Optimisation de champ moyen régularisé par l'information de Fisher

Julien CLAISSE [1] Giovanni CONFORTI [2] Zhenjie REN [1] Songbo WANG [2]

[1]CEREMADE, Université Paris-Dauphine

[2]CMAP, École Polytechnique

01/06/2022

1. Le problème
2. Condition du premier ordre
3. La dynamique
4. HJB CM
5. Décroissance d'energie
6. Convergence
7. Descente de gradient

1 Le problème

2 Condition du premier ordre

3 La dynamique

4 HJB CM

5 Décroissance d'energie

6 Convergence

7 Descente de gradient

Motivations

On s'intéresse aux réseaux de neurones.

Cas : une seule couche cachée.

$i = 1, \ldots, n$ – neurones.

$\varphi : \mathbb{R} \to \mathbb{R}$ – fonctions d'activation, e.g. $\varphi(x) = x_+$ (ReLU).

Perte quadratique.

Problème : minimiser

$$F(n, a, b, c) = \mathbf{E}\left[\left|f(Z) - \frac{1}{n}\sum_{i=1}^{n} c_k \varphi(a_k Z + b_k)\right|^2\right].$$

Quand $n \to \infty$,

$$F \to \mathbf{E}\left[\left|f(Z) - \mathbb{E}_m\left[C\varphi(AZ + B)\right]\right|^2\right]$$

où $(A, B, C) \sim m$.

Remarques : F est convexe en m. Ce n'est plus vrai pour réseaux profonds ($\#$ couche ≥ 2).

Champ-moyen

On considère une fonction “champ-moyen” générale :

$$F : \mathcal{P}\left(\mathbb{R}^d\right) \to \mathbb{R}.$$

On suppose souvent que F est continue par apport à la distance de Wasserstein.

Régularisation entropique

On fixe une mesure de base. $H(m)$ est l'entropique relative.
On considère

$$\inf_m F(m) + \frac{\sigma^2}{2} H(m) .$$

[Hu, Ren, Šiška, Szpruch, 2019] : la descente de gradient (par apport à Wasserstein) donne la loi marginale de Langevin de champ moyen

$$dX_t = -DF(m_t, X_t)\, dt + \sigma dW_t, m_t \sim X_t.$$

m_t converges vers l'unique minimiseur de $F(m) + \frac{\sigma^2}{2} H(m)$.

Régularisation de Fisher

L'information de Fisher :

$$I(m) = \int |\nabla \log m|^2 \, m dx = 4 \int \left|\nabla \sqrt{m}\right|^2 dx.$$

On considère

$$\inf_m F(m) + \frac{\sigma^2}{4} I(m) = \inf_m F^\sigma(m) \, .$$

L'espace de probas dont l'information de Fisher est fini :

$$\mathcal{P}_H = \left\{ m \in \mathcal{P}\left(\mathbb{R}^d\right) : \sqrt{m} \in H^1 \right\}.$$

On suppose toujours que F est convexe dans la suite.

1 Le problème

2 Condition du premier ordre

3 La dynamique

4 HJB CM

5 Décroissance d'energie

6 Convergence

7 Descente de gradient

Différentiabilité d'une fonction CM

Définition

On dit que $F : \mathcal{P}\left(\mathbb{R}^d\right) \to \mathbb{R}$ est C^1 s'il existe une continue $\frac{\delta F}{\delta m} : \mathcal{P}\left(\mathbb{R}^d\right) \times \mathbb{R}^d \to \mathbb{R}$ t.q. pour tout $m_0, m_1 \in \mathcal{P}$

$$F(m_1) - F(m_0) = \int_0^1 \int \frac{\delta F}{\delta m}(m_t, x)\, d(m_1 - m_0)(x)\, dt$$

où $m_t = (1-t)\, m_0 + t m_1, t \in (0,1)$.

Remarques :

1. $\frac{\delta F}{\delta m}$ est définie à cste près.
2. (F est convexe). Si m minimise F, alors $\frac{\delta F}{\delta m}(m, \cdot)$ est cste.

Calculs différentiels de l'information de Fisher

Rappel :

$$I(m) = \int \frac{|\nabla m|^2}{m}.$$

On calcule formellement:

$$\begin{aligned}
\delta I(m) &= \int \frac{2\nabla m \cdot \nabla \delta m}{m} - \frac{|\nabla m|^2}{m^2}\delta m \\
&= \int \left(-2\nabla \cdot \left(\frac{\nabla m}{m} \right) - \frac{|\nabla m|^2}{m^2} \right) \delta m.
\end{aligned}$$

Donc

$$\frac{\delta F^\sigma}{\delta m} = \frac{\delta F}{\delta m} - \frac{\sigma^2}{2} \nabla \cdot \left(\frac{\nabla m}{m} \right) - \frac{\sigma^2}{4} \frac{|\nabla m|^2}{m^2}.$$

Pour tout rendre rigoureux...

Théorème

Si $m \in \mathcal{P}_H$ est t.q. $m \in L^\infty$, $m^{-1} \in L^\infty_{loc}$, et

$$\frac{\delta F^\sigma}{\delta m}(m,\cdot) = \frac{\delta F}{\delta m}(m,\cdot) - \frac{\sigma^2}{2}\nabla\cdot\left(\frac{\nabla m}{m}\right) - \frac{\sigma^2}{4}\frac{|\nabla m|^2}{m^2} = cste$$

au sens de distribution, alors m est l'unique minimiseur de F^σ.

Théorème

Si $m, m' \in \mathcal{P}_H \cap L^\infty$ et que $m^{-1} \in L^\infty_{loc}$. (Une certaine croissance de m, m'.) Alors

$$F^\sigma(m') - F^\sigma(m) \geq \int \frac{\delta F^\sigma}{\delta m}(m,x)\,(m'-m)\,dx.$$

1 Le problème

2 Condition du premier ordre

3 La dynamique

4 HJB CM

5 Décroissance d'energie

6 Convergence

7 Descente de gradient

Observations

Notons $\psi = \sqrt{m}$. La condition du premier ordre est équivalente à

$$\begin{aligned} cste &= \frac{\delta F}{\delta m} - \sigma^2 \nabla \cdot \left(\frac{\nabla \psi}{\psi} \right) - \sigma^2 \frac{|\nabla \psi|^2}{\psi^2} = \frac{\delta F}{\delta m} - \sigma^2 \frac{\Delta \psi}{\psi} \\ \Leftrightarrow \quad cste \cdot \psi &= \frac{\delta F}{\delta m} \psi - \sigma^2 \Delta \psi. \end{aligned}$$

ψ est une fonction propre de l'opérateur de Schrödinger CM

$$\sigma^2 \Delta - \frac{\delta F}{\delta m} (m, \cdot) .$$

Observations

Notons $u = -\log m$. La condition du premier ordre est équivalente à

$$cste = \frac{\delta F}{\delta m} + \frac{\sigma^2}{2}\Delta u - \frac{\sigma^2}{4}\left|\nabla u\right|^2.$$

C'est une équation HJB CM associée à un problème de contrôle ergodique.

Définition de la dynamique

On définit la dynamique :

$$\partial_t m_t = -\frac{\delta F^\sigma}{\delta m}(m_t, \cdot)\, m_t$$

où $\frac{\delta F}{\delta m}$ est choisie t.q. $\int \frac{\delta F^\sigma}{\delta m}(m, x)\, dm = 0$.
"Sanity check": $\partial_t \langle \mathbf{1}, m_t \rangle = 0$. Conservation de masse.
Formellement F^σ est décroissante :

$$\frac{dF^\sigma(m_t)}{dt} = -\int \left| \frac{\delta F^\sigma}{\delta m}(m_t, \cdot) \right|^2 dm_t$$

On peut espérer que $m_t \to$ minimiseur.

Formulations équivalentes

La dynamique en ψ :

$$\partial_t \psi_t = \frac{\sigma^2}{2} \Delta \psi_t - \frac{1}{2} \frac{\delta F}{\delta m}(m_t, \cdot)\, \psi_t$$

– Schrödinger dynamique CM

La dynamique en u:

$$\partial_t u = \frac{\sigma^2}{2} \Delta u - \frac{\sigma^2}{4} |\nabla u|^2 + \frac{\delta F}{\delta m}(m_t, \cdot)$$

– HJB dynamique CM

1 Le problème

2 Condition du premier ordre

3 La dynamique

4 HJB CM

5 Décroissance d'energie

6 Convergence

7 Descente de gradient

Hypothèses

F est continue par apport à $\mathcal{W}_1$ et convexe.
$F \in C^1$ et sa dérivée $\frac{\delta F}{\delta m}$ peut se décomposer

$$\frac{\delta F}{\delta m}(m, x) = g(x) + G(m, x)$$

où

1. $\kappa \, \mathrm{id} \leq \nabla^2 g \leq C \, \mathrm{id}$;
2. G est uniformément Lipschitzienne en x : $\sup_m \|\nabla G(m, \cdot)\|_\infty \leq L_G$.
3. ∇G est Lipschitzienne en m, x : $\forall m, m', x, x'$

$$\left|\nabla G(m, x) - \nabla G\left(m', x'\right)\right| \leq L_G \left(\mathcal{W}_1\left(m, m'\right) + \left|x - x'\right|\right).$$

Décomposition

$$\partial_t u = \frac{\sigma^2}{2}\Delta u - \frac{\sigma^2}{4}|\nabla u|^2 + \frac{\delta F}{\delta m}(m_t, \cdot)$$
$$= \frac{\sigma^2}{2}\Delta u - \frac{\sigma^2}{4}|\nabla u|^2 + g + G(m_t, \cdot)$$

On veut décomposer la fonction valeur $u = v + w$ où v, w résolvent résp.

$$\partial_t v = \frac{\sigma^2}{2}\Delta v - \frac{\sigma^2}{4}|\nabla v|^2 + g$$
$$\partial_t w = \frac{\sigma^2}{2}\Delta w - \frac{\sigma^2}{2}\nabla v \cdot \nabla w - \frac{\sigma^2}{4}|\nabla w|^2 + G(m_t, \cdot)$$

Convexité de v

$$\partial_t v = \frac{\sigma^2}{2}\Delta v - \frac{\sigma^2}{4}\left|\nabla v\right|^2 + g.$$

L'équation est classique. On a l'existence et l'unicité du problème de valeur initiale. De plus, on a

Proposition

Si $v_0 = v(0,\cdot)$ est θ_0-convexe, alors $v_t = v(t,\cdot)$ est θ_t-convexe où θ_t résout

$$\frac{d\theta_t}{dt} = \kappa - \frac{\sigma^2}{2}\theta_t^2$$

Heuristiques :

$$\partial_t \nabla^2 v = \frac{\sigma^2}{2}\Delta\nabla^2 v - \frac{\sigma^2}{2}\nabla v \cdot \nabla\nabla^2 v - \frac{\sigma^2}{2}\nabla^2 v \nabla^2 v + \nabla^2 g$$

Estimée a priori de w

Proposition

On suppose que w résout classiquement sur $[0, T]$

$$\partial_t w = \frac{\sigma^2}{2}\Delta w - \frac{\sigma^2}{2}\nabla v \cdot \nabla w - \frac{\sigma^2}{4}\left|\nabla w\right|^2 + L\,(t, x)$$

où L est uniformément Lipschitzienne en x et la valeur initiale $w_0 = w(0, \cdot)$ est aussi Lipschitzienne. On suppose de plus que $w, \nabla w$ sont à croissance polynomiale. Alors w est uniformément Lipschitzienne en x, dont la constante peut être indépendante de T.

Estimée a priori de w : preuve

$$\partial_t w = \frac{\sigma^2}{2}\Delta w - \frac{\sigma^2}{2}\nabla v \cdot \nabla w + \frac{\sigma^2}{2}\inf_{\alpha}\left\{-\alpha\cdot\nabla w + \frac{1}{2}|\alpha|^2\right\} + L(t,x)$$

On applique le théorème de vérification pour retrouver la rep. contrôle sto :

$$w(t,x) = \inf_{\alpha}\mathbb{E}\left[\int_0^t L(t-s,X_s) + \frac{\sigma^2}{4}|\alpha_s|^2\, ds + w(0,X_t)\right]$$

$$dX_s = -\frac{\sigma^2}{2}(\alpha_s + \nabla v_{t-s}(X_s))\, ds + \sigma dW_s, \quad X_0 = x$$

L'infimum est atteint par $\alpha_s^* = \nabla w_{t-s}(X_s)$.
On définit X' partant de x', avec le contrôle optimal pour x, en utilisant le même BM :

$$w\left(t,x'\right) \leq \mathbb{E}\left[\int_0^t L\left(t-s,X'_s\right) + \frac{\sigma^2}{4}|\alpha_s|^2\, ds + w\left(0,X'_t\right)\right]$$

$$dX'_s = -\frac{\sigma^2}{2}\left(\alpha_s + \nabla v_{t-s}\left(X'_s\right)\right) ds + \sigma dW_s, \quad X'_0 = x'$$

Estimée a priori de w : preuve

$$d\left(X_s - X'_s\right) = -\frac{\sigma^2}{2}\left(\nabla v_{t-s}\left(X_s\right) - \nabla v_{t-s}\left(X'_s\right)\right) ds$$

v_s est θ_s-convexe. Alors

$$\left|X_s - X'_s\right| \leq Ce^{-cs}\left|x - x'\right|$$

Soustraire les reps. de $w(t,x)$, $(w(t,x')$:

$$\begin{aligned} w(t,x) - w\left(t,x'\right) &\leq \mathbb{E}\left[\int_0^t L\left(t-s, X_s\right) - L\left(t-s, X'_s\right) ds \right. \\ &\quad \left. + w(0,X_t) - w\left(0, X'_t\right)\right] \\ &\leq \mathbb{E}\left[\int_0^t \|L\|_{\text{Lip}}\left|X_s - X'_s\right| ds + \|\nabla w_0\|_\infty \left|X_t - X'_t\right|\right] \\ &\leq C\left|x - x'\right| \end{aligned}$$

Couplage de réflexion

Théorème (Eberle, 2011)

Soient $b_1, b_2 : \mathbb{R}^d \to \mathbb{R}$ *où* b_1 *est strictement décroissant:*

$$(x-y)\cdot(b_1(x)-b_1(y)) \leq -\theta\,|x-y|^2$$

et b_2 *est borné.* $b = b_1 + b_2$. *Si la diffusion* $dX_t = b(X_t)\,dt + dW_t$ *n'explose pas, alors il existe cstes* c, C *t.q. pour toutes lois marginales* m_t, m'_t *de la diffusion dont* $m_0 = \delta_x, m'_0 = \delta_{x'}$, *on a*

$$\mathcal{W}_1\left(m_t, m'_t\right) \leq Ce^{-ct}\left|x-x'\right|.$$

Caractère bien posé

$$\partial_t w = \frac{\sigma^2}{2}\Delta w - \frac{\sigma^2}{2}\nabla v \cdot \nabla w - \frac{\sigma^2}{4}\left|\nabla w\right|^2 + G\left(m_t, \cdot\right)$$

Par contre, cette équation n'est pas classique. On doit montrer à la main qu'elle est bien posée.
Considérons la suite d'applications $w \mapsto m \mapsto \bar{w}$, où

$$\begin{aligned} m_t &= \exp\left(-v_t - w_t\right) \\ \partial_t \bar{w} &= \frac{\sigma^2}{2}\Delta \bar{w} - \frac{\sigma^2}{2}\nabla v \cdot \nabla \bar{w} - \left|\nabla \bar{w}\right|^2 + G\left(m_t, \cdot\right) \end{aligned}$$

On mesure la différence entre w, w' par $\sup_{t\in[0,T]} \left\|\nabla w - \nabla w'\right\|$. Avec la normalisation c'est une vraie norme.
On veut que l'application composée $w \mapsto \bar{w}$ soient une contraction pour T petit.

$w \mapsto m$

Soient $\bar{m}_t = \exp\left(-v_t - w_t\right), \bar{m}' = \exp\left(-v_t - w_t'\right)$. $\bar{m}_t$ est la mesure invariante de diffusion

$$dX_h = -\nabla v_t\left(X_h\right) dh - \nabla w_t\left(X_h\right) dh + \sqrt{2} dW_t$$

et pareil pour $\bar{m}'$. On utilise encore la méthode de réflexion pour construire un couplage t.q.

$$\begin{aligned}\mathcal{W}_1\left(\bar{m}_t, \bar{m}_t'\right) &\leq C\mathbb{E}\left[\int_0^\infty e^{-ch}\left|\nabla w_t\left(X_h\right) - \nabla w_t'\left(X_h\right)\right|\right] dh \\ &\leq C\left\|\nabla w_t - \nabla w_t'\right\|.\end{aligned}$$

$m \mapsto \bar{w}$

Soient $m, m' \in C([0,T]; \mathcal{W}_1)$. Notons $\delta\bar{w} = \bar{w} - \bar{w}'$. $\delta\bar{w}$ satisfait

$$\partial_t \delta\bar{w} = \frac{\sigma^2}{2}\Delta\delta\bar{w} - \frac{\sigma^2}{2}\nabla v \cdot \nabla\delta\bar{w} - \frac{\sigma^2}{2}\left(\nabla\bar{w} + \nabla\bar{w}'\right) \cdot \nabla\delta\bar{w} + G - G'$$

On a représentation probabiliste

$$\delta\bar{w}(t,x) = \mathbb{E}\left[\int_0^t \left(G - G'\right)(t-s, X_s)\, ds + \delta\bar{w}(0, X_t)\right]$$

$$dX_s = -\frac{\sigma^2}{2}\nabla v(t-s, X_s) - \frac{\sigma^2}{2}\left(\nabla\bar{w} + \nabla\bar{w}'\right)(t-s, X_s) + \sigma dW_s, s \le t$$

$$X_0 = x.$$

Observons que le drift de $X = \theta_{t-s}$-monotone + borné. On utilise le *couplage de réflexion* pour trouver un X' partant de x', satisfaisant l'équation de diffusion t.q. $\mathbb{E}\,|X_s - X'_s| \le C|x - x'|$.

$$\sup_{t\in[0,T]} \|\delta\nabla\bar{w}(t,\cdot)\|_\infty \le CT \sup_{t\in[0,T]} \mathcal{W}_1\left(m_t, m'_t\right) + C\|\delta\nabla\bar{w}(0,\cdot)\|_\infty .$$

1 Le problème

2 Condition du premier ordre

3 La dynamique

4 HJB CM

5 Décroissance d'energie

6 Convergence

7 Descente de gradient

Preuve rigoureuse

On veut rendre rigoureux la formule suivante

$$\frac{dF^{\sigma}\left(m_t\right)}{dt} = -\int \left|\frac{\delta F^{\sigma}}{\delta m}\left(m_t, x\right)\right|^2 m_t dx.$$

Outil: convexité, convergence dominée.
Convexité:

$$\begin{aligned}\int \frac{\delta F^{\sigma}}{\delta m}\left(m_{t+h}, x\right)\left(m_{t+h} - m_t\right) dx \geq F^{\sigma}\left(m_{t+h}\right) - F^{\sigma}\left(m_t\right) \\ \geq \int \frac{\delta F^{\sigma}}{\delta m}\left(m_t, x\right)\left(m_{t+h} - m_t\right) dx\end{aligned}$$

où m_t résout la dynamique classiquement, i.e.

$$m_{t+h} - m_t = -\int_0^h \int \frac{\delta F^{\sigma}}{\delta m}\left(m_{t+r}, x\right) m_{t+r} dx dr.$$

Intégrabilité

Pour appliquer convergence dominée, on a besoin de

1. $\sup_t \left| \frac{\delta F^\sigma}{\delta m}(t,x) \right| \leq C\left(1+|x|^2\right)$;
2. $\sup_t \int |x|^{4+\delta} m_t dx < +\infty$.

Rappel :

$$\frac{\delta F^\sigma}{\delta m} = \frac{\delta F}{\delta m} + \frac{\sigma^2}{2}\Delta u - \frac{\sigma^2}{4}|\nabla u|^2$$

A noter que :

1. On peut prouver classiquement (Bernstein, “turnpike”) $\sup_t |\nabla v(x)| \leq C(1+|x|)$;
2. $\nabla u = \nabla v + \nabla w$ dont ∇v est à croissant linéaire, ∇w borné;
3. $\sup_t \left\|\nabla^2 v_t\right\|_\infty < +\infty$;
4. $m_t = \exp(-v_t - w_t)$. On peut utiliser concentration pour montrer que $\int |x|^p m_t dx < C_p$ pour tout $p \geq 1$.

Il nous reste à borner Δu.

Estimée sur $\nabla^2 u$: couplage de réflexion

∇u résout

$$\partial_t \nabla u = \frac{\sigma^2}{2}\Delta \nabla u - \frac{\sigma^2}{2}\nabla u \cdot \nabla^2 u + \nabla \frac{\delta F}{\delta m}$$

Représentation probabiliste:

$$\nabla u(t,x) = \mathbb{E}\left[\int_0^t \nabla \frac{\delta F}{\delta m}(m_{t-s}, X_s) + \nabla u(0, X_t)\right]$$

$$\begin{aligned} dX_s &= -\sigma^2 \nabla u(t-s, X_s)\, ds + \sigma dW_s \\ &= -\sigma^2 (\nabla v + \nabla w)(t-s, X_s)\, ds + \sigma dW_s \end{aligned}$$

Drift = monotone + borné. On utilise le couplage de réflexion pour trouver une proba t.q. (X' la même diffusion dont le point de départ est x')

$$\mathbb{E}\left|X_s - X'_s\right| \le Ce^{-cs}\left|x - x'\right|.$$

Donc ∇u est uniformément Lipschitz en x, i.e. $\sup_t \left\|\nabla^2 u_t\right\|_\infty < +\infty$.

1 Le problème

2 Condition du premier ordre

3 La dynamique

4 HJB CM

5 Décroissance d'energie

6 Convergence

7 Descente de gradient

Preuve

$$\mathcal{K} := \left\{ \left(m, \sqrt{m}\right) : \int |x|^2 m + \left|\nabla \sqrt{m}\right|^2 dx \leq C \right\}.$$

Il est connu que $\mathcal{K}$ est $\mathcal{W}_1 \otimes$ faible-H^1 compact.
$\left(m_t, \sqrt{m_t}\right) \in \mathcal{K}$ pour un certain C.
$\exists \left(\hat{m}, \sqrt{\hat{m}}\right)$ un point d'accumulation de $\left(m_{t_k}, \sqrt{m_{t_k}}\right)$ pour $t_k \to \infty$.
L'énergie est décroissante et finie. Elle ne peut pas décroître infiniment.

$$\frac{dF^\sigma(m_t)}{dt} = -\int \left|\frac{\delta F^\sigma}{\delta m}(m_t, x)\right|^2 m_t dx.$$

Il existe une suite $h_j \downarrow 0$ t.q.

$$\int \left|\frac{\delta F^\sigma}{\delta m}\left(m_{t_k+h_j}, x\right)\right|^2 m_{t_k+h_j} dx \to 0.$$

On prétends que $h_j = 0$ dans la suite.

Preuve (suite)

Pour tout $m \in L^\infty$,

$$F^\sigma(m_{t_k}) \leq F^\sigma(m) + \int \frac{\delta F^\sigma}{\delta m}(m_{t_k}, x)(m - m_{t_k})\, dx$$
$$\leq F^\sigma(m) + \int_K \frac{\delta F^\sigma}{\delta m}(m_{t_k}, x)(m - m_{t_k})\, dx + \varepsilon$$
$$\leq F^\sigma(m) + \varepsilon +$$
$$\left(\int_K \left|\frac{\delta F^\sigma}{\delta m}(m_{t_k}, x)\right|^2 m_{t_k} dx\right)^{1/2} \left(\int_K \frac{(m - m_{t_k})^2}{m_{t_k}}\right)^{1/2}$$

On peut prouver que $m_{t_k}^{-1} \in L^\infty_{\text{loc}}$. Donc

$$F(\hat{m}) \leq \liminf F^\sigma(m_{t_k}) \leq F^\sigma(m).$$

Puis approcher m quelconque par des mesure borné et conclure.

1 Le problème

2 Condition du premier ordre

3 La dynamique

4 HJB CM

5 Décroissance d’energie

6 Convergence

7 Descente de gradient

GD à la JKO (heuristiques)

On mesure la distance entre mesures par l'entropie relative.
A chaque étape,

$$m_{k+1} = \arg\min_m h^{-1} H(m|m_k) + F^\sigma(m)$$

Calculs du premier ordre formels:

$$\begin{aligned} 0 &= h^{-1}\delta \int \log \frac{m}{m_k} m + \delta F^\sigma(m) \\ &= h^{-1} \int \log \frac{m}{m_k} \delta m + \int \frac{\delta F^\sigma}{\delta m}(m,\cdot)\,\delta m \end{aligned}$$

de sorte que

$$m_{k+1} = \frac{m_k}{Z_k} \exp\left(-h \frac{\delta F^\sigma}{\delta m}(m_{k+1},\cdot)\right) \approx m_k \left(1 - h \frac{\delta F^\sigma}{\delta m}(m_{k+1},\cdot)\right).$$

On espère $m_{kh} \to m_t$ quand $h \to 0$.

Conclusions

1. Problème d'optimisation champ-moyen avec Fisher, FOC
2. Dynamique (Schrödinger CM, HJB CM, GD entropie-Fisher)
3. Convergence (pas de vitesse, encore...)
4. Pas du tout de numérique