Conception d'un PINN a partir d'une EDO stochastique avec potentiel, diffusion et memoire longue

Table des matières

Résumé 3

1. Pourquoi un PINN classique ne suffit pas 4

2. Premiere architecture : un PINN latent a deux etages 4

3. Pourquoi il ne faut pas imposer la derivee trajectorielle brute 5

4. Premiere voie robuste : les moments conditionnels des increments 5

5. Deuxieme voie : apprendre la densite plutot que la trajectoire 5

6. Troisieme voie : imposer explicitement la memoire longue 5

7. Forme generale de la fonction de cout 6

8. Approximation d'etat etendu pour rendre la memoire apprenable 6

9. Recommandation pratique 7

10. Conclusion 7

Synthese des trois cadres possibles 7

Résumé

Ce chapitre s'intéresse à la conception d'un PINN adapté à une équation différentielle stochastique avec potentiel, diffusion et mémoire longue. L'idée principale est qu'un PINN classique, conçu pour des équations différentielles déterministes, ne suffit pas ici. Le modèle étudié mélange en effet trois difficultés : une dérive non linéaire issue d'un potentiel, un bruit brownien, et surtout une volatilité latente à mémoire longue. Le système n'est donc pas markovien si l'on observe seulement la variable principale x(t)x(t)x(t). Pour rendre le problème apprenable, il faut introduire une variable latente z(t)z(t)z(t), liée à la volatilité, et modéliser explicitement sa mémoire longue.

Le texte propose d'abord une architecture à deux étages. Un premier réseau approxime la variable principale x(t)x(t)x(t), tandis qu'un second réseau approxime la variable latente z(t)=ln(1+σ(t)2)z(t)=\ln(1+\sigma(t)^2)z(t)=ln(1+σ(t)2). Le potentiel peut être imposé, paramétré, ou appris. Cette structure permet de séparer la géométrie du paysage dynamique, portée par le potentiel, de la texture temporelle des fluctuations, portée par la volatilité latente. Le chapitre insiste toutefois sur un point fondamental : il serait incorrect d'imposer directement la dérivée temporelle d'une trajectoire stochastique brute, car celle-ci n'existe pas au sens classique en présence de bruit brownien.

Pour contourner cette difficulté, plusieurs voies robustes sont proposées. La première consiste à travailler sur les moments conditionnels des incréments. On n'impose plus l'équation trajectorielle instantanée, mais la bonne moyenne locale et la bonne variance locale des incréments. Cette approche permet de construire une perte physics-informed fondée soit sur la vraisemblance conditionnelle, soit sur les résidus de moments, ce qui est particulièrement naturel pour les SDE. La deuxième voie consiste à apprendre directement la densité de probabilité p(x,t)p(x,t)p(x,t) via une équation de Fokker-Planck, approche élégante lorsque l'on s'intéresse davantage aux lois stationnaires qu'aux trajectoires individuelles.

La troisième voie, la plus proche de l'esprit du modèle, consiste à imposer explicitement la mémoire longue. Comme la variable z(t)z(t)z(t) est définie par une convolution avec un noyau en loi de puissance, le problème rejoint celui des opérateurs non locaux ou fractionnaires, donc l'esprit des fPINN. Le chapitre propose alors une fonction de coût combinant plusieurs termes : adéquation aux données, cohérence de la dérive, cohérence de la diffusion, respect de la mémoire longue et contraintes structurelles sur la positivité de la variance, la forme du potentiel ou la régularité des champs. Le modèle devient ainsi un PINN latent structurel, à la frontière entre PINN, neural SDE et modèles fractionnaires.

Le texte montre ensuite qu'en pratique, la mémoire longue est plus facile à apprendre si l'on utilise une approximation d'état étendu. L'idée consiste à approcher le noyau en loi de puissance par une somme d'exponentielles, ce qui permet d'introduire des variables auxiliaires vérifiant des équations différentielles ordinaires simples. Le système redevient alors markovien dans un espace d'état élargi, ce qui le rend beaucoup plus stable et apprenable numériquement. Cette reformulation conserve l'intuition physique de la cascade d'échelles tout en facilitant l'implémentation.

Enfin, le chapitre recommande une stratégie pragmatique : commencer par un prototype discret, où l'on discrétise l'équation de xxx et où l'on modélise znz_nzn par un filtre de mémoire longue ou une approximation autorégressive, avant d'envelopper le tout dans une architecture physics-informed. Le texte conclut que le bon cadre n'est donc pas un PINN standard, mais un PINN latent structurel combinant

potentiel, diffusion stochastique et mémoire longue. Ce cadre est présenté comme particulièrement pertinent pour des systèmes complexes intermittents à barrières de potentiel, comme les marchés financiers, la turbulence ou le clutter de mer.

1. Pourquoi un PINN classique ne suffit pas

On peut concevoir un PINN a partir du modele considere, mais il faut le faire sous une forme plus generale qu'un PINN classique pour EDO deterministe. L'equation melange en effet trois difficultes : une derivee non lineaire issue d'un potentiel, un forcage brownien, et surtout une volatilite latente a memoire longue.

Un PINN standard sait tres bien imposer une contrainte de type x point egal f(x,t), mais ici la vraie physique est stochastique et non markovienne si l'on regarde seulement x. Il faut donc construire un modele physics-informed elargi, a la frontiere entre PINN, neural SDE et fPINN.

Le point cle est le suivant. Si l'on ecrit simplement le modele en temps continu, alors x(t) seul n'est pas un etat ferme au sens markovien. En revanche, le systeme devient beaucoup plus traitable si l'on introduit une variable latente z(t) liee a la volatilite et si l'on impose a z une loi de memoire longue par convolution, ou une approximation equivalente en espace d'etat. C'est cette augmentation d'etat qui rend le probleme apprenable.

$$dX_t = -V'(X_t)dt + \sigma(t)dW_t^{(1)} + sdW_t^{(2)}$$

$$z(t) = ln(1 + \sigma(t)^2)$$

$$\sigma(t) = \sqrt{exp(z(t)) - 1}$$

$$z(t) = (K_\alpha * \xi)(t), K_\alpha(t) \sim t^{-\alpha}$$

2. Premiere architecture : un PINN latent a deux etages

La premiere architecture possible, et a mon sens la plus fidele au modele, est un PINN latent a deux etages. On prend un reseau x_theta(t) pour approximer la trajectoire moyenne ou l'etat latent de x, et un second reseau z_phi(t) pour approximer la dynamique de la volatilite log-transformee.

Le potentiel V peut etre connu, parametre, ou lui-meme appris. Si l'on garde un potentiel quartique pedagogique, la derivee du potentiel definit la derivee deterministe du systeme. On peut alors apprendre les coefficients du potentiel afin d'identifier le paysage dynamique a partir des donnees.

$$x_\theta : t \mapsto x_\theta(t), z_\phi : t \mapsto z_\phi(t)$$

$$V(x) = ax^4 - bx^2 + cx + d$$

$$f_\psi(x) = -V_\psi'(x) = -(4ax^3 - 2bx + c)$$

3. Pourquoi il ne faut pas imposer la derivee trajectorielle brute

Il faut toutefois faire attention a un point fondamental. Pour une trajectoire stochastique brute, la derivee temporelle classique de x(t) n'existe pas au sens usuel a cause du bruit brownien. Un PINN qui essaierait de forcer directement une egalite differentielle trajectorielle serait mal pose.

Il faut donc passer par une formulation plus robuste. La bonne idee consiste a imposer non pas la trajectoire differentielle instantanee, mais les proprietes statistiques des increments, ou bien une equation de densite, ou encore une contrainte non locale sur la variable latente de volatilite.

$$\dot{x}_{\theta}(t) + V'\left(x_{\theta}(t)\right) = \sigma(t)\dot{W}_1(t) + s\dot{W}_2(t)$$

4. Premiere voie robuste : les moments conditionnels des increments

La premiere voie consiste a travailler sur les moments conditionnels. En pas de temps Delta t, on discretise l'equation et l'on impose la bonne moyenne locale ainsi que la bonne variance locale des increments. La physique n'est donc plus imposee point par point, mais en loi conditionnelle.

Cette formulation conduit a une perte physics-informed tres naturelle, fondee soit sur une vraisemblance gaussienne conditionnelle, soit sur des residus de moments. En pratique, cette voie est souvent la meilleure pour les SDEs, parce qu'elle respecte directement la nature probabiliste du modele.

$$\Delta X_n \simeq - V'\left(X_n\right)\Delta t + \sqrt{\left(\sigma_n^2 + s^2\right)\Delta t}\varepsilon_n, \varepsilon_n \sim N(0,1)$$

$$E[\Delta X_n | X_n, z_n) \simeq - V'\left(X_n\right)\Delta t$$

$$Var[\Delta X_n | X_n, z_n) \simeq \left(\sigma_n^2 + s^2\right)\Delta t$$

5. Deuxieme voie : apprendre la densite plutot que la trajectoire

La deuxieme voie consiste a travailler non sur la trajectoire x(t), mais sur la densite de probabilite p(x,t). Si sigma(t) etait deterministe ou suffisamment lente, la densite conditionnelle satisferait une equation de Fokker-Planck comportant une derivee de convection issue du potentiel et une diffusion dependant de la variance instantanee.

On pourrait alors construire un PINN sur p_theta(x,t), avec une perte issue du residu de Fokker-Planck. Cette voie est elegante si l'objectif est d'apprendre des densites ou des lois stationnaires. Toutefois, la memoire longue de sigma complique la fermeture exacte sur p(x,t) seul. Il faut donc soit conditionner par z, soit passer a une equation effective non locale.

$$\partial_t p = \partial_x(V'(x)p) + \frac{1}{2}\left(\sigma(t)^2 + s^2\right)\partial_{xx}p$$

6. Troisieme voie : imposer explicitement la memoire longue

La troisieme voie consiste a modeliser explicitement la memoire longue avec une contrainte fractionnaire ou convolutionnelle. Votre formule pour z(t) evoque un operateur non local en temps. Cela

correspond tres bien a l'esprit des fPINNs, qui sont justement concus pour traiter des operateurs fractionnaires ou non locaux.

Dans ce cadre, on peut soit conserver une contrainte integrale explicite de type convolution, soit reformuler la memoire longue de z sous forme d'une equation fractionnaire approchante. Le reseau n'apprend alors pas seulement une trajectoire, mais une dynamique soumise a une memoire inherente.

$$z(t) = \left(K_\alpha * \xi\right)(t), K_\alpha(t) \sim t^{-\alpha}$$

7. Forme generale de la fonction de cout

Concretement, on peut definir deux reseaux x_theta et z_phi, puis construire une perte totale qui combine les donnees, la derivee moyenne, la diffusion, la memoire longue et les contraintes structurelles. Cette decomposition est tres naturelle et rend explicite le partage entre ce qui est impose par la physique et ce qui est appris a partir des donnees.

Le terme de donnees compare la sortie du reseau x_theta aux observations. Le terme de derivee impose la bonne moyenne locale des increments. Le terme de diffusion impose la bonne variance locale. Le terme de memoire impose la structure a longue portee de z_phi. Enfin, un terme de prior permet d'imposer la positivite de sigma^2, une symetrie eventuelle du potentiel, la regularite des champs et une valeur plausible du parametre s.

$$\sigma_\phi(t) = \sqrt{exp\left(z_\phi(t)\right) - 1}$$

$$L = \lambda_{data} L_{data} + \lambda_{drift} L_{drift} + \lambda_{diff} L_{diff} + \lambda_{mem} L_{mem} + \lambda_{prior} L_{prior}$$

$$L_{drift} = \frac{1}{N}\sum_n \left[\frac{\Delta x_n}{\Delta t} + V'\left(x_\theta\left(t_n\right)\right)\right)^2$$

$$L_{diff} = \frac{1}{N}\sum_n \left[\left(\Delta x_n\right)^2 - \left(\sigma_\phi\left(t_n\right)^2 + s^2\right)\Delta t\right)^2$$

$$L_{mem} = \frac{1}{N}\sum_n \left[z_\phi\left(t_n\right) - \sum_{k<n} w_{n-k}\xi_k\right)^2, w_j \propto j^{-\alpha}$$

8. Approximation d'etat etendu pour rendre la memoire apprenable

Il y a cependant une subtilite importante. Dans l'ecriture initiale, z(t) est une convolution d'un noyau singulier t puissance moins alpha avec un bruit blanc gaussien. Si l'on veut l'apprendre tel quel, un PINN pur n'est pas toujours la meilleure brique numerique.

En pratique, il est souvent plus stable d'utiliser une approximation d'etat etendu de la memoire longue. On approche le noyau puissance par une somme d'exponentielles. Cela permet d'introduire des variables auxiliaires verifiant chacune une ODE simple. Le modele redevient alors markovien dans un espace d'etat etendu, ce qui est beaucoup plus favorable a l'apprentissage. Cette idee conserve l'intuition physique de la cascade d'echelles tout en rendant le calcul numeriquement faisable.

$$t^{-\alpha} \simeq \sum_{m=1}^{M} c_m exp\left(-\lambda_m t\right)$$

$$r^{\cdot}_m = -\lambda_m r_m + \xi_m(t)$$

$$z(t) \simeq \sum_{m=1}^{M} c_m r_m(t)$$

$$état\ étendu : \left(x, r_1, \ldots, r_M\right)$$

9. Recommandation pratique

La recommandation la plus solide consiste a commencer par un prototype discret simple. On discretise l'equation pour x, puis l'on modelise z_n soit par un filtre a memoire longue de type convolution discrete, soit par une approximation AR de longue memoire. Ensuite seulement, on enveloppe le tout dans une architecture physics-informed ou l'on apprend le potentiel V, l'exposant alpha, eventuellement le parametre s, et les etats latents z_n.

Cette version sera plus stable qu'un PINN continu construit d'emblee, tout en restant deja tres puissante scientifiquement. Elle est particulierement adaptee a des donnees du type CAC 40, turbulence, clutter de mer, ou plus generalement a des systemes complexes intermittents a barriere de potentiel et agitation multi-echelle.

$$X_{n+1} = X_n - V'\left(X_n\right)\Delta t + \sqrt{\left(exp\left(z_n\right) - 1 + s^2\right)\Delta t}\varepsilon_n$$

$$z_n = \sum_{k=0}^{n} a_k \eta_{n-k}, a_k \sim k^{-\alpha}$$

10. Conclusion

Autrement dit, l'EDO consideree conduit moins a un PINN classique qu'a un PINN latent structurel combinant potentiel, diffusion stochastique et memoire longue. C'est meme un tres beau cas d'ecole, parce que le modele est a la fois assez simple pour etre identifiable et assez riche pour depasser les PINNs standards.

Un tel dispositif peut servir de demonstrateur transdisciplinaire pour la finance, la turbulence, le clutter de mer et, plus generalement, pour les systemes complexes ou une dynamique lente de paysage se combine a une agitation rapide et a une variabilite de longue portee.

Synthese des trois cadres possibles

Cadre	Point fort	Limite principale
PINN sur moments locaux	Simple, interpretable, proche de l'EDA	Ne reconstruit pas toute la loi trajectorielle a lui seul

PINN / Fokker-Planck	Naturel pour les densites et les lois stationnaires	Fermeture delicate quand sigma a une memoire longue
Neural SDE / fPINN latent	Le plus fidele au modele complet	Plus exigeant numeriquement et conceptuellement

Note de transcription. Les equations ont ete saisies sous une forme textuelle mathematique propre, directement editable dans Word et sans insertion d'images, afin de rester compatibles avec une reprise ulterieure dans l'editeur d'equations de Word.