SUPPLEMENTARY MATERIAL OF "SEQUENTIAL MONTE CARLO WITH GAUSSIAN MIXTURE APPROXIMATION FOR INFINITE-DIMENSIONAL STATISTICAL INVERSE PROBLEMS"

HAOYU LU, JUNXIONG JIA, AND DEYU MENG

ABSTRACT. In this supplementary material, we provide an introduction to the transition kernels used in the text, the derivation of the pCN-GM transfer kernel, some details of numerical experiments, discussions of the main text, and all of the proof details for the theorems presented in the main text.

1. FOUR TRANSITION KERNELS

In this section, we introduce four methods for the mutation step in SMC, each with a corresponding transition kernel $Q(u, dv)$ and acceptance rate $a(u, v)$. Here, u denotes the current state, and v denotes the next state. The general framework is the Metropolis-Hastings method, a renowned Markov chain Monte Carlo approach, which is described in Algorithm 1. Note that the Metropolis-Hastings method

Algorithm 1 Metropolis-Hastings method

> **Input**: Specify the initial state u_0 and the length of the Markov chain I.
> **For** $i = 1, \cdots, I$, **do**:
>   Draw v_i from a transition kernel $Q(u_{i-1}, dv)$;
>   Let $u_i = v_i$ with probability $a(u_i, v_i)$;
> **Output**: $\{u_i\}_{i=1}^I$.

can be specified by $Q(u, dv)$ and $a(u, v)$. The acceptance rate function is always calculated by

$$a(u,v) = \min\left\{1, \frac{\mu^{\mathrm{d}}(du)Q(u,dv)}{\mu^{\mathrm{d}}(dv)Q(v,du)}\right\}. \tag{1.1}$$

Thus, we only provide examples of proposal distributions:

- RW: $Q_{\mathrm{RW}}(u, dv) = u + \beta\mathcal{N}(0, \mathcal{C})$.
- pCN: $Q_{\mathrm{pCN}}(u, dv) = \sqrt{1-\beta^2}u + \beta\mathcal{N}(0, \mathcal{C})$.
- pCN-GM: $Q_{\text{pCN-GM}}(u, dv) = \sum_{i=1}^M w_i m_i + \sqrt{1-\beta^2}\sum_{i=1}^M w_i(u_i - m_i) + \beta\sum_{i=1}^M w_i\mathcal{N}(0, \mathcal{C}_i)$.

Here, β represents the step size of the transition, $\mathcal{C}$ is the covariance operator of the prior, and the parameters of the Gaussian mixture measure, $\{w_i, m_i, \mathcal{C}_i\}_{i=1}^M$, should be determined to approximate the posterior with $\sum_{i=1}^M w_i\mathcal{N}(m_i, \mathcal{C}_i)$. The final proposal using in the mutation is given by

$$Q_{\mathrm{GM}}(u, dv) = \sum_{i=1}^M w_i\mathcal{N}(m_i, \mathcal{C}_i), \ a(u, v) = 1,$$

which is an approximation of $Q_{\text{pCN-GM}}$. The four transition kernels are closely related; Q_{pCN} is a generalization of Q_{RW} to the infinite-dimensional setting, $Q_{\text{pCN-GM}}$ extends Q_{pCN} for multimodal distributions, and Q_{GM} is a special case of $Q_{\text{pCN-GM}}$ ($\beta = 1$). Note that the acceptance rate function defined by (1.1) may not be well-defined in the infinite-dimensional setting. For instance, the acceptance rate function derived from Q_{RW} is ill-defined, as discussed in Section 4.6, which outlines a necessary condition for the well-definedness of the acceptance rate function.

2. From Langevin system to pCN-GM proposal

In this section, we detail the process of deriving the pCN-GM proposal from a system of Langevin equations. Consider the Langevin equation [3]

$$\frac{du}{ds} = -\mathcal{K}(\mathcal{L}u + \gamma D\Phi(u)) + \sqrt{2\mathcal{K}}\frac{db}{ds},$$

where u represents velocity, b is a Brownian motion in $\mathcal{H}$ with the covariance operator being the identity, $\mathcal{L} = \mathcal{C}^{-1}$, and $\mathcal{K} = \mathcal{C}$ acts as a preconditioner. Set $\gamma = 0$ and consider a dynamical system of M particles with distinct distribution perturbations:

$$\begin{cases} \dfrac{du_1}{ds} = -u_1 + \sqrt{2\mathcal{C}_1}\dfrac{db}{ds}, \\ \dfrac{du_2}{ds} = -u_2 + \sqrt{2\mathcal{C}_2}\dfrac{db}{ds}, \\ \qquad \vdots \\ \dfrac{du_M}{ds} = -u_M + \sqrt{2\mathcal{C}_M}\dfrac{db}{ds}. \end{cases}$$

Let w_j denote mass, and we will examine the evolution of the average velocity

$$u = \frac{\sum_{j=1}^M w_j u_j}{\sum_{j=1}^M w_j}.$$

By multiplying each equation by its respective mass and summing, we obtain

$$\sum_{j=1}^M w_j \frac{du_j}{ds} = -\sum_{j=1}^M w_j u_j + \sum_{j=1}^M w_j \sqrt{2\mathcal{C}_j}\frac{db}{ds}.$$

Discretize the equation using the Crank-Nicolson method, it follows that

$$\begin{aligned} \sum_{j=1}^M w_j \frac{v_j - u_j}{\delta} &= -\sum_{j=1}^M w_j \frac{u_j + v_j}{2} + \sum_{j=1}^M w_j \sqrt{2\mathcal{C}_j}\frac{b(s+\delta) - b(s)}{\delta}, \\ \sum_{j=1}^M w_j (v_j - u_j) &= -\delta \sum_{j=1}^M w_j \frac{u_j + v_j}{2} + \sum_{j=1}^M w_j \sqrt{2\mathcal{C}_j}\xi, \ \xi \sim \mathcal{N}(0, \delta I). \end{aligned}$$

Rearrange the terms to obtain the following equation:

$$\begin{aligned}\sum_{j=1}^{M} w_j \left(1+\frac{1}{2}\delta\right) v_j =& \sum_{j=1}^{M} w_j \left(1-\frac{1}{2}\delta\right) u_j + \sum_{j=1}^{M} w_j \sqrt{2\delta\mathcal{C}_j}\xi,\ \xi \sim \mathcal{N}(0,\delta I),\\ (2+\delta)\sum_{j=1}^{M} w_j v_j =& (2-\delta)\sum_{j=1}^{M} w_j u_j + \sqrt{8\delta}\sum_{j=1}^{M} w_j \xi_j,\ \xi_j \sim \mathcal{N}(0,\mathcal{C}_j),\\ \sum_{j=1}^{M} w_j v_j =& \frac{2-\delta}{2+\delta}\sum_{j=1}^{M} w_j u_j + \frac{\sqrt{8\delta}}{2+\delta}\sum_{j=1}^{M} w_j \xi_j,\ \xi_j \sim \mathcal{N}(0,\mathcal{C}_j).\end{aligned}$$

Let $\beta := \frac{\sqrt{8\delta}}{2+\delta}$, then

$$\sum_{j=1}^{M} w_j v_j = \sqrt{1-\beta^2}\sum_{j=1}^{M} w_j u_j + \beta\sum_{j=1}^{M} w_j \xi_j,\ \xi_j \sim \mathcal{N}(0,\mathcal{C}_j). \tag{2.1}$$

Divide by the total mass, but continue to denote it as w_j, ensuring that $\sum_{j=1}^{M} w_j = 1$. Now, w_j has the same meaning as w_j in the main text. This approach can yield a mixed version of the pCN, i.e.,

$$v = \sqrt{1-\beta^2}u + \beta\sum_{j=1}^{M} w_j \xi_j,\ \xi_j \sim \mathcal{N}(0,\mathcal{C}_j).$$

In other words, we obtain a pCN method with mean-free Gaussian mixture measure. Subtract the mean from (2.1), and we obtain that

$$\sum_{j=1}^{M} w_j (v_j - m_j) = \sqrt{1-\beta^2}\sum_{j=1}^{M} w_j (u_j - m_j) + \beta\sum_{j=1}^{M} w_j \xi_j,\ \xi_j \sim \mathcal{N}(0,\mathcal{C}_j).$$

The evolution of the average velocity is

$$v = \sqrt{1-\beta^2}u + (1-\sqrt{1-\beta^2})\sum_{j=1}^{M} w_j m_j + \beta\sum_{j=1}^{M} w_j \xi_j,\ \xi_j \sim \mathcal{N}(0,\mathcal{C}_j),$$

which is precisely the pCN-GM transition kernel, corresponding to a Gaussian mixture $\sum_{j=1}^{M} w_j \mathcal{N}(m_j, C_j)$.

3. Some numerical details

Determine the temperatures: We need to choose the temperatures $0 = h_1 \leq h_2 \leq \cdots \leq h_{J-1} < h_J = 1$. If these temperatures are too dense, the SMC will run slowly; if they are too sparse, the adjacent posteriors μ_j and μ_{j+1} will differ significantly, which is not conducive to importance sampling [4]. We can use the effective sample size (ESS) to select appropriate temperatures:

$$\mathrm{ESS} := \left(\sum_{i=1}^{N} w_i^2\right)^{-1},$$

where w_i is the weight of the i-th particle. A small h_j can result in a large ESS. An excessively large value of h can cause the loss of a large number of samples during the resampling step, thereby leading to a small ESS. Thus, we can use a simple bisection method to find an h_j such that $\mathrm{ESS} > N_{\text{thresh}} := 0.6N$ [1].

Settings of pCN transition kernel: We should determine the step size parameter β in the proposal (2.20) of the main text to target average acceptance rates in a neighborhood of 0.2 [5, 2]. Let the average acceptance rate be denoted as α_j^N, where N represents the number of particles, and j is the index of the layer within the SMC algorithm. Then, we can employ an adaptive strategy as follows [1]:

$$\beta_{j+1} = \begin{cases} 2\beta_j, & \text{if } \alpha_j^N > 0.3, \\ 0.5\beta_j, & \text{if } \alpha_j^N < 0.15, \\ \beta_j, & \text{if } 0.15 \leq \alpha_j^N \leq 0.3. \end{cases}$$

This update formula adjusts the step size of the next layer based on the average acceptance rate of each layer, thereby maintaining an average acceptance rate as close as possible to 0.2.

4. Details of the main text

In this section, we offer detailed proofs and some discussions for all the theorems and lemmas that were mentioned in the main text. These proofs are essential for a rigorous understanding of the mathematical framework and the theoretical underpinnings of the algorithms discussed.

4.1. **Discussion of total variation distance.**

In Section 2 we use the following distance:

$$d(\mu, \nu) = \sup_{|f|_\infty \leq 1} \sqrt{\mathbb{E}_\omega \left| \int f d\mu - \int f d\nu \right|^2}. \tag{4.1}$$

Now we prove that $d(\mu, \nu)$ is a random version of total variation distance defined by

$$d_{\mathrm{TV}}(\mu, \nu) = \int \left| \frac{d\mu}{d\eta} - \frac{d\nu}{d\eta} \right| d\eta,$$

where η is a measure satisfying $\mu \ll \eta$ and $\nu \ll \eta$.

If μ and ν are not random, it follows from formula (4.1) that

$$d(\mu, \nu) = \sup_{|f|_\infty \leq 1} \left| \int f d\mu - \int f d\nu \right|.$$

Let $\eta = \frac{\mu+\nu}{2}$. Then, $\mu \ll \eta$ and $\nu \ll \eta$. Thus

$$d(\mu, \nu) = \sup_{|f|_\infty \leq 1} \left| \int f \frac{d\mu}{d\eta} d\eta - \int f \frac{d\nu}{d\eta} d\eta \right| = \sup_{|f|_\infty \leq 1} \left| \int f \left(\frac{d\mu}{d\eta} - \frac{d\nu}{d\eta} \right) d\eta \right|.$$

It is obvious that

$$d(\mu, \nu) \leq \sup_{|f|_\infty \leq 1} \int \left| f \left(\frac{d\mu}{d\eta} - \frac{d\nu}{d\eta} \right) \right| d\eta \leq \int \left| \frac{d\mu}{d\eta} - \frac{d\nu}{d\eta} \right| d\eta = d_{\mathrm{TV}}(\mu, \nu).$$

On the other hand, let $f_0 = \mathrm{sgn}(\frac{d\mu}{d\eta} - \frac{d\nu}{d\eta})$, where $\mathrm{sgn}(x)$ is the sign function. Now we have $|f_0|_\infty \leq 1$ and

$$\left| \int f_0 \left(\frac{d\mu}{d\eta} - \frac{d\nu}{d\eta} \right) d\eta \right| = \int \left| \frac{d\mu}{d\eta} - \frac{d\nu}{d\eta} \right| d\eta = d_{\mathrm{TV}}(\mu, \nu)$$

Thus, $d(\mu, \nu) = d_{\mathrm{TV}}(\mu, \nu)$ holds for non-random measures μ and ν, and we can regard $d(\mu, \nu)$ as a random version of $d_{\mathrm{TV}}(\mu, \nu)$.

4.2. **Proof of Lemma 2.5(c) (error of operator L).** Under Assumptions 2.2, let $\nu_j = P_j\mu_j$, then we have

$$d(L_j\nu_{j-1}, L_j\mu) \leq 2\kappa_1^{-1}\kappa_2^{-1} d(\nu_{j-1}, \mu).$$

Proof. Let $g_j(v) := e^{-\Phi_j(u)}$, then it follows from Assumptions 2.2(a) that $|\kappa_1 g_j(v)| \leq 1$. For simplicity we denote $g' = \kappa_1 g_j$ with $|g'| \leq 1$. Moreover, with Assumptions 2.2(c) we know that

$$\begin{aligned}
\nu_{j-1}(g_j) &= \int g_j P_{j-1}\mu_{j-1}(dm) \\
&= \int g_j \mu_{j-1}(dm) \\
&= \int e^{-\Phi_j(m)} e^{-\Phi_1(m)-\cdots-\Phi_{j-1}(m)}\mu_0(dm) \\
&\geq \int e^{-\sum_{j=1}^J \Phi_j(m)}\mu_0(dm) \\
&= \int e^{-\Phi(m)}\mu_0(dm) \\
&> \kappa_2.
\end{aligned} \tag{4.2}$$

The aim is to measure the distance between $L_j\nu_{j-1}$ and $L_j\mu$. Hence we need to compute the difference between the integral of f with respect to two measures, i.e.,

$$\begin{aligned}
(L_j\nu_{j-1})(f) - (L_j\mu)(f) &= \frac{\nu_{j-1}(fg_j)}{\nu_{j-1}(g_j)} - \frac{\mu(fg_j)}{\mu(g_j)} \\
&= \frac{\nu_{j-1}(fg_j)}{\nu_{j-1}(g_j)} - \frac{\mu(fg_j)}{\nu_{j-1}(g_j)} + \frac{\mu(fg_j)}{\nu_{j-1}(g_j)} - \frac{\mu(fg_j)}{\mu(g_j)} \\
&= \frac{\kappa_1^{-1}}{\nu_{j-1}(g_j)}[\nu_{j-1}(fg') - \mu(fg')] + \frac{\mu(fg_j)}{\mu(g_j)}\frac{\kappa_1^{-1}}{\nu_{j-1}(g_j)}[\mu(g') - \nu_{j-1}(g')].
\end{aligned}$$

Take square of both sides, and using $(a+b)^2 \leq 2(a^2+b^2)$ we obtain

$$\begin{aligned}
|(L_j\nu_{j-1})(f) - (L_j\mu)(f)|^2 \leq & \frac{2\kappa_1^{-2}}{\nu_{j-1}(g_j)^2}|\nu_{j-1}(fg') - \mu(fg')|^2 \\
& + 2\frac{\mu(fg_j)^2}{\mu(g_j)^2}\frac{\kappa_1^{-2}}{\nu_{j-1}(g_j)^2}|\mu(g') - \nu_{j-1}(g')|^2.
\end{aligned}$$

By applying inequality (4.2), we have

$$|(L_j\nu_{j-1})(f) - (L_j\mu)(f)|^2 \leq 2\kappa_1^{-2}\kappa_2^{-2}|\nu_{j-1}(fg') - \mu(fg')|^2 + 2|f|_\infty^2\kappa_1^{-2}\kappa_2^{-2}|\mu(g') - \nu_{j-1}(g')|^2.$$

Take the supremum of both sides for $|f|_\infty \le 1$. Then, it follows that

$$\begin{aligned}
d(L_j\nu_{j-1}, L_j\mu) &= \sup_{|f|_\infty \le 1} \sqrt{\mathbb{E}_\omega |(L_j\nu_{j-1})(f) - (L_j\mu)(f)|^2} \\
&\le \sup_{|f|_\infty \le 1} \sqrt{2\kappa_1^{-2}\kappa_2^{-2}\mathbb{E}_\omega|\nu_{j-1}(fg') - \mu(fg')|^2 + 2|f|_\infty^2 \kappa_1^{-2}\kappa_2^{-2}\mathbb{E}_\omega|\mu(g') - \nu_{j-1}(g')|^2} \\
&\le \sqrt{2\kappa_1^{-2}\kappa_2^{-2} \sup_{|f|_\infty \le 1} \mathbb{E}_\omega|\nu_{j-1}(fg') - \mu(fg')|^2 + 2\kappa_1^{-2}\kappa_2^{-2} d(\mu, \nu_{j-1})^2} \\
&= \sqrt{2\kappa_1^{-2}\kappa_2^{-2} d(\mu, \nu_{j-1})^2 + 2\kappa_1^{-2}\kappa_2^{-2} d(\mu, \nu_{j-1})^2} \\
&= \frac{2}{\kappa_1\kappa_2} d(\mu, \nu_{j-1}),
\end{aligned}$$

which completes the proof. □

4.3. **Proof of Theorem 2.7 (convergence of SMC).**

Proof. First we give an iterative error estimate:

$$\begin{aligned}
d(\mu_{j+1}^N, \mu_{j+1}) &= d(L_{j+1}S^N P_j \mu_j^N, L_{j+1}\mu_j) \\
\text{(Lemma 2.5}(c)\text{ in the main text)} \quad &\le 2\kappa_1^{-1}\kappa_2^{-1} d(S^N P_j \mu_j^N, \mu_j) \\
\text{(triangle inequality)} \quad &\le 2\kappa_1^{-1}\kappa_2^{-1}[d(S^N P_j \mu_j^N, P_j\mu_j^N) + d(P_j\mu_j^N, \mu_j)] \\
\text{(Lemma 2.5}(a), (b)\text{ in the main text)} \quad &\le 2\kappa_1^{-1}\kappa_2^{-1}\left[\frac{1}{\sqrt{N}} + d(\mu_j^N, \mu_j)\right].
\end{aligned}$$

Notice that $\mu_1^N = L_1 S^N P_0 \mu_0$ and the aforementioned iteration will terminate at

$$d(\mu_1^N, \mu_1) \le 2\kappa_1^{-1}\kappa_2^{-1}\left[\frac{1}{\sqrt{N}} + d(\mu_0, \mu_0)\right] = 2\kappa_1^{-1}\kappa_2^{-1}\frac{1}{\sqrt{N}}.$$

After iterating, we deduce that

$$d(\mu_J^N, \mu_J) \le \frac{1}{\sqrt{N}} \sum_{j=1}^{J} \left(\frac{2}{\kappa_1\kappa_2}\right)^j.$$

This completes the proof. □

4.4. **Proof of Theorem 2.9 (the well-definedness of $\mu_0(dv)P(v, du)$).**

Proof. The measure $\mu(du)P(u,dv)$ is defined on $\mathcal{H}\times\mathcal{H}$. Consider its characteristic function

$$\begin{aligned}
&\int_{\mathcal{H}\times\mathcal{H}} e^{i(u,\xi)}e^{i(v,\eta)}\mu(du)P(u,dv)\\
=&\int_{\mathcal{H}} e^{i(u,\xi)}\mu(du)\int_{\mathcal{H}} e^{i(v,\eta)}P(u,dv)\\
=&\int_{\mathcal{H}} e^{i(u,\xi)}\mu(du)\sum_{j=1}^{M} w_j e^{i(\gamma u+(1-\gamma)m_j,\eta)-\frac{1}{2}(\eta,\beta^2\mathcal{C}_j\eta)}\\
=&\sum_{j=1}^{M} w_j e^{i((1-\gamma)m_j,\eta)-\frac{1}{2}(\eta,\beta^2\mathcal{C}_j\eta)}\int_{\mathcal{H}} e^{i(u,\xi+\gamma\eta)}\mu(du)\\
=&\sum_{j=1}^{M} w_j e^{i((1-\gamma)m_j,\eta)-\frac{1}{2}(\eta,\beta^2\mathcal{C}_j\eta)}e^{-\frac{1}{2}(\xi+\gamma\eta,\mathcal{C}(\xi+\gamma\eta))}\\
=&\sum_{j=1}^{M} w_j e^{i((1-\gamma)m_j,\eta)-\frac{\beta^2}{2}(\eta,\mathcal{C}_j\eta)-\frac{1}{2}(\xi,\mathcal{C}\xi)-\frac{\gamma^2}{2}(\eta,\mathcal{C}\eta)-\gamma(\eta,\mathcal{C}\xi)},
\end{aligned}$$

which is the characteristic function of an M-component Gaussian mixture measure in $\mathcal{H}\times\mathcal{H}$. The mean of j-th component is $(0,(1-\gamma)m_j)$, and the covariance operator is

$$\mathcal{V}_j=\begin{bmatrix}\mathcal{C} & \gamma\mathcal{C}\\ \gamma\mathcal{C} & \beta^2\mathcal{C}_j+\gamma^2\mathcal{C}\end{bmatrix}.$$

In order to show that a Gaussian mixture measure is well-defined, we must show that each component is well-defined, that is, each $\mathcal{V}_j$ is a positive definite, self-adjoint, and trace-class operator.

It is obvious that $\mathcal{V}_j$ is self-adjoint. Then it follows from

$$((\xi,\eta),\mathcal{V}_j(\xi,\eta))=\frac{1}{2}(\eta,\beta^2\mathcal{C}_j\eta)+\frac{1}{2}(\xi+\gamma\eta,\mathcal{C}(\xi+\gamma\eta))$$

that $\mathcal{V}_j$ is positive. In order to show that $\mathcal{V}_j$ is a trace-class operator, we need to compute the eigen pairs of $\mathcal{V}_j$. Note that $\mathcal{C}_j$ shares the same eigenfunctions, we claim that $\mathcal{V}_j$ has eigenfunctions of the form $(\phi_i,t\phi_i)$, i.e.,

$$\begin{bmatrix}\mathcal{C} & \gamma\mathcal{C}\\ \gamma\mathcal{C} & \beta^2\mathcal{C}_j+\gamma^2\mathcal{C}\end{bmatrix}\begin{bmatrix}\phi_i\\ t\phi_i\end{bmatrix}=\begin{bmatrix}(1+\gamma t)\lambda_i\phi_i\\ (\gamma\lambda_i+\beta^2\lambda_{ji}t+\gamma^2\lambda_i t)\phi_i\end{bmatrix}=\begin{bmatrix}(1+\gamma t)\lambda_i\phi_i\\ (\frac{\gamma}{t}\lambda_i+\beta^2\lambda_{ji}+\gamma^2\lambda_i)t\phi_i\end{bmatrix}.$$

Hence we have

$$\lambda_i+\gamma t\lambda_i=\frac{\gamma}{t}\lambda_i+\beta^2\lambda_{ji}+\gamma^2\lambda_i, \tag{4.3}$$

which is a quadratic equation in t. Note that $\beta^2+\gamma^2=1$, the solution of equation (4.3) is

$$t^{\pm}=\frac{\beta^2}{2\gamma\lambda_i}(\lambda_{ji}-\lambda_i)\pm\sqrt{\frac{\beta^4}{4\gamma^2\lambda_i^2}(\lambda_{ji}-\lambda_i)^2+1}. \tag{4.4}$$

Thus $\mathcal{V}_j$ have eigenfunction $[\phi_i,t^{\pm}\phi_i]$, and the corresponding eigenvalue $(1+\gamma t^{\pm})\lambda_i$. Note that

$$span\{[\phi_i,t^{\pm}\phi_i]_{i=1}^{\infty}\}=span\{[\phi_i,0]_{i=1}^{\infty},[0,\phi_i]_{i=1}^{\infty}\}=\mathcal{H}\times\mathcal{H},$$

which means we have found all eigen pairs.

Now we turn to prove that the operator $\mathcal{V}_j$ is of trace class. Denote ℓ^2 as the space of square-summable sequences. It follows from the equivalence among $\{\mathcal{N}(m_j, \mathcal{C}_j)\}_{j=1}^M$ that $\mathcal{C}^{-1/2}\mathcal{C}_j\mathcal{C}^{-1/2} - I$ is a Hilbert-Schmidt operator, which means $\frac{\lambda_{ji}}{\lambda_i} - 1 \in \ell^2$. For simplicity, we denote $l_{ji} = \frac{\lambda_{ji}}{\lambda_i} - 1 \in \ell^2$, and the solutions (4.4) become

$$t^{\pm} = \frac{\beta^2}{2\gamma} l_{ji} \pm \sqrt{\frac{\beta^4}{4\gamma^2} l_{ji}^2 + 1}.$$

Moreover, the sum of eigenvalues is given by

$$\begin{aligned}
\mathrm{tr}(\mathcal{V}_j) =& \sum_i (1+\gamma t^+)\lambda_i + \sum_i (1+\gamma t^-)\lambda_i \\
=& 2\sum_i \lambda_i + \gamma \sum_i (t^+ + t^-)\lambda_i \\
=& 2\mathrm{tr}(\mathcal{C}) + \gamma \sum_i \frac{\beta^2}{\gamma} \lambda_{ji}\lambda_i \\
=& 2\mathrm{tr}(\mathcal{C}) + \beta^2 \sum_i \lambda_{ji}\lambda_i < \infty,
\end{aligned}$$

which completes the proof. □

Using the same method, it can be deduced that $\mu(dv)P(v, du)$ is a Gaussian mixture measure. As a summary, we have that they are both M-component Gaussian mixture measures, i.e.,

$$\mu_0(du)P(u, dv) = \sum_{j=1}^M w_j \mu_j, \quad \mu_j = \mathcal{N}(m_j, \mathcal{V}_j), \tag{4.5}$$

$$\mu_0(dv)P(v, du) = \sum_{j=1}^M w_j \mu_j', \quad \mu_j' = \mathcal{N}\left(m_j', \mathcal{V}_j'\right), \tag{4.6}$$

where the mean functions and covariance operators are as follows:

$$m_j = [0, (1-\gamma))m_j], \qquad \mathcal{V}_j = \begin{bmatrix} \mathcal{C} & \gamma\mathcal{C} \\ \gamma\mathcal{C} & \beta^2\mathcal{C}_j + \gamma^2\mathcal{C} \end{bmatrix}, \tag{4.7}$$

$$m_j' = [(1-\gamma)m_j, 0], \qquad \mathcal{V}_j' = \begin{bmatrix} \beta^2\mathcal{C}_j + \gamma^2\mathcal{C} & \gamma\mathcal{C} \\ \gamma\mathcal{C} & \mathcal{C} \end{bmatrix}. \tag{4.8}$$

We claim that the eigenfunction of $\mathcal{V}_j'$ is $[t\phi_i, \phi_i]^T$, then we have

$$\begin{bmatrix} \beta^2\mathcal{C}_j + \gamma^2\mathcal{C} & \gamma\mathcal{C} \\ \gamma\mathcal{C} & \mathcal{C} \end{bmatrix} \begin{bmatrix} t\phi_i \\ \phi_i \end{bmatrix} = \begin{bmatrix} (\gamma\lambda_i + \beta^2\lambda_{ji}t + \gamma^2\lambda_i t)\phi_i \\ (1+\gamma t)\lambda_i\phi_i \end{bmatrix} = \begin{bmatrix} (\frac{\gamma}{t}\lambda_i + \beta^2\lambda_{ji} + \gamma^2\lambda_i)t\phi_i \\ (1+\gamma t)\lambda_i\phi_i \end{bmatrix}.$$

Hence the eigenvalue λ_i should satisfy

$$\frac{\gamma}{t}\lambda_i + \beta^2\lambda_{ji} + \gamma^2\lambda_i = \lambda_i + \gamma t\lambda_i.$$

which is the same as equation (4.3). Therefore, we find that the eigenvalues of $\mathcal{V}_j'$ are identical to those of $\mathcal{V}_j$. As a summary, it follows that

- The eigen pairs of $\mathcal{V}_j$ are $1+\gamma t^{\pm}$ and $[\phi_i, t^{\pm}\phi_i]$.
- The eigen pairs of $\mathcal{V}_j'$ are: $1+\gamma t^{\pm}, [t^{\pm}\phi_i, \phi_i]$.

4.5. **Proof of Theorem 2.10 (the well-definedness of pCN-GM method).**

Proof. Our goal is to prove that $\mathcal{V}^{-1/2}\mathcal{V}'\mathcal{V}^{-1/2} - I$ is a Hilbert-Schmidt operator, i.e., its eigenvalues $\{\lambda_i\}_{i=1}^{\infty} \in \ell^2$. Thus, we compute the eigenvalues first. We only need to compute the eigen pairs of $\mathcal{V}^{-1/2}\mathcal{V}'\mathcal{V}^{-1/2}$, i.e., to solve the following equation:

$$\mathcal{V}'x = \eta\mathcal{V}x. \tag{4.9}$$

Note that $\mathcal{V}$ and $\mathcal{V}'$ have the same two-dimensional invariant subspace

$$\mathrm{span}\{[\phi_i, t^{\pm}\phi_i]\} = \mathrm{span}\{[\phi_i, 0], [0, \phi_i]\} = \mathrm{span}\{[t^{\pm}\phi_i, \phi_i]\}.$$

Hence, we claim that the solution to (4.9) is of the form $[a\phi_i, b\phi_i]$, and then substitute it back. It follows that

$$\begin{bmatrix} \beta^2\mathcal{C}_{j_2} + \gamma^2\mathcal{C} & \gamma\mathcal{C} \\ \gamma\mathcal{C} & \mathcal{C} \end{bmatrix} \begin{bmatrix} a\phi_i \\ b\phi_i \end{bmatrix} = \eta \begin{bmatrix} \mathcal{C} & \gamma\mathcal{C} \\ \gamma\mathcal{C} & \beta^2\mathcal{C}_{j_1} + \gamma^2\mathcal{C} \end{bmatrix} \begin{bmatrix} a\phi_i \\ b\phi_i \end{bmatrix}.$$

Currently, we have a system of equations

$$(\beta^2\lambda_{j_2 i} + \gamma^2\lambda_i)a + \gamma\lambda_i b = \eta(\lambda_i a + \gamma\lambda_i b), \tag{4.10}$$

$$\gamma\lambda_i a + \lambda_i b = \eta[\gamma\lambda_i a + (\beta^2\lambda_{j_1 i}\gamma^2\lambda_i)b]. \tag{4.11}$$

It is apparent that $b \neq 0$, thus we let $b = \gamma$ for simplicity. Moreover, let

$$l_{j_1 i} = \frac{\lambda_{j_1 i} - \lambda_i}{\lambda_i}, \quad l'_{j_2 i} = \frac{\lambda_{j_2 i} - \lambda_i}{\lambda_i},$$

then we have $l_{j_1 i}, l'_{j_2 i} \in \ell^2$ due to the equivalence among the components of the Gaussian mixture measure. Then it follows from (4.10) that

$$\begin{aligned} (\beta^2\lambda_{j_2 i} + \gamma^2\lambda_i)a + \gamma^2\lambda_i =& \eta(\lambda_i a + \gamma^2\lambda_i), \\ (\beta^2(\lambda_{j_2 i} - \lambda_i) + \lambda_i)a + \gamma^2\lambda_i =& \eta\lambda_i(a + \gamma^2), \\ (\beta^2 l'_{j_2 i} + 1)a + \gamma^2 =& \eta(a + \gamma^2), \\ (\beta^2 l'_{j_2 i} + 1 - \eta)a =& \gamma^2(\eta - 1). \end{aligned} \tag{4.12}$$

It follows from (4.11) that

$$\begin{aligned} \lambda_i a + \lambda_i =& \eta[\lambda_i a + (\beta^2\lambda_{j_1 i} + \gamma^2\lambda_i)], \\ \lambda_i(a + 1) =& \eta[\lambda_i a + (\beta^2(\lambda_{j_1 i} - \lambda_i) + \lambda_i)], \\ a + 1 =& \eta[a + (\beta^2 l_{j_1 i} + 1)], \\ (1 - \eta)a =& \eta(\beta^2 l_{j_1 i} + 1) - 1. \end{aligned} \tag{4.13}$$

Combining equations (4.12),(4.13) and eliminating a, it follows that

$$\begin{aligned} (\beta^2 l'_{j_2 i} + 1 - \eta)[\eta(\beta^2 l_{j_1 i} + 1) - 1] =& -\gamma^2(\eta - 1)^2, \\ [\eta - (\beta^2 l'_{j_2 i} + 1)][\eta(\beta^2 l_{j_1 i} + 1) - 1] =& \gamma^2(\eta - 1)^2, \end{aligned}$$

By organizing this equation, we obtain

$$\begin{aligned} (\beta^2 l_{j_1 i} + 1 - \gamma^2)\eta^2 + [-(\beta^2 l'_{j_2 i} + 1)(\beta^2 l_{j_1 i} + 1) - 1 + 2\gamma^2]\eta + \beta^2 l'_{j_2 i} + 1 - \gamma^2 =& 0, \\ \beta^2(l_{j_1 i} + 1)\eta^2 - [\beta^4 l_{j_1 i} l'_{j_2 i} + \beta^2(l_{j_1 i} + l'_{j_2 i}) + 2\beta^2]\eta + \beta^2(l'_{j_2 i} + 1) =& 0, \\ (l_{j_1 i} + 1)\eta^2 - [\beta^2 l_{j_1 i} l'_{j_2 i} + (l_{j_1 i} + l'_{j_2 i}) + 2]\eta + l'_{j_2 i} + 1 =& 0. \end{aligned} \tag{4.14}$$

Denote the solutions of equation (4.14) as η_{i+}, η_{i-}. For simplicity we compute the sum and product of the solutions. Recall that $l_{j_1 i}, l'_{j_2 i} \in \ell^2$ for $1 \leq j_1, j_2 \leq M$, thus $l_{j_1 i}, l'_{j_2 i} \to 0, i \to \infty$.

- The sum is
$$\eta_{i+} + \eta_{i-} = \frac{\beta^2 l_{j_1 i} l'_{j_2 i} + (l_{j_1 i} + l'_{j_2 i}) + 2}{l_{j_1 i} + 1} \to 2,\ i \to \infty.$$
Denote
$$l_{3i} := \eta_{i+} + \eta_{i-} - 2 = \frac{\beta^2 l_{j_1 i} l'_{j_2 i} + (l'_{j_2 i} - l_{j_1 i})}{l_{j_1 i} + 1},$$
then we have $l_{3i} \in \ell^2$.
- The product is
$$\eta_{i+}\eta_{i-} = \frac{l'_{j_2 i} + 1}{l_{j_1 i} + 1} \to 1.$$
Denote
$$l_{4i} := \eta_{i+}\eta_{i-} - 1 = \frac{l'_{j_2 i} - l_{j_1 i}}{l_{j_1 i} + 1},$$
then we have $l_{2i} \in \ell^2$.

We need to show that $\mathcal{V}^{-1/2}\mathcal{V}'\mathcal{V}^{-1/2} - I$ is a Hilbert-Schmidt operator, i.e. $\{\eta_{i+} - 1, \eta_{i-} - 1\}_{i=1}^{\infty} \in \ell^2$. Let's compute the sum in the subspace firstly:
$$\begin{aligned}(\eta_{i+} - 1)^2 + (\eta_{i-} - 1)^2 =& (\eta_{i+}^2 + \eta_{i-}^2) + 2 - 2(\eta_{i+} + \eta_{i-}) \\ =& (\eta_{i+} + \eta_{i-})^2 - 2(\eta_{i+}\eta_{i-} - 1) - 2(\eta_{i+} + \eta_{i-}) \\ =& (\eta_{i+} + \eta_{i-})(\eta_{i+} + \eta_{i-} - 2) - 2(\eta_{i+}\eta_{i-} - 1),\end{aligned}$$
Using $\eta_{i+} + \eta_{i-} = l_{3i} + 2$ and $\eta_{i+}\eta_{i-} = l_{4i} + 1$, it follows that
$$\begin{aligned}(\eta_{i+} - 1)^2 + (\eta_{i-} - 1)^2 =& l_{3i}^2 + 2l_{3i} - 2l_{4i} \\ =& l_{3i}^2 + 2\left[\frac{\beta^2 l_{j_1 i} l'_{j_2 i} + (l'_{j_2 i} - l_{j_1 i})}{l_{j_1 i} + 1} - \frac{l'_{j_2 i} - l_{j_1 i}}{l_{j_1 i} + 1}\right] \\ =& l_{3i}^2 + 2\beta^2 \frac{l_{j_1 i} l'_{j_2 i}}{l_{j_1 i} + 1},\end{aligned}$$
which is summable because $l_{j_1 i}, l'_{j_2 i}, l_{3i} \in \ell^2$, and the proof is completed. □

4.6. **Discussion of Remark 2.8 (uniqueness of pCN-GM proposal).**

Proof. This proof process is similar to that described in Section 4.5. Let $\beta^2 + \gamma^2 = I$, and we will show that I must be 1 to ensure the equivalence. The condition is that $\mathcal{V}^{-1/2}\mathcal{V}'\mathcal{V}^{-1/2} - I$ is a Hilbert-Schmidt operator, i.e., its eigenvalue $\lambda_i \in \ell^2$. To compute the eigenvalue, we solve
$$\mathcal{V}'x = \eta\mathcal{V}x.$$
We claim that the solution is of the form $[a\phi_i, b\phi_i]$. It follows that
$$\begin{bmatrix} \beta^2\mathcal{C}_{j_2} + \gamma^2\mathcal{C} & \gamma\mathcal{C} \\ \gamma\mathcal{C} & \mathcal{C} \end{bmatrix}\begin{bmatrix} a\phi_i \\ b\phi_i \end{bmatrix} = \eta\begin{bmatrix} \mathcal{C} & \gamma\mathcal{C} \\ \gamma\mathcal{C} & \beta^2\mathcal{C}_{j_1} + \gamma^2\mathcal{C} \end{bmatrix}\begin{bmatrix} a\phi_i \\ b\phi_i \end{bmatrix}.$$

Currently, we have a system of equations:

$$(\beta^2\lambda_{j_2 i}+\gamma^2\lambda_i)a+\gamma\lambda_i b=\eta(\lambda_i a+\gamma\lambda_i b), \tag{4.15}$$

$$\gamma\lambda_i a+\lambda_i b=\eta[\gamma\lambda_i a+(\beta^2\lambda_{j_1 i}\gamma^2\lambda_i)b]. \tag{4.16}$$

It is apparent that $b\neq 0$, thus we let $b=\gamma$ for simplicity. Moreover, let

$$l_{j_1 i}=\frac{\lambda_{j_1 i}-\lambda_i}{\lambda_i},\quad l'_{j_2 i}=\frac{\lambda_{j_2 i}-\lambda_i}{\lambda_i},$$

then $l_{j_1 i}, l'_{j_2 i}\in \ell^2$ due to the equivalence among components of the Gaussian mixture measure. From this step, the calculation process differs from Section 4.5. It follows from equation (4.15) that

$$\begin{aligned}
(\beta^2\lambda_{j_2 i}+\gamma^2\lambda_i)a+\gamma^2\lambda_i =&\eta(\lambda_i a+\gamma^2\lambda_i),\\
(\beta^2(\lambda_{j_2 i}-\lambda_i)+I\lambda_i)a+\gamma^2\lambda_i =&\eta\lambda_i(a+\gamma^2),\\
(\beta^2 l'_{j_2 i}+I)a+\gamma^2 =&\eta(a+\gamma^2),\\
(\beta^2 l'_{j_2 i}+I-\eta)a =&\gamma^2(\eta-1).
\end{aligned} \tag{4.17}$$

It follows from equation (4.16) that

$$\begin{aligned}
\lambda_i a+\lambda_i =&\eta[\lambda_i a+(\beta^2\lambda_{j_1 i}+\gamma^2\lambda_i)],\\
\lambda_i(a+1) =&\eta[\lambda_i a+(\beta^2(\lambda_{j_1 i}-\lambda_i)+I\lambda_i)],\\
a+1 =&\eta[a+(\beta^2 l_{j_1 i}+I)],\\
(1-\eta)a =&\eta(\beta^2 l_{j_1 i}+I)-1.
\end{aligned} \tag{4.18}$$

Combining equations (4.17) and (4.18), we then eliminate a and find that

$$\begin{aligned}
(\beta^2 l'_{j_2 i}+I-\eta)[\eta(\beta^2 l_{j_1 i}+I)-1] =&-\gamma^2(\eta-1)^2,\\
(\eta-(\beta^2 l'_{j_2 i}+I))[\eta(\beta^2 l_{j_1 i}+I)-1] =&\gamma^2(\eta-1)^2.
\end{aligned}$$

By organizing this equation, we obtain

$$\begin{aligned}
(\beta^2 l_{j_1 i}+I-\gamma^2)\eta^2+[-(\beta^2 l'_{j_2 i}+I)(\beta^2 l_{j_1 i}+I)-1+2\gamma^2]\eta+\beta^2 l'_{j_2 i}+I-\gamma^2 =&0,\\
\beta^2(l_{j_1 i}+1)\eta^2-[\beta^4 l_{j_1 i}l'_{j_2 i}+\beta^2(l_{j_1 i}+l'_{j_2 i})+I^2+1-2\gamma^2]\eta+\beta^2(l'_{j_2 i}+1) =&0,\\
(l_{j_1 i}+1)\eta^2-\left[\beta^2 l_{j_1 i}l'_{j_2 i}+(l_{j_1 i}+l'_{j_2 i})+\frac{I^2+1-2\gamma^2}{\beta^2}\right]\eta+l'_{j_2 i}+1 =&0,\\
(l_{j_1 i}+1)\eta^2-\left[\beta^2 l_{j_1 i}l'_{j_2 i}+(l_{j_1 i}+l'_{j_2 i})+\frac{(I-1)^2}{\beta^2}+2\right]\eta+l'_{j_2 i}+1 =&0.
\end{aligned} \tag{4.19}$$

Denote the solutions to equation (4.19) as η_1, η_2. We still compute the sum and product of the solutions first.

- The sum is $\eta_1+\eta_2=\frac{\beta^2 ll'+(l+l')+2+\frac{(I-1)^2}{\beta^2}}{l+1}\to 2+\frac{(I-1)^2}{\beta^2}$, and let

$$l_1 := \eta_1+\eta_2-2=\frac{\beta^2 ll'+(l'-l)+\frac{(I-1)^2}{\beta^2}}{l+1}\to\frac{(I-1)^2}{\beta^2}.$$

- The product is $\eta_1\eta_2=\frac{l'+1}{l+1}\to 1$, and let

$$l_2 := \eta_1\eta_2-1=\frac{l'-l}{l+1}\in\ell^2.$$

Denote the solutions to equation (4.19) as η_{i+}, η_{i-}. We still compute the sum and product of the solutions first.

- The sum is

$$\eta_{i+} + \eta_{i-} = \frac{\beta^2 l_{j_1 i} l'_{j_2 i} + (l_{j_1 i} + l'_{j_2 i}) + 2 + \frac{(I-1)^2}{\beta^2}}{l_{j_1 i} + 1} \to 2 + \frac{(I-1)^2}{\beta^2},$$

then let

$$l_{3i} := \eta_{i+} + \eta_{i-} - 2 = \frac{\beta^2 l_{j_1 i} l'_{j_2 i} + (l'_{j_2 i} - l_{j_1 i}) + \frac{(I-1)^2}{\beta^2}}{l_{j_1 i} + 1} \to \frac{(I-1)^2}{\beta^2}.$$

- The product is

$$\eta_{i+}\eta_{i-} = \frac{l'_{j_2 i} + 1}{l_{j_1 i} + 1} \to 1,$$

then let

$$l_{4i} := \eta_{i+}\eta_{i-} - 1 = \frac{l'_{j_2 i} - l_{j_1 i}}{l_{j_1 i} + 1} \in \ell^2.$$

To see if $\mathcal{V}^{-1/2}\mathcal{V}'\mathcal{V}^{-1/2} - I$ is a Hilbert-Schmidt operator, let's compute the sum of $(\eta - 1)^2$ in the subspace firstly:

$$\begin{aligned}
(\eta_{i+} - 1)^2 + (\eta_{i-} - 1)^2 =& (\eta_{i+} + \eta_{i-})(\eta_{i+} + \eta_{i-} - 2) - 2(\eta_{i+}\eta_{i-} - 1) \\
=& l_{3i}^2 + 2l_{3i} - 2l_{4i} \\
=& l_{3i}^2 + 2\left[\frac{\beta^2 l_{j_1 i} l'_{j_2 i} + (l'_{j_2 i} - l_{j_1 i})}{l_{j_1 i} + 1} - \frac{l'_{j_2 i} - l_{j_1 i}}{l_{j_1 i} + 1}\right] \\
=& l_{3i}^2 + 2\beta^2 \frac{l_{j_1 i} l'_{j_2 i}}{l_{j_1 i} + 1}.
\end{aligned}$$

If $I = 1$, then $l_{3i} \in \ell^2$ and the above sum is finite because $l_{j_1 i}, l'_{j_2 i} \in \ell^2$. If $I \neq 1$, then $l_{3i} \to \frac{(I-1)^2}{\beta^2} \neq 0$, the above sum is infinite.

In conclusion, $\mathcal{V}^{-1/2}\mathcal{V}'\mathcal{V}^{-1/2} - I$ is a Hilbert-Schmidt operator if and only if $\beta^2 + \gamma^2 = 1$. Consequently, if we want to use a positive-coefficient linear combination of the current state u and a Gaussian mixture, then the only choice is our pCN-GM proposal. □

4.7. **Proof of Theorem 2.12 (denseness of Gaussian mixture measures).**

Proof. We use the total variation distance

$$d(\mu, \nu) = \sup_{|f|_\infty \le 1} \left| \int f d\mu - \int f d\nu \right|.$$

First we can approximate $\int f d\mu_{post} = \int f(u) \frac{1}{Z} e^{-\Phi(u)} \mu_0(du)$ by replacing the likelihood with a finite-dimensional version. Let $Z_N = \int e^{-\Phi(u_N)} d\mu_0^N(u_N)$, it follows that

$$\begin{aligned}|Z_N - Z| &= \left|\int e^{-\Phi(u_N)} d\mu_0^N(u_N) - \int e^{-\Phi(u)} d\mu_0(u)\right| \\ &\leq \left|\int e^{-\Phi(u_N)} d\mu_0^N(u_N) - \int e^{-\Phi(u_N)} d\mu_0(u)\right| + \int |e^{-\Phi(u_N)} - e^{-\Phi(u)}| d\mu_0(u) \\ &= \int |e^{-\Phi(u_N)} - e^{-\Phi(u)}| d\mu_0(u) \to 0, N \to \infty.\end{aligned}$$

Here we need Φ to be continuous. Then we can measure the distance between μ_{post} and the posterior derived from the finite-dimensional likelihood. Using Dominated Convergence Theorem, we have

$$\begin{aligned}&\left|\int f d\mu_{post} - \int f(u)\frac{1}{Z_N}e^{-\Phi(u_N)}\mu_0(du)\right| \\ =&\left|\int f(u)\frac{1}{Z}e^{-\Phi(u)}\mu_0(du) - \int f(u)\frac{1}{Z_N}e^{-\Phi(u_N)}\mu_0(du)\right| \\ \leq&\int |f(u)| \left|\frac{1}{Z}e^{-\Phi(u)} - \frac{1}{Z_N}e^{-\Phi(u_N)}\right| \mu_0(du) \\ \leq&\int |f(u)| \left|\frac{1}{Z}(e^{-\Phi(u)} - e^{-\Phi(u_N)}) + \left(\frac{1}{Z} - \frac{1}{Z_N}\right) e^{-\Phi(u_N)}\right| \mu_0(du) \\ \leq&|f|_\infty \epsilon_1. \qquad (4.20)\end{aligned}$$

By decomposing the prior into its first N dimensions and the remaining dimensions, we obtain

$$\begin{aligned}\int f(u)\frac{1}{Z_N}e^{-\Phi(u_N)}\mu_0(du) &= \iint f(u)\frac{1}{Z_N}e^{-\Phi(u_N)}\mu_0^N(du_N)\mu_0^\perp(du_\perp) \\ &= \iint f(u)\frac{1}{Z_N}e^{-\Phi(u_N)}p(u_N)\lambda(du_N)\mu_0^\perp(du_\perp),\end{aligned}$$

where λ is the Lebesgue measure on $\mathbb{R}^N$. Note that $Z_N = \int e^{-\Phi(u_N)}p(u_N)\lambda(du_N)$, hence $\frac{1}{Z_N}e^{-\Phi(u_N)}p(u_N)$ is a density function defined on $\mathbb{R}^N$, which can be approximated by a Gaussian mixture density function. More precisely, denote the sequence of Gaussian mixture density functions as follows:

$$\begin{aligned}g_k(u_N) =&\sum_{i=1}^{I_k} w_{ki}p(u_N; m_{ki}, \Sigma_{ki}), \\ p(u; m, \Sigma) =&\frac{1}{(2\pi)^{N/2}|\det(\Sigma)|^{1/2}} \exp\left\{-\frac{1}{2}(u-m)^T\Sigma^{-1}(u-m)\right\}.\end{aligned}$$

Then it follows from Theorem 5.d in [6] that the density function can be approximated by g_k, i.e.,

$$\lim_{k\to\infty} g_k(u_N) = \frac{1}{Z_N}e^{-\Phi(u_N)}p(u_N), a.e.$$

According to the Dominated Convergence Theorem, we have

$$\begin{aligned}&\lim_{k\to\infty}\iint f(u)\left[g_k(u_N)-\frac{1}{Z_N}e^{-\Phi(u_N)}p(u_N)\right]\lambda(du_N)\mu_0^{\perp}(du_{\perp})\\ =&\iint\lim_{k\to\infty}f(u)\left[g_k(u_N)-\frac{1}{Z_N}e^{-\Phi(u_N)}p(u_N)\right]\lambda(du_N)\mu_0^{\perp}(du_{\perp})\\ =&0.\end{aligned}$$

Hence there exists a g_K such that

$$\iint f(u)\left[g_K(u_N)-\frac{1}{Z_N}e^{-\Phi(u_N)}p(u_N)\right]\lambda(du_N)\mu_0^{\perp}(du_{\perp})<\epsilon. \tag{4.21}$$

Construct a measure $\tilde{\mu}_{post}^N$ in the first N dimensions based on the density function g_K and the Lebesgue measure, and construct the product measure of $\tilde{\mu}_{post}^N$ and $\mu_0^{\perp}$, i.e.,

$$\begin{aligned}\tilde{\mu}_{post}^N(du_N):=&g_K(u_N)\lambda(du_N),\\ \tilde{\mu}_{post}(du):=&\tilde{\mu}_{post}^N(du_N)\times\mu_0^{\perp}(du_{\perp}).\end{aligned}$$

Then it follows from (4.21) that

$$\left|\int f(u)\frac{1}{Z_N}e^{-\Phi(u_N)}\mu_0(du)-\int f(u)\tilde{\mu}_{post}(du)\right|<\epsilon_2. \tag{4.22}$$

Combining the formulas (4.20) and (4.22), we can deduce that

$$\begin{aligned}&\left|\int fd\mu_{post}-\int fd\tilde{\mu}_{post}\right|\\ \le&\left|\int fd\mu_{post}-\int f(u)\frac{1}{Z_N}e^{-\Phi(u_N)}\mu_0(du)\right|+\left|\int f(u)\frac{1}{Z_N}e^{-\Phi(u_N)}\mu_0(du)-\int fd\tilde{\mu}_{post}\right|\\ <&|f|_{\infty}\epsilon_1+\epsilon_2.\end{aligned}$$

Note that ϵ_1 and ϵ_2 do not depend on f, therefore we can take the supremum of f. Consequently, it follows that

$$d(\mu_{\text{post}},\tilde{\mu}_{\text{post}})<\epsilon_1+\epsilon_2.$$

Now we find a Gaussian mixture measure, $\tilde{\mu}_{\text{post}}$, which approximates the posterior measure. Furthermore, since we only modified the first N dimensions when constructing $\tilde{\mu}_{\text{post}}$, this measure is naturally equivalent to the prior measure. □

4.8. **Proof of Theorem 2.14 (error of Gaussian mixture approximation).**

Proof. Note that the transition kernel Q is part of P. In fact we have

$$P(u,dv)=Q(u,dv)a(u,v)+\delta_u(dv)\int(1-a(u,w))Q(u,dw), \tag{4.23}$$

where

$$a(u,v)=\min\left\{1,\frac{\mu_{post}(dv)Q(v,du)}{Q(u,dv)\mu_{post}(du)}\right\}=\min\left\{1,\frac{\mu_{post}(dv)}{Q(dv)}\frac{Q(du)}{\mu_{post}(du)}\right\}.$$

For simplicity we use $Q(dv)$ instead of $Q(u,dv)$:

$$P(u,dv)=Q(dv)a(u,v)+\delta_u(dv)\int(1-a(u,w))Q(dw). \tag{4.24}$$

Now we compute the total variation distance between $Q\mu^N$ and $P\mu^N$:

$$
\begin{aligned}
&\int_U f(u)Q\mu^N(du) - \int_U f(u)P\mu^N(du)\\
=&\int_U f(u)\int_V Q(du)\mu^N(dv) - \int_U f(u)\int_V P(v,du)\mu^N(dv)\\
=&\iint_{UV} f(u)\left[1-a(v,u)\right]\mu^N(dv)Q(du) - \int_V f(v)\mu^N(dv)\int_W [1-a(v,w)]Q(dw)\\
=&\iint f(u)\left[1-a(v,u)\right]\mu^N(dv)Q(du) - \iint f(v)[1-a(v,u)]Q(du)\mu^N(dv)\\
=&\iint [f(u)-f(v)][1-a(v,u)]\mu^N(dv)Q(du).
\end{aligned}
$$

If $a = 1$, then $d_{\mathrm{TV}} = 0$ and the result is obvious; otherwise $\frac{\mu_{post}(du)}{Q(du)}\frac{Q(dv)}{\mu_{post}(dv)} < 1$ and

$$
\begin{aligned}
d(Q\mu^N, P\mu^N) &= \sup_{|f|_\infty\le 1}\left|\iint [f(u)-f(v)][1-a(v,u)]\mu^N(dv)Q(du)\right|\\
&\le \iint 2|1-a(v,u)|\left|\mu^N(dv)-\mu_{post}(dv)\right|Q(du)\\
&\quad+\sup_{|f|_\infty\le 1}\left|\iint (f(u)-f(v))(1-a(v,u))\mu_{post}(dv)Q(du)\right|\\
&\le 2d(\mu^N,\mu_{post}) + \sup_{|f|_\infty\le 1}\left|\iint (f(u)-f(v))(1-a(v,u))\mu_{post}(dv)Q(du)\right|
\end{aligned}
$$

where the boundedness of $a(u, v)$ is used. Finally, a straight computation complete the proof:

$$
\begin{aligned}
&\sup_{|f|_\infty\le 1}\left|\iint [f(u)-f(v)]\left[1-\frac{\mu_{post}(du)}{Q(du)}\frac{Q(dv)}{\mu_{post}(dv)}\right]\mu_{post}(dv)Q(du)\right|\\
=&\sup_{|f|_\infty\le 1}\left|\iint [f(u)-f(v)]\left[\mu_{post}(dv)Q(du)-\mu_{post}(du)Q(dv)\right]\right|\\
=&\sup_{|f|_\infty\le 1}\left|\int f dQ - \int f d\mu_{post} - \int f d\mu_{post} + \int f dQ\right|\\
=&2d(Q,\mu_{post}).
\end{aligned}
$$

□

References

[1] A. Beskos, A. Jasra, E. A. Muzaffer, and A. M. Stuart. Sequential monte carlo methods for bayesian elliptic inverse problems. *Statistics and Computing*, 25:727–737, 2015.

[2] Alexandros Beskos, Gareth Roberts, and Andrew Stuart. Optimal scalings for local metropolis–hastings chains on nonproduct targets in high dimensions. *The Annals of Applied Probability*, 19(3):863 – 898, 2009.

[3] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster. *Statistical Science*, 28(3):424–446, 2013.

[4] C. Dai, J. Heng, P. E. Jacob, and N. Whiteley. An invitation to sequential monte carlo samplers. *Journal of the American Statistical Association*, 117(539):1587–1600, 2022.

[5] A. Gelman, W. R. Gilks, and G. O. Roberts. Weak Convergence and Optimal Scaling of Random Walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110 – 120, 1997.

[6] T. T. Nguyen, H. D. Nguyen, F. Chamroukhi, and G. J. McLachlan. Approximation by finite mixtures of continuous density functions that vanish at infinity. *Cogent Mathematics & Statistics*, 7(1):1750861, 2020.

School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, 710049, China
Email address: luhaoyu@stu.xjtu.edu.cn

School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, 710049, China
Email address: jjx323@xjtu.edu.cn

School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, 710049, China
Email address: dymeng@mail.xjtu.edu.cn