Topological Time Series Analysis. *Notices of the American Mathematical Society*, 66(05):1.

Perez, D. (2022a). On C0-persistent homology and trees. arXiv, arXiv:2012.02634v3.

Perez, D. (2022b). *Persistent Homology of Stochastic Processes and Their Zeta Functions*. PhD thesis, ENS.

Perez, D. (2023). On the persistent homology of almost surely $C\hat{0}$ stochastic processes. *Journal of Applied and Computational Topology*.

Perng, C.-S., Wang, H., Zhang, S., and Parker, D. (2000). Landmarks: A new model for similarity-based pattern querying in time series databases. In *Proceedings of 16th International Conference on Data Engineering (Cat. No.00CB37073)*, pages 33–42, San Diego, CA, USA. IEEE Comput. Soc.

Plonka, G. and Zheng, Y. (2016). Relation between total variation and persistence distance and its application in signal processing. *Advances in Computational Mathematics*, 42(3):651–674.

Qigao, F., Hai, Z., Yan, S., Yixin, Z., Xiangpeng, Z., Jie, J., and Pengsong, Z. (2018). An Optimal Enhanced Kalman Filter for a ZUPT-Aided Pedestrian Positioning Coupling Model. *Sensors (Basel, Switzerland)*, 18(5).

Radulović, D. (1996). The bootstrap for empirical processes based on stationary observations. *Stochastic Processes and their Applications*, 65(2):259–279.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer Series in Statistics. Springer, New York.

Rankine, L., Mesbah, M., and Boashash, B. (2007). IF estimation for multicomponent signals using image processing techniques in the time–frequency domain. *Signal Processing*, 87(6):1234–1250.

Reininghaus, J., Huber, S., Bauer, U., and Kwitt, R. (2014). A Stable Multi-Scale Kernel for Topological Machine Learning. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4741–4748.

Roycraft, B., Krebs, J., and Polonik, W. (2023). Bootstrapping Persistent Betti Numbers and Other Stabilizing Statistics. *(to appear in) Annals of Statistics*.

Ruiz-Ortiz, M. A., Gómez-Larrañaga, J. C., and Rodriguez-Viorato, J. (2022). Some applications of TDA on financial markets. arXiv, arXiv:2203.05603.

Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):43–49.

Sebastiano Cultrera di Montesano, Herbert Edelsbrunner, Monika Henzinger, and Lara Ost (2023). Dynamically Maintaining the Persistent Homology of Time Series.

Shevchenko, G. (2017). Kolmogorov continuity theorem and Holder norm. MathOverflow.

Srivastava, A., Klassen, E., Joshi, S. H., and Jermyn, I. H. (2011). Shape Analysis of Elastic Curves in Euclidean Spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(7):1415–1428.

Stefan, A., Athitsos, V., and Das, G. (2013). The Move-Split-Merge Metric for Time Series. *IEEE Transactions on Knowledge and Data Engineering*, 25(6):1425–1438.

Steinwart, I. (2022). Mathematics of Gaussian Processes for Machine Learning. 3-Day Meeting of Statisticians in Paris.

Stovall, S. H. (1997). Basic Inertial Navigation. Technical Report NAWCWPNS TM 8128, Naval Air Warfare Center Weapons Division, China Lake, California.

Su, J., Kurtek, S., Klassen, E., and Srivastava, A. (2014). Statistical analysis of trajectories on Riemannian manifolds: Bird migration, hurricane tracking and video surveillance. *The Annals of Applied Statistics*, 8(1):530–552.

Takens, F. (1981). Detecting strange attractors in turbulence. In Rand, D. and Young, L.-S., editors, *Dynamical Systems and Turbulence, Warwick 1980*, Lecture Notes in Mathematics, pages 366–381, Berlin, Heidelberg. Springer.

Tang, R. and Muller, H.-G. (2008). Pairwise curve synchronization for functional data. *Biometrika*, 95(4):875–889.

Tazaki, S., Sato, H., and Kawase, M.Method and Apparatus for detecting tire revolution using magnetic field. Patent. US 6,246,226 B1.

Vishwanath, S., Sriperumbudur, B. K., Fukumizu, K., and Kuriki, S. (2022). Robust Topological Inference in the Presence of Outliers. arXiv, arXiv:2206.01795.

Zabulon, J., Frank, D., and Fallet, G.Procede pour mesurer la vitesse de rotation d'une roue de vehicule. Patent. 15 58072.

Zielinski, B., Lipinski, M., Juda, M., Zeppelzauer, M., and Dlotko, P. (2021). Persistence Codebooks for Topological Data Analysis. *Artificial Intelligence Review*, 54(3):1969–2009.