How to use Sequential Monte Carlo for optimization

Haozhe Sun

November 6, 2023

Introduction to Sequential Monte Carlo (SMC)

- Sequential Monte Carlo (SMC) adapts sequentially to target densities, concentrating samples on desirable areas.
- Originated from particle filtering (Gordon et al., 1993) and used in Bayesian inference (Chopin, 2002).
- Easily applied in Bayesian statistics: transitioning from prior to posterior distribution.
- Less known in optimization contexts;

Optimization Using Sequential Monte Carlo (SMC) Sampling

- **SMC and Bayesian Inference:** SMC originally for fixed-parameter filtering, extended for Bayesian inference. Particles represent posterior distribution, leading to optimization solutions.
- **Generalizing Objectives:** Non-probability related functions transformed into positive functions, allowing SMC to sample without known norming constants.
- **Advantages of SMC:** Generic method applicable to all optimization problems, effective in high dimensions, derivative-free, and suitable for global optimization.
- **Applications and Comparisons:** SMC applied in various scenarios including offline/online optimizations, constrained/discrete problems. Upcoming comparison with MCMC methods.

Density-tempered SMC for Optimization

Importance Sampling and Resampling

- **Estimating Mode:** Using importance sampling to estimate mode of $f(\boldsymbol{\theta} \mid \mathcal{D})$, a distribution short of a norming constant.
- **Sampling Method:** Draws sample from a simple distribution $g(\boldsymbol{\theta})$. Importance weights $w_i = f\left(\boldsymbol{\theta}_i \mid \mathcal{D}\right) / g\left(\boldsymbol{\theta}_i\right)$ used to represent empirical distribution.
- **Self-Normalization:** Removes need for norming constant. Target function only needs to be nonnegative, proportional to a density or probability function.
- **Sequential Targets:** Sequence of target distributions $\left\{f_{\delta_p}(\boldsymbol{\theta} \mid \mathcal{D}), p = 0, 1, 2, \ldots\right\}$. Task is to move system from $f_{\delta_p}(\boldsymbol{\theta} \mid \mathcal{D})$ to $f(\boldsymbol{\theta} \mid \mathcal{D})$ sequentially.

Density-tempered SMC for optimization

Introduction and Setup

- Origin from Del Moral et al. (2006), and later Duan and Fulop (2013, 2015).
- Initial particle cloud $\{\boldsymbol{\theta}_i, i = 1, 2, \ldots, N\}$ from an easy-to-sample density $I(\boldsymbol{\theta})$.
- Moving to target distribution $f(\boldsymbol{\theta} \mid \mathcal{D})$ in one step is challenging.
- Use of intermediate target distributions $\left\{f_{\delta_p}(\boldsymbol{\theta} \mid \mathcal{D}), p = 0, 1, 2, \ldots\right\}$ for controlled moves.

Density Tempering Formula

- $f_{\delta_p}(\boldsymbol{\theta} \mid \mathcal{D}) \propto f(\boldsymbol{\theta} \mid \mathcal{D})^{\delta_p} I(\boldsymbol{\theta})^{1-\delta_p}$
- Sequence $\delta_0 < \delta_1 < \delta_2, \ldots$ in $[0, 1]$ can be self-adaptively chosen.

Density-tempered SMC for optimization

Effective Sample Size (ESS)

- Selecting δ_{p+1} via grid search to maintain pre-specified ESS.
- $\mathrm{ESS} = \frac{\left(\sum_{i=1}^{N} w_i\right)^2}{\sum_{i=1}^{N} w_i^2}$.

Reweighting Sequential Samples

- Reweighting to reflect added importance weights.
- $w_i^{(p+1)} = w_i^{(p)} \frac{f_{\delta_{p+1}}\left(\boldsymbol{\theta}_i^{(p)} | \mathcal{D}\right)}{f_{\delta_p}\left(\boldsymbol{\theta}_i^{(p)} | \mathcal{D}\right)}$.

Density-tempered SMC for optimization

Resampling to Reduce Weight Imbalance

- Resampling introduced by Gordon et al. (1993).
- Approaches: multinomial, residual, stratified, systematic.

Rejuvenation Step and Particle Diversity

- Resampling doesn't fundamentally solve particle degeneracy; decreases distinct particles for balanced weights.
- Rejuvenation step needed to restore particle diversity.

Introduction to Support Boosting

- First proposed by Gilks and Berzuini (2001).
- A move step added after resampling to rejuvenate particle set and boost empirical support.
- New particles proposed via a Markov chain transition kernel conditional on resampled particles.

Support Boosting in Particle Filtering

Maintaining the Underlying Target Distribution

- Markov kernel boosts sample variety without altering the target distribution.
- Input sample drawn from the same stationary distribution.

Choice of Markov Kernel and Efficiency

- The choice of Markov kernel is crucial for algorithm efficiency.
- Metropolis-Hastings kernel (Hastings, 1970; Metropolis et al., 1953) is most commonly used.
- Works as detailed in Algorithm 1.

Support Boosting in Particle Filtering

ALGORITHM 1 The Metropolis–Hastings algorithm

Given the system's current particles $\boldsymbol{\theta}$ and at the tempering value of δ_p,

Step 1. Propose $\boldsymbol{\theta}^* \sim Q(\cdot|\boldsymbol{\theta})$, where Q is a proposal sampler's distribution. It together with $f_{\delta_p}(\boldsymbol{\theta}|\mathcal{D}) \propto f(\boldsymbol{\theta}|\mathcal{D})^{\delta_p} I(\boldsymbol{\theta})^{1-\delta_p}$ satisfies the reversibility condition: $f_{\delta_p}(\boldsymbol{\theta}^*|\mathcal{D})Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = f_{\delta_p}(\boldsymbol{\theta}|\mathcal{D})Q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$.

Step 2. Compute the acceptance rate α,

$$\alpha = \min\left(1, \frac{f_{\delta_p}(\boldsymbol{\theta}^*|\mathcal{D})Q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{f_{\delta_p}(\boldsymbol{\theta}|\mathcal{D})Q(\boldsymbol{\theta}^*|\boldsymbol{\theta})}\right). \tag{4}$$

Step 3. With probability α, accept $\boldsymbol{\theta}^*$, otherwise keep the old particle.

Step 4. Repeat Steps 1–3 until some criteria are met (e.g., reaching a threshold level of cumulative acceptance rate).

Proposal Samplers and Key Steps of Density-Tempered SMC

Choice of Proposal Sampler's Density Q

- Gaussian distribution centered at current location $\boldsymbol{\theta}_i$ (random walk).
- High acceptance rate with small variance but leads to similar values (artificial boosting).
- Chopin (2002): Independent proposal sampler with/without cross correlations.
- A mixture of independent and random walk samplers beneficial for global and local exploration.

Targeting Particle Replacement

- Target at a random subvector of $\boldsymbol{\theta}$ to increase acceptance rate in high dimensions.

Key Steps of Density-Tempered SMC

ALGORITHM 2 The density-tempered SMC algorithm

Step 0. *Initialization*: generate a particle set $\boldsymbol{\theta}^{(0)}$ from an initialization distribution $I(\boldsymbol{\theta})$. The initial weight $w_i^{(0)} = 1/N$ is associated with the initial tempering factor $\delta_0 = 0$.

Step 1. *Reweighting and determining* δ_p: set $p = 1$ and search for next δ_p over a predefined grid over [0, 1] such that its corresponding ESS, computed for the reweighted particles with the incremental importance weights $w_i^{(p)} = w_i^{(p-1)} \left[\frac{f(\boldsymbol{\theta}_i|\mathcal{D})}{I(\boldsymbol{\theta}_i)}\right]^{\delta_p - \delta_{p-1}}$, is greater than $50\% \times N$. Denote the corresponding reweighted particle set as $\left\{\boldsymbol{\theta}_i^{(p)}, w_i^{(p)}\right\}_{i=1}^{N}$.

Step 2. *Resampling*: randomly draw N particles according to $\boldsymbol{w}^{(p)}$ to produce an equally weighted particle set. Denote the resampled particles as $\left\{\boldsymbol{\theta}_i^{(p),r}, 1/N\right\}_{i=1}^{N}$.

Step 3. *Support-boosting move*: propose N independent particles $\boldsymbol{\theta}^*$ and deploy the Metropolis–Hastings kernel as described in Algorithm 1 to replace $\boldsymbol{\theta}^{(p),r}$. The Markov chain transition kernel targets the density-tempered intermediate function $f_{\delta_p}(\boldsymbol{\theta}|\mathcal{D}) \propto f(\boldsymbol{\theta}|\mathcal{D})^{\delta_p} I(\boldsymbol{\theta})^{1-\delta_p}$. Denote the rejuvenated particle set as $\left\{\boldsymbol{\theta}_i^{(p),*}, 1/N\right\}_{i=1}^{N}$.

Step 4. *Loop*: if $\delta_p < 1$, set $p = p + 1$ and return to Step 1.

Step 5. The SMC optimal solution is the particle corresponding to the maximum functional value.

A Nonstatistical Example of Density-Tempered SMC

Objective

- Illustrate the use of density-tempered SMC in nonconvex optimization without involving data.

Optimization Problem

Maximize $f(\mathbf{x})$ over $\mathbf{x} \in \mathbb{R}^2$, where:

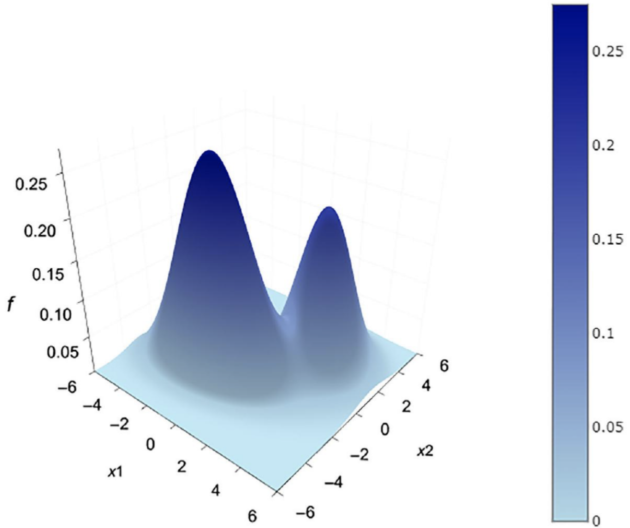
$$f(\mathbf{x}) \equiv \phi\left(\mathbf{x}; \begin{bmatrix} -1 \\ -2 \end{bmatrix}, \begin{bmatrix} 4 & 0.6 \\ 0.6 & 1 \end{bmatrix}\right) + \phi\left(\mathbf{x}; \begin{bmatrix} 2.5 \\ 2 \end{bmatrix}, \begin{bmatrix} 2.25 & -0.45 \\ -0.45 & 2.25 \end{bmatrix}\right)$$

Function Definition

Function $\phi(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is defined as:

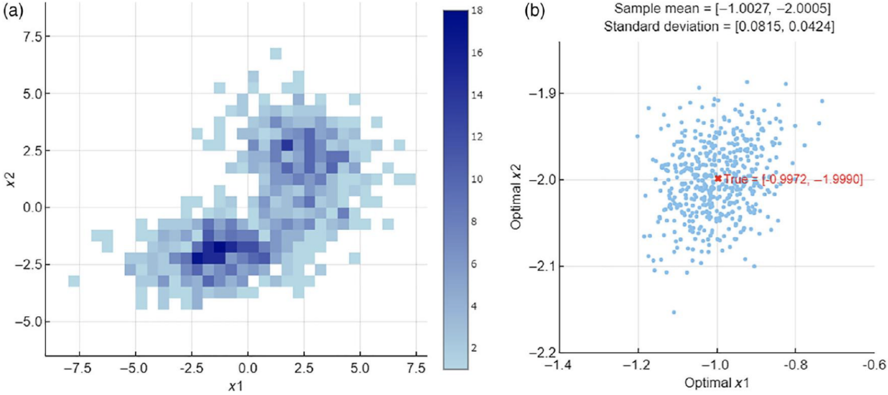
$$\phi(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{|\boldsymbol{\Sigma}|} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

Visualization



The bi-modal target function in Equation (5)

Result


(a)
7.5
5.0
2.5
0.0
−2.5
−5.0
x2
−7.5
−5.0
-2.5
0.0
2.5
5.0
7.5
x1
18
16
14
12
10
8
6
4
2
(b)
Sample mean = [−1.0027, −2.0005]
Standard deviation = [0.0815, 0.0424]
−1.9
−2.0
−2.1
−2.2
Optimal x2
True = [-0.9972, −1.9990]
−1.4
−1.2
−1.0
−0.8
-0.6
Optimal x1

Constrained Optimization Using SMC

Context

- Constrained optimization problems are prevalent in real-world applications.
- SMC techniques can handle various constraints in a straightforward manner.

Approaches to Handle Constraints

- Simple bounds: Truncated sampling distributions.
- Complex constraints: Introduce an indicator function $\chi(\mathbf{x} \in \mathcal{C})$ into the objective function.
- Algorithmic adaptation: Check for constraint satisfaction when proposing new particles.

Optimization for Discontinuous Functions

Context

- Discontinuous functions, such as multidimensional step functions, present challenges for gradient-based methods.
- SMC is well-suited for optimizing such functions.

Example: Credit Rating Optimization

- Objective: Find cutoff values for mapping probabilities of default into implied credit ratings.
- Method: Duan and Li (2021) used SMC for optimization in a credit rating context.

SMC in Credit Rating Optimization

Data and Model

- Data: S&P credit migration matrices and NUS-CRI database of PDs.
- Model: Defines eight cutoff values linked to rating buffer zones and modifiers.

Optimization Task

- Task: Optimize cutoff values to match model-generated credit migration matrices with observed data.
- Note: The solution is a nonsingleton set due to the step function nature of the objective.

k-Fold Duplication Method

Objective: Increase the efficiency and precision of SMC sampling without additional computational costs.

Method Overview:

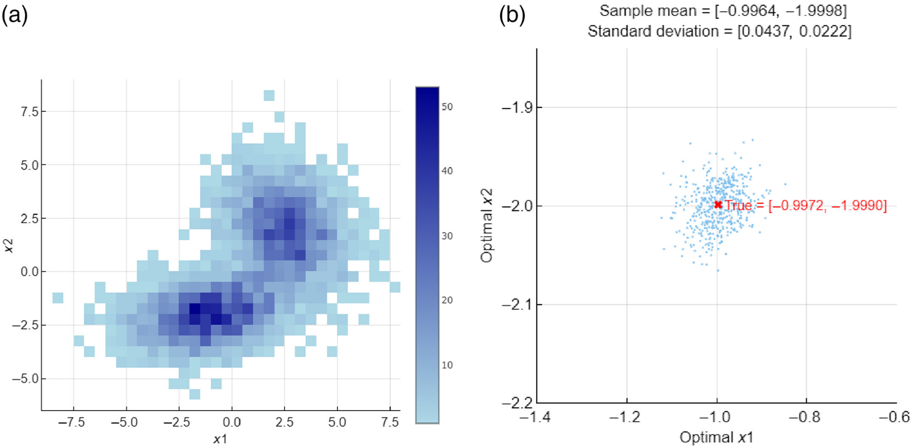
- Initially, obtain a representative SMC sample of size N.
- Duplicate this sample k times to create kN particles.
- Perform "duplicate-and-boost" steps to maintain diversity and coverage of the sample.

Advantages:

- Bypasses the need for density-tempering steps, saving computational time.
- Enhances the accuracy of the SMC sample with minimal extra computational burden.

Application:

- 4-fold duplication of a 1,000-particle SMC sample demonstrated a clear increase in the density and precision of the SMC solution.


(a)
x2
x1
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
50
40
30
20
10
(b)
Sample mean = [−0.9964, −1.9998]
Standard deviation = [0.0437, 0.0222]
Optimal x2
Optimal x1
True = [−0.9972, −1.9990]
−1.9
−2.0
−2.1
−2.2
−1.4
−1.2
−1.0
−0.8
−0.6

References

Doucet, A., de Freitas, N., & Gordon, N. (2001).
An Introduction to Sequential Monte Carlo Methods.
In *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag.

N. Kantas, A. Doucet, S.S. Singh, J.M. Maciejowski (2009).
An Overview of Sequential Monte Carlo Methods for Parameter Estimation in General State-Space Models.
IFAC Proceedings Volumes, 42(10), 774-785.
https://doi.org/10.3182/20090706-3-FR-2004.00129

Duan, J.-C., Li, S., & Xu, Y. (2023).
Sequential Monte Carlo Optimization and Statistical Inference.
WIREs Computational Statistics, 15(3), e1598.
https://doi.org/10.1002/wics.1598