Appendix

Roadmap.
- Section A discusses the limitations of the paper.
- Section B discusses the societal impacts of the paper.
- Section C provides the preliminary for the paper.
- Section D provides the case when we consider a continuous time score function, specifically a single Gaussian.
- Section E provides the case when we consider the score function to be 2 mixtures of Gaussians.
- Section F provides the case when we consider the score function to be k mixture of Gaussians.
- Section G provides the tools that we use from previous papers.
- Section H provides lemmas that we use for a more concrete calculation for theorems in Section G.
- Section I provides our main results when we consider the data distribution is k mixture of Gaussians.

A. Limitations

This work has not directly addressed the practical applications of our results. Additionally, we did not provide a sample complexity bound for our settings. Future research could explore how these findings might be implemented in real-world scenarios and work on improving these limitations.

B. Societal Impacts

We explore and provide a deeper understanding of the diffusion models and also explicitly give the Lipschitz constant for k-mixture of Gaussians, which may inspire a better algorithm design.

Our theoretical results have several important implications: (1) Architecture Design: Our finding that the Lipschitz constant is independent of the number of mixture components, but inversely proportional to $\sigma_{\min}$, suggests model architectures with implicit regularization on the minimum singular values could be more stable. (2) Training Guidance: As demonstrated in our experiments, understanding the relationship between covariance structures and Lipschitz constants can guide hyperparameter selection.

On the other hand, our paper is purely theoretical in nature, so we foresee no immediate negative ethical impact.

C. Preliminary

This section provides some preliminary knowledge and is organized as below:
- Section C.1 provides the facts we use.
- Section C.2 provides the property of exp function we use.
- Section C.3 provides the Lipschitz multiplication property we use.

C.1. Facts

We provide several basic facts from calculus and linear algebra that are used in the proofs.

Fact C.1 (Calculus)**.** *For $x \in \mathbb{R}$, $y \in \mathbb{R}$, $t \in \mathbb{R}$, $u \in \mathbb{R}^n$, $v \in \mathbb{R}^n$, it is well-known that*
- $\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}x}{\mathrm{d}y}\frac{\mathrm{d}y}{\mathrm{d}t}$ *(chain rule)*
- $\frac{\mathrm{d}xy}{\mathrm{d}t} = \frac{\mathrm{d}x}{\mathrm{d}t}y + \frac{\mathrm{d}y}{\mathrm{d}t}x$ *(product rule)*
- $\frac{\mathrm{d}x^n}{\mathrm{d}x} = nx^{n-1}$ *(power rule)*
- $\frac{\mathrm{d}\langle u,v\rangle}{\mathrm{d}u} = v$ *(derivative of the inner product)*
- $\frac{\mathrm{d}\exp(x)}{\mathrm{d}x} = \exp(x)$ *(derivative of exponential function)*
- $\frac{\mathrm{d}\log x}{\mathrm{d}x} = \frac{1}{x}$ *(derivative of logarithm function)*
- $\frac{\mathrm{d}}{\mathrm{d}u}\|u\|_2^2 = 2u$ *(derivative of ℓ_2 norm)*
- $\frac{\mathrm{d}y}{\mathrm{d}x} = 0$ *if y is independent from x. (derivative of independent variables)*

Fact C.2 (Norm Bounds)**.** *For $a \in \mathbb{R}$, $b \in \mathbb{R}$, $u \in \mathbb{R}^n$, $v \in \mathbb{R}^n$, $A \in \mathbb{R}^{k\times n}$, $W \in \mathbb{R}^{n\times n}$ is symmetric and p.s.d., we have*
- $\|au\|_2 = |a| \cdot \|u\|_2$ *(absolute homogeneity)*
- $\|u+v\|_2 \le \|u\|_2 + \|v\|_2$ *(triangle inequality)*
- $|u^\top v| \le \|u\|_2 \cdot \|v\|_2$ *(Cauchy–Schwarz inequality)*

- $\|u^\top\|_2 = \|u\|_2$
- $\|Au\|_2 \leq \|A\| \cdot \|u\|_2$
- $\|aA\| = |a| \cdot \|A\|$
- $\|A\| = \sigma_{\max}(A)$
- $\|A^{-1}\| = \frac{1}{\sigma_{\min}(A)}$
- $u^\top W u \geq \|u\|_2^2 \cdot \sigma_{\min}(W)$.
- $\sigma_{\min}(W^{-1}) = \frac{1}{\sigma_{\max}(W)}$.

Fact C.3 (Matrix Calculus)**.** *Let $W \in \mathbb{R}^{n \times n}$ denote a symmetric matrix. Let $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$. Suppose that s is independent of x. Then, we know*
- $\frac{\mathrm{d}}{\mathrm{d}x}(x-s)^\top W(x-s) = 2W(x-s)$

C.2. Properties of exp functions

During the course of proving the Lipschitz continuity for mixtures of Gaussians, we found that we need to use the following bound for the exp function.

Fact C.4. *For $|a-b| \leq 0.1$, where $a \in \mathbb{R}$, $b \in \mathbb{R}$, we have*

$$|\exp(a) - \exp(b)| \leq |\exp(a)| \cdot 2|a-b|$$

Proof. We have

$$\begin{aligned} |\exp(a) - \exp(b)| &= |\exp(a) \cdot (1 - \exp(b-a))| \\ &= |\exp(a)| \cdot |(1 - \exp(b-a))| \\ &\leq |\exp(a)| \cdot 2|a-b| \end{aligned}$$

where the first step follows from simple algebra, the second step follows from $|a \cdot b| = |a| \cdot |b|$, and the last step follows from $|\exp(x) - 1| \leq 2x$ for all $x \in (0, 0.1)$. □

Fact C.5. *For $\|u - v\|_\infty \leq 0.1$, where $u, v \in \mathbb{R}^n$, we have*

$$\|\exp(u) - \exp(v)\|_2 \leq \|\exp(u)\|_2 \cdot 2\|u - v\|_\infty$$

Proof. We have

$$\begin{aligned} \|\exp(u) - \exp(v)\|_2 &= \|\exp(u) \circ (\mathbf{1}_n - \exp(v-u))\|_2 \\ &\leq \|\exp(u)\|_2 \cdot \|\mathbf{1}_n - \exp(v-u)\|_\infty \\ &\leq \|\exp(u)\|_2 \cdot 2\|u - v\|_\infty \end{aligned}$$

where the first step follows from notation of Hardamard product, the second step follows from $\|u \circ v\|_2 \leq \|u\|_\infty \cdot \|v\|_2$, and the last step follows from $|\exp(x) - 1| \leq 2x$ for all $x \in (0, 0.1)$. □

Fact C.6 (Mean value theorem for vector function)**.** *For vector $x, y \in C \subset \mathbb{R}^n$, vector function $f(x) : C \to \mathbb{R}$, $g(x) : C \to \mathbb{R}^m$, let f, g be differentiable on open convex domain C, we have*
- *Part 1: $f(y) - f(x) = \nabla f(x + t(y-x))^\top (y - x)$*
- *Part 2: $\|g(y) - g(x)\|_2 \leq \|g'(x + t(y-x))\| \cdot \|y - x\|_2$ for some $t \in (0, 1)$, where $g'(a)$ denotes a matrix which its (i,j)-th term is $\frac{\mathrm{d}g(a)_j}{\mathrm{d}a_i}$.*
- *Part 3: If $\|g'(a)\| \leq M$ for all $a \in C$, then $\|g(y) - g(x)\|_2 \leq M\|y - x\|_2$ for all $x, y \in C$.*

Proof. **Proof of Part 1**

Part 1 can be verified by applying Mean Value Theorem of 1-variable function on $\gamma(c) = f(x + c(y-x))$.

$$f(y) - f(x) = \gamma(1) - \gamma(0) = \gamma'(t)(1-0) = \nabla f(x + t(y-x))^\top (y - x)$$

where $t \in (0, 1)$.

Proof of Part 2

Let $G(c) := (g(y) - g(x))^\top g(c)$, we have

$$\begin{aligned}\|g(y) - g(x)\|_2^2 &= G(y) - G(x) \\ &= \nabla G(x + t(y-x))^\top (y - x) \\ &= (\underbrace{g'(x + t(y-x))}_{n \times m} \cdot \underbrace{(g(y) - g(x))}_{m \times 1})^\top \cdot \underbrace{(y - x)}_{n \times 1} \\ &\leq \|g'(x + t(y-x))\| \cdot \|g(y) - g(x)\|_2 \cdot \|y - x\|_2\end{aligned}$$

the initial step is by basic calculation, the second step is from **Part 1**, the third step uses chain rule, the 4th step is due to Cauchy-Schwartz inequality. Removing $\|g(y) - g(x)\|_2$ on both sides gives the result.

Proof of Part 3

Part 3 directly follows from **Part 2**. □

We show the upper bound of $\exp'$ below, assuming input is bounded.

Fact C.7. *Let $g'(a)$ denotes a matrix whose (i, j)-th term is $\frac{\mathrm{d} g(a)_j}{\mathrm{d} a_i}$. For $u \in \mathbb{R}^n$, $v \in \mathbb{R}^n$, $\|u\|_2, \|v\|_2 \leq R$, where $R \geq 0$, $t \in (0, 1)$, we have*

$$\|\exp'(u + t(v - u))\| \leq \exp(R)$$

Proof. We can show

$$\begin{aligned}\|\exp'(u + t(v - u))\| &= \|\mathrm{diag}(\exp(u + t(v - u)))\| \\ &\leq \sigma_{\max}(\mathrm{diag}(\exp(u + t(v - u)))) \\ &= \max_{i \in [n]} \exp(u_i + t(v_i - u_i)) \\ &\leq \max_{i \in [n]} \max\{\exp(v_i), \exp(u_i)\} \\ &\leq \exp(R)\end{aligned}$$

where the first step follows from $\frac{\mathrm{d}\exp(x)}{\mathrm{d}x} = \mathrm{diag}(\exp(x))$, the second step follows from Fact C.2, the third step follows from spectral norm of a diagonal matrix is the absolute value of its largest entry, the fourth step follows from $t \in (0, 1)$, and the last step follows from $\|\exp(v)\|_\infty \leq \exp(\|v\|_\infty) \leq \exp(\|v\|_2)$. □

Fact C.8. *For $u \in \mathbb{R}^n$, $v \in \mathbb{R}^n$, $\|u\|_2, \|v\|_2 \leq R$, where $R \geq 0$, we have*

$$\|\exp(u) - \exp(v)\|_2 \leq \exp(R)\|u - v\|_2$$

Proof. We can show, for $t \in (0, 1)$,

$$\begin{aligned}\|\exp(u) - \exp(v)\|_2 &\leq \|\exp'(u + t(v - u))\| \cdot \|u - v\|_2 \\ &\leq \exp(R)\|u - v\|_2\end{aligned}$$

where the first step follows from Fact C.6, the second step follows from Fact C.7. □

Fact C.9. *For $a \in \mathbb{R}$, $b \in \mathbb{R}$, $a, b \leq R$, where $R \geq 0$, we have*

$$|\exp(a) - \exp(b)| \leq \exp(R)|a - b|$$

Proof. We can show, for $t \in (0, 1)$,

$$\begin{aligned}|\exp(a) - \exp(b)| &= |\exp'(a + t(b - a))| \cdot |a - b| \\ &= |\exp(a + t(b - a))| \cdot |a - b| \\ &\leq \max\{\exp(a), \exp(b)\} \cdot |a - b| \\ &\leq \exp(R)|u - v|\end{aligned}$$

where the first step follows from Mean Value Theorem, the second step follows from Fact C.1, the third step follows from $t \in (0, 1)$, and the last step follows from $a, b \leq R$.

□

C.3. Lipschitz multiplication property

Our overall proofs of Lipschitz constant for k-mixture of Gaussians follow the idea from Fact below.

Fact C.10. *If the following conditions hold*
- $\|f_i(x) - f_i(y)\|_2 \leq L \cdot \|x - y\|_2$
- $R := \max_{i \in [n], x} |f_i(x)|$

Then, we have

$$|\prod_{i=1}^k f_i(x) - \prod_{i=1}^k f_i(y)| \leq k \cdot R^{k-1} \cdot L \cdot \|x - y\|_2$$

Proof. We can show

$$\begin{aligned}
& |\prod_{i=1}^k f_i(x) - \prod_{i=1}^k f_i(y)| \\
= & |f_k(x) \prod_{i=1}^{k-1} f_i(x) - f_k(y) \prod_{i=1}^{k-1} f_i(y)| \\
\leq & |f_k(x) \prod_{i=1}^{k-1} f_i(x) - f_k(y) \prod_{i=1}^{k-1} f_i(x)| + |f_k(y) \prod_{i=1}^{k-1} f_i(x) - f_k(y) \prod_{i=1}^{k-1} f_i(y)| \\
= & |(f_k(x) - f_k(y)) \prod_{i=1}^{k-1} f_i(x)| + |f_k(y)(\prod_{i=1}^{k-1} f_i(x) - \prod_{i=1}^{k-1} f_i(y))| \\
\leq & L \cdot \|x - y\|_2 \cdot R^{k-1} + R \cdot |\prod_{i=1}^{k-1} f_i(x) - \prod_{i=1}^{k-1} f_i(y)| \\
\leq & L \cdot \|x - y\|_2 \cdot R^{k-1} + R \cdot (|L \cdot \|x - y\|_2 \cdot R^{k-2} + R \cdot |\prod_{i=1}^{k-2} f_i(x) - \prod_{i=1}^{k-2} f_i(y)|) \\
= & 2 \cdot L \cdot \|x - y\|_2 \cdot R^{k-1} + R^2 \cdot |\prod_{i=1}^{k-2} f_i(x) - \prod_{i=1}^{k-2} f_i(y)| \\
\leq & k \cdot R^{k-1} \cdot L \cdot \|x - y\|_2
\end{aligned}$$

where the first step follows from simple algebra, the second step follows from Fact C.2, the third step follows from rearranging terms, the fourth step follows from the assumptions of the lemma, the fifth step follows from the same logic of above, the sixth step follows from simple algebra, and the last step follows from the recursive process. □

D. Single Gaussian Case

In this section, we consider the continuous case of $p_t(x)$, which is the probability density function (pdf) of the input data x, and also a function of time t. More specifically, we consider the cases when $p_t(x)$ is: (1) a single Gaussian where either the mean is a function of time (Section D.1) or the covariance is a function of time (Section D.2), (2) a single Gaussian where both the mean and the covariance are a function of time (Section D.3). And then, we compute the upper bound and Lipschitz constant for the score function i.e. the gradient of log pdf $\frac{\mathrm{d} \log p_t(x)}{\mathrm{d} x}$.

D.1. Case when the mean of $p_t(x)$ is a function of time

We start our calculation by a simple case. Consider p_t such that

$$p_t(x) = \Pr_{x' \sim \mathcal{N}(t \mathbf{1}_d, I_d)} [x' = x]$$

Let pdf is $\mathbb{R}^d \to \mathbb{R}$ denote p_t. We have $\log(\mathsf{pdf}())$ is $\mathbb{R}^d \to \mathbb{R}$. Then, we can get gradient $\nabla \log(\mathsf{pdf}())$ is a function of t because of $\mathcal{N}(t\mathbf{1}_d, I_d)$. Inject x and y into the gradient function, then we are done.

Below we define the pdf for the continues case when the mean is a function of time.

Definition D.1. *If the following conditions hold*
- *Let* $x \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.

We define

$$p_t(x) := \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2}\|x - t\mathbf{1}_d\|_2^2)$$

Further, we have

$$\log p_t(x) = -\frac{d}{2}\log(2\pi) - \frac{1}{2}\|x - t\mathbf{1}_d\|_2^2$$

Below we calculate the score function of pdf for the continuous case when the mean is a function of time.

Lemma D.2. *If the following conditions hold*
- *Let* $x \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.

Then,

$$\frac{\mathrm{d}\log p_t(x)}{\mathrm{d}x} = t\mathbf{1}_d - x$$

Proof. We can show

$$\begin{aligned}
\frac{\mathrm{d}\log p_t(x)}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}(-\frac{d}{2}\log(2\pi) - \frac{1}{2}\|x - t\mathbf{1}_d\|_2^2) \\
&= -\frac{1}{2} \cdot \frac{\mathrm{d}}{\mathrm{d}x}\|x - t\mathbf{1}_d\|_2^2 \\
&= -\frac{1}{2} \cdot 2(x - t\mathbf{1}_d) \\
&= t\mathbf{1}_d - x
\end{aligned}$$

where the first step follows from Definition D.1, the second step follows from variables are independent, the third step follows from Fact C.1, and the last step follows from simple algebra. $\square$

Below we calculate the upper bound for the score function of pdf for continuous case when the mean is a function of time.

Lemma D.3 (Linear growth)**.** *If the following conditions hold*
- *Let* $x \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.

Then,

$$\|\frac{\mathrm{d}\log p_t(x)}{\mathrm{d}x}\|_2 \leq t + \|x\|_2$$

Proof. We can show

$$\begin{aligned}
\|\frac{\mathrm{d}\log p_t(x)}{\mathrm{d}x}\|_2 &= \|t\mathbf{1}_d - x\|_2 \\
&\leq \|t\mathbf{1}_d\|_2 + \|-x\|_2 \\
&\leq t + \|x\|_2
\end{aligned}$$

where the first step follows from Lemma D.2, the second step follows from Fact C.2, and the last step follows from simple algebra. $\square$

Below we calculate the Lipschitz constant for the score function of pdf for continuous case when the mean is a function of time.

Lemma D.4 (Lipschitz)**.** *If the following conditions hold*
- *Let* $x, \widetilde{x} \in \mathbb{R}^d$*.*
- *Let* $t \in \mathbb{R}$*, and* $t \geq 0$*.*

Then,

$$\|\frac{\mathrm{d}\log p_t(x)}{\mathrm{d}x} - \frac{\mathrm{d}\log p_t(\widetilde{x})}{\mathrm{d}\widetilde{x}}\|_2 = \|\widetilde{x} - x\|_2$$

Proof. We can show

$$\begin{aligned}\|\frac{\mathrm{d}\log p_t(x)}{\mathrm{d}x} - \frac{\mathrm{d}\log p_t(\widetilde{x})}{\mathrm{d}\widetilde{x}}\|_2 &= \|t\mathbf{1}_d - x - (t\mathbf{1}_d - \widetilde{x})\|_2 \\ &= \|\widetilde{x} - x\|_2\end{aligned}$$

where the first step follows from Lemma D.2, and the last step follows from simple algebra. □

D.2. Case when the covariance of $p_t(x)$ is a function of time

Below we define the pdf for continuous case when the covariance is a function of time.

Definition D.5. *If the following conditions hold*
- *Let* $x \in \mathbb{R}^d$*.*
- *Let* $t \in \mathbb{R}$*, and* $t \geq 0$*.*

We define

$$p_t(x) := \frac{1}{t^{1/2}(2\pi)^{d/2}} \exp(-\frac{1}{2t}\|x - \mathbf{1}_d\|_2^2)$$

Further, we have

$$\log p_t(x) = -\frac{1}{2}\log t - \frac{d}{2}\log(2\pi) - \frac{1}{2t}\|x - \mathbf{1}_d\|_2^2$$

Below we calculate the score function of pdf for continuous case when the covariance is a function of time.

Lemma D.6. *If the following conditions hold*
- *Let* $x \in \mathbb{R}^d$*.*
- *Let* $t \in \mathbb{R}$*, and* $t \geq 0$*.*

Then,

$$\frac{\mathrm{d}\log p_t(x)}{\mathrm{d}x} = \frac{1}{t}(\mathbf{1}_d - x)$$

Proof. We can show

$$\begin{aligned}\frac{\mathrm{d}\log p_t(x)}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}(-\frac{1}{2}\log t - \frac{d}{2}\log(2\pi) - \frac{1}{2t}\|x - \mathbf{1}_d\|_2^2) \\ &= -\frac{1}{2t} \cdot \frac{\mathrm{d}}{\mathrm{d}x}\|x - \mathbf{1}_d\|_2^2 \\ &= -\frac{1}{2t} \cdot 2(x - \mathbf{1}_d) \\ &= \frac{1}{t}(\mathbf{1}_d - x)\end{aligned}$$

where the first step follows from Definition D.5, the second step follows from variables are independent, the third step follows from Fact C.1, and the last step follows from simple algebra. □

Below we calculate the upper bound of the score function of pdf for the continuous case when the covariance is a function of time.

Lemma D.7 (Linear growth)**.** *If the following conditions hold*
- *Let* $x \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.

Then,

$$\|\frac{\mathrm{d} \log p_t(x)}{\mathrm{d}x}\|_2 \leq \frac{1}{t}(1 + \|x\|_2)$$

Proof. We can show

$$\begin{aligned}
\|\frac{\mathrm{d} \log p_t(x)}{\mathrm{d}x}\|_2 &= \|\frac{1}{t}(\mathbf{1}_d - x)\|_2 \\
&= |\frac{1}{t}| \cdot \|\mathbf{1}_d - x\|_2 \\
&= \frac{1}{t}\|\mathbf{1}_d - x\|_2 \\
&\leq \frac{1}{t}(\|\mathbf{1}_d\|_2 + \| - x\|_2) \\
&= \frac{1}{t}(1 + \|x\|_2)
\end{aligned}$$

where the first step follows from Lemma D.6, the second step follows from Fact C.2, the third step follows from $t \geq 0$, the fourth step follows from Fact C.2, and the last step follows from simple algebra. □

Below we calculate the Lipschitz constant of the score function of pdf for continuous case when the covariance is a function of time.

Lemma D.8 (Lipschitz)**.** *If the following conditions hold*
- *Let* $x, \widetilde{x} \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.

Then,

$$\|\frac{\mathrm{d} \log p_t(x)}{\mathrm{d}x} - \frac{\mathrm{d} \log p_t(\widetilde{x})}{\mathrm{d}\widetilde{x}}\|_2 = \frac{1}{t}\|x - \widetilde{x}\|_2$$

Proof. We can show

$$\begin{aligned}
\|\frac{\mathrm{d} \log p_t(x)}{\mathrm{d}x} - \frac{\mathrm{d} \log p_t(\widetilde{x})}{\mathrm{d}\widetilde{x}}\|_2 &= \|\frac{1}{t}(\mathbf{1}_d - x) - \frac{1}{t}(\mathbf{1}_d - \widetilde{x})\|_2 \\
&= \|\frac{1}{t}(\widetilde{x} - x)\|_2 \\
&= \frac{1}{t}\|x - \widetilde{x}\|_2
\end{aligned}$$

where the first step follows from Lemma D.6, the second step follows from simple algebra, the third step follows from Fact C.2. □

D.3. A general version for single Gaussian

Now we combine the previous results by calculate a slightly more complex case. Consider p_t such that

$$p_t(x) = \Pr_{x' \sim \mathcal{N}(\mu(t), \Sigma(t))}[x' = x]$$

where $\mu(t) \in \mathbb{R}^d$, $\Sigma(t) \in \mathbb{R}^{d \times d}$ and they are derivative to t and $\Sigma(t)$ is a symmetric p.s.d. matrix whose the smallest singular value is always larger than a fixed $\sigma_{\min} > 0$.

Definition D.9. *If the following conditions hold*
- *Let* $x \in \mathbb{R}^d$.

- *Let $t \in \mathbb{R}$, and $t \geq 0$.*
  We define

$$p_t(x) := \frac{1}{(2\pi)^{d/2}\det(\Sigma(t))^{1/2}} \exp(-\frac{1}{2}(x-\mu(t))^\top \Sigma(t)^{-1}(x-\mu(t))).$$

  Further, we have

$$\log p_t(x) = -\frac{d}{2}\log(2\pi) - \frac{1}{2}\log\det(\Sigma(t)) - \frac{1}{2}(x-\mu(t))^\top \Sigma(t)^{-1}(x-\mu(t))$$

Below we calculate the score function of pdf for continuous case when both the mean and covariance are a function of time.

Lemma D.10. *If the following conditions hold*
- *Let $x \in \mathbb{R}^d$.*
- *Let $t \in \mathbb{R}$, and $t \geq 0$.*

Then,

$$\frac{\mathrm{d}\log p_t(x)}{\mathrm{d}x} = -\Sigma(t)^{-1}(x-\mu(t))$$

Proof. We can show

$$\begin{aligned}
\frac{\mathrm{d}\log p_t(x)}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}(-\frac{d}{2}\log(2\pi) - \frac{1}{2}\log\det(\Sigma(t)) - \frac{1}{2}(x-\mu(t))^\top \Sigma(t)^{-1}(x-\mu(t))) \\
&= -\frac{1}{2}\cdot\frac{\mathrm{d}}{\mathrm{d}x}(x-\mu(t))^\top \Sigma(t)^{-1}(x-\mu(t)) \\
&= -\frac{1}{2}\cdot 2\Sigma(t)^{-1}(x-\mu(t)) \\
&= -\Sigma(t)^{-1}(x-\mu(t))
\end{aligned}$$

where the first step follows from Definition D.9, the second step follows from Fact C.1, the third step follows from Fact C.3, and the last step follows from simple algebra. □

Below we calculate the upper bound of the score function of pdf for continuous case when both the mean and covariance is a function of time.

Lemma D.11 (Linear growth)**.** *If the following conditions hold*
- *Let $x \in \mathbb{R}^d$.*
- *Let $t \in \mathbb{R}$, and $t \geq 0$.*

Then,

$$\|\frac{\mathrm{d}\log p_t(x)}{\mathrm{d}x}\|_2 \leq \frac{1}{\sigma_{\min}(\Sigma(t))}\cdot(\|\mu(t)\|_2 + \|x\|_2)$$

Proof. We can show

$$\begin{aligned}
\|\frac{\mathrm{d}\log p_t(x)}{\mathrm{d}x}\|_2 &= \|-\Sigma(t)^{-1}(x-\mu(t))\|_2 \\
&\leq \|-\Sigma(t)^{-1}\|\cdot\|x-\mu(t)\|_2 \\
&= \|\Sigma(t)^{-1}\|\cdot\|x-\mu(t)\|_2 \\
&= \frac{1}{\sigma_{\min}(\Sigma(t))}\cdot\|x-\mu(t)\|_2 \\
&\leq \frac{1}{\sigma_{\min}(\Sigma(t))}\cdot(\|x\|_2 + \|-\mu(t)\|_2) \\
&= \frac{1}{\sigma_{\min}(\Sigma(t))}\cdot(\|\mu(t)\|_2 + \|x\|_2)
\end{aligned}$$

where the first step follows from Lemma D.10, the second step follows from Fact C.2, the third step follows from Fact C.2, the fourth step follows from Fact C.2, the fifth step follows from Fact C.2, and the last step follows from Fact C.2. □

Below we calculate the Lipschitz constant of the score function of pdf for continuous case when both the mean and covariance are a function of time.

Lemma D.12 (Lipschitz)**.** *If the following conditions hold*
- *Let* $x, \widetilde{x} \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.

Then,

$$\| \frac{\mathrm{d} \log p_t(x)}{\mathrm{d} x} - \frac{\mathrm{d} \log p_t(\widetilde{x})}{\mathrm{d} \widetilde{x}} \|_2 \leq \frac{1}{\sigma_{\min}(\Sigma(t))} \cdot \|x - \widetilde{x}\|_2$$

Proof. We can show

$$\begin{aligned}
\| \frac{\mathrm{d} \log p_t(x)}{\mathrm{d} x} - \frac{\mathrm{d} \log p_t(\widetilde{x})}{\mathrm{d} \widetilde{x}} \|_2 &= \| -\Sigma(t)^{-1}(x - \mu(t)) - (-\Sigma(t)^{-1}(\widetilde{x} - \mu(t))) \|_2 \\
&= \| -\Sigma(t)^{-1}(x - \widetilde{x}) \|_2 \\
&\leq \| -\Sigma(t)^{-1} \| \cdot \|x - \widetilde{x}\|_2 \\
&= \frac{1}{\sigma_{\min}(\Sigma(t))} \cdot \|x - \widetilde{x}\|_2
\end{aligned}$$

where the first step follows from Lemma D.10, the second step follows from simple algebra, the third step follows from Fact C.2, and the last step follows from Fact C.2. □

E. A General Version for Two Gaussian

In this section, we compute the linear growth and Lipschitz constant for a mixture of 2 Gaussian where both the mean and covariance are a function of time. The organization of this section is as follows:
- Section E.1 defines the probability density function (pdf) $p_t(x)$ that we use, which is a mixture of 2 Gaussian.
- Section E.2 provides lemmas that are used for calculation of the score function i.e. gradient of the log pdf $\frac{\mathrm{d} \log p_t(x)}{\mathrm{d} x}$.
- Section E.3 provides the expression of the score function.
- Section E.4 provides lemmas that are used for calculation of the upper bound of the score function.
- Section E.5 provides the expression of the upper bound of the score function.
- Section E.6 provides lemmas of upper bound for some base functions that are used for calculation of the Lipschitz constant of the score function.
- Section E.7 provides lemmas of Lipschitz constant for some base functions that are used for calculation of the Lipschitz constant of the score function.
- Section E.8 provides lemmas of Lipschitz constant for $f(x)$ that are used for calculation of the Lipschitz constant of the score function.
- Section E.9 provides lemmas of Lipschitz constant for $g(x)$ that are used for calculation of the Lipschitz constant of the score function.
- Section E.10 provides the expression of the Lipschitz constant of the score function.

First, we define the following. Let $\alpha(t) \in (0, 1)$ and also is a function of time t. Consider p_t such that

$$p_t(x) = \Pr_{x' \sim \alpha(t)\mathcal{N}(\mu_1(t), \Sigma_1(t)) + (1-\alpha(t)))\mathcal{N}(\mu_2(t), \Sigma_2(t))}[x' = x]$$

where $\mu_1(t), \mu_2(t) \in \mathbb{R}^d$, $\Sigma_1(t), \Sigma_2(t) \in \mathbb{R}^{d \times d}$ and they are derivative to t and $\Sigma_1(t), \Sigma_2(t)$ is a symmetric p.s.d. matrix whose the smallest singular value is always larger than a fixed value $\sigma_{\min} > 0$.

For further simplicity of calculation, we denote $\alpha(t)$ to be α.

E.1. Definitions

Below we define function N_1 and N_2.

Definition E.1. *If the following conditions hold*
- *Let* $x \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.

We define

$$N_1(x) := \frac{1}{(2\pi)^{d/2}\det(\Sigma_1(t))^{1/2}} \exp(-\frac{1}{2}(x-\mu_1(t))^\top \Sigma_1(t)^{-1}(x-\mu_1(t)))$$

and

$$N_2(x) := \frac{1}{(2\pi)^{d/2}\det(\Sigma_2(t))^{1/2}} \exp(-\frac{1}{2}(x-\mu_2(t))^\top \Sigma_2(t)^{-1}(x-\mu_2(t)))$$

It's clearly to see that $N_i \leq \frac{1}{(2\pi)^{d/2}\det(\Sigma_i(t))^{1/2}}$ *since* $N_i(x)$ *takes maximum when* $x = \mu_i$.

Below we define the pdf for 2 mixtures of Gaussians.

Definition E.2. *If the following conditions hold*

- *Let* $x \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.
- *Let* $\alpha \in \mathbb{R}$ *and* $\alpha \in (0,1)$.
- *Let* $N_1(x), N_2(x)$ *be defined as Definition E.1.*

We define

$$\begin{aligned} p_t(x) :=& \frac{\alpha}{(2\pi)^{d/2}\det(\Sigma_1(t))^{1/2}} \exp(-\frac{1}{2}(x-\mu_1(t))^\top \Sigma_1(t)^{-1}(x-\mu_1(t))) + \\ & \frac{1-\alpha}{(2\pi)^{d/2}\det(\Sigma_2(t))^{1/2}} \exp(-\frac{1}{2}(x-\mu_2(t))^\top \Sigma_2(t)^{-1}(x-\mu_2(t))). \end{aligned}$$

This can be further rewritten as follows:

$$p_t(x) = \alpha N_1(x) + (1-\alpha) N_2(x)$$

Further, we have

$$\log p_t(x) = \log(\alpha N_1(x) + (1-\alpha) N_2(x))$$

E.2. Lemmas for calculation of the score function

This subsection describes lemmas that are used for further calculation of the score function.

This lemma calculates the gradient of function N_i.

Lemma E.3. *If the following conditions hold*

- *Let* $x \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.
- *Let* $\alpha \in \mathbb{R}$ *and* $\alpha \in (0,1)$.
- *Let* $N_1(x), N_2(x)$ *be defined as Definition E.1.*

Then, for $i \in \{1,2\}$, *we have*

$$\frac{\mathrm{d}N_i(x)}{\mathrm{d}x} = N_i(x)(-\Sigma_i(t)^{-1}(x-\mu_i(t)))$$

Proof. We can show

$$\begin{aligned} \frac{\mathrm{d}N_i(x)}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}(\frac{1}{(2\pi)^{d/2}\det(\Sigma_i(t))^{1/2}} \exp(-\frac{1}{2}(x-\mu_i(t))^\top \Sigma_i(t)^{-1}(x-\mu_i(t)))) \\ &= N_i(x) \cdot \frac{\mathrm{d}}{\mathrm{d}x}(-\frac{1}{2}(x-\mu_i(t))^\top \Sigma_i(t)^{-1}(x-\mu_i(t))) \\ &= N_i(x)(-\frac{1}{2} \cdot 2\Sigma_i(t)^{-1}(x-\mu_i(t))) \\ &= N_i(x)(-\Sigma_i(t)^{-1}(x-\mu_i(t))) \end{aligned}$$

where the first step follows from Definition E.1, the second step follows from Fact C.1, the third step follows from Fact C.3, and the last step follows from simple algebra. ☐

This lemma calculates the gradient of function $p_t(x)$.

Lemma E.4. *If the following conditions hold*

- *Let* $x \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$*, and* $t \geq 0$.
- *Let* $\alpha \in \mathbb{R}$ *and* $\alpha \in (0, 1)$.
- *Let* $p_t(x)$ *be defined as Definition E.2.*
- *Let* $N_1(x), N_2(x)$ *be defined as Definition E.1.*

Then,

$$\frac{\mathrm{d}p_t(x)}{\mathrm{d}x} = \alpha N_1(x)(-\Sigma_1(t)^{-1}(x - \mu_1(t))) + (1 - \alpha)N_2(x)(-\Sigma_2(t)^{-1}(x - \mu_2(t)))$$

Proof. We can show

$$\begin{aligned}\frac{\mathrm{d}p_t(x)}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}(\alpha N_1(x) + (1 - \alpha)N_2(x)) \\ &= \alpha \frac{\mathrm{d}}{\mathrm{d}x} N_1(x) + (1 - \alpha)\frac{\mathrm{d}}{\mathrm{d}x} N_2(x) \\ &= \alpha N_1(x)(-\Sigma_1(t)^{-1}(x - \mu_1(t))) + (1 - \alpha)N_2(x)(-\Sigma_2(t)^{-1}(x - \mu_2(t)))\end{aligned}$$

where the first step follows from Definition E.2, the second step follows from Fact C.1, and the last step follows from Lemma E.3. □

E.3. Calculation of the score function

Below we define $f(x)$ and $g(x)$ that simplify further calculation.

Definition E.5. *For further simplicity, we define the following functions:*

If the following conditions hold

- *Let* $x \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$*, and* $t \geq 0$.
- *Let* $\alpha \in \mathbb{R}$ *and* $\alpha \in (0, 1)$.
- *Let* $N_1(x), N_2(x)$ *be defined as Definition E.1.*

We define

$$f(x) := \frac{\alpha N_1(x)}{\alpha N_1(x) + (1 - \alpha)N_2(x)}$$

and

$$g(x) := \frac{(1 - \alpha)N_2(x)}{\alpha N_1(x) + (1 - \alpha)N_2(x)}$$

And it's clearly to see that $0 \leq f(x) \leq 1$*,* $0 \leq g(x) \leq 1$ *and* $f(x) + g(x) = 1$.

This lemma calculates the score function.

Lemma E.6. *If the following conditions hold*

- *Let* $x \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$*, and* $t \geq 0$.
- *Let* $\alpha \in \mathbb{R}$ *and* $\alpha \in (0, 1)$.
- *Let* $p_t(x)$ *be defined as Definition E.2.*
- *Let* $N_1(x), N_2(x)$ *be defined as Definition E.1.*
- *Let* $f(x), g(x)$ *be defined as Definition E.5.*

Then,

$$\frac{\mathrm{d}\log p_t(x)}{\mathrm{d}x} = \frac{\alpha N_1(x)(-\Sigma_1(t)^{-1}(x - \mu_1(t)))}{\alpha N_1(x) + (1 - \alpha)N_2(x)} + \frac{(1 - \alpha)N_2(x)(-\Sigma_2(t)^{-1}(x - \mu_2(t)))}{\alpha N_1(x) + (1 - \alpha)N_2(x)}$$

Proof. We can show

$$\begin{aligned}
\frac{\mathrm{d}\log p_t(x)}{\mathrm{d}x} &= \frac{\mathrm{d}\log p_t(x)}{\mathrm{d}p_t(x)}\frac{\mathrm{d}p_t(x)}{\mathrm{d}x} \\
&= \frac{1}{p_t(x)}\frac{\mathrm{d}p_t(x)}{\mathrm{d}x} \\
&= \frac{1}{p_t(x)}(\alpha N_1(x)(-\Sigma_1(t)^{-1}(x-\mu_1(t))) + (1-\alpha)N_2(x)(-\Sigma_2(t)^{-1}(x-\mu_2(t)))) \\
&= \frac{\alpha N_1(x)(-\Sigma_1(t)^{-1}(x-\mu_1(t)))}{\alpha N_1(x)+(1-\alpha)N_2(x)} + \frac{(1-\alpha)N_2(x)(-\Sigma_2(t)^{-1}(x-\mu_2(t)))}{\alpha N_1(x)+(1-\alpha)N_2(x)} \\
&= f(x)(-\Sigma_1(t)^{-1}(x-\mu_1(t)) + g(x)(-\Sigma_2(t)^{-1}(x-\mu_2(t))
\end{aligned}$$

where the first step follows from Fact C.1, the second step follows from Fact C.1, the third step follows from Lemma E.4, the fourth step follows from Definition E.2 and the last step follows from Definition E.5.

□

E.4. Lemmas for the calculation of the upper bound of the score function

This section provides lemmas that are used in calculation of upper bound of the score function.

This lemma calculates the upper bound of function $\|-\Sigma_i(t)^{-1}(x-\mu_i(t))\|_2$.

Lemma E.7. *If the following conditions hold*

- *Let $x \in \mathbb{R}^d$.*
- *Let $t \in \mathbb{R}$, and $t \geq 0$.*

Then, for each $i \in \{1,2\}$, we have

$$\|-\Sigma_i(t)^{-1}(x-\mu_i(t))\|_2 \leq \frac{1}{\sigma_{\min}(\Sigma_i(t))}\cdot(\|x\|_2 + \|\mu_i(t)\|_2)$$

Proof. We can show

$$\begin{aligned}
\|-\Sigma_i(t)^{-1}(x-\mu_i(t))\|_2 &\leq \|-\Sigma_i(t)^{-1}\|\cdot\|x-\mu_i(t)\|_2 \\
&= \|\Sigma_i(t)^{-1}\|\cdot\|x-\mu_i(t)\|_2 \\
&= \frac{1}{\sigma_{\min}(\Sigma_i(t))}\cdot\|x-\mu(t)\|_2 \\
&\leq \frac{1}{\sigma_{\min}(\Sigma_i(t))}\cdot(\|x\|_2 + \|-\mu_i(t)\|_2) \\
&= \frac{1}{\sigma_{\min}(\Sigma_i(t))}\cdot(\|x\|_2 + \|\mu_i(t)\|_2)
\end{aligned}$$

where the first step follows from Fact C.2, the second step follows from Fact C.2, the third step follows from Fact C.2, the fourth step follows from Fact C.2, and the last step follows from simple algebra. □

E.5. Upper bound of the score function

This lemma calculates the upper bound of the score function.

Lemma E.8 (Linear growth)**.** *If the following conditions hold*

- *Let $x \in \mathbb{R}^d$.*
- *Let $t \in \mathbb{R}$, and $t \geq 0$.*
- *Let $\alpha \in \mathbb{R}$ and $\alpha \in (0,1)$.*
- *Let $p_t(x)$ be defined as Definition E.2.*
- *Let $f(x), g(x)$ be defined as Definition E.5.*
- *Let $\sigma_{\min} := \min\{\sigma_{\min}(\Sigma_1(t)), \sigma_{\min}(\Sigma_2(t))\}$.*
- *Let $\mu_{\max} := \max\{\|\mu_1(t)\|_2, \|\mu_2(t)\|_2, 1\}$.*

Then,

$$\|\frac{\mathrm{d}\log p_t(x)}{\mathrm{d}x}\|_2 \leq \sigma_{\min}^{-1} \cdot \mu_{\max} \cdot (1 + \|x\|_2)$$

Proof. We can show

$$\begin{aligned}
\|\frac{\mathrm{d}\log p_t(x)}{\mathrm{d}x}\|_2 &= \|f(x)(-\Sigma_1(t)^{-1}(x-\mu_1(t)) + g(x)(-\Sigma_2(t)^{-1}(x-\mu_2(t))\|_2 \\
&\leq \|f(x)(-\Sigma_1(t)^{-1}(x-\mu_1(t))\|_2 + \|g(x)(-\Sigma_2(t)^{-1}(x-\mu_2(t))_2\|_2 \\
&\leq \max_{i\in[2]} \| - \Sigma_i(t)^{-1}(x-\mu_i(t))\|_2 \\
&\leq \max_{i\in[2]}(\frac{1}{\sigma_{\min}(\Sigma_i(t))} \cdot (\|x\|_2 + \|\mu_i(t)\|_2)) \\
&\leq \sigma_{\min}^{-1}(\mu_{\max} + \|x\|_2) \\
&\leq \sigma_{\min}^{-1} \cdot \mu_{\max} \cdot (1 + \|x\|_2)
\end{aligned}$$

where the first step follows from Lemma E.6, the second step follows from Fact C.2, the third step follows from $f(x)+g(x) = 1$ and $f(x), g(x) \geq 0$, the fourth step follows from Lemma E.7, the fifth step follows from definition of $\mu_{\max}$ and $\sigma_{\min}$, and the last step follows from $\mu_{\max} \geq 1$. □

E.6. Lemmas for Lipschitz calculation: upper bound of base functions

This section provides the lemmas of bounds of base functions that are used in calculation of Lipschitz of the score function.

This lemma calculate the upper bound of the function $\| - \Sigma_i(t)^{-1}(x-\mu_i(t))\|_2$.

Lemma E.9. *If the following conditions hold*

- *Let* $x \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$*, and* $t \geq 0$.
- *Let* $\|x - \mu_i(t)\|_2 \leq R$*, where* $R \geq 1$*, for each* $i \in \{1, 2\}$.

Then, for each $i \in \{1, 2\}$*, we have*

$$\| - \Sigma_i(t)^{-1}(x-\mu_i(t))\|_2 \leq \frac{R}{\sigma_{\min}(\Sigma_i(t))}$$

Proof. We can show

$$\begin{aligned}
\| - \Sigma_i(t)^{-1}(x-\mu_i(t))\|_2 &\leq \| - \Sigma_i(t)^{-1}\| \cdot \|x - \mu_i(t)\|_2 \\
&= \frac{1}{\sigma_{\min}(\Sigma_i(t))} \cdot \|x - \mu_i(t)\|_2 \\
&\leq \frac{R}{\sigma_{\min}(\Sigma_i(t))}
\end{aligned}$$

where the first step follows from Fact C.2, the second step follows from Fact C.2, and the last step follows from $\|x-\mu_i(t)\|_2 \leq R$. □

This lemma calculate the lower bound of the function $(x-\mu_i(t))^\top \Sigma_i(t)^{-1}(x-\mu_i(t))$.

Lemma E.10. *If the following conditions hold*

- *Let* $x \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$*, and* $t \geq 0$.
- *Let* $\|x - \mu_i(t)\|_2 \leq R$*, where* $R \geq 1$*, for each* $i \in \{1, 2\}$.
- *Let* $\|x - \mu_i(t)\|_2 \geq \beta$*, where* $\beta \in (0, 0.1)$*, for each* $i \in \{1, 2\}$.

Then,

$$(x-\mu_i(t))^\top \Sigma_i(t)^{-1}(x-\mu_i(t)) \geq \frac{\beta^2}{\sigma_{\max}(\Sigma_i(t))}$$

Proof. We can show

$$\begin{aligned}\text{LHS} &\geq \|x-\mu_i(t)\|_2^2 \cdot \sigma_{\min}(\Sigma_i(t)^{-1}) \\ &= \|x-\mu_i(t)\|_2^2 \cdot \frac{1}{\sigma_{\max}(\Sigma_i(t))} \\ &\geq \frac{\beta^2}{\sigma_{\max}(\Sigma_i(t))}\end{aligned}$$

where the first step follows from Fact C.2, the second step follows from Fact C.2, and the last step follows from $\|x-\mu_i(t)\|_2 \geq \beta$. □

This lemma calculate the upper bound of the function $\exp(-\frac{1}{2}(x-\mu_i(t))^\top \Sigma_i(t)^{-1}(x-\mu_i(t)))$.

Lemma E.11. *If the following conditions hold*

- *Let* $x \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.
- *Let* $\|x-\mu_i(t)\|_2 \leq R$, *where* $R \geq 1$, *for each* $i \in \{1,2\}$.
- *Let* $\|x-\mu_i(t)\|_2 \geq \beta$, *where* $\beta \in (0, 0.1)$, *for each* $i \in \{1,2\}$.

Then,

$$\exp(-\frac{1}{2}(x-\mu_i(t))^\top \Sigma_i(t)^{-1}(x-\mu_i(t))) \leq \exp(-\frac{\beta^2}{2\sigma_{\max}(\Sigma_i(t))})$$

Proof. We can show

$$\begin{aligned}\text{LHS} = &\ \exp(-\frac{1}{2}(x-\mu_i(t))^\top \Sigma_i(t)^{-1}(x-\mu_i(t))) \\ \leq &\ \exp(-\frac{\beta^2}{2\sigma_{\max}(\Sigma_i(t))})\end{aligned}$$

where the first step follows from Fact C.2, the second step follows from Lemma E.10. □

E.7. Lemmas for Lipschitz calculation: Lipschitz constant of base functions

This section provides the lemmas of Lipschitz constant of base functions that are used in calculation of Lipschitz of the score function.

This lemma calculates Lipschitz constant of function $\|-\Sigma_i(t)^{-1}(x-\mu_i(t)) - (-\Sigma_i(t)^{-1}(\widetilde{x}-\mu_i(t)))\|_2$.

Lemma E.12. *If the following conditions hold*

- *Let* $x, \widetilde{x} \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.

Then, for $i \in \{1,2\}$, *we have*

$$\|-\Sigma_i(t)^{-1}(x-\mu_i(t)) - (-\Sigma_i(t)^{-1}(\widetilde{x}-\mu_i(t)))\|_2 \leq \frac{1}{\sigma_{\min}(\Sigma_i(t))} \cdot \|x-\widetilde{x}\|_2$$

Proof. We can show

$$\begin{aligned}\text{LHS} &= \|-\Sigma_i(t)^{-1}(x-\widetilde{x})\|_2 \\ &\leq \|-\Sigma_i(t)^{-1}\| \cdot \|x-\widetilde{x}\|_2 \\ &= \frac{1}{\sigma_{\min}(\Sigma_i(t))} \cdot \|x-\widetilde{x}\|_2\end{aligned}$$

where the first step follows from simple algebra, the second step follows from Fact C.2, and the last step follows from Fact C.2. □

This lemma calculates Lipschitz constant of function $|-\frac{1}{2}(x-\mu_i(t))^\top \Sigma_i(t)^{-1}(x-\mu_i(t)) - (-\frac{1}{2}(\widetilde{x}-\mu_i(t))^\top \Sigma_i(t)^{-1}(\widetilde{x}-\mu_i(t)))|$.

Lemma E.13. *If the following conditions hold*
- *Let* $x, \widetilde{x} \in \mathbb{R}^d$*.*
- *Let* $t \in \mathbb{R}$*, and* $t \geq 0$*.*
- *Let* $\|x - \mu_i(t)\|_2 \leq R$*,* $\|\widetilde{x} - \mu_i(t)\|_2 \leq R$*, where* $R \geq 1$*, for each* $i \in \{1, 2\}$*.*

Then, for each $i \in \{1, 2\}$*, we have*

$$|-\frac{1}{2}(x - \mu_i(t))^\top \Sigma_i(t)^{-1}(x - \mu_i(t)) - (-\frac{1}{2}(\widetilde{x} - \mu_i(t))^\top \Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t)))| \leq \frac{R}{\sigma_{\min}(\Sigma_i(t))} \cdot \|x - \widetilde{x}\|_2$$

Proof. We can show

$$\begin{aligned}
\text{LHS} \leq & |-\frac{1}{2}(x - \mu_i(t))^\top \Sigma_i(t)^{-1}(x - \mu_i(t)) - (-\frac{1}{2}(x - \mu_i(t))^\top \Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t)))| \\
& + |-\frac{1}{2}(x - \mu_i(t))^\top \Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t)) - (-\frac{1}{2}(\widetilde{x} - \mu_i(t))^\top \Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t)))| \\
\leq & |-\frac{1}{2}(x - \mu_i(t))^\top \Sigma_i(t)^{-1}(x - \widetilde{x})| + |-\frac{1}{2}(x - \widetilde{x})^\top \Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t))| \\
\leq & \frac{1}{2} \cdot \|\Sigma_i(t)^{-1}(x - \mu_i(t))\|_2 \cdot \|x - \widetilde{x}\|_2 + \frac{1}{2} \cdot \|\Sigma_i(t)^{-1}(x - \widetilde{x})\|_2 \cdot \|\widetilde{x} - \mu_i(t)\|_2 \\
\leq & \frac{1}{2} \cdot \frac{1}{\sigma_{\min}(\Sigma_i(t))} \cdot R \cdot \|x - \widetilde{x}\|_2 + \frac{1}{2} \cdot \|\Sigma_i(t)^{-1}(x - \widetilde{x})\|_2 \cdot R \\
\leq & \frac{1}{2} \cdot \frac{1}{\sigma_{\min}(\Sigma_i(t))} \cdot R \cdot \|x - \widetilde{x}\|_2 + \frac{1}{2} \cdot \frac{1}{\sigma_{\min}(\Sigma_i(t))} \cdot \|x - \widetilde{x}\|_2 \cdot R \\
= & \frac{R}{\sigma_{\min}(\Sigma_i(t))} \cdot \|x - \widetilde{x}\|_2
\end{aligned}$$

where the first step follows from Fact C.2, the second step follows from simple algebra, the third step follows from Fact C.2, the fourth step follows from $\|x - \mu_i(t)\|_2 \leq R$, $\|\widetilde{x} - \mu_i(t)\|_2 \leq R$, the fifth step follows from Lemma E.12, and the last step follows from simple algebra. □

This lemma calculates Lipschitz constant of function $|N_i(x) - N_i(\widetilde{x})|$.

Lemma E.14. *If the following conditions hold*
- *Let* $x, \widetilde{x} \in \mathbb{R}^d$*.*
- *Let* $t \in \mathbb{R}$*, and* $t \geq 0$*.*
- *Let* $N_1(x), N_2(x)$ *be defined as Definition E.1.*
- *Let* $\|x - \mu_i(t)\|_2 \leq R$*, where* $R \geq 1$*, for each* $i \in \{1, 2\}$*.*
- *Let* $\|x - \mu_i(t)\|_2 \geq \beta$*, where* $\beta \in (0, 0.1)$*, for each* $i \in \{1, 2\}$*.*

Then, for each $i \in \{1, 2\}$*, we have*

$$|N_i(x) - N_i(\widetilde{x})| \leq \frac{1}{(2\pi)^{d/2} \det(\Sigma_i(t))^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}(\Sigma_i(t))}) \cdot \frac{R}{\sigma_{\min}(\Sigma_i(t))} \cdot \|x - \widetilde{x}\|_2$$

Proof. We can show

$$\begin{aligned}
|N_i(x) - N_i(\widetilde{x})| = & |\frac{1}{(2\pi)^{d/2} \det(\Sigma_i(t))^{1/2}} \exp(-\frac{1}{2}(x - \mu_i(t))^\top \Sigma_i(t)^{-1}(x - \mu_i(t))) \\
& - \frac{1}{(2\pi)^{d/2} \det(\Sigma_i(t))^{1/2}} \exp(-\frac{1}{2}(\widetilde{x} - \mu_i(t))^\top \Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t)))| \\
= & \frac{1}{(2\pi)^{d/2} \det(\Sigma_i(t))^{1/2}} \cdot |\exp(-\frac{1}{2}(x - \mu_i(t))^\top \Sigma_i(t)^{-1}(x - \mu_i(t))) \\
& - \exp(-\frac{1}{2}(\widetilde{x} - \mu_i(t))^\top \Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t)))| \\
\leq & \frac{1}{(2\pi)^{d/2} \det(\Sigma_i(t))^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}(\Sigma_i(t))})
\end{aligned}$$

$$\cdot \, | -\frac{1}{2}(x-\mu_i(t))^\top \Sigma_i(t)^{-1}(x-\mu_i(t)) - (-\frac{1}{2}(\widetilde{x}-\mu_i(t))^\top \Sigma_i(t)^{-1}(\widetilde{x}-\mu_i(t)))|$$
$$\leq \frac{1}{(2\pi)^{d/2}\det(\Sigma_i(t))^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}(\Sigma_i(t))}) \cdot \frac{R}{\sigma_{\min}(\Sigma_i(t))} \cdot \|x-\widetilde{x}\|_2$$

where the first step follows from Definition E.1, the second step follows from simple algebra, the third step follows from Fact C.9, and the last step follows from Lemma E.11. □

This lemma calculates Lipschitz constant of function $|\alpha N_1(x) + (1-\alpha)N_2(x) - (\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x}))|$.

Lemma E.15. *If the following conditions hold*
- *Let* $x, \widetilde{x} \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.
- *Let* $N_1(x), N_2(x)$ *be defined as Definition E.1.*
- *Let* $\alpha \in \mathbb{R}$ *and* $\alpha \in (0,1)$.
- *Let* $\|x-\mu_i(t)\|_2 \leq R$, *where* $R \geq 1$, *for each* $i \in \{1,2\}$.
- *Let* $\|x-\mu_i(t)\|_2 \geq \beta$, *where* $\beta \in (0, 0.1)$, *for each* $i \in \{1,2\}$.
- *Let* $\sigma_{\min} := \min\{\sigma_{\min}(\Sigma_1(t)), \sigma_{\min}(\Sigma_2(t))\}$.
- *Let* $\sigma_{\max} := \max\{\sigma_{\max}(\Sigma_1(t)), \sigma_{\max}(\Sigma_2(t))\}$.
- *Let* $\det_{\min} := \min\{\det(\Sigma_1(t)), \det(\Sigma_2(t))\}$.

Then, we have

$$|\alpha N_1(x) + (1-\alpha)N_2(x) - (\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x}))| \leq \frac{1}{(2\pi)^{d/2}\det_{\min}^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x-\widetilde{x}\|_2$$

Proof. We can show

$$\begin{aligned}
\mathrm{LHS} &= |\alpha N_1(x) - \alpha N_1(\widetilde{x}) + (1-\alpha)N_2(x) - (1-\alpha)N_2(\widetilde{x})| \\
&\leq \alpha|N_1(x) - N_1(\widetilde{x})| + (1-\alpha)|N_2(x) - N_2(\widetilde{x})| \\
&\leq \frac{\alpha}{(2\pi)^{d/2}\det(\Sigma_1(t))^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}(\Sigma_1(t))}) \cdot \frac{R}{\sigma_{\min}(\Sigma_1(t))} \cdot \|x-\widetilde{x}\|_2 \\
&\quad + \frac{1-\alpha}{(2\pi)^{d/2}\det(\Sigma_2(t))^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}(\Sigma_2(t))}) \cdot \frac{R}{\sigma_{\min}(\Sigma_2(t))} \cdot \|x-\widetilde{x}\|_2 \\
&\leq \frac{1}{(2\pi)^{d/2}} \max_{i\in[2]} \frac{1}{\det(\Sigma_i(t))^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}(\Sigma_i(t))}) \cdot \frac{R}{\sigma_{\min}(\Sigma_i(t))} \cdot \|x-\widetilde{x}\|_2 \\
&\leq \frac{1}{(2\pi)^{d/2}\det_{\min}^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x-\widetilde{x}\|_2
\end{aligned}$$

where the first step follows from simple algebra, the second step follows from Fact C.2, the third step follows from Lemma E.14, the fourth step follows from $\alpha \in (0,1)$, and the last step follows from the definition of $\det_{\min}, \sigma_{\max}, \sigma_{\min}$. □

This lemma calculates Lipschitz constant of function $|(\alpha N_1(x) + (1-\alpha)N_2(x))^{-1} - (\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x}))^{-1}|$.

Lemma E.16. *If the following conditions hold*
- *Let* $x, \widetilde{x} \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.
- *Let* $N_1(x), N_2(x)$ *be defined as Definition E.1.*
- *Let* $\alpha \in \mathbb{R}$ *and* $\alpha \in (0,1)$.
- *Let* $\|x-\mu_i(t)\|_2 \leq R$, *where* $R \geq 1$, *for each* $i \in \{1,2\}$.
- *Let* $\|x-\mu_i(t)\|_2 \geq \beta$, *where* $\beta \in (0, 0.1)$, *for each* $i \in \{1,2\}$.
- *Let* $\alpha N_1(x) + (1-\alpha)N_2(x) \geq \gamma$, *where* $\gamma \in (0, 0.1)$.
- *Let* $\sigma_{\min} := \min\{\sigma_{\min}(\Sigma_1(t)), \sigma_{\min}(\Sigma_2(t))\}$.
- *Let* $\sigma_{\max} := \max\{\sigma_{\max}(\Sigma_1(t)), \sigma_{\max}(\Sigma_2(t))\}$.
- *Let* $\det_{\min} := \min\{\det(\Sigma_1(t)), \det(\Sigma_2(t))\}$.

Then,

$$\begin{aligned} & |(\alpha N_1(x) + (1-\alpha)N_2(x))^{-1} - (\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x}))^{-1}| \\ \leq & \gamma^{-2} \cdot \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2 \end{aligned}$$

Proof. We can show

$$\begin{aligned} \text{LHS} \leq & (\alpha N_1(x) + (1-\alpha)N_2(x))^{-1} \cdot (\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x}))^{-1} \\ & \cdot |\alpha N_1(x) + (1-\alpha)N_2(x) - (\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x}))| \\ \leq & \gamma^{-2} \cdot |\alpha N_1(x) + (1-\alpha)N_2(x) - (\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x}))| \\ \leq & \gamma^{-2} \cdot \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2 \end{aligned}$$

where the first step follows from simple algebra, the second step follows from $\alpha N_1(x) + (1-\alpha)N_2(x) \geq \gamma$, and the last step follows from Lemma E.15. □

E.8. Lemmas for Lipschitz calculation: $f(x)$

This lemma calculates Lipschitz constant of function $|f(x) - f(\widetilde{x})|$.

Lemma E.17. *If the following conditions hold*

- *Let $x, \widetilde{x} \in \mathbb{R}^d$.*
- *Let $t \in \mathbb{R}$, and $t \geq 0$.*
- *Let $N_1(x), N_2(x)$ be defined as Definition E.1.*
- *Let $f(x)$ be defined as Definition E.5.*
- *Let $\alpha \in \mathbb{R}$ and $\alpha \in (0, 1)$.*
- *Let $\|x - \mu_i(t)\|_2 \leq R$, where $R \geq 1$, for each $i \in \{1, 2\}$.*
- *Let $\|x - \mu_i(t)\|_2 \geq \beta$, where $\beta \in (0, 0.1)$, for each $i \in \{1, 2\}$.*
- *Let $\alpha N_1(x) + (1-\alpha)N_2(x) \geq \gamma$, where $\gamma \in (0, 0.1)$.*
- *Let $\sigma_{\min} := \min\{\sigma_{\min}(\Sigma_1(t)), \sigma_{\min}(\Sigma_2(t))\}$.*
- *Let $\sigma_{\max} := \max\{\sigma_{\max}(\Sigma_1(t)), \sigma_{\max}(\Sigma_2(t))\}$.*
- *Let $\det_{\min} := \min\{\det(\Sigma_1(t)), \det(\Sigma_2(t))\}$.*

Then,

$$|f(x) - f(\widetilde{x})| \leq 2\alpha \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2$$

Proof. We can show

$$\begin{aligned} |f(x) - f(\widetilde{x})| = & |\frac{\alpha N_1(x)}{\alpha N_1(x) + (1-\alpha)N_2(x)} - \frac{\alpha N_1(\widetilde{x})}{\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x})}| \\ \leq & |\frac{\alpha N_1(x)}{\alpha N_1(x) + (1-\alpha)N_2(x)} - \frac{\alpha N_1(x)}{\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x})}| \\ & + |\frac{\alpha N_1(x)}{\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x})} - \frac{\alpha N_1(\widetilde{x})}{\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x})}| \\ = & \alpha \cdot |N_1(x)| \cdot |(\alpha N_1(x) + (1-\alpha)N_2(x))^{-1} - (\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x}))^{-1}| \\ & + \alpha \cdot |N_1(x) - N_1(\widetilde{x})| \cdot |(\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x}))^{-1}| \end{aligned}$$

where the first step follows from Definition E.5, the second step follows from Fact C.2, and the last step follows from simple algebra.

For the first term in the above, we have

$$\alpha \cdot |N_1(x)| \cdot |(\alpha N_1(x) + (1-\alpha)N_2(x))^{-1} - (\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x}))^{-1}|$$

$$\leq \alpha \cdot \frac{1}{(2\pi)^{d/2}\det(\Sigma_1(t))^{1/2}} \cdot |(\alpha N_1(x) + (1-\alpha)N_2(x))^{-1} - (\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x}))^{-1}|$$
$$\leq \alpha \cdot \frac{1}{(2\pi)^{d/2}\det(\Sigma_1(t))^{1/2}} \cdot \gamma^{-2} \cdot \frac{1}{(2\pi)^{d/2}\det_{\min}^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2$$
$$\leq \alpha \cdot \gamma^{-2} \cdot \frac{1}{(2\pi)^d \det_{\min}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2 \tag{11}$$

where the first step follows from $N_1(x) \leq \frac{1}{(2\pi)^{d/2}\det(\Sigma_1(t))^{1/2}}$, the second step follows from Lemma E.16 and the last step follows from definition of $\det_{\min}$.

For the second term in the above, we have

$$\alpha \cdot |N_1(x) - N_1(\widetilde{x})| \cdot |(\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x}))^{-1}|$$
$$\leq \alpha \cdot \gamma^{-1} \cdot |N_1(x) - N_1(\widetilde{x})|$$
$$\leq \alpha \cdot \gamma^{-1} \cdot \frac{1}{(2\pi)^{d/2}\det(\Sigma_1(t))^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}(\Sigma_1(t))}) \cdot \frac{R}{\sigma_{\min}(\Sigma_1(t))} \cdot \|x - \widetilde{x}\|_2 \tag{12}$$

where the first step follows from $\alpha N_1(x) + (1-\alpha)N_2(x) \geq \gamma$, the second step follows from Lemma E.15.

Combining Eq. (11) and Eq. (12) together, we have

$$|f(x) - f(\widetilde{x})| \leq \alpha \cdot \gamma^{-2} \cdot \frac{1}{(2\pi)^d \det_{\min}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2$$
$$+ \alpha \cdot \gamma^{-1} \cdot \frac{1}{(2\pi)^{d/2}\det(\Sigma_1(t))^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}(\Sigma_1(t))}) \cdot \frac{R}{\sigma_{\min}(\Sigma_1(t))} \cdot \|x - \widetilde{x}\|_2$$
$$\leq \alpha \cdot \gamma^{-2} \cdot \frac{1}{(2\pi)^d \det_{\min}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2$$
$$+ \alpha \cdot \gamma^{-1} \cdot \frac{1}{(2\pi)^{d/2}\det_{\min}^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2$$
$$\leq 2\alpha \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2}\det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2$$

where the first step follows from the bound of the first term and the second term, the second step follows from the definition of $\det_{\min}, \sigma_{\max}, \sigma_{\min}$, and the third step follows from $\gamma < 0.1$. □

This lemma calculates Lipschitz constant of function $\|f(x)(-\Sigma_1(t)^{-1}(x - \mu_1(t))) - f(\widetilde{x})(-\Sigma_1(t)^{-1}(\widetilde{x} - \mu_1(t)))\|_2$.

Lemma E.18. *If the following conditions hold*
- *Let* $x, \widetilde{x} \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.
- *Let* $N_1(x), N_2(x)$ *be defined as Definition E.1.*
- *Let* $f(x)$ *be defined as Definition E.5.*
- *Let* $\alpha \in \mathbb{R}$ *and* $\alpha \in (0, 1)$.
- *Let* $\|x - \mu_i(t)\|_2 \leq R$, *where* $R \geq 1$, *for each* $i \in \{1, 2\}$.
- *Let* $\|x - \mu_i(t)\|_2 \geq \beta$, *where* $\beta \in (0, 0.1)$, *for each* $i \in \{1, 2\}$.
- *Let* $\alpha N_1(x) + (1-\alpha)N_2(x) \geq \gamma$, *where* $\gamma \in (0, 0.1)$.
- *Let* $\sigma_{\min} := \min\{\sigma_{\min}(\Sigma_1(t)), \sigma_{\min}(\Sigma_2(t))\}$.
- *Let* $\sigma_{\max} := \max\{\sigma_{\max}(\Sigma_1(t)), \sigma_{\max}(\Sigma_2(t))\}$.
- *Let* $\det_{\min} := \min\{\det(\Sigma_1(t)), \det(\Sigma_2(t))\}$.

Then, we have

$$\|f(x)(-\Sigma_1(t)^{-1}(x - \mu_1(t))) - f(\widetilde{x})(-\Sigma_1(t)^{-1}(\widetilde{x} - \mu_1(t)))\|_2$$
$$\leq (\frac{1}{\sigma_{\min}} + 2\alpha \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2}\det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot (\frac{R}{\sigma_{\min}})^2) \cdot \|x - \widetilde{x}\|_2$$

Proof. We can show

$$\begin{aligned}
\mathrm{LHS} \leq & \|f(x)(-\Sigma_1(t)^{-1}(x-\mu_1(t))) - f(x)(-\Sigma_1(t)^{-1}(\widetilde{x}-\mu_1(t)))\|_2 \\
& + \|f(x)(-\Sigma_1(t)^{-1}(\widetilde{x}-\mu_1(t))) - f(\widetilde{x})(-\Sigma_1(t)^{-1}(\widetilde{x}-\mu_1(t)))\|_2 \\
\leq & |f(x)| \cdot \|(-\Sigma_1(t)^{-1}(x-\mu_1(t))) - (-\Sigma_1(t)^{-1}(\widetilde{x}-\mu_1(t)))\|_2 \\
& + |f(x) - f(\widetilde{x})| \cdot \| -\Sigma_1(t)^{-1}(\widetilde{x}-\mu_1(t))\|_2
\end{aligned}$$

where the first step follows from Fact C.2, the second step follows from Fact C.2.

For the first term in the above, we have

$$\begin{aligned}
& |f(x)| \cdot \|(-\Sigma_1(t)^{-1}(x-\mu_1(t))) - (-\Sigma_1(t)^{-1}(\widetilde{x}-\mu_1(t)))\|_2 \\
\leq & \|(-\Sigma_1(t)^{-1}(x-\mu_1(t))) - (-\Sigma_1(t)^{-1}(\widetilde{x}-\mu_1(t)))\|_2 \\
\leq & \frac{1}{\sigma_{\min}(\Sigma_1(t))} \cdot \|x-\widetilde{x}\|_2 \qquad (13)
\end{aligned}$$

where the first step follows from $f(x) \leq 1$, the second step follows from Lemma E.12.

For the second term in the above, we have

$$\begin{aligned}
& |f(x) - f(\widetilde{x})| \cdot \| - \Sigma_1(t)^{-1}(\widetilde{x}-\mu_1(t))\|_2 \\
\leq & \frac{R}{\sigma_{\min}(\Sigma_1(t))} \cdot |f(x) - f(\widetilde{x})| \\
\leq & \frac{R}{\sigma_{\min}(\Sigma_1(t))} \cdot 2\alpha \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x-\widetilde{x}\|_2 \qquad (14)
\end{aligned}$$

where the first step follows from Lemma E.9, the second step follows from Lemma E.17.

Combining Eq. (13) and Eq. (14) together, we have

$$\begin{aligned}
& \|f(x)(-\Sigma_1(t)^{-1}(x-\mu_1(t))) - f(\widetilde{x})(-\Sigma_1(t)^{-1}(\widetilde{x}-\mu_1(t)))\|_2 \\
\leq & \frac{1}{\sigma_{\min}(\Sigma_1(t))} \cdot \|x-\widetilde{x}\|_2 \\
& + \frac{R}{\sigma_{\min}(\Sigma_1(t))} \cdot 2\alpha \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x-\widetilde{x}\|_2 \\
\leq & \frac{1}{\sigma_{\min}} \cdot \|x-\widetilde{x}\|_2 \\
& + \frac{R}{\sigma_{\min}} \cdot 2\alpha \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x-\widetilde{x}\|_2 \\
= & (\frac{1}{\sigma_{\min}} + 2\alpha \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot (\frac{R}{\sigma_{\min}})^2) \cdot \|x-\widetilde{x}\|_2
\end{aligned}$$

where the first step follows from the bound of the first term and the second term, the second step follows from the definition of $\det_{\min}, \sigma_{\max}, \sigma_{\min}$, and the last step follows from simple algebra. □

E.9. Lemmas for Lipschitz calculation: $g(x)$

This lemma calculates Lipschitz constant of function $|g(x) - g(\widetilde{x})|$.

Lemma E.19. *If the following conditions hold*

- *Let* $x, \widetilde{x} \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.
- *Let* $N_1(x), N_2(x)$ *be defined as Definition E.1.*
- *Let* $g(x)$ *be defined as Definition E.5.*
- *Let* $\alpha \in \mathbb{R}$ *and* $\alpha \in (0, 1)$.

- *Let* $\|x-\mu_i(t)\|_2 \leq R$, *where* $R \geq 1$, *for each* $i \in \{1,2\}$.
- *Let* $\|x-\mu_i(t)\|_2 \geq \beta$, *where* $\beta \in (0, 0.1)$, *for each* $i \in \{1,2\}$.
- *Let* $\alpha N_1(x) + (1-\alpha) N_2(x) \geq \gamma$, *where* $\gamma \in (0, 0.1)$.
- *Let* $\sigma_{\min} := \min\{\sigma_{\min}(\Sigma_1(t)), \sigma_{\min}(\Sigma_2(t))\}$.
- *Let* $\sigma_{\max} := \max\{\sigma_{\max}(\Sigma_1(t)), \sigma_{\max}(\Sigma_2(t))\}$.
- *Let* $\det_{\min} := \min\{\det(\Sigma_1(t)), \det(\Sigma_2(t))\}$.

Then,

$$|g(x) - g(\widetilde{x})| \leq 2(1-\alpha) \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2$$

Proof. We can show

$$\begin{aligned}
|g(x) - g(\widetilde{x})| = & |\frac{(1-\alpha)N_2(x)}{\alpha N_1(x) + (1-\alpha)N_2(x)} - \frac{(1-\alpha)N_2(\widetilde{x})}{\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x})}| \\
\leq & |\frac{(1-\alpha)N_2(x)}{\alpha N_1(x) + (1-\alpha)N_2(x)} - \frac{(1-\alpha)N_2(x)}{\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x})}| \\
& + |\frac{(1-\alpha)N_2(x)}{\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x})} - \frac{(1-\alpha)N_2(\widetilde{x})}{\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x})}| \\
= & (1-\alpha) \cdot |N_2(x)| \cdot |(\alpha N_1(x) + (1-\alpha)N_2(x))^{-1} - (\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x}))^{-1}| \\
& + (1-\alpha) \cdot |N_2(x) - N_2(\widetilde{x})| \cdot |(\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x}))^{-1}|
\end{aligned}$$

where the first step follows from Definition E.5, the second step follows from Fact C.2, and the last step follows from simple algebra.

For the first term in the above, we have

$$\begin{aligned}
& (1-\alpha) \cdot |N_2(x)| \cdot |(\alpha N_1(x) + (1-\alpha)N_2(x))^{-1} - (\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x}))^{-1}| \\
\leq & (1-\alpha) \cdot \frac{1}{(2\pi)^{d/2} \det(\Sigma_2(t))^{1/2}} \cdot |(\alpha N_1(x) + (1-\alpha)N_2(x))^{-1} - (\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x}))^{-1}| \\
\leq & (1-\alpha) \cdot \frac{1}{(2\pi)^{d/2} \det(\Sigma_2(t))^{1/2}} \cdot \gamma^{-2} \cdot \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2 \\
\leq & (1-\alpha) \cdot \gamma^{-2} \cdot \frac{1}{(2\pi)^d \det_{\min}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2
\end{aligned} \tag{15}$$

where the first step follows from $N_2(x) \leq \frac{1}{(2\pi)^{d/2} \det(\Sigma_2(t))^{1/2}}$, the second step follows from Lemma E.16.

For the second term in the above, we have

$$\begin{aligned}
& (1-\alpha) \cdot |N_2(x) - N_2(\widetilde{x})| \cdot |(\alpha N_1(\widetilde{x}) + (1-\alpha)N_2(\widetilde{x}))^{-1}| \\
\leq & (1-\alpha) \cdot \gamma^{-1} \cdot |N_2(x) - N_2(\widetilde{x})| \\
\leq & (1-\alpha) \cdot \gamma^{-1} \cdot \frac{1}{(2\pi)^{d/2} \det(\Sigma_2(t))^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}(\Sigma_2(t))}) \cdot \frac{R}{\sigma_{\min}(\Sigma_2(t))} \cdot \|x - \widetilde{x}\|_2
\end{aligned} \tag{16}$$

where the first step follows from $\alpha N_1(x) + (1-\alpha)N_2(x) \geq \gamma$, the second step follows from Lemma E.15.

Combining Eq. (15) and Eq. (16) together, we have

$$\begin{aligned}
|g(x) - g(\widetilde{x})| \leq & (1-\alpha) \cdot \gamma^{-2} \cdot \frac{1}{(2\pi)^d \det_{\min}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2 \\
& + (1-\alpha) \cdot \gamma^{-1} \cdot \frac{1}{(2\pi)^{d/2} \det(\Sigma_2(t))^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}(\Sigma_2(t))}) \cdot \frac{R}{\sigma_{\min}(\Sigma_2(t))} \cdot \|x - \widetilde{x}\|_2 \\
\leq & (1-\alpha) \cdot \gamma^{-2} \cdot \frac{1}{(2\pi)^d \det_{\min}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2
\end{aligned}$$

$$+ (1-\alpha) \cdot \gamma^{-1} \cdot \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2$$
$$\leq 2(1-\alpha) \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^{d} \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2$$

where the first step follows from the bound of the first term and the second term, the second step follows from the definition of $\det_{\min}, \sigma_{\max}, \sigma_{\min}$, and the last step follows from $\gamma < 0.1$. □

This lemma calculates Lipschitz constant of function $\|g(x)(-\Sigma_2(t)^{-1}(x - \mu_2(t))) - g(\widetilde{x})(-\Sigma_2(t)^{-1}(\widetilde{x} - \mu_2(t)))\|_2$.

Lemma E.20. *If the following conditions hold*

- *Let* $x, \widetilde{x} \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.
- *Let* $N_1(x), N_2(x)$ *be defined as Definition E.1.*
- *Let* $g(x)$ *be defined as Definition E.5.*
- *Let* $\alpha \in \mathbb{R}$ *and* $\alpha \in (0, 1)$.
- *Let* $\|x - \mu_i(t)\|_2 \leq R$, *where* $R \geq 1$, *for each* $i \in \{1, 2\}$.
- *Let* $\|x - \mu_i(t)\|_2 \geq \beta$, *where* $\beta \in (0, 0.1)$, *for each* $i \in \{1, 2\}$.
- *Let* $\alpha N_1(x) + (1-\alpha) N_2(x) \geq \gamma$, *where* $\gamma \in (0, 0.1)$.
- *Let* $\sigma_{\min} := \min\{\sigma_{\min}(\Sigma_1(t)), \sigma_{\min}(\Sigma_2(t))\}$.
- *Let* $\sigma_{\max} := \max\{\sigma_{\max}(\Sigma_1(t)), \sigma_{\max}(\Sigma_2(t))\}$.
- *Let* $\det_{\min} := \min\{\det(\Sigma_1(t)), \det(\Sigma_2(t))\}$.

Then, we have

$$\|g(x)(-\Sigma_2(t)^{-1}(x - \mu_2(t))) - g(\widetilde{x})(-\Sigma_2(t)^{-1}(\widetilde{x} - \mu_2(t)))\|_2$$
$$\leq (\frac{1}{\sigma_{\min}} + 2(1-\alpha) \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^{d} \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot (\frac{R}{\sigma_{\min}})^2) \cdot \|x - \widetilde{x}\|_2$$

Proof. We can show

$$\begin{aligned}
\mathrm{LHS} \leq & \|g(x)(-\Sigma_2(t)^{-1}(x - \mu_2(t))) - g(x)(-\Sigma_2(t)^{-1}(\widetilde{x} - \mu_2(t)))\|_2 \\
& + \|g(x)(-\Sigma_2(t)^{-1}(\widetilde{x} - \mu_2(t))) - f(\widetilde{x})(-\Sigma_2(t)^{-1}(\widetilde{x} - \mu_2(t)))\|_2 \\
\leq & |g(x)| \cdot \|(-\Sigma_2(t)^{-1}(x - \mu_2(t))) - (-\Sigma_2(t)^{-1}(\widetilde{x} - \mu_2(t)))\|_2 \\
& + |g(x) - g(\widetilde{x})| \cdot \| - \Sigma_2(t)^{-1}(\widetilde{x} - \mu_2(t))\|_2
\end{aligned}$$

where the first step follows from Fact C.2, the second step follows from Fact C.2.

For the first term in the above, we have

$$\begin{aligned}
& |g(x)| \cdot \|(-\Sigma_2(t)^{-1}(x - \mu_2(t))) - (-\Sigma_2(t)^{-1}(\widetilde{x} - \mu_2(t)))\|_2 \\
\leq & \|(-\Sigma_2(t)^{-1}(x - \mu_2(t))) - (-\Sigma_2(t)^{-1}(\widetilde{x} - \mu_2(t)))\|_2 \\
\leq & \frac{1}{\sigma_{\min}(\Sigma_2(t))} \cdot \|x - \widetilde{x}\|_2 \qquad (17)
\end{aligned}$$

where the first step follows from $g(x) \leq 1$, the second step follows from Lemma E.12.

For the second term in the above, we have

$$\begin{aligned}
& |g(x) - g(\widetilde{x})| \cdot \| - \Sigma_2(t)^{-1}(\widetilde{x} - \mu_2(t))\|_2 \\
\leq & \frac{R}{\sigma_{\min}(\Sigma_2(t))} \cdot |g(x) - g(\widetilde{x})| \\
\leq & \frac{R}{\sigma_{\min}(\Sigma_2(t))} \cdot 2(1-\alpha) \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^{d} \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2 \qquad (18)
\end{aligned}$$

where the first step follows from Lemma E.9, the second step follows from Lemma E.19.

Combining Eq. (17) and Eq. (18) together, we have

$$
\begin{aligned}
& \|g(x)(-\Sigma_2(t)^{-1}(x-\mu_2(t))) - g(\widetilde{x})(-\Sigma_2(t)^{-1}(\widetilde{x}-\mu_2(t)))\|_2 \\
\leq & \frac{1}{\sigma_{\min}(\Sigma_2(t))} \cdot \|x-\widetilde{x}\|_2 \\
& + \frac{R}{\sigma_{\min}(\Sigma_2(t))} \cdot 2(1-\alpha) \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x-\widetilde{x}\|_2 \\
\leq & \frac{1}{\sigma_{\min}} \cdot \|x-\widetilde{x}\|_2 \\
& + \frac{R}{\sigma_{\min}} \cdot 2(1-\alpha) \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x-\widetilde{x}\|_2 \\
= & (\frac{1}{\sigma_{\min}} + 2(1-\alpha) \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot (\frac{R}{\sigma_{\min}})^2) \cdot \|x-\widetilde{x}\|_2
\end{aligned}
$$

where the first step follows from the bound of the first term and the second term, the second step follows from the definition of $\det_{\min}, \sigma_{\max}, \sigma_{\min}$, and the last step follows from simple algebra. □

E.10. Lipschitz constant of the score function

This lemma calculates the Lipschitz constant of the score funciton.

Lemma E.21 (Lipschitz)**.** *If the following conditions hold*

- *Let* $x, \widetilde{x} \in \mathbb{R}^d$.
- *Let* $t \in \mathbb{R}$*, and* $t \geq 0$.
- *Let* $N_1(x), N_2(x)$ *be defined as Definition E.1.*
- *Let* $\alpha \in \mathbb{R}$ *and* $\alpha \in (0,1)$.
- *Let* $p_t(x)$ *be defined as Definition E.2.*
- *Let* $f(x), g(x)$ *be defined as Definition E.5.*
- *Let* $\|x-\mu_i(t)\|_2 \leq R$*, where* $R \geq 1$*, for each* $i \in \{1,2\}$.
- *Let* $\|x-\mu_i(t)\|_2 \geq \beta$*, where* $\beta \in (0, 0.1)$*, for each* $i \in \{1,2\}$.
- *Let* $\alpha N_1(x) + (1-\alpha) N_2(x) \geq \gamma$*, where* $\gamma \in (0, 0.1)$.
- *Let* $\sigma_{\min} := \min\{\sigma_{\min}(\Sigma_1(t)), \sigma_{\min}(\Sigma_2(t))\}$.
- *Let* $\sigma_{\max} := \max\{\sigma_{\max}(\Sigma_1(t)), \sigma_{\max}(\Sigma_2(t))\}$.
- *Let* $\det_{\min} := \min\{\det(\Sigma_1(t)), \det(\Sigma_2(t))\}$.

Then,

$$
\|\frac{\mathrm{d}\log p_t(x)}{\mathrm{d}x} - \frac{\mathrm{d}\log p_t(\widetilde{x})}{\mathrm{d}\widetilde{x}}\|_2 \leq (\frac{2}{\sigma_{\min}} + \frac{2R^2}{\gamma^2 \sigma_{\min}^2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}})) \cdot \|x-\widetilde{x}\|_2
$$

Proof. We can show

$$
\begin{aligned}
\mathrm{LHS} = & \|f(x)(-\Sigma_1(t)^{-1}(x-\mu_1(t))) + g(x)(-\Sigma_2(t)^{-1}(x-\mu_2(t))) \\
& - (f(\widetilde{x})(-\Sigma_1(t)^{-1}(\widetilde{x}-\mu_1(t))) + g(\widetilde{x})(-\Sigma_2(t)^{-1}(\widetilde{x}-\mu_2(t))))\|_2 \\
\leq & \|f(x)(-\Sigma_1(t)^{-1}(x-\mu_1(t))) - f(\widetilde{x})(-\Sigma_1(t)^{-1}(\widetilde{x}-\mu_1(t)))\|_2 \\
& + \|g(x)(-\Sigma_2(t)^{-1}(x-\mu_2(t))) - g(\widetilde{x})(-\Sigma_2(t)^{-1}(\widetilde{x}-\mu_2(t)))\|_2 \\
\leq & (\frac{1}{\sigma_{\min}} + 2\alpha \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot (\frac{R}{\sigma_{\min}})^2) \cdot \|x-\widetilde{x}\|_2 \\
& + \|g(x)(-\Sigma_2(t)^{-1}(x-\mu_2(t))) - g(\widetilde{x})(-\Sigma_2(t)^{-1}(\widetilde{x}-\mu_2(t)))\|_2 \\
\leq & (\frac{1}{\sigma_{\min}} + 2\alpha \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot (\frac{R}{\sigma_{\min}})^2) \cdot \|x-\widetilde{x}\|_2
\end{aligned}
$$

$$+ (\frac{1}{\sigma_{\min}} + 2(1-\alpha) \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot (\frac{R}{\sigma_{\min}})^2) \cdot \|x - \widetilde{x}\|_2$$

$$= (\frac{2}{\sigma_{\min}} + \frac{2R^2}{\gamma^2 \sigma_{\min}^2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}})) \cdot \|x - \widetilde{x}\|_2$$

where the first step follows from Lemma E.6, the second step follows from Fact C.2, the third step follows from Lemma E.18, the fourth step follows from Lemma E.20, and the last step follows from simple algebra. □

F. A General Version for k Gaussian

In this section we consider a more general case of k mixture of Gaussians.

- Section F.1 provides the definition for k mixture of Gaussians.
- Section F.2 provides the expression of the score function.
- Section F.3 provides the upper bound of the score function.
- Section F.4 provides lemmas that are used in further calculation of Lipschitz constant.
- Section F.5 provides the Lipschitz constant for k mixture of Gaussians.

F.1. Definitions

Let $i \in [k]$. Let $\alpha_i(t) \in (0,1)$, $\sum_{i=1}^k \alpha_i(t) = 1$, and is a function of time t. Consider p_t such that

$$p_t(x) = \Pr_{x' \sim \sum_{i=1}^k \alpha_i(t) \mathcal{N}(\mu_i(t), \Sigma_i(t))} [x' = x]$$

where $\mu_i(t) \in \mathbb{R}^d$, $\Sigma_i(t) \in \mathbb{R}^{d \times d}$ and they are derivative to t and $\Sigma_i(t)$ is a symmetric p.s.d. matrix whose the smallest singular value is always larger than a fixed value $\sigma_{\min} > 0$.

Below we define the pdf for a single multivariate Gaussian.

Definition F.1. *If the following conditions hold*

- *Let $x \in \mathbb{R}^d$.*
- *Let $i \in [k]$.*
- *Let $t \in \mathbb{R}$, and $t \geq 0$.*

We define

$$N_i(x) := \frac{1}{(2\pi)^{d/2} \det(\Sigma_i(t))^{1/2}} \exp(-\frac{1}{2}(x - \mu_i(t))^\top \Sigma_i(t)^{-1} (x - \mu_1(t)))$$

This is the pdf of a single Gaussian so it's clearly to see that $0 \leq N_i \leq \frac{1}{(2\pi)^{d/2} \det(\Sigma_i(t))^{1/2}}$ since $N_i(x)$ takes maximum when $x = \mu_i$.

Below we define the pdf for k mixtures of Gaussians.

Definition F.2. *If the following conditions hold*

- *Let $x \in \mathbb{R}^d$.*
- *Let $i \in [k]$.*
- *Let $t \in \mathbb{R}$, and $t \geq 0$.*
- *Let $\alpha_i(t) \in \mathbb{R}$, $\sum_{i=1}^k \alpha_i(t) = 1$, and $\alpha_i(t) \in (0,1)$.*
- *Let $N_i(x)$ be defined as Definition F.1.*

We define

$$p_t(x) := \sum_{i=1}^k \frac{\alpha_i(t)}{(2\pi)^{d/2} \det(\Sigma_i(t))^{1/2}} \exp(-\frac{1}{2}(x - \mu_i(t))^\top \Sigma_i(t)^{-1} (x - \mu_i(t)))$$

This can be further rewritten as follows:

$$p_t(x) = \sum_{i=1}^k \alpha_i(t) N_i(x)$$

Further, we have

$$\log p_t(x) = \log(\sum_{i=1}^k \alpha_i(t) N_i(x))$$

This lemma calculates the gradient of pdf for k mixture of Gaussians.

Lemma F.3. *If the following conditions hold*
- *Let $x \in \mathbb{R}^d$.*
- *Let $i \in [k]$.*
- *Let $t \in \mathbb{R}$, and $t \geq 0$.*
- *Let $\alpha_i(t) \in \mathbb{R}$, $\sum_{i=1}^k \alpha_i(t) = 1$, and $\alpha_i(t) \in (0, 1)$.*
- *Let $N_i(x)$ be defined as Definition F.1.*
- *Let $p_t(x)$ be defined as Definition F.2*

We have

$$\frac{\mathrm{d}p_t(x)}{\mathrm{d}x} = \sum_{i=1}^k \alpha_i(t) N_i(x)(-\Sigma_i(t)^{-1}(x - \mu_i(t)))$$

Proof. We can show

$$\begin{aligned}
\frac{\mathrm{d}p_t(x)}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x} \sum_{i=1}^k \alpha_i(t) N_i(x) \\
&= \sum_{i=1}^k \alpha_i(t) \frac{\mathrm{d}N_i(x)}{\mathrm{d}x} \\
&= \sum_{i=1}^k \alpha_i(t) N_i(x)(-\Sigma_i(t)^{-1}(x - \mu_i(t)))
\end{aligned}$$

where the first step follows from Definition F.2, the second step follows from Fact C.1, and the last step follows from Lemma E.3.

□

Below we define f_i that simplifies further calculation.

Definition F.4. *If the following conditions hold*
- *Let $x \in \mathbb{R}^d$.*
- *Let $i \in [k]$.*
- *Let $t \in \mathbb{R}$, and $t \geq 0$.*
- *Let $\alpha_i(t) \in \mathbb{R}$, $\sum_{i=1}^k \alpha_i(t) = 1$, and $\alpha_i(t) \in (0, 1)$.*
- *Let $N_i(x)$ be defined as Definition F.1.*

For further simplicity, we define

$$f_i(x) := \frac{\alpha_i(t) N_i(x)}{\sum_{i=1}^k \alpha_i(t) N_i(x)}$$

It's clearly to see that $0 \leq f_i(x) \leq 1$ and $\sum_{i=1}^k f_i(x) = 1$

F.2. Calculation of the score function

This lemma calculates the score function for k mixture of Gaussians.

Lemma F.5. *If the following conditions hold*
- *Let $x \in \mathbb{R}^d$.*
- *Let $i \in [k]$.*

- *Let $t \in \mathbb{R}$, and $t \geq 0$.*
- *Let $\alpha_i(t) \in \mathbb{R}$, $\sum_{i=1}^k \alpha_i(t) = 1$, and $\alpha_i(t) \in (0, 1)$.*
- *Let $N_i(x)$ be defined as Definition F.1.*
- *Let $p_t(x)$ be defined as Definition F.2.*
- *Let $f_i(x)$ be defined as Definition F.4.*

We have

$$\frac{\mathrm{d} \log p_t(x)}{\mathrm{d} x} = \sum_{i=1}^k f_i(x)(-\Sigma_i(t)^{-1}(x - \mu_i(t)))$$

Proof. We can show

$$\begin{aligned}
\frac{\mathrm{d} \log p_t(x)}{\mathrm{d} x} &= \frac{\mathrm{d} \log p_t(x)}{\mathrm{d} p_t(x)} \frac{\mathrm{d} p_t(x)}{\mathrm{d} x} \\
&= \frac{1}{p_t(x)} \frac{\mathrm{d} p_t(x)}{\mathrm{d} x} \\
&= \frac{1}{p_t(x)} \sum_{i=1}^k \alpha_i(t) N_i(x)(-\Sigma_i(t)^{-1}(x - \mu_i(t))) \\
&= \frac{\sum_{i=1}^k \alpha_i(t) N_i(x)(-\Sigma_i(t)^{-1}(x - \mu_i(t)))}{\sum_{i=1}^k \alpha_i(t) N_i(x)} \\
&= \sum_{i=1}^k f_i(x)(-\Sigma_i(t)^{-1}(x - \mu_i(t)))
\end{aligned}$$

where the first step follows from Fact C.1, the second step follows from Fact C.1, the third step follows from Lemma F.3, the fourth step follows from Definition F.2, and the last step follows from Definition F.4. □

F.3. Upper bound of the score function

This lemma calculates upper bound of the score function for k mixture of Gaussians.

Lemma F.6. *If the following conditions hold*

- *Let $x \in \mathbb{R}^d$.*
- *Let $i \in [k]$.*
- *Let $t \in \mathbb{R}$, and $t \geq 0$.*
- *Let $\alpha_i(t) \in \mathbb{R}$, $\sum_{i=1}^k \alpha_i(t) = 1$, and $\alpha_i(t) \in (0, 1)$.*
- *Let $p_t(x)$ be defined as Definition F.2.*
- *Let $f_i(x)$ be defined as Definition F.4.*
- *Let $\sigma_{\min} := \min\{\sigma_{\min}(\Sigma_1(t)), \sigma_{\min}(\Sigma_2(t)), \dots, \sigma_{\min}(\Sigma_k(t))\}$.*
- *Let $\mu_{\max} := \max\{1, \|\mu_1(t)\|_2, \|\mu_2(t)\|_2, \dots, \|\mu_k(t)\|_2\}$.*

Then, we have

$$\|\frac{\mathrm{d} \log p_t(x)}{\mathrm{d} x}\|_2 \leq \sigma_{\min}^{-1} \cdot \mu_{\max} \cdot (1 + \|x\|_2)$$

Proof. We can show

$$\begin{aligned}
\|\frac{\mathrm{d} \log p_t(x)}{\mathrm{d} x}\|_2 &= \|\sum_{i=1}^k f_i(x)(-\Sigma_i(t)^{-1}(x - \mu_i(t)))\|_2 \\
&\leq \sum_{i=1}^k f_i(x)\| - \Sigma_i(t)^{-1}(x - \mu_i(t))\|_2 \\
&\leq \max_{i \in [k]} \| - \Sigma_i(t)^{-1}(x - \mu_i(t))\|_2
\end{aligned}$$

$$\leq \max_{i\in[k]} \frac{1}{\sigma_{\min}(\Sigma_i(t))} \cdot (\|x\|_2 + \|\mu_i(t)\|_2)$$
$$\leq \sigma_{\min}^{-1}(\mu_{\max} + \|x\|_2)$$
$$\leq \sigma_{\min}^{-1} \cdot \mu_{\max} \cdot (1 + \|x\|_2)$$

where the first step follows from Lemma F.5, the second step follows from triangle inequality, the third step follows from $\sum_{i=1}^k f_i(x) = 1$ and $f_i(x) \geq 0$, the fourth step follows from Lemma E.7, the fifth step follows from definition of $\mu_{\max}$ and $\sigma_{\min}$, and the last step follows from $\mu_{\max} \geq 1$. □

F.4. Lemmas for Lipshitz calculation

This section provides lemmas for calculation of Lipschitz constant of the score function for k mixture of Gaussians.

This lemma calculates Lipschitz constant of function $|\sum_{i=1}^k \alpha_i(t)N_i(x) - \sum_{i=1}^k \alpha_i(t)N_i(\widetilde{x})|$.

Lemma F.7. *If the following conditions hold*

- *Let* $x, \widetilde{x} \in \mathbb{R}^d$.
- *Let* $i \in [k]$.
- *Let* $t \in \mathbb{R}$*, and* $t \geq 0$.
- *Let* $\alpha_i(t) \in \mathbb{R}$, $\sum_{i=1}^k \alpha_i(t) = 1$*, and* $\alpha_i(t) \in (0, 1)$.
- *Let* $N_i(x)$ *be defined as Definition F.1.*
- *Let* $\|x - \mu_i(t)\|_2 \leq R$*, where* $R \geq 1$*, for each* $i \in [k]$.
- *Let* $\|x - \mu_i(t)\|_2 \geq \beta$*, where* $\beta \in (0, 0.1)$*, for each* $i \in [k]$.
- *Let* $\sigma_{\min} := \min\{\sigma_{\min}(\Sigma_1(t)), \sigma_{\min}(\Sigma_2(t)), \ldots, \sigma_{\min}(\Sigma_k(t))\}$.
- *Let* $\sigma_{\max} := \max\{\sigma_{\max}(\Sigma_1(t)), \sigma_{\max}(\Sigma_2(t)), \ldots, \sigma_{\max}(\Sigma_k(t))\}$.
- *Let* $\det_{\min} := \min\{\det(\Sigma_1(t)), \det(\Sigma_2(t)), \ldots, \det(\Sigma_k(t))\}$.

Then, we have

$$|\sum_{i=1}^k \alpha_i(t)N_i(x) - \sum_{i=1}^k \alpha_i(t)N_i(\widetilde{x})| \leq \frac{1}{(2\pi)^{d/2}\det_{\min}^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2$$

Proof. We can show

$$\text{LHS} = |\sum_{i=1}^k \alpha_i(t)(N_i(x) - N_i(\widetilde{x}))|$$
$$\leq \sum_{i=1}^k \alpha_i(t)|N_i(x) - N_i(\widetilde{x})|$$
$$\leq \sum_{i=1}^k \alpha_i(t) \frac{1}{(2\pi)^{d/2}\det(\Sigma_i(t))^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}(\Sigma_i(t))}) \cdot \frac{R}{\sigma_{\min}(\Sigma_i(t))} \cdot \|x - \widetilde{x}\|_2$$
$$\leq \max_{i\in[k]} \frac{1}{(2\pi)^{d/2}\det(\Sigma_i(t))^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}(\Sigma_i(t))}) \cdot \frac{R}{\sigma_{\min}(\Sigma_i(t))} \cdot \|x - \widetilde{x}\|_2$$
$$\leq \frac{1}{(2\pi)^{d/2}\det_{\min}^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2$$

where the first step follows from simple algebra, the second step follows from Fact C.2, the third step follows from Lemma E.14, the fourth step follows from $\sum_{i=1}^k \alpha_i(t) = 1$, and $\alpha_i(t) \in (0, 1)$, and the last step follows from the definition of $\det_{\min}, \sigma_{\max}, \sigma_{\min}$. □

This lemma calculates Lipschitz constant of function $|(\sum_{i=1}^k \alpha_i(t)N_i(x))^{-1} - (\sum_{i=1}^k \alpha_i(t)N_i(\widetilde{x}))^{-1}|$.

Lemma F.8. *If the following conditions hold*

- *Let* $x, \widetilde{x} \in \mathbb{R}^d$.
- *Let* $i \in [k]$.

- *Let* $t \in \mathbb{R}$*, and* $t \geq 0$.
- *Let* $\alpha_i(t) \in \mathbb{R}$, $\sum_{i=1}^k \alpha_i(t) = 1$*, and* $\alpha_i(t) \in (0,1)$.
- *Let* $N_i(x)$ *be defined as Definition F.1.*
- *Let* $\|x - \mu_i(t)\|_2 \leq R$*, where* $R \geq 1$*, for each* $i \in [k]$.
- *Let* $\|x - \mu_i(t)\|_2 \geq \beta$*, where* $\beta \in (0, 0.1)$*, for each* $i \in [k]$.
- *Let* $\sum_{i=1}^k \alpha_i(t)N_i(x) \geq \gamma$*, where* $\gamma \in (0, 0.1)$.
- *Let* $\sigma_{\min} := \min\{\sigma_{\min}(\Sigma_1(t)), \sigma_{\min}(\Sigma_2(t)), \dots, \sigma_{\min}(\Sigma_k(t))\}$.
- *Let* $\sigma_{\max} := \max\{\sigma_{\max}(\Sigma_1(t)), \sigma_{\max}(\Sigma_2(t)), \dots, \sigma_{\max}(\Sigma_k(t))\}$.
- *Let* $\det_{\min} := \min\{\det(\Sigma_1(t)), \det(\Sigma_2(t)), \dots, \det(\Sigma_k(t))\}$.

Then, we have

$$|(\sum_{i=1}^k \alpha_i(t)N_i(x))^{-1} - (\sum_{i=1}^k \alpha_i(t)N_i(\widetilde{x}))^{-1}| \leq \gamma^{-2} \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2$$

Proof. We can show

$$\begin{aligned}
\mathrm{LHS} &= (\sum_{i=1}^k \alpha_i(t)N_i(x))^{-1} \cdot (\sum_{i=1}^k \alpha_i(t)N_i(\widetilde{x}))^{-1} \cdot |\sum_{i=1}^k \alpha_i(t)N_i(x) - \sum_{i=1}^k \alpha_i(t)N_i(\widetilde{x})| \\
&\leq \gamma^{-2} \cdot |\sum_{i=1}^k \alpha_i(t)N_i(x) - \sum_{i=1}^k \alpha_i(t)N_i(\widetilde{x})| \\
&\leq \gamma^{-2} \cdot \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2
\end{aligned}$$

where the first step follows from simple algebra, the second step follows from $\sum_{i=1}^k \alpha_i(t)N_i(x) \geq \gamma$, and the last step follows from Lemma F.7. ☐

This lemma calculates Lipschitz constant of function $|f_i(x) - f_i(\widetilde{x})|$.

Lemma F.9. *If the following conditions hold*

- *Let* $x, \widetilde{x} \in \mathbb{R}^d$.
- *Let* $i \in [k]$.
- *Let* $t \in \mathbb{R}$*, and* $t \geq 0$.
- *Let* $\alpha_i(t) \in \mathbb{R}$, $\sum_{i=1}^k \alpha_i(t) = 1$*, and* $\alpha_i(t) \in (0,1)$.
- *Let* $N_i(x)$ *be defined as Definition F.1.*
- *Let* $f_i(x)$ *be defined as Definition F.4.*
- *Let* $\|x - \mu_i(t)\|_2 \leq R$*, where* $R \geq 1$*, for each* $i \in [k]$.
- *Let* $\|x - \mu_i(t)\|_2 \geq \beta$*, where* $\beta \in (0, 0.1)$*, for each* $i \in [k]$.
- *Let* $\sum_{i=1}^k \alpha_i(t)N_i(x) \geq \gamma$*, where* $\gamma \in (0, 0.1)$.
- *Let* $\sigma_{\min} := \min\{\sigma_{\min}(\Sigma_1(t)), \sigma_{\min}(\Sigma_2(t)), \dots, \sigma_{\min}(\Sigma_k(t))\}$.
- *Let* $\sigma_{\max} := \max\{\sigma_{\max}(\Sigma_1(t)), \sigma_{\max}(\Sigma_2(t)), \dots, \sigma_{\max}(\Sigma_k(t))\}$.
- *Let* $\det_{\min} := \min\{\det(\Sigma_1(t)), \det(\Sigma_2(t)), \dots, \det(\Sigma_k(t))\}$.

Then, for each $i \in [k]$*, we have*

$$|f_i(x) - f_i(\widetilde{x})| \leq 2\alpha_i(t) \cdot \gamma^{-2} \cdot \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2$$

Proof. We can show

$$\begin{aligned}
|f_i(x) - f_i(\widetilde{x})| &= |\alpha_i(t)N_i(x) \cdot (\sum_{i=1}^k \alpha_i(t)N_i(x))^{-1} - \alpha_i(t)N_i(\widetilde{x}) \cdot (\sum_{i=1}^k \alpha_i(t)N_i(\widetilde{x}))^{-1}| \\
&\leq |\alpha_i(t)N_i(x) \cdot (\sum_{i=1}^k \alpha_i(t)N_i(x))^{-1} - \alpha_i(t)N_i(x) \cdot (\sum_{i=1}^k \alpha_i(t)N_i(\widetilde{x}))^{-1}|
\end{aligned}$$

$$
\begin{aligned}
&+ |\alpha_i(t) N_i(x) \cdot (\sum_{i=1}^k \alpha_i(t) N_i(\widetilde{x}))^{-1} - \alpha_i(t) N_i(\widetilde{x}) \cdot (\sum_{i=1}^k \alpha_i(t) N_i(\widetilde{x}))^{-1}| \\
\leq\ & \alpha_i(t) N_i(x) \cdot |(\sum_{i=1}^k \alpha_i(t) N_i(x))^{-1} - (\sum_{i=1}^k \alpha_i(t) N_i(\widetilde{x}))^{-1}| \\
&+ \alpha_i(t) (\sum_{i=1}^k \alpha_i(t) N_i(\widetilde{x}))^{-1} |N_i(x) - N_i(\widetilde{x})|
\end{aligned}
$$

where the first step follows from Definition F.4, the second step follows from Fact C.2, and the last step follows from simple algebra.

For the first term in the above, we have

$$
\begin{aligned}
& \alpha_i(t) N_i(x) \cdot |(\sum_{i=1}^k \alpha_i(t) N_i(x))^{-1} - (\sum_{i=1}^k \alpha_i(t) N_i(\widetilde{x}))^{-1}| \\
\leq\ & \alpha_i(t) \cdot \frac{1}{(2\pi)^{d/2} \det(\Sigma_i(t))^{1/2}} \cdot |(\sum_{i=1}^k \alpha_i(t) N_i(x))^{-1} - (\sum_{i=1}^k \alpha_i(t) N_i(\widetilde{x}))^{-1}| \\
\leq\ & \alpha_i(t) \cdot \frac{1}{(2\pi)^{d/2} \det(\Sigma_i(t))^{1/2}} \cdot \gamma^{-2} \cdot \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2 \\
\leq\ & \alpha_i(t) \cdot \gamma^{-2} \cdot \frac{1}{(2\pi)^d \det_{\min}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2
\end{aligned} \tag{19}
$$

where the first step follows from $N_i(x) \leq \frac{1}{(2\pi)^{d/2} \det(\Sigma_i(t))^{1/2}}$, the second step follows from Lemma F.8, and the last step follows from definition of $\det_{\min}$.

For the second term in the above, we have

$$
\begin{aligned}
& \alpha_i(t) (\sum_{i=1}^k \alpha_i(t) N_i(\widetilde{x}))^{-1} |N_i(x) - N_i(\widetilde{x})| \\
\leq\ & \alpha_i(t) \cdot \gamma^{-1} \cdot |N_i(x) - N_i(\widetilde{x})| \\
\leq\ & \alpha_i(t) \cdot \gamma^{-1} \cdot \frac{1}{(2\pi)^{d/2} \det(\Sigma_i(t))^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}(\Sigma_i(t))}) \cdot \frac{R}{\sigma_{\min}(\Sigma_i(t))} \cdot \|x - \widetilde{x}\|_2
\end{aligned} \tag{20}
$$

where the first step follows from $\sum_{i=1}^k \alpha_i(t) N_i(x) \geq \gamma$, the second step follows from Lemma F.7.

Combining Eq. (19) and Eq. (20) together, we have

$$
\begin{aligned}
|f_i(x) - f_i(\widetilde{x})| \leq\ & \alpha_i(t) \cdot \gamma^{-2} \cdot \frac{1}{(2\pi)^d \det_{\min}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2 \\
&+ \alpha_i(t) \cdot \gamma^{-1} \cdot \frac{1}{(2\pi)^{d/2} \det(\Sigma_i(t))^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}(\Sigma_i(t))}) \cdot \frac{R}{\sigma_{\min}(\Sigma_i(t))} \cdot \|x - \widetilde{x}\|_2 \\
\leq\ & \alpha_i(t) \cdot \gamma^{-2} \cdot \frac{1}{(2\pi)^d \det_{\min}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2 \\
&+ \alpha_i(t) \cdot \gamma^{-1} \cdot \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}} \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2 \\
\leq\ & 2\alpha_i(t) \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2
\end{aligned}
$$

where the first step follows from the bound of the first term and the second term, the second step follows from the definition of $\det_{\min}, \sigma_{\max}, \sigma_{\min}$, and the last step follows from $\gamma < 0.1$. □

This lemma calculates Lipschitz constant of function $\|f_i(x)(-\Sigma_i(t)^{-1}(x - \mu_i(t))) - f_i(\widetilde{x})(-\Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t)))\|_2$.

Lemma F.10. *If the following conditions hold*

- *Let* $x, \widetilde{x} \in \mathbb{R}^d$.
- *Let* $i \in [k]$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.
- *Let* $\alpha_i(t) \in \mathbb{R}$, $\sum_{i=1}^k \alpha_i(t) = 1$, *and* $\alpha_i(t) \in (0,1)$.
- *Let* $N_i(x)$ *be defined as Definition F.1.*
- *Let* $f_i(x)$ *be defined as Definition F.4.*
- *Let* $\|x - \mu_i(t)\|_2 \leq R$, *where* $R \geq 1$, *for each* $i \in [k]$.
- *Let* $\|x - \mu_i(t)\|_2 \geq \beta$, *where* $\beta \in (0, 0.1)$, *for each* $i \in [k]$.
- *Let* $\sum_{i=1}^k \alpha_i(t) N_i(x) \geq \gamma$, *where* $\gamma \in (0, 0.1)$.
- *Let* $\sigma_{\min} := \min\{\sigma_{\min}(\Sigma_1(t)), \sigma_{\min}(\Sigma_2(t)), \dots, \sigma_{\min}(\Sigma_k(t))\}$.
- *Let* $\sigma_{\max} := \max\{\sigma_{\max}(\Sigma_1(t)), \sigma_{\max}(\Sigma_2(t)), \dots, \sigma_{\max}(\Sigma_k(t))\}$.
- *Let* $\det_{\min} := \min\{\det(\Sigma_1(t)), \det(\Sigma_2(t)), \dots, \det(\Sigma_k(t))\}$.

Then, for each $i \in [k]$, *we have*

$$\begin{aligned}
&\|f_i(x)(-\Sigma_i(t)^{-1}(x - \mu_i(t))) - f_i(\widetilde{x})(-\Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t)))\|_2 \\
\leq & (\frac{|f_i(x)|}{\sigma_{\min}} + 2\alpha_i(t) \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot (\frac{R}{\sigma_{\min}})^2) \cdot \|x - \widetilde{x}\|_2
\end{aligned}$$

Proof. We can show

$$\begin{aligned}
\text{LHS} \leq & \|f_i(x)(-\Sigma_i(t)^{-1}(x - \mu_i(t))) - f_i(x)(-\Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t)))\|_2 \\
& + \|f_i(x)(-\Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t))) - f_i(\widetilde{x})(-\Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t)))\|_2 \\
\leq & |f_i(x)| \cdot \|(-\Sigma_i(t)^{-1}(x - \mu_i(t))) - (-\Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t)))\|_2 \\
& + |f_i(x) - f_i(\widetilde{x})| \cdot \| - \Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t))\|_2
\end{aligned}$$

where the first step follows from Fact C.2, the second step follows from Fact C.2.

For the first term in the above, we have

$$\begin{aligned}
& |f_i(x)| \cdot \|(-\Sigma_i(t)^{-1}(x - \mu_i(t))) - (-\Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t)))\|_2 \\
\leq & \frac{|f_i(x)|}{\sigma_{\min}(\Sigma_i(t))} \cdot \|x - \widetilde{x}\|_2 \qquad (21)
\end{aligned}$$

where the first step follows from Lemma E.12.

For the second term in the above, we have

$$\begin{aligned}
& |f_i(x) - f_i(\widetilde{x})| \cdot \| - \Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t))\|_2 \\
\leq & \frac{R}{\sigma_{\min}(\Sigma_i(t))} \cdot |f_i(x) - f_i(\widetilde{x})| \\
\leq & \frac{R}{\sigma_{\min}(\Sigma_i(t))} \cdot 2\alpha_i(t) \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2 \qquad (22)
\end{aligned}$$

where the first step follows from Lemma E.9, the second step follows from Lemma F.9.

Combining Eq. (21) and Eq. (22) together, we have

$$\begin{aligned}
& \|f_i(x)(-\Sigma_i(t)^{-1}(x - \mu_i(t))) - f_i(\widetilde{x})(-\Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t)))\|_2 \\
\leq & \frac{|f_i(x)|}{\sigma_{\min}(\Sigma_i(t))} \cdot \|x - \widetilde{x}\|_2 \\
& + \frac{R}{\sigma_{\min}(\Sigma_i(t))} \cdot 2\alpha_i(t) \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2 \\
\leq & \frac{|f_i(x)|}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2
\end{aligned}$$

$$+ \frac{R}{\sigma_{\min}} \cdot 2\alpha_i(t) \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot \frac{R}{\sigma_{\min}} \cdot \|x - \widetilde{x}\|_2$$
$$= (\frac{|f_i(x)|}{\sigma_{\min}} + 2\alpha_i(t) \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot (\frac{R}{\sigma_{\min}})^2) \cdot \|x - \widetilde{x}\|_2$$

where the first step follows from the bound of the first term and the second term, the second step follows from the definition of $\det_{\min}, \sigma_{\max}, \sigma_{\min}$, and the last step follows from simple algebra. □

F.5. Lipschitz constant of the score function

This lemma calculates Lipschitz constant of the score function for k mixture of Gaussians.

Lemma F.11. *If the following conditions hold*

- *Let* $x, \widetilde{x} \in \mathbb{R}^d$.
- *Let* $i \in [k]$.
- *Let* $t \in \mathbb{R}$, *and* $t \geq 0$.
- *Let* $\alpha_i(t) \in \mathbb{R}$, $\sum_{i=1}^k \alpha_i(t) = 1$, *and* $\alpha_i(t) \in (0, 1)$.
- *Let* $N_i(x)$ *be defined as Definition F.1.*
- *Let* $p_t(x)$ *be defined as Definition F.2.*
- *Let* $f_i(x)$ *be defined as Definition F.4.*
- *Let* $\|x - \mu_i(t)\|_2 \leq R$, *where* $R \geq 1$, *for each* $i \in [k]$.
- *Let* $\|x - \mu_i(t)\|_2 \geq \beta$, *where* $\beta \in (0, 0.1)$, *for each* $i \in [k]$.
- *Let* $\sum_{i=1}^k \alpha_i(t) N_i(x) \geq \gamma$, *where* $\gamma \in (0, 0.1)$.
- *Let* $\sigma_{\min} := \min\{\sigma_{\min}(\Sigma_1(t)), \sigma_{\min}(\Sigma_2(t)), \ldots, \sigma_{\min}(\Sigma_k(t))\}$.
- *Let* $\sigma_{\max} := \max\{\sigma_{\max}(\Sigma_1(t)), \sigma_{\max}(\Sigma_2(t)), \ldots, \sigma_{\max}(\Sigma_k(t))\}$.
- *Let* $\det_{\min} := \min\{\det(\Sigma_1(t)), \det(\Sigma_2(t)), \ldots, \det(\Sigma_k(t))\}$.

Then, we have

$$\|\frac{\mathrm{d} \log p_t(x)}{\mathrm{d} x} - \frac{\mathrm{d} \log p_t(\widetilde{x})}{\mathrm{d} \widetilde{x}}\|_2$$
$$\leq (\frac{1}{\sigma_{\min}} + \frac{2R^2}{\gamma^2 \sigma_{\min}^2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}})) \cdot \|x - \widetilde{x}\|_2$$

Proof. We can show

$$\mathrm{LHS}$$
$$= \|\sum_{i=1}^k f_i(x)(-\Sigma_i(t)^{-1}(x - \mu_i(t))) - \sum_{i=1}^k f_i(\widetilde{x})(-\Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t)))\|_2$$
$$\leq \sum_{i=1}^k \|f_i(x)(-\Sigma_i(t)^{-1}(x - \mu_i(t))) - f_i(\widetilde{x})(-\Sigma_i(t)^{-1}(\widetilde{x} - \mu_i(t)))\|_2$$
$$\leq \sum_{i=1}^k (\frac{|f_i(x)|}{\sigma_{\min}} + 2\alpha_i(t) \cdot \gamma^{-2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}}) \cdot (\frac{R}{\sigma_{\min}})^2) \cdot \|x - \widetilde{x}\|_2$$
$$= (\frac{1}{\sigma_{\min}} + \frac{2R^2}{\gamma^2 \sigma_{\min}^2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2} \det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}})) \cdot \|x - \widetilde{x}\|_2$$

where the first step follows from Lemma F.5, the second step follows from triangle inequality, the third step follows from Lemma F.10, and the last step follows from $\sum_{i=1}^k |f_i(x)| = \sum_{i=1}^k f_i(x) = 1$ and $\sum_{i=1}^k \alpha_i(t) = 1$. □

G. Tools From Previous Work

In this section, we present several key theoretical results from previous work that serve as building blocks for our analysis. We begin with important assumptions about score-based diffusion models, followed by theorems establishing error bounds for different numerical solvers.

Assumption G.1 (Lipschitz score, Assumption 1 in [13], page 6)**.** *For all $t \geq 0$, the score $\nabla \log p_t$ is L-Lipschitz.*

Assumption G.2 (Second momentum bound, Assumption 2 in [13], page 6 and Assumption 2 in [10], page 6)**.** *We assume that $m_2^2 := M_2 := \mathbb{E}_{p_0}[\|x\|_2^2] < \infty$.*

Assumption G.3 (Smooth data distributions, Assumption 4 in [10], page 10)**.** *The data distribution admits a density $p_0 \in C^2(\mathbb{R}^d)$ and $\nabla \log p_0$ is L-Lipschitz, where C^2 means second-order differentiable.*

Remark G.4. *We can notice $M_2 = m_2^2$. The theorems from [10] use M_2 for KL divergence. The theorems form [13] and [14] use $m_2 = \sqrt{m_2^2} = \sqrt{M_2}$ for total variance, because of Pinsker's inequality (Lemma 5.3).*

Each theorem presented below provides different perspectives on bounding the error between the learned distribution and the target distribution. Theorem G.5 focuses on total variation distance for DDPM, while Theorems G.6 and G.7 analyze KL divergence under different conditions. Theorems G.8 and G.9 establish bounds for the ODE-based solvers DPOM and DPUM respectively. We first state a tool from previous work [13].

Theorem G.5 (DDPM, Theorem 2 in [13], page 7)**.** *Suppose that Assumptions G.1, G.2 and 5.1 hold. Let $\widehat{q}_T$ be the output of DDPM algorithm at times T, and suppose that the step size $h := T/N$ satisfies $h \lesssim 1/L$, where $L \geq 1$. Then, it holds that*

$$\mathrm{TV}(\widehat{q}_T, p_0) \lesssim \underbrace{\sqrt{\mathrm{KL}(p_0 \| \mathcal{N}(0, I))} \exp(-T)}_{\text{convergence of forward process}} + \underbrace{(L\sqrt{dh} + Lm_2 h)\sqrt{T}}_{\text{discretization error}} + \underbrace{\epsilon_0 \sqrt{T}}_{\text{score estimation error}}.$$

Then, we state a tool from previous work [10].

Theorem G.6 (Theorem 1 in [10])**.** *Suppose that Assumptions G.1, G.2, 5.1 hold. If $L \geq 1$, $h_k \leq 1$ for $k \in [N]$ and $T \geq 1$, using uniform discretization points yields the following: (1) Using Exponential Integrator scheme (8), we have $\mathrm{KL}(p_0 \| \widehat{q}_T) \lesssim (M_2 + d)\exp(-T) + T\epsilon_0^2 + \frac{dT^2L^2}{N}$. In particular, choosing $T = \Theta(\log(dM_2/\epsilon_0^2))$ and $N = \Theta(dT^2L^2/\epsilon_0^2)$ makes this $\widetilde{O}(\epsilon_0^2)$. (2) Using the Euler-Maruyama scheme (7), we have $\mathrm{KL}(p_0 \| \widehat{q}_T) \lesssim (M_2 + d)\exp(-T) + T\epsilon_0^2 + \frac{dT^2L^2}{N} + \frac{T^3 M_2}{N^2}$.*

Theorem G.7 (Theorem 5 in [10], page 10)**.** *There is a universal constant K such that the following holds. Under Assumptions G.2, 5.1, and G.3 hold, by using the exponentially decreasing (then constant) step size $h_k = c\min\{\max\{t_k, \frac{1}{L}\}, 1\}, c = \frac{T + \log L}{N} \leq \frac{1}{Kd}$, the sampling dynamic (8) results in a distribution $\widehat{q}_T$ such that $\mathrm{KL}(p_0 \| \widehat{q}_T) \lesssim (M_2 + d)\exp(-T) + T\epsilon_0^2 + \frac{d^2(T + \log L)^2}{N}$. Choosing $T = \Theta(\log(dM_2/\epsilon_0^2))$ and $N = \Theta(d^2(T + \log L)^2/\epsilon_0^2)$ makes this $\widetilde{O}(\epsilon_0^2)$. In addition, for the Euler-Maruyama scheme (7), the same bounds hold with an additional $M_2 \sum_{k=1}^N h_k^3$ term.*

Finally, we state a tool from previous work [14].

Theorem G.8 (DPOM, Theorem 2 in [14], page 6)**.** *Suppose that Assumptions G.1, G.2 and 5.1 hold. If $\widehat{q}_T$ denotes the output of DPOM (see Algorithm 1 in [14]) with early stopping. Then, it holds that*

$$\mathrm{TV}(\widehat{q}_T, p_0) \lesssim (\sqrt{d} + m_2)\exp(-T) + L^2 T d^{1/2} h_{\mathrm{pred}} + L^{3/2} T d^{1/2} h_{\mathrm{corr}}^{1/2} + L^{1/2} T \epsilon_0 + \epsilon.$$

In particular, if we set $T = \Theta(\log(dm_2^2/\epsilon^2))$, $h_{\mathrm{pred}} = \widetilde{\Theta}(\frac{\epsilon}{L^2 d^{1/2}})$, $h_{\mathrm{corr}} = \widetilde{\Theta}(\frac{\epsilon}{L^3 d})$, and if the score estimation error satisfies $\epsilon_0 \leq \widetilde{O}(\frac{\epsilon}{\sqrt{L}})$, then we can obtain TV error ϵ with a total iteration complexity of $\widetilde{\Theta}(L^3 d/\epsilon^2)$ steps.

Theorem G.9 (DPUM, Theorem 3 in [14], page 7)**.** *Suppose that Assumptions G.1, G.2 and 5.1 hold. If $\widehat{q}_T$ denotes the output of DPUM (see Algorithm 2 in [14]) with early stopping. Then, it holds that*

$$\mathrm{TV}(\widehat{q}_T, p_0) \lesssim (\sqrt{d} + m_2)\exp(-T) + L^2 T d^{1/2} h_{\mathrm{pred}} + L^{3/2} T d^{1/2} h_{\mathrm{corr}}^{1/2} + L^{1/2} T \epsilon_0 + \epsilon.$$

In particular, if we set $T = \Theta(\log(dm_2^2/\epsilon^2))$, $h_{\mathrm{pred}} = \widetilde{\Theta}(\frac{\epsilon}{L^2 d^{1/2}})$, $h_{\mathrm{corr}} = \widetilde{\Theta}(\frac{\epsilon}{L^{3/2} d^{1/2}})$, and if the score estimation error satisfies $\epsilon_0 \leq \widetilde{O}(\frac{\epsilon}{\sqrt{L}})$, then we can obtain TV error ϵ with a total iteration complexity of $\widetilde{\Theta}(L^2 d^{1/2}/\epsilon)$ steps.

H. Putting It All Together

Our overall goal is that we want to provide a more concrete calculation for theorems in Section G by assuming the data distribution is a k mixture of Gaussian. Now we provide lemmas that are used in further calculation.

Now we provide the lemma for k-mixtue of Gaussians which states that if p_0 is mixture Gaussians, then all the pdfs in the diffusion process are also mixtures of Gaussians.

Proposition H.1 (Formal version of Proposition 3.2)**.** *Let $a, b \in \mathbb{R}$. Let $\mathcal{D}$ be a k-mixture of Gaussian distribution, and let p be its* pdf*, i.e.,*

$$p(x) := \sum_{i=1}^{k} \frac{\alpha_i}{(2\pi)^{d/2} \det(\Sigma_i)^{1/2}} \exp(-\frac{1}{2}(x-\mu_i)^\top \Sigma_i^{-1}(x-\mu_i))$$

Let $x \in \mathbb{R}^d$ sample from $\mathcal{D}$. Let $z \in \mathbb{R}^d$ and $z \sim \mathcal{N}(0, I)$, which is independent from x. Then we have a new random variable $y = ax + bz$ which is also a k-mixture of Gaussian distribution $\widetilde{\mathcal{D}}$, whose pdf *is*

$$\widetilde{p}(x) := \sum_{i=1}^{k} \frac{\alpha_i}{(2\pi)^{d/2} \det(\widetilde{\Sigma}_i)^{1/2}} \exp(-\frac{1}{2}(x-\widetilde{\mu}_i)^\top \widetilde{\Sigma}_i^{-1}(x-\widetilde{\mu}_i)),$$

where $\widetilde{\mu}_i = a\mu_i, \widetilde{\Sigma}_i = a^2\Sigma_i + b^2 I$.

Proof. First, we know that the pdf of the sum of two independent random variables is the convolution of their pdf.

From [67] we know that the convolution of 2 Gaussians is another Gaussian, i.e. $\mathcal{N}(\mu_1, \Sigma_1) * \mathcal{N}(\mu_2, \Sigma_2) = \mathcal{N}(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2)$, where $*$ is the convolution operator.

And we know the pdf of a linear transformation of a random variable $x \in \mathbb{R}^d$, let's say $Ax + b$ where $A \in \mathbb{R}^{d\times d}, b \in \mathbb{R}^d$, is $\frac{1}{|\det(A)|} p(A^{-1}(x-b))$.

If we consider the transformation ax where $a \in \mathbb{R}, x \in \mathbb{R}^d$, this transformation can be written as aIx. Therefore the pdf of ax is $\frac{1}{|\det(aI)|} p((aI)^{-1}x) = \frac{1}{|a^d|} p(x/a)$.

Now we prove the lemma. To find the pdf of ax, where $x \sim p(x)$, we can show

$$\begin{aligned}
|\frac{1}{|a^d|} p(x/a)| &= \frac{1}{|a^d|} \sum_{i=1}^{k} \frac{\alpha_i}{(2\pi)^{d/2} \det(\Sigma_i)^{1/2}} \exp(-\frac{1}{2}(\frac{x}{a}-\mu_i)^\top \Sigma_i^{-1}(\frac{x}{a}-\mu_i)) \\
&= \sum_{i=1}^{k} \frac{\alpha_i}{(2\pi)^{d/2} \det(a^2\Sigma_i)^{1/2}} \exp(-\frac{1}{2}(x-a\mu_i)^\top a^{-2}\Sigma_i^{-1}(a-a\mu_i)) \\
&= \sum_{i=1}^{k} \alpha_i \mathcal{N}(a\mu_i, a^2\Sigma_i)
\end{aligned}$$

where the first step follows from the definition of $p(x)$, the second step follows from $a^{2d}\det(\Sigma_i) = \det(a^2\Sigma_i)$, and the last step follows from definition of Gaussian distribution.

For a single standard Gaussian random variable z, the pdf of bz will simply be $\mathcal{N}(0, b^2 I)$.

To find the pdf of $y = ax + bz$, we can show

$$\begin{aligned}
\widetilde{p}(x) &= \frac{1}{|a^d|} p(x/a) * \mathcal{N}(0, b^2 I) \\
&= (\sum_{i=1}^{k} \alpha_i \mathcal{N}(a\mu_i, a^2\Sigma_i)) * \mathcal{N}(0, b^2 I) \\
&= \sum_{i=1}^{k} (\alpha_i \mathcal{N}(a\mu_i, a^2\Sigma_i) * \mathcal{N}(0, b^2 I)) \\
&= \sum_{i=1}^{k} \alpha_i \mathcal{N}(a\mu_i, a^2\Sigma_i + b^2 I)
\end{aligned}$$

where the first step follows from the pdf of the sum of 2 independent random variables is the convolution of their pdf, the second step follows from the pdf of $\frac{1}{|a^d|}p(x/a)$, the third step follows from the distributive property of convolution, and the last step follows from $\mathcal{N}(a\mu_i, a^2\Sigma_i) * \mathcal{N}(0, b^2 I) = \mathcal{N}(a\mu_i, a^2\Sigma_i + b^2 I)$.

Thus, the pdf of y can be written as a mixture of k Gaussians:

$$\widetilde{p}(x) := \sum_{i=1}^{k} \frac{\alpha_i}{(2\pi)^{d/2}\det(\widetilde{\Sigma}_i)^{1/2}} \exp(-\frac{1}{2}(x-\widetilde{\mu}_i)^\top \widetilde{\Sigma}_i^{-1}(x-\widetilde{\mu}_i)),$$

where $\widetilde{\mu}_i = a\mu_i, \widetilde{\Sigma}_i = a^2\Sigma_i + b^2 I$. □

Now we provide the lemma for the second momentum of k-mixtue of Gaussians.

Lemma H.2 (Formal version of Lemma 3.6)**.** *If the following conditions hold:*

- *$x_0 \sim p_0$, where p_0 is defined by Eq.* (9).

Then, we have

$$m_2^2 := \mathop{\mathbb{E}}_{x_0\sim p_0}[\|x_0\|_2^2] = \sum_{i=1}^{k} \alpha_i(\|\mu_i\|_2 + \mathrm{tr}[\Sigma_i])$$

Proof. From [46], we know the second momentum of data distribution $p_0(x)$ is given by:

$$\mathbb{E}[x_0 x_0^\top] = \sum_{i=1}^{k} \alpha_i \cdot (\|\mu_i\|_2^2 + \Sigma_i) \tag{23}$$

Then, we can show

$$\begin{aligned}
\mathbb{E}[\|x_0\|_2^2] &= \mathbb{E}[x_0^\top x_0] \\
&= \mathbb{E}[\mathrm{tr}[x_0 x_0^\top]] \\
&= \mathrm{tr}[\mathbb{E}[x_0 x_0^\top]] \\
&= \mathrm{tr}[\sum_{i=1}^{k} \alpha_i(\mu_i\mu_i^\top + \Sigma_i)] \\
&= \sum_{i=1}^{k} \alpha_i(\|\mu_i\|_2^2 + \mathrm{tr}[\Sigma_i])
\end{aligned}$$

where the first step follows from definition of ℓ_2-norm, the second step follows from $\mathrm{tr}[aa^\top] = a^\top a$ where a is a vector, the third step follows from the linearity of the trace operator, the fourth step follows from Eq. (23), and the last step follows from $\mathrm{tr}[aa^\top] = \|a\|_2^2$. □

Now we give the Lipschitz constant explicitly.

Lemma H.3 (Formal version of Lemma 3.5)**.** *If the following conditions hold*

- *Let $\|x - a_t\mu_i\|_2 \le R$, where $R \ge 1$, for each $i \in [k]$.*
- *Let $\|x - a_t\mu_i\|_2 \ge \beta$, where $\beta \in (0, 0.1)$, for each $i \in [k]$.*
- *Let $p_t(x)$ be defined as Eq.* (10) *and $p_t(x) \ge \gamma$, where $\gamma \in (0, 0.1)$.*
- *Let $\sigma_{\min} := \min_{i\in[k]}\{\sigma_{\min}(a_t^2\Sigma_i + b_t^2 I)\}$.*
- *Let $\sigma_{\max} := \max_{i\in[k]}\{\sigma_{\max}(a_t^2\Sigma_i + b_t^2 I)\}$.*
- *Let $\det_{\min} := \min_{i\in[k]}\{\det(a_t^2\Sigma_i + b_t^2 I)\}$.*

The Lipschitz constant for the score function $\frac{\mathrm{d}\log(p_t(x))}{\mathrm{d}x}$ is given by:

$$L = \frac{1}{\sigma_{\min}} + \frac{2R^2}{\gamma^2\sigma_{\min}^2} \cdot (\frac{1}{(2\pi)^d \det_{\min}} + \frac{1}{(2\pi)^{d/2}\det_{\min}^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max}})$$

Proof. Using Lemma F.11 and Proposition H.1, we can get the result. □

I. More Calculation for Application

In this section, we will provide a more concrete calculation for Theorem G.5, Theorem G.6, Theorem G.7, Theorem G.8 and Theorem G.9.

I.1. Concrete calculation of Theorem G.5

Theorem I.1 (DDPM, total variation, formal version of Theorem 5.4)**.** *If the following conditions hold:*

- *Condition 3.4 and Assumption 5.1.*
- *The step size $h_k := T/N$ satisfies $h_k = O(1/L)$ and $L \geq 1$ for $k \in [N]$.*
- *Let $\widehat{q}$ denote the density of the output of the* EulerMaruyama *defined by Definition 4.3.*

We have

$$\mathrm{TV}(\widehat{q}, p_0) \lesssim \underbrace{\sqrt{\mathrm{KL}(p_0 \| \mathcal{N}(0,I))} \exp(-T)}_{\text{convergence of forward process}} + \underbrace{(L\sqrt{dh} + Lm_2 h)\sqrt{T}}_{\text{discretization error}} + \underbrace{\epsilon_0 \sqrt{T}}_{\text{score estimation error}}.$$

where

- $L = \frac{1}{\sigma_{\min(p_t)}} + \frac{2R^2}{\gamma^2 \sigma^2_{\min(p_t)}} \cdot (\frac{1}{(2\pi)^d \det_{\min(p_t)}} + \frac{1}{(2\pi)^{d/2} \det^{1/2}_{\min(p_t)}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max(p_t)}})$,
- $m_2 = (\sum_{i=1}^k \alpha_i (\|\mu_i\|_2^2 + \mathrm{tr}[\Sigma_i]))^{1/2}$,
- $\mathrm{KL}(p_0(x) \| \mathcal{N}(0,I)) \leq \frac{1}{2}(-\log(\det_{\min(p_0)}) + d\sigma_{\max(p_0)} + \mu_{\max(p_0)} - d)$.

Proof. Now we want to find a more concrete L in Assumption G.1. Notice that from Proposition H.1, we know that at any time between $0 \leq t \leq T$, p_t is also a k-mixture of Gaussian, except that the mean and covariance change with time.

Using Lemma H.3, we can get L.

Now we want to find the second momentum in Assumption G.2. Using Lemma H.2, we know that $m_2 = (\sum_{i=1}^k \alpha_i (\|\mu_i\|_2^2 + \mathrm{tr}[\Sigma_i]))^{1/2}$.

In Assumption 5.1, we also assume the same thing.

Now we want to have a more concrete setting for Theorem G.5 by calculating each term directly. Notice that now we have all the quantities except for the KL divergence term. Thus, we calculate $\mathrm{KL}(p_0 \| \mathcal{N}(0,I))$, which means the KL divergence of data distribution and standard Gaussian.

In our notation, we have

$$\begin{aligned} \mathrm{KL}(p_0(x) \| \mathcal{N}(0,I)) &= \mathrm{KL}(\sum_{i=1}^k \alpha_i \mathcal{N}(\mu_i, \Sigma_i) \| \mathcal{N}(0,I)) \\ &= \int \sum_{i=1}^k \alpha_i \mathcal{N}(\mu_i, \Sigma_i) \log(\frac{\sum_{i=1}^k \alpha_i \mathcal{N}(\mu_i, \Sigma_i)}{\mathcal{N}(0,I)}) \mathrm{d}x. \end{aligned}$$

However, this integral has no close form, but we can find an upper bound for this KL divergence instead.

We know the KL divergence of 2 normal distribution is given by:

$$\begin{aligned} &\mathrm{KL}(\mathcal{N}(\mu_1, \Sigma_1) \| \mathcal{N}(\mu_2, \Sigma_2)) \\ = &-\frac{1}{2}\log(\frac{\det(\Sigma_1)}{\det(\Sigma_2)}) + \frac{1}{2}\mathrm{tr}[(\Sigma_2)^{-1}\Sigma_1] + \frac{1}{2}(\mu_1 - \mu_2)^\top (\Sigma_2)^{-1}(\mu_1 - \mu_2) - \frac{d}{2} \end{aligned}$$

We define $\sigma_{\max(p_0)} = \max_{i \in [k]}\{\sigma_{\max}(\Sigma_i)\}$, $\det_{\min(p_0)} = \min_{i \in [k]}\{\det(\Sigma_i)\}$, $\mu_{\max(p_0)} = \max_{i \in [k]}\{\|\mu_i\|_2^2\}$. From [29], we can show

$$\begin{aligned} \mathrm{KL}(\sum_{i=1}^k \alpha_i \mathcal{N}(\mu_i, \Sigma_i) \| \mathcal{N}(0,I)) &\leq \sum_{i=1}^k \alpha_i \mathrm{KL}(\mathcal{N}(\mu_i, \Sigma_i) \| \mathcal{N}(0,I)) \\ &= \sum_{i=1}^k \frac{\alpha_i}{2}(-\log(\det(\Sigma_i)) + \mathrm{tr}[\Sigma_i] + \|\mu_i\|_2^2 - d) \\ &\leq \max_{i \in [k]} \frac{1}{2}(-\log(\det(\Sigma_i)) + \mathrm{tr}[\Sigma_i] + \|\mu_i\|_2^2 - d) \end{aligned}$$

$$\leq \frac{1}{2}(-\log(\det_{\min(p_0)}) + d\sigma_{\max(p_0)} + \mu_{\max(p_0)} - d)$$

where the first step follows from the convexity of KL divergence, the second step follows from KL divergence of 2 normal distribution, the third step follows from $\sum_{i=1}^k \alpha_i = 1$ and $0 \leq \alpha_i \leq 1$, and the last step follows from the definition of $\det_{\min(p_0)}, \sigma_{\max(p_0)}, \mu_{\max(p_0)}$.

Then we have all the quantities in Theorem G.5. After directly applying the theorem, we finish the proof. □

I.2. Concrete calculation for Theorem G.6

Theorem I.2 (DDPM, KL divergence, formal version of Theorem 5.5)**.** *If the following conditions hold:*

- *Condition 3.4 and Assumption 5.1.*
- *We use uniform discretization points.*

We have

- *Let $\widehat{q}$ denote the density of the output of the* ExponentialIntegrator *defined by Definition 4.4, we have*

$$\mathrm{KL}(p_0 \| \widehat{q}) \lesssim (M_2 + d)e^{-T} + T\epsilon_0^2 + \frac{dT^2L^2}{N}.$$

In particular, choosing $T = \Theta(\log(M_2 d/\epsilon_0))$ and $N = \Theta(dT^2L^2/\epsilon_0^2)$, then we can show that

$$\mathrm{KL}(p_0 \| \widehat{q}) = \widetilde{O}(\epsilon_0^2)$$

- *Let $\widehat{q}$ denote the density of the output of the* EulerMaruyama *defined by Definition 4.3, we have*

$$\mathrm{KL}(p_0 \| \widehat{q}) \lesssim (M_2 + d)e^{-T} + T\epsilon_0^2 + \frac{dT^2L^2}{N} + \frac{T^3M_2}{N^2}.$$

where

- $L = \frac{1}{\sigma_{\min(p_t)}} + \frac{2R^2}{\gamma^2\sigma^2_{\min(p_t)}} \cdot (\frac{1}{(2\pi)^d \det_{\min(p_t)}} + \frac{1}{(2\pi)^{d/2}\det^{1/2}_{\min(p_t)}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max(p_t)}}),$
- $M_2 = \sum_{i=1}^k \alpha_i(\|\mu_i\|_2^2 + \mathrm{tr}[\Sigma_i]).$

Proof. Using Lemma H.3, we can get L. Using Lemma H.2, we can get m_2. Then we directly apply Theorem G.6. □

I.3. Concrete calculation for Theorem G.7

Theorem I.3 (DDPM, KL divergence for smooth data distribution, formal version of Theorem 5.6)**.** *If the following conditions hold:*

- *Condition 3.4 and Assumption 5.1.*
- *We use the exponentially decreasing (then constant) step size $h_k = c\min\{\max\{t_k, \frac{1}{L}\}, 1\}, c = \frac{T+\log L}{N} \leq \frac{1}{Kd}$.*
- *Let $\widehat{q}$ denote the density of the output of the* ExponentialIntegrator *defined by Definition 4.4.*

We have

$$\mathrm{KL}(p_0 \| \widehat{q}) \lesssim (M_2 + d)\exp(-T) + T\epsilon_0^2 + \frac{d^2(T + \log L)^2}{N},$$

where

- $L = \frac{1}{\sigma_{\min(p_0)}} + \frac{2R^2}{\gamma^2(\sigma_{\min(p_0)})^2} \cdot (\frac{1}{(2\pi)^d \det_{\min(p_0)}} + \frac{1}{(2\pi)^{d/2}(\det_{\min(p_0)})^{1/2}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max(p_0)}}),$
- $M_2 = \sum_{i=1}^k \alpha_i(\|\mu_i\|_2^2 + \mathrm{tr}[\Sigma_i]).$

Furthermore, if we choosing $T = \Theta(\log(M_2 d/\epsilon_0))$ and $N = \Theta(d^2(T + \log L)^2/\epsilon_0^2)$, then we can show

$$\mathrm{KL}(p_0 \| \widehat{q}) \leq \widetilde{O}(\epsilon_0^2)$$

In addition, for Euler-Maruyama scheme defined in Definition 4.3, the same bounds hold with an additional $M_2 \sum_{k=1}^N h_k^3$ term.

Proof. Clearly, p_0 is second-order differentiable. Using Lemma H.3, we can get L. Using Lemma H.2, we can get m_2. Then we directly apply Theorem G.7. □

I.4. Concrete calculation for Theorem G.8

Theorem I.4 (DPOM, formal version of Theorem 5.7)**.** *If the following conditions hold:*

- *Condition 3.4 and Assumption 5.1.*
- *We use the* DPOM *algorithm defined in Definition 4.5, and let $\widehat{q}$ be the output density of it.*

We have

$$\mathrm{TV}(\widehat{q}, p_0) \lesssim (\sqrt{d} + m_2)\exp(-T) + L^2 T d^{1/2} h_{\mathrm{pred}} + L^{3/2} T d^{1/2} h_{\mathrm{corr}}^{1/2} + L^{1/2} T \epsilon_0 + \epsilon.$$

where

- $L = \frac{1}{\sigma_{\min(p_t)}} + \frac{2R^2}{\gamma^2 \sigma^2_{\min(p_t)}} \cdot (\frac{1}{(2\pi)^d \det_{\min(p_t)}} + \frac{1}{(2\pi)^{d/2} \det^{1/2}_{\min(p_t)}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max(p_t)}})$,
- $m_2 = (\sum_{i=1}^k \alpha_i (\|\mu_i\|_2^2 + \mathrm{tr}[\Sigma_i]))^{1/2}$.

In particular, if we set $T = \Theta(\log(d m_2/\epsilon))$, $h_{\mathrm{pred}} = \widetilde{\Theta}(\frac{\epsilon}{L^2 d^{1/2}})$, $h_{\mathrm{corr}} = \widetilde{\Theta}(\frac{\epsilon}{L^3 d})$, and if the score estimation error satisfies $\epsilon_0 \le \widetilde{O}(\frac{\epsilon}{\sqrt{L}})$, then we can obtain TV error ϵ with a total iteration complexity of $\widetilde{\Theta}(L^3 d/\epsilon^2)$ steps.

Proof. Using Lemma H.3, we can get L. Using Lemma H.2, we can get m_2. Then we directly apply Theorem G.8. □

I.5. Concrete calculation for Theorem G.9

Theorem I.5 (DPUM, formal version of Theorem 5.8)**.** *If the following conditions hold:*

- *Condition 3.4 and Assumption 5.1.*
- *We use the* DPUM *algorithm defined in Definition 4.6, and let $\widehat{q}$ be the output density of it.*

We have

$$\mathrm{TV}(\widehat{q}, p_0) \lesssim (\sqrt{d} + m_2)\exp(-T) + L^2 T d^{1/2} h_{\mathrm{pred}} + L^{3/2} T d^{1/2} h_{\mathrm{corr}}^{1/2} + L^{1/2} T \epsilon_0 + \epsilon.$$

where

- $L = \frac{1}{\sigma_{\min(p_t)}} + \frac{2R^2}{\gamma^2 \sigma^2_{\min(p_t)}} \cdot (\frac{1}{(2\pi)^d \det_{\min(p_t)}} + \frac{1}{(2\pi)^{d/2} \det^{1/2}_{\min(p_t)}}) \cdot \exp(-\frac{\beta^2}{2\sigma_{\max(p_t)}})$,
- $m_2 = (\sum_{i=1}^k \alpha_i (\|\mu_i\|_2^2 + \mathrm{tr}[\Sigma_i]))^{1/2}$.

In particular, if we set $T = \Theta(\log(d m_2/\epsilon))$, $h_{\mathrm{pred}} = \widetilde{\Theta}(\frac{\epsilon}{L^2 d^{1/2}})$, $h_{\mathrm{corr}} = \widetilde{\Theta}(\frac{\epsilon}{L^{3/2} d^{1/2}})$, and if the score estimation error satisfies $\epsilon_0 \le \widetilde{O}(\frac{\epsilon}{\sqrt{L}})$, then we can obtain TV error ϵ with a total iteration complexity of $\widetilde{\Theta}(L^2 d^{1/2}/\epsilon)$ steps.

Proof. Using Lemma H.3, we can get L. Using Lemma H.2, we can get m_2. Then we directly apply Theorem G.9. □