

Notes

François Rousseau

2025-10-12

Table of contents

Préface	3
1 Introduction	4
2 Résumé	5
3 Systèmes dynamiques	6
3.1 Équations différentielles	6
3.2 Flot de gradient	7
3.3 Évolution d'une densité	7
3.4 Flot de gradient de Wasserstein	7
3.5 Flow matching	9
Références	10

Préface

Quelques notes de travail.

1 Introduction

Intro.

2 Résumé

En résumé, ce livre ne contient pas encore grand-chose.

3 Systèmes dynamiques

Un système dynamique est une formalisation mathématique permettant de décrire l'évolution de l'état d'un système dans le temps. Les trois composantes fondamentales sont :

1. X : l'espace d'états, c'est-à-dire l'ensemble de tous les états possibles du système (également appelé espace de phase),
2. T : l'ensemble de temps,
3. ϕ_t : l'opérateur d'évolution (appelé flot pour les systèmes à temps continu) qui transforme un état initial x_0 en l'état x_t à l'instant t .

3.1 Équations différentielles

Si l'opérateur d'évolution ne dépend que du temps, on peut écrire :

$$\dot{x}_t = f(t, x_t)$$

qui est une équation différentielle¹ ordinaire (EDO) du premier ordre. L'EDO définit la trajectoire d'un point dans un espace d'états de dimension finie.

Habituellement, on considère que la fonction f , appelée champ de vecteurs, est connue et l'objectif est alors de calculer les trajectoires des points dans l'espace d'états (la fonction x est l'inconnue). Il s'agit de faire de la prédiction à partir de conditions initiales. Se posent alors les questions de l'existence, d'unicité ou de stabilité des trajectoires. En apprentissage, la configuration est typiquement inversée : on considère un ensemble de points d'apprentissage et on souhaite estimer une fonction f qui prédit au mieux les observations.

Dans le cas où la fonction inconnue x dépend de plusieurs variables (comme le temps et les coordonnées spatiales), on

¹ Les équations différentielles apparaissent avec l'invention du calcul infinitésimal par Isaac Newton et Gottfried Leibniz.

parlera d'équation différentielle partielle (EDP). C'est le cas lorsque l'on souhaite modéliser un système qui varie à la fois dans le temps et dans l'espace. L'état du système x_t est alors une fonction définie sur le domaine spatial. L'ensemble de toutes les fonctions possibles est appelé espace fonctionnel. Une EDP décrit alors l'évolution d'un système dynamique en dimension infinie.

Transport optimal : la carte de transport optimal est le gradient d'une fonction convexe.

3.2 Flot de gradient

Un flot de gradient est un système dynamique dissipatif défini par une EDO de la forme $\dot{x}_t = -\nabla V(x_t)$. La trajectoire suit alors la direction de la plus forte pente descendante de la fonction potentielle V (qui est une fonction de Lyapunov). La discrétisation numérique via un schéma d'Euler explicite correspond à l'algorithme de descente de gradient.

3.3 Évolution d'une densité

L'évolution d'une densité de probabilité ρ_t transportée par un champ de vecteurs v_t est décrite par l'équation de continuité (EDP) :

$$\frac{\partial \rho_t}{\partial t} + \operatorname{div}(\rho_t v_t) = 0 .$$

Dans le cadre du transport optimal, la formule de Benamou-Brenier montre que le chemin le plus court entre deux densités est celui qui minimise l'énergie cinétique totale tout en satisfaisant cette EDP.

3.4 Flot de gradient de Wasserstein

Si on souhaite étendre la notion de flot de gradient (définie pour un point) à une densité de probabilité ρ_t , on obtient le flot de gradient de Wasserstein. Dans ce contexte, on cherche

l'évolution de ρ_t qui va minimiser une fonctionnelle d'énergie $F[\rho]$, c'est à dire le champ de vitesse suivant :

$$v_t = -\nabla \left(\frac{\delta F[\rho_t]}{\delta \rho} \right) .$$

En considérant la contrainte de l'équation de continuité, on obtient l'EDP suivante :

$$\frac{\partial \rho_t}{\partial t} = \operatorname{div} \left(\rho_t \nabla \left(\frac{\delta F[\rho_t]}{\delta \rho} \right) \right) .$$

Exemple : On peut par exemple choisir pour $F(\rho)$ l'intégrale de l'entropie de Boltzmann-Shannon :

$$F(\rho) = \int_{\mathbb{R}^d} U(\rho(x)) dx$$

avec $U(s) = s \ln(s)$. La variation première (notée $\frac{\delta F}{\delta \rho}$) est égale à la dérivée de la fonction de densité U :

$$\frac{\delta F}{\delta \rho} = U'(\rho) = \ln(\rho) + 1 .$$

On calcule ensuite le gradient de ce potentiel dans l'espace physique \mathbb{R}^d :

$$\nabla \left(\frac{\delta F}{\delta \rho} \right) = \nabla(\ln \rho + 1) = \frac{\nabla \rho}{\rho} .$$

Ce terme représente la vitesse à laquelle les particules se déplacent pour minimiser l'entropie. Injecté dans l'équation de continuité, on obtient l'équation de la chaleur :

$$\frac{\partial \rho_t}{\partial t} = \operatorname{div}(\nabla \rho) = \Delta \rho .$$

La diffusion de la chaleur correspond donc à une descente de gradient de l'entropie dans l'espace de Wasserstein. Le flux de chaleur est donc le chemin le plus court (géodésique) pour dissiper l'énergie libre².

L'EDP de flot de gradient de Wasserstein est souvent difficile à résoudre directement, on utilise en pratique le schéma de Jordan-Kinderlehrer-Otto (JKO), qui est l'analogue de la discrétisation

² Ce résultat a été démontré par Richard Jordan, David Kinderlehrer et Felix Otto dans l'article intitulé « The variational formulation of the Fokker-Planck equation », publié en 1998.

par schéma d'Euler implicite pour les flots de gradient classiques. On construit donc une suite de densité (ρ_k) pour des pas de temps discrets. À chaque étape k , connaissant la densité précédente ρ_{k-1} , la nouvelle densité ρ_k est obtenue en résolvant le problème de minimisation variationnelle suivant :

$$\rho_k = \arg \min_{\rho} \left(F(\rho) + \frac{1}{2\tau} W_2^2(\rho, \rho_{k-1}) \right).$$

3.5 Flow matching

Le Flow Matching s'inscrit dans une perspective lagrangienne du transport de densité en s'intéressant au flot qui déplace chaque particule dont l'évolution est décrite par une EDO.

Références