Consensus Based Sampling

Xuda Ye

Peking University

August 24, 2021

Problem Setting

- Basic goal: optimization & sampling without the use of gradient.
- Inverse problem: find $\theta \in \mathbb{R}^d$ from $y \in \mathbb{R}^K$ where

$$y = G(\theta) + \eta \tag{1.1}$$

Here G is the forward mapping and η is the observational noise.

- Assume $\theta \sim \mathsf{N}(0, \Sigma)$ and $\eta \sim \mathsf{N}(0, \Gamma)$. The posterior distribution is

$$\rho(\theta) = \frac{\exp(-f(\theta))}{\int \exp(-f(\theta))\mathrm{d}\theta} \tag{1.2}$$

where

$$f(\theta) = \frac{1}{2}|y - G(\theta)|_\Gamma^2 + \frac{1}{2}|\theta|_\Sigma^2$$

Problem Setting

- Sampling: sample the posterior distribution $\rho(\theta) \propto \exp(-f(\theta))$.
- Optimization: find the minimal point θ^* of $f(\theta)$.

Evaluation of $f(\theta)$ is a black box, and the gradient of $f(\theta)$ is not available. Therefore, a gradient-free optimization or sampling method is required.

Previous Work

1. Sequential Monte Carlo (Dashti & Stuart, 2015)
   - Target distribution approximated by a Dirac distribution.
   - Dirac distribution evolved by weighting and resampling.
   - Requires additional proposal step.
2. Ensemble Kalman inversion (Kovachki & Stuart, 2018)
   Ensemble Kalman sampling (Hoffmann & Stuart, 2019)
   - Approximate the gradient by the difference in the ensemble.
   - Sampling is accurate only for linear problems.
3. Consensus-Based Optimization (Carrillo & Jin, 2020)
   - The ensemble attracted by the consensus.

Consensus-Based Sampling (CBS)

CBS (Carrillo, Hoffmann & Stuart, 2021) has the following properties:

1. The mean-field equation is exact in the linear case (linear mapping & Gaussian prior), and the explicit convergence rate is obtained.
2. When G is nonlinear, the mean-field equation admits a Gaussian distribution as the steady state, whose bias from the exact posterior distribution is estimated.
3. Numerical experiments show that CBS is competitive with EKI/EKS.

Consensus-Based Sampling (CBS)

CBS begins with the McKean difference equation[1].

Given tha parameters $\lambda > 0$, $\beta > 0$ and $\alpha \in (0,1)$,

$$\begin{cases} \theta_{n+1} = \mathcal{M}_\beta(\rho_n) + \alpha(\theta_n - \mathcal{M}_\beta(\rho_n)) + \sqrt{(1-\alpha^2)\lambda^{-1}\mathcal{C}_\beta(\rho_n)}\boldsymbol{\xi}_n \\ \rho_n = \mathrm{Law}(\theta_n) \end{cases} \tag{2.1}$$

where $\boldsymbol{\xi}_n \sim \mathsf{N}(0,1)$ and $\mathcal{M}_\beta, \mathcal{C}_\beta$ denote the mean and covariance of the β-reweighted distribution:

$$\mathcal{M}_\beta : \rho \mapsto \mathcal{M}(L_\beta \rho), \quad \mathcal{C}_\beta : \rho \mapsto \mathcal{C}(L_\beta \rho), \quad L_\beta : \rho \mapsto \frac{\rho e^{-\beta f}}{\int \rho e^{-\beta f}} \tag{2.2}$$

[1] Coefficients depend on the solution itself.

Consensus-Based Sampling (CBS)

Parameters in the equation (2.1):

- $\beta > 0$: inverse temperature;
- $\lambda > 0$: controlling the diffusion;
- $\alpha \in [0, 1)$: how much θ_n is attracted by the consensus $\mathcal{M}_\beta(\rho)$.

Understanding the equation (2.1):

- β-reweighted distribution: $\rho(\theta)$ is weighted by $e^{-\beta f(\theta)}$. Positions with low potential (small $f(\theta)$) is assigned with larger weight.
- $M_\beta(\rho_n)$ serves as the consensus of the distribution $\rho_n(\theta)$.

Consensus-Based Sampling (CBS)

The continuous-time limit $(\alpha = 1)$ of (2.1) is the McKean SDE:

$$\begin{cases} \mathrm{d}\theta_t = -(\theta_t - \mathcal{M}_\beta(\rho_t))\mathrm{d}t + \sqrt{2\lambda^{-1}\mathcal{C}_\beta(\rho_t)}\mathrm{d}\mathbf{W}_t \\ \rho_t = \mathrm{Law}(\theta_t) \end{cases} \tag{2.3}$$

Use $(2.1)(2.3)$ to solve the sampling/optimization problem:

- Sampling: $\lambda = (1+\beta)^{-1}$.
- Optimization: $\lambda = 1$.

The reason for the choices of λ will be stated heuristically.

Consensus-Based Sampling (CBS)

For convenience, let

$$g(\theta; \mathbf{m}, C) = \frac{1}{\sqrt{(2\pi)^d \det(C)}} \exp\left(-\frac{1}{2}|\theta - \mathbf{m}|_C^2 \right) \tag{2.4}$$

be the probability density of the Gaussian distribution $\mathsf{N}(\mathbf{m}, C)$.

$$\mathbf{m}_\beta(\mathbf{m}, C) := \mathcal{M}_\beta(\mathsf{N}(\mathbf{m}, C)), \qquad C_\beta(\mathbf{m}, C) := \mathcal{C}_\beta(\mathsf{N}(\mathbf{m}, C))$$

are the mean and covariance of the β-reweighted distribution for $\mathsf{N}(\mathbf{m}, C)$.

Consensus-Based Sampling (CBS)

Taking expectation and covariance in (2.1), we obtain[2]

$$\begin{aligned} \mathcal{M}(\rho_{n+1}) &= \alpha\mathcal{M}(\rho_n) + (1-\alpha)\mathcal{M}_\beta(\rho_n) \\ \mathcal{C}(\rho_{n+1}) &= \alpha^2\mathcal{C}(\rho_n) + \lambda^{-1}(1-\alpha^2)\mathcal{C}_\beta(\rho_n) \end{aligned} \tag{2.6}$$

If ρ_∞ is the steady state, we have

$$\mathcal{M}(\rho_\infty) = \mathcal{M}_\beta(\rho_\infty), \quad \mathcal{C}(\rho_\infty) = \lambda^{-1}\mathcal{C}_\beta(\rho_\infty) \tag{$*$}$$

For the linear problem, assume the posterior distribution is $\mathsf{N}(\mathbf{a}, A)$, i.e.,

$$f(\theta) = \frac{1}{2}|\theta - \mathbf{a}|_A^2$$

then the steady state $\rho_\infty = \mathsf{N}(\mathbf{m}_\infty, C_\infty)$ is Gaussian.

[2] (2.6) is independent of the parameter α.

Consensus-Based Sampling (CBS)

For the linear mapping G, we have

$$\begin{aligned}\mathbf{m}_\beta(\mathbf{m}, C) &= (C^{-1} + \beta A^{-1})^{-1}(\beta A^{-1}\mathbf{a} + C^{-1}\mathbf{m}) \\ C_\beta(\mathbf{m}, C) &= (C^{-1} + \beta A^{-1})^{-1}\end{aligned} \tag{2.7}$$

Insert $\rho_\infty = \mathsf{N}(\mathbf{m}_\infty, C_\infty)$ and we obtain

$$\begin{aligned}\mathbf{m}_\infty &= (C_\infty^{-1} + \beta A^{-1})^{-1}(\beta A^{-1}\mathbf{a} + C^{-1}\mathbf{m}) \\ C_\infty &= \lambda^{-1}(C_\infty^{-1} + \beta A^{-1})^{-1}\end{aligned}$$

whose solution is

$$\boxed{\mathbf{m}_\infty = \mathbf{a}, \quad C_\infty = \frac{1-\lambda}{\lambda\beta}A}$$

- Optimization: $\lambda = 1$.
- Sampling: $\lambda = (1+\beta)^{-1}$.

Consensus-Based Sampling (CBS)

CBS directly inspires from the CBO.

1. Original version (Pinnau & Totzeck, 2017):

$$\mathrm{d}\theta = -(\theta - \mathcal{M}_\beta(\rho_t))\mathrm{d}t + |\theta - \mathcal{M}_\beta(\rho_t)|\mathrm{d}\mathbf{W}_t$$

2. Modified version (Carrillo & Jin, 2020):

$$\mathrm{d}\theta = -\lambda(\theta - \mathcal{M}_\beta(\rho_t)) + \sigma \sum_{i=1}^{d} e_i(\theta - \mathcal{M}_\beta(\rho_t))_i \mathrm{d}\mathbf{W}_t^i$$

3. This version:

$$\mathrm{d}\theta_t = -(\theta_t - \mathcal{M}_\beta(\rho_t))\mathrm{d}t + \sqrt{2\mathcal{C}_\beta(\rho_t)}\mathrm{d}\mathbf{W}_t$$

To optimize the target function, the diffusion coefficient vanishes as the ensemble collapses.

Key Properties

The equations (2.1) and (2.3) are essentially the evolution of the probability denisty ρ.

- (2.1) is governed by

$$\rho_{n+1}(\theta) = \int_{\mathbb{R}^d} g\Big(\theta; \mathcal{M}_\beta(\theta_n) + \alpha(u - \mathcal{M}_\beta(\rho_n)), (1-\alpha^2)\lambda^{-1}\mathcal{C}_\beta(\rho_n)\Big)\rho_n(u)\mathrm{d}u \tag{2.11}$$

- (2.3) is governed by

$$\frac{\partial \rho}{\partial t} = \nabla \cdot \Big((\theta - \mathcal{M}_\beta(\rho))\rho + \lambda^{-1}\mathcal{C}_\beta(\rho)\nabla\rho\Big) \tag{2.13}$$

Key Properties

Lemma 2.1 (Gaussian steady state)

Let probability distribution ρ_∞ have finite second moment and be a steady-state of (2.11) or (2.13). Then

$$\rho_\infty(\theta) = g(\theta; \mathcal{M}_\beta(\rho_\infty), \lambda^{-1}\mathcal{C}_\beta(\rho_\infty)) \tag{2.15}$$

Conversely, all probability distributions solving (2.15) are steady states of (2.11) or (2.13).

The lemma implies the steady state ρ_∞ must be Gaussian, and

$$\mathcal{M}(\rho_\infty) = \mathcal{M}_\beta(\rho_\infty), \quad \mathcal{C}(\rho_\infty) = \lambda^{-1}\mathcal{C}_\beta(\rho_\infty)$$

Convergence Analysis (Linear)

- In the linear case, assume the posterior distribution is $\mathsf{N}(\mathbf{a}, A)$, i.e., the corresponding potential is $f(\theta) = \frac{1}{2}|\theta - \mathbf{a}|_A^2$.
- Let $\rho_0 = \mathsf{N}(\mathbf{m}_0, C_0)$ be the initial distribution and define the constant

$$k_0 = \left\|C_0^{-1}\right\|_{A^{-1}} = \left\|A^{\frac{1}{2}} C_0^{-1} A^{\frac{1}{2}}\right\|$$

The convergence rate is shown in the following table:

	Sampling		Optimization	
	Mean	Covariance	Mean	Covariance
$\alpha = 0$	$\left(\frac{1}{1+\beta}\right)^n$	$\left(\frac{1}{1+\beta}\right)^n$	$\frac{k_0}{k_0+\beta n}$	$\frac{k_0}{k_0+\beta n}$
$\alpha \in (0,1)$	$\left(\frac{1+\alpha\beta}{1+\beta}\right)^n$	$\left(\frac{1+\alpha^2\beta}{1+\beta}\right)^n$	$\left(\frac{k_0+\beta}{k_0+\beta+\beta(1-\alpha^2)n}\right)^{\frac{1}{1+\alpha}}$	$\frac{k_0+\beta}{k_0+\beta+\beta(1-\alpha^2)n}$
$\alpha = 1$	$\mathrm{e}^{-\left(\frac{\beta}{1+\beta}\right)t}$	$\mathrm{e}^{-\left(\frac{2\beta}{1+\beta}\right)t}$	$\left(\frac{k_0+\beta}{k_0+\beta+2\beta t}\right)^{\frac{1}{2}}$	$\frac{k_0+\beta}{k_0+\beta+2\beta t}$

TABLE 1: Convergence rates for CBS in sampling and optimization modes, in the case of a Gaussian target distribution and a Gaussian initial condition with $C_0 \in \mathcal{S}_{++}^d$. This table summarizes the results in Propositions 2.4 to 2.6. All rates are sharp, see Remark 2.4.

Convergence Analysis (Linear)

The convergence rate is deduced from the update formula of $(\mathbf{m}_n, C_n)$:

$$m_{n+1} - \mathbf{a} = [\alpha I_d + (1-\alpha)A(A+\beta C_n)^{-1}](\mathbf{m}_n - \mathbf{a})$$
$$C_{n+1} = [\alpha^2 I_d + (1-\alpha^2)\lambda^{-1}A(A+\beta C_n)^{-1}]C_n$$

In the limit $\alpha \to 1$, $(\mathbf{m}(t), C(t))$ is evolved by

$$\dot{\mathbf{m}} = -\beta C(A+\beta C)^{-1}(\mathbf{m} - \mathbf{a})$$
$$\dot{C} = -2\beta C(A+\beta C)^{-1}\left(C - \left(\frac{1-\lambda}{\beta\lambda}\right)A\right)$$

Convergence Analysis (Nonlinear)

In the nonliear case, there are several assumptions on the potential function $f(\theta)$.

Assumption 1, 2 (Convexity & Boundedness of Hessian)

$f \in C^2(\mathbb{R}^d)$ and

$$lI_d \leqslant L \leqslant \nabla^2 f(\theta) \leqslant U \leqslant uI_d$$

for some $l, u > 0$ and $L, U \in \mathbb{S}^d$.

Convergence Analysis (Nonlinear)

The convergence rate is shown in the following table:

	Sampling		Optimization	
	Mean ($d=1$)	Covariance ($d=1$)	Mean ($d=1$)	Covariance (any d)
$\alpha=0$	$\left(\frac{k}{\beta}\right)^n$	$\left(\frac{k}{\beta}\right)^n$	$\lesssim \frac{\log(n)}{n}$	$\frac{\widetilde{k}_0}{\widetilde{k}_0+\beta n}$
$\alpha\in(0,1)$	$\left(\alpha+(1-\alpha^2)\frac{k}{\beta}\right)^n$	$\left(\alpha+(1-\alpha^2)\frac{k}{\beta}\right)^n$	$\lesssim n^{-1/q}$ (not optimal)	$\frac{\widetilde{k}_0+\beta}{\widetilde{k}_0+\beta+\beta(1-\alpha^2)n}$
$\alpha=1$	$\mathrm{e}^{-\left(1-\frac{2k}{\beta}\right)t}$	$\mathrm{e}^{-\left(1-\frac{2k}{\beta}\right)t}$	$\lesssim t^{-1/q}$ (not optimal)	$\frac{\widetilde{k}_0+\beta}{\widetilde{k}_0+\beta+2\beta t}$

TABLE 2: Sharp upper bounds on the convergence rates for CBS in sampling and optimization modes, in the case of a non-Gaussian target distribution and a Gaussian initial condition with strictly positive definite covariance matrix C_0. Here k is a positive constant independent of n, t, α and β, and $\widetilde{k}_0 := \left\|L^{1/2}C_0^{-1}L^{1/2}\right\|$, where L is the symmetric positive definite matrix from Assumption 1, and q is any constant strictly greater than $2\max(2,u/\ell)$, where ℓ and u are the constants from Assumption 1 and Assumption 2, respectively. Obtaining sharp convergence rates for the mean in the non-Gaussian case for $\alpha\neq 0$ in optimization mode is an open problem.

Convergence Analysis (Nonlinear)

Theorem 3.5. *Let* $\lambda = 1$, $\beta > 0$, $C_0 \in \mathcal{S}^d_{++}$, *and suppose that Assumptions 1 and 2 hold. If there exists* $\hat{\theta} \in \mathbf{R}^d$ *such that* $\mathbf{m}_n \xrightarrow[n\to\infty]{} \hat{\theta}$ *for some* $\alpha \in [0,1)$ *or* $\mathbf{m}(t) \xrightarrow[t\to\infty]{} \hat{\theta}$ *for* $\alpha = 1$, *then* $\hat{\theta} = \theta_*$ *is the minimizer of* f.

Proposition 3.7 (Convergence in the one-dimensional case)**.** *Let* $d = 1$, $\lambda = 1$, $\beta > 0$, $C_0 \in \mathcal{S}^d_{++}$, *and suppose that Assumptions 1 and 2 are satisfied. Then it holds that* $m_n \xrightarrow[n\to\infty]{} \theta_*$ *for* $\alpha \in [0,1)$ *and, likewise,* $m(t) \xrightarrow[t\to\infty]{} \theta_*$ *for* $\alpha = 1$.

Proposition 3.8 (Rate of convergence)**.** *Let* $d = 1$, $\lambda = 1$, $\beta > 0$, $\alpha = 0$, $C_0 \in \mathcal{S}^d_{++}$ *and suppose that Assumptions 1 and 2 are satisfied. Suppose additionally that* $\mathrm{e}^{-\beta f}$ *is, together with all its derivatives, bounded from above uniformly in* $\mathbf{R}$. *Then there exists a positive constant* $k = k(m_0, C_0)$ *such that, for sufficiently large* n,

$$|m_n - \theta_*| \leqslant k \left(\frac{\log n}{n} \right).$$

The convergence result holds when $f(\theta)$ is convex and $\beta > 0$ is a fixed constant.

Convergence Analysis (Nonlinear)

The proof of this result relies on the fact that the ensemble collapse to a Dirac distribution as time evolves.

Proposition 3.4 (Collapse of the ensemble in optimization mode). *Let $\lambda = 1$ and $\beta > 0$ and assume that Assumption 1 holds. Then we have*

(i) *Discrete time $\alpha = 0$. If $C_0 \in \mathcal{S}^d_{++}$, then*

$$C_n \preccurlyeq \left(\frac{\left\| L^{-1/2} C_0^{-1} L^{-1/2} \right\|}{\left\| L^{-1/2} C_0^{-1} L^{-1/2} \right\| + \beta n} \right) C_0. \tag{3.1}$$

(ii) *Discrete time $\alpha \in (0,1)$. If $C_0 \in \mathcal{S}^d_{++}$, then*

$$C_n \preccurlyeq \left(\frac{\left\| L^{-1/2} C_0^{-1} L^{-1/2} \right\| + \beta}{\left\| L^{-1/2} C_0^{-1} L^{-1/2} \right\| + \beta + \beta(1-\alpha^2) n} \right) C_0. \tag{3.2}$$

(iii) *Continuous time $\alpha = 1$. If $C(0) \in \mathcal{S}^d_{++}$, then*

$$C(t) \preccurlyeq \left(\frac{\left\| L^{-1/2} C(0)^{-1} L^{-1/2} \right\| + \beta}{\left\| L^{-1/2} C(0)^{-1} L^{-1/2} \right\| + \beta + 2\beta t} \right) C(0). \tag{3.3}$$

Convergence Analysis (Nonlinear)

Comparison with Jin's result:

Theorem 3.1. *If β, λ, σ and the initial distribution are chosen such that*

$$\begin{aligned} \mu &:= 2\lambda - \sigma^2 - 2\sigma^2 \frac{e^{-\beta L_m}}{M_L(0)} > 0, \\ \nu &:= \frac{2V(0)}{\mu M_L^2(0)} \beta e^{-2\beta L_m} c_L (2\lambda + \sigma^2) \leq \frac{3}{4}, \end{aligned} \tag{3.4}$$

then $V(t) \to 0$ exponentially fast and there exists $\tilde{x}$ such that $\bar{x}^*(t) \to \tilde{x}$, $\mathbb{E}X \to \tilde{x}$ exponentially fast. Moreover, it holds that

$$\begin{aligned} L(\tilde{x}) &\leq -\frac{1}{\beta} \log M_L(0) - \frac{1}{2\beta} \log(1-\nu) \\ &\leq L_m + r(\beta) + \frac{\log 2}{\beta}, \end{aligned}$$

where

$$r(\beta) := -\frac{1}{\beta} \log M_L(0) - L_m \to 0, \ \ \beta \to \infty.$$

The result does not require the convexity of $f(\theta)$ but requires $\beta \to \infty$.

Convergence Analysis (Nonlinear)

Theorem 3.9 (Existence of steady states). *Let* $\lambda = (1+\beta)^{-1}$, $\beta > 0$ *and* $\alpha \in [0,1]$. *Suppose Assumptions 1 and 2 are satisfied. Then there exists* $\underline{\beta}$ *such that, for all* $\beta \geqslant \underline{\beta}$, *the dynamics* (2.11) *and* (2.13) *admit a Gaussian steady state* $g\big(\bullet; \mathbf{m}_\infty(\beta), C_\infty(\beta)\big)$ *satisfying*

$$U^{-1} \preccurlyeq C_\infty(\beta) \preccurlyeq L^{-1} \quad \text{and} \quad |\mathbf{m}_\infty(\beta) - \theta_*| = \mathcal{O}\left(\frac{1}{\sqrt{\beta}}\right).$$

Convergence Analysis (Nonlinear)

- In the nonlinear case, the posterior distribution is non-Gaussian in general, but the steady state of CBS is always Gaussian. Therefore, we cannot expect the steady state approaching the exact posterior as we adjust the parameters β, α.
- In the 1D case $(d = 1)$, we are able to estimate the difference of mean and covariance between the steady state and the exact posterior distribution.

Theorem 3.10 (Convergence to the steady state). *Let $d = 1$ and $\lambda = (1+\beta)^{-1}$, and suppose Assumptions 1 and 4 hold. For any $R \in (0, C_*)$, there exists $\underline{\beta} = \underline{\beta}(f, R)$ and $k = k(f, R)$ such that the following statements hold for all $\beta \geqslant \underline{\beta}$:*

- ***Steady state.*** *There exists a pair $(m_\infty(\beta), C_\infty(\beta))$, unique in $B_R(\theta_*, C_*)$, such that the Gaussian density $\rho_\infty = g(\bullet; m_\infty, C_\infty)$ satisfies (2.15), and this pair satisfies*

$$\left| \begin{pmatrix} m_\infty(\beta) \\ C_\infty(\beta) \end{pmatrix} - \begin{pmatrix} m_* \\ C_0 \end{pmatrix} \right| \leqslant \frac{k}{\beta}.$$

  By Lemma 2.1, the density ρ_∞ is a steady state of both the iterative scheme (2.11) with any $\alpha \in [0, 1)$ and the nonlinear Fokker–Planck equation (2.13), corresponding to $\alpha = 1$.

- ***Discrete time*** $\alpha \in [0, 1)$. *If Assumption 3 holds and the moments of the initial (Gaussian) law satisfy $(m_0, C_0) \in B_R(\theta_*, C_*)$, then the solution to the iterative scheme (2.11)*

converges geometrically to the steady state ρ_∞ provided that $\alpha + (1 - \alpha^2)\frac{k}{\beta} < 1$. More precisely,

$$\forall n \in \mathbf{N}, \qquad \left| \begin{pmatrix} m_n \\ C_n \end{pmatrix} - \begin{pmatrix} m_\infty(\beta) \\ C_\infty(\beta) \end{pmatrix} \right| \leqslant \left(\alpha + (1-\alpha^2)\frac{k}{\beta} \right)^n \left| \begin{pmatrix} m_0 \\ C_0 \end{pmatrix} - \begin{pmatrix} m_\infty(\beta) \\ C_\infty(\beta) \end{pmatrix} \right|.$$

- ***Continuous time*** $\alpha = 1$. *If Assumption 3 holds and the moments of the initial (Gaussian) law satisfy $(m_0, C_0) \in B_R(\theta_*, C_*)$, then the solution to the mean field Fokker Planck equation (2.13) converges exponentially to the steady state ρ_∞ provided that $1 - \frac{2k}{\beta} > 0$. More precisely,*

$$\forall t \geqslant 0, \qquad \left| \begin{pmatrix} m(t) \\ C(t) \end{pmatrix} - \begin{pmatrix} m_\infty(\beta) \\ C_\infty(\beta) \end{pmatrix} \right| \leqslant \exp\left(-\left(1 - \frac{2k}{\beta}\right) t \right) \left| \begin{pmatrix} m_0 \\ C_0 \end{pmatrix} - \begin{pmatrix} m_\infty(\beta) \\ C_\infty(\beta) \end{pmatrix} \right|.$$

Convergence Analysis (Nonlinear)

The proof is based on the update formula of mean and covariance:

$$\begin{aligned} \mathbf{m}_{n+1} &= \alpha \mathbf{m}_n + (1-\alpha) m_\beta(\mathbf{m}_n, C_n) \\ C_{n+1} &= \alpha^2 C_n + \lambda^{-1}(1-\alpha^2) C_\beta(\mathbf{m}_n, C_n) \end{aligned} \tag{2.6}$$

where $\rho_n = \mathsf{N}(\mathbf{m}_n, C_n)$ is the distribution at the n-th timestep. Now it's useful to introduce the mapping

$$\Phi_\beta : \begin{pmatrix} m \\ C \end{pmatrix} \mapsto \begin{pmatrix} m_\beta(m, C) \\ \lambda^{-1} C_\beta(m, C) \end{pmatrix}, \qquad \lambda = (1+\beta)^{-1}$$

The convergence of $(\mathbf{m}_n, C_n)$ now relies on the contractivity of Φ_β.

Convergence Analysis (Nonlinear)

In the 1D case ($d = 1$), existence of the fixed point and the local contractivity.

Proposition 5.6 (Existence of a fixed point of Φ_β). *Let $d = 1$ and assume that Assumptions 1 and 4 hold. Then there exist $\widetilde{k} = \widetilde{k}(f)$ and $\widetilde{\beta} = \widetilde{\beta}(f)$ such that, for all $\beta \geqslant \widetilde{\beta}$, there exists a fixed point $\big(m_\infty(\beta), C_\infty(\beta)\big)$ of Φ_β satisfying*

$$|m_\infty(\beta) - \theta_*|^2 + |C_\infty(\beta) - C_*|^2 \leqslant \left|\frac{\widetilde{k}}{\beta}\right|^2.$$

Proposition 5.7 (Φ_β is a contraction). *Under the same assumptions as in Proposition 5.6 and for any $R \in (0, C_*)$, there exists a constant $\widehat{\beta} = \widehat{\beta}(f, R)$ and $\widehat{k} = \widehat{k}(f, R)$ such that, for all $\beta \geqslant \widehat{\beta}$, the map Φ_β is a contraction with constant $\widehat{k}/\beta$ for the Euclidean norm over the closed ball of radius R centered at (θ_*, C_*): for all (m_1, C_1) and (m_2, C_2) in $B_R(\theta_*, C_*)$, it holds that*

$$|\Phi_\beta(m_1, C_1) - \Phi_\beta(m_2, C_2)| \leqslant \frac{\widehat{k}}{\beta}\left|\begin{pmatrix} m_2 \\ C_2 \end{pmatrix} - \begin{pmatrix} m_1 \\ C_1 \end{pmatrix}\right|.$$

Particle Approximation

In practice, the distribution ρ_n is approxiamted as a Dirac distribution of J particles $\{\theta_n^{(j)}\}_{j=1}^J$, and the equation (2.1) becomes

$$\theta_{n+1}^{(j)} = \mathcal{M}_\beta(\rho_n^J) + \alpha(\theta_n^{(j)} - \mathcal{M}_\beta(\rho_n^J)) + \sqrt{(1-\alpha^2)\lambda^{-1}\mathcal{C}_\beta(\rho_n^J)}\boldsymbol{\xi}_n^{(j)}, \quad j = 1, \cdots, J$$

where

$$\rho_n^J := \frac{1}{J}\sum_{j=1}^J \delta_{\theta_n^{(j)}}$$

is the Dirac distribution. The continuous time dynamics is approximated as

$$\dot{\theta}^{(j)} = -(\theta^{(j)} - \mathcal{M}_\beta(\rho_t^J)) + \sqrt{2\lambda^{-1}\mathcal{C}_\beta(\rho_t)}\dot{\mathbf{W}}_t^{(j)}$$

where $\{\mathbf{W}_t^{(j)}\}_{j=1}^J$ are indepedent Brownian motions in $\mathbb{R}^d$.