Identifier une variable d'ordre stationnaire : potentiel, volatilité lente et bruit rapide

Méthode de reconstruction d'une EDO stochastique effective à partir d'un signal $X(t)$

Ce texte propose une méthode de travail pour analyser un signal temporel $X(t)$ supposé stationnaire, déjà interprété comme variable d'ordre d'un système complexe. L'objectif n'est pas de retrouver toutes les variables microscopiques du système, mais de construire une représentation réduite qui sépare trois contributions : un potentiel effectif, une volatilité lente et un bruit rapide. Cette démarche prolonge l'idée de Landau, où une variable d'ordre permet de lire une transition dans une énergie libre, mais elle la transpose dans un cadre empirique : on part d'un signal mesuré et l'on cherche à reconstruire, à partir des données, la structure dynamique qui l'organise.

1. Hypothèse de départ : $X(t)$ est une variable d'ordre stationnaire

On suppose que le système complexe a déjà été réduit à une variable globale $X(t)$. Cette variable peut provenir d'une ACP appliquée à un réseau de capteurs, d'une observable physique pertinente, d'un indice financier, d'un indicateur de projet ou de toute autre grandeur condensant le comportement collectif du système. L'hypothèse importante est la stationnarité : les propriétés statistiques de $X(t)$, notamment sa densité de probabilité, ne changent pas significativement sur la fenêtre d'observation.

$$X(t) stationnaire$$

Cette hypothèse autorise une lecture statistique stable du signal. Si le système visite régulièrement certaines valeurs de X et en évite d'autres, ces préférences d'occupation peuvent être interprétées comme la trace d'un paysage effectif. Ce paysage ne doit pas être confondu avec une énergie thermodynamique stricte. Il s'agit plutôt d'un potentiel effectif, construit à partir des observations, qui résume la structure moyenne de la dynamique.

La représentation visée est une équation différentielle stochastique réduite de la forme suivante :

$$dX_t = -V'(X_t)dt + \sigma(t)dW_{1,t} + sdW_{2,t}$$

Dans cette écriture, le premier terme représente la dérive structurante. Il ramène le système vers ses régimes préférentiels. Le second terme représente une volatilité lente, c'est-à-dire une modulation temporelle de l'intensité des fluctuations. Le troisième terme représente un bruit rapide résiduel, plus proche d'une agitation locale ou d'une innovation blanche.

2. Première étape : estimer le potentiel effectif à partir de la densité

La première étape consiste à estimer la densité empirique $p(X)$ du signal. Comme $X(t)$ est supposé stationnaire, l'histogramme observé sur une fenêtre suffisamment longue peut être interprété comme une approximation de cette densité stationnaire. En pratique, il est préférable d'utiliser un histogramme lissé ou une estimation par noyau, car la suite de la méthode demande de dériver le potentiel, et les dérivées amplifient fortement le bruit d'échantillonnage.

On définit alors le potentiel effectif par analogie avec la physique statistique :

$$V_{eff}(X) = -\ logp(X) + C$$

La constante C est arbitraire, car seule la pente du potentiel intervient dans la dynamique. Cette définition signifie que les zones très probables correspondent à des vallées du potentiel. Le système y séjourne souvent. À l'inverse, les zones peu probables correspondent à des crêtes ou des barrières. Le potentiel est donc une manière de transformer une information statistique, la densité d'occupation, en une information dynamique, le paysage des régimes.

Cette étape est proche de l'intuition de Landau. Dans la théorie de Landau, l'énergie libre est développée en fonction d'une variable d'ordre et ses minima indiquent les phases stables. Ici, on inverse la démarche : au lieu de postuler une énergie libre, on la reconstruit empiriquement à partir de la densité observée de la variable d'ordre.

Une fois V_{eff} estimé, on calcule la force moyenne associée :

$$F(X) = -\ V'_{eff}(X)$$

Cette force représente la dérive moyenne du système. Si X est proche d'un bassin stable, $F(X)$ tend à le ramener vers le minimum du potentiel. Si X est proche d'une barrière, la force indique dans quel sens le système est repoussé. Cette dérive n'est donc pas un simple ajustement statistique : elle donne une lecture physique du signal en termes d'attracteurs, de seuils et de transitions possibles.

3. Deuxième étape : retirer la dérive structurante

Le signal observé contient à la fois une tendance dynamique moyenne et des fluctuations. Pour analyser correctement la volatilité et le bruit, il faut d'abord retirer la dérive associée au potentiel. On travaille en temps discret, car les données sont généralement échantillonnées à pas Δt. L'incrément du signal est :

$$\Delta X_t = X_{t+\Delta t} - X_t$$

Si l'équation continue est approximativement valable, alors l'incrément discret peut s'écrire :

$$\Delta X_t \approx - V'(X_t)\Delta t + \eta_t$$

Le terme η_t est le résidu après retrait de la dérive. On le calcule donc de la manière suivante :

$$\eta_t = \Delta X_t + V'(X_t)\Delta t$$

Cette opération est essentielle. Si elle n'est pas faite, on mélange deux phénomènes différents : d'un côté la structure moyenne du système, de l'autre ses fluctuations. Le résidu η_t correspond à ce qui reste lorsque l'on a retiré la force de rappel moyenne. Il contient encore la volatilité lente et le bruit rapide, mais il est désormais débarrassé du mouvement déterministe moyen dans le paysage effectif.

À ce stade, un premier contrôle est nécessaire. Le résidu η_t doit être à peu près centré. Si sa moyenne reste fortement différente de zéro, cela signifie que le potentiel est mal estimé, que le signal n'est pas suffisamment stationnaire, ou que la dérive ne dépend pas seulement de X mais aussi d'autres variables cachées ou d'un effet temporel non pris en compte.

4. Troisième étape : distinguer volatilité lente et bruit rapide

Après retrait de la dérive, on suppose que le résidu se décompose en deux composantes :

$$\eta_t \approx \sigma(t)\Delta W_{1,t} + s\Delta W_{2,t}$$

Ces deux termes sont tous deux aléatoires, mais ils n'ont pas le même statut. Le bruit rapide correspond à l'innovation instantanée, faiblement corrélée dans le temps. Il est associé aux hautes fréquences et aux fluctuations locales résiduelles. La volatilité lente, au contraire, ne se manifeste pas directement par une valeur moyenne de η_t, mais par l'intensité variable des fluctuations. Elle représente le fait que le système connaît des périodes calmes et des périodes agitées.

La difficulté méthodologique est donc la suivante : on ne sépare pas directement deux bruits en regardant seulement η_t. On sépare une innovation rapide et une enveloppe lente de cette innovation. C'est pourquoi il faut regarder non seulement η_t, mais aussi son amplitude ou son énergie locale :

$$|\eta_t), \eta_t^2, log\left(\eta_t^2 + \varepsilon\right)$$

Si η_t est peu autocorrélé, mais que η_t^2 ou $log\left(\eta_t^2 + \varepsilon\right)$ présente une autocorrélation lente, alors on observe une signature classique de volatilité lente. Les chocs instantanés ne sont pas prévisibles, mais leur intensité est organisée dans le temps. C'est le mécanisme que l'on appelle souvent *clustering* de volatilité : les grandes fluctuations tendent à être suivies d'autres grandes fluctuations, et les petites fluctuations d'autres petites fluctuations.

Cette propriété est centrale dans les séries financières, mais elle apparaît plus généralement dans les signaux issus de systèmes complexes : turbulence, clutter de mer, endommagement, données de projet, signaux biologiques ou organisationnels. Elle traduit l'influence des variables cachées éliminées lors de la réduction. Ces variables ne sont plus décrites explicitement, mais elles modulent l'intensité de l'agitation de la variable d'ordre.

5. Estimer la volatilité lente

Une première méthode simple consiste à estimer une variance locale par fenêtre glissante. On calcule une moyenne locale des carrés du résidu :

$$\sigma\hat{}^2(t) = \frac{1}{m\Delta t} \sum_{k=t-m/2}^{t+m/2} \eta_k^2$$

La largeur m de la fenêtre joue un rôle important. Si m est trop petit, l'estimation suit trop fortement les fluctuations rapides et ne représente plus une volatilité lente. Si m est trop grand, elle gomme les variations pertinentes du milieu. Le choix de m doit donc être relié à l'échelle de temps que l'on cherche à séparer : la volatilité doit évoluer plus lentement que les innovations rapides, mais assez vite pour représenter les changements d'état du système.

Une méthode plus fine consiste à utiliser des ondelettes ou un filtrage fréquentiel. L'idée est de travailler sur l'énergie locale du résidu, par exemple η_t^2 ou $log\left(\eta_t^2 + \varepsilon\right)$, puis d'en extraire les composantes basses fréquences. Cette procédure est souvent plus adaptée aux systèmes complexes, car la volatilité peut ne pas être régulière. Elle peut présenter des intermittences, des changements de régime et des comportements multi-échelles.

La séparation fréquentielle est donc possible, mais elle doit être formulée correctement. On ne filtre pas directement $X(t)$, car $X(t)$ mélange la structure, la dérive et les fluctuations. On filtre d'abord le résidu corrigé η_t, puis surtout son énergie locale. La volatilité lente correspond alors à la composante basse fréquence de l'enveloppe des fluctuations.

6. Quatrième étape : normaliser le résidu et tester le bruit rapide

Une fois la volatilité lente estimée, on définit une innovation normalisée :

$$z_t = \frac{\eta_t}{\sigma\hat{}(t)}$$

Si la séparation est pertinente, z_t doit avoir perdu l'essentiel de la mémoire présente dans η_t^2. Il doit être centré, faiblement autocorrélé, et son carré doit présenter beaucoup moins de mémoire que celui du résidu initial. Autrement dit, la volatilité lente doit avoir absorbé

l'organisation temporelle de l'intensité des fluctuations. Ce qui reste est alors le bruit rapide, ou innovation résiduelle.

Cette étape est un test de cohérence. Si z_t reste fortement autocorrélé, cela signifie que la dérive n'a pas été correctement retirée, que la volatilité estimée est trop grossière, ou que la dynamique de X nécessite des variables supplémentaires. Si z_t est décorrélé mais possède des queues très lourdes, il peut être nécessaire de remplacer le bruit gaussien par un bruit de Student, un bruit à sauts ou une loi plus adaptée aux extrêmes.

On dispose donc d'un critère de validation très opérationnel : la bonne décomposition est celle qui transforme le résidu brut en une innovation aussi proche que possible d'un bruit rapide, tout en isolant une volatilité lente interprétable.

7. Justification théorique : de Landau à Mori-Zwanzig

La méthode proposée rassemble plusieurs traditions théoriques. Le premier appui est la théorie de Landau. Dans cette théorie, une variable d'ordre permet de décrire le changement collectif d'un système, et l'énergie libre écrite en fonction de cette variable indique les états stables. L'idée transposée ici est que la variable $X(t)$, construite ou observée, permet de lire le système dans un potentiel effectif. Les minima de ce potentiel correspondent aux régimes privilégiés du système.

Le deuxième appui est la dynamique de Langevin. Une variable macroscopique peut être représentée par une dérive déterministe et un bruit. La dérive exprime la structure moyenne, tandis que le bruit exprime les fluctuations dues aux degrés de liberté non explicitement décrits.

Le troisième appui est la reconstruction de Langevin par les coefficients de Kramers-Moyal. Dans cette approche, on estime directement, à partir d'une série temporelle, les coefficients de dérive et de diffusion. Cela justifie l'idée qu'une dynamique stochastique effective peut être reconstruite à partir des données, sous certaines conditions de stationnarité, de résolution temporelle et de validité markovienne ou quasi-markovienne.

Le quatrième appui est le formalisme de Mori-Zwanzig. Lorsqu'un système de grande dimension est réduit à quelques variables résolues, les variables éliminées ne disparaissent pas purement et simplement. Elles réapparaissent sous forme de mémoire, de bruit effectif et parfois de dépendance d'état. Cette idée donne une justification profonde à l'introduction d'une volatilité lente : elle représente une manière compacte de porter l'influence des variables cachées sur l'intensité des fluctuations de la variable d'ordre.

Enfin, les modèles modernes de volatilité, notamment en finance, montrent que la volatilité elle-même peut être un processus structuré, persistant, multi-échelle ou rugueux. Cela donne un cadre moderne à l'idée que l'intensité du bruit n'est pas constante, mais porte une information importante sur l'état du milieu.

8. Synthèse de la méthode

La méthode complète peut se résumer comme une chaîne d'identification :

$$X(t) \to p(X) \to V(X) \to \eta_t \to \sigma\hat{}(t) \to z_t$$

On commence par le signal stationnaire $X(t)$. On estime sa densité $p(X)$, puis on construit un potentiel effectif $V(X) = -\ logp(X)$. On dérive ce potentiel pour obtenir la force moyenne $F(X) = -\ V'(X)$. On retire ensuite cette dérive des incréments du signal afin d'obtenir un résidu η_t. Ce résidu contient encore l'ensemble des fluctuations. On estime alors l'enveloppe lente de son énergie pour obtenir une volatilité $\sigma(t)$. Enfin, on normalise le résidu par cette volatilité afin d'obtenir une innovation rapide z_t.

Le résultat n'est pas seulement une procédure statistique. C'est une interprétation physique du signal. Le potentiel décrit les régimes et les barrières. La volatilité lente décrit l'état du milieu, l'effet des variables cachées et la mémoire de la complexité éliminée. Le bruit rapide décrit les innovations résiduelles, ce qui reste imprévisible à l'échelle d'échantillonnage.

Dans cette perspective, l'EDO stochastique devient un modèle générique de représentation d'un signal issu d'un système complexe réduit à une variable d'ordre. Elle ne remplace pas les équations complètes du système, mais elle donne une forme normale exploitable :

$$dX_t = -\ V'\left(X_t\right)dt\ +\ \sigma(t)dW_{1,t} + sdW_{2,t}$$

Cette forme normale est particulièrement intéressante parce qu'elle relie une donnée mesurée à une structure interprétable. Elle permet de passer d'un signal à un paysage, d'un paysage à un résidu, d'un résidu à une volatilité, et d'une volatilité à une lecture de la complexité cachée du système.

9. Points de vigilance

Plusieurs précautions sont nécessaires. La première concerne la stationnarité. Si le signal change de régime au cours du temps, la densité $p(X)$ mélange plusieurs états et le potentiel reconstruit devient ambigu. Il faut alors travailler par fenêtres ou introduire explicitement un paramètre de régime.

La deuxième concerne l'estimation du potentiel. La relation $V\ = -\ logp$ donne un potentiel effectif, mais elle ne suffit pas toujours à identifier la dérive réelle si la diffusion dépend fortement de X. Dans ce cas, il est préférable de compléter l'approche par une estimation conditionnelle des coefficients de Kramers-Moyal.

La troisième concerne la séparation volatilité/bruit. Cette séparation repose sur une hypothèse d'échelle : la volatilité doit varier plus lentement que le bruit rapide. Si les deux phénomènes se superposent fortement en fréquence, la séparation devient moins nette et doit être interprétée comme une approximation.

La quatrième concerne la nature du bruit. Le bruit rapide n'est pas nécessairement gaussien. Dans les systèmes complexes, il peut présenter des queues lourdes, des sauts ou des asymétries. Le test final ne doit donc pas seulement vérifier la blancheur, mais aussi la forme de la distribution des innovations.

Bibliographie restreinte

Landau, L. D., & Lifshitz, E. M. *Statistical Physics*. Référence classique pour la théorie de Landau, la variable d'ordre et l'énergie libre.

Hohenberg, P. C., & Halperin, B. I. (1977). *Theory of dynamic critical phenomena*. Reviews of Modern Physics, 49(3), 435-479. Référence majeure sur les dynamiques de variables d'ordre près des transitions critiques.

Friedrich, R., Peinke, J., Sahimi, M., & Tabar, M. R. R. (2011). *Approaching complexity by stochastic methods: From biological systems to turbulence*. Physics Reports, 506(5), 87-162. Référence utile pour la reconstruction de Langevin, les coefficients de Kramers-Moyal et les séries complexes.

Honisch, C., & Friedrich, R. (2011). *Estimation of Kramers-Moyal coefficients at low sampling rates*. Physical Review E, 83, 066701. Référence méthodologique sur les difficultés d'estimation à pas d'échantillonnage fini.

Mori, H. (1965). *Transport, collective motion, and Brownian motion*. Progress of Theoretical Physics, 33(3), 423-455. Article fondateur du formalisme de projection et de la mémoire.

Zwanzig, R. (2001). *Nonequilibrium Statistical Mechanics*. Oxford University Press. Présentation de référence du formalisme de projection et des équations généralisées de Langevin.

Bollerslev, T. (1986). *Generalized autoregressive conditional heteroskedasticity*. Journal of Econometrics, 31(3), 307-327. Référence fondatrice des modèles GARCH et de la volatilité conditionnelle.

Gatheral, J., Jaisson, T., & Rosenbaum, M. (2018). *Volatility is rough*. Quantitative Finance, 18(6), 933-949. Référence récente sur la rugosité de la volatilité et la modélisation fractionnaire de la log-volatilité.