Notations of Mirror Particle-Based Variational Inference

1 Notations

Target distribution and potential function.

- Target distribution and potential function in primal space: $\pi(\mathbf{x}) \propto \exp(-V(\mathbf{x}))$.
- Target distribution and potential function in dual space: $\varpi(\mathbf{y}) \propto \exp(-W(\mathbf{y}))$.
- $\mathbf{x}$ is on a d-dimensional simplex $\Delta_d := \{\mathbf{x} \in \mathbb{R}^d | \sum_i^d x_i \leq 1, x_i \geq 0\}$, and $\mathbf{y} \in \mathbb{R}^d$.

Mirror map, take the entropy function as an example.

- Entropic mirror map: $\phi(\mathbf{x}) = \sum_i^d x_i \log(x_i) + (1 - \sum_i^d x_i) \log(1 - \sum_i^d x_i)$.
- The convex conjugate of $\phi(\mathbf{x})$: $\phi^*(\mathbf{y}) = \sup_{\mathbf{x}\in\Delta_d} \mathbf{y}^T\mathbf{x} - \phi(\mathbf{x}) = \log(1 + \sum_i^d e^{y_i})$.
- Map $\mathbf{x} \to \mathbf{y}$: $\nabla\phi(\mathbf{x}) = \log(\mathbf{x}) - \log(1 - \sum_i^d x_i)\mathbf{1}$.
- Map $\mathbf{y} \to \mathbf{x}$: $\nabla\phi^*(\mathbf{y}) = e^{\mathbf{y}}/(1 + \sum_i^d e^{y_i})$.
- $\nabla\phi$ is a bijection from Δ_d to $\mathbb{R}^d$ with inverse map $(\nabla\phi)^{-1} = \nabla\phi^*$.
- $\nabla^2\phi(\mathbf{x}) = \mathrm{diag}(1/\mathbf{x}) + \mathbf{1}\mathbf{1}^T/(1 - \sum_i^d x_i)$.
- $\nabla^2\phi(\mathbf{x})^{-1} = \mathrm{diag}(\mathbf{x}) - \mathbf{x}\mathbf{x}^T$.
- $\log\det(\nabla^2\phi(\mathbf{x})) = -\log\left(1 - \sum_{i=1}^d x_i\right) - \sum_{i=1}^d \log(x_i)$.
- $\log\det(\nabla^2\phi^*(\mathbf{y})) = \sum_{i=1}^d y_i - (d+1)\phi^*(\mathbf{y})$.

Relation between π and ϖ Hsieh et al. [2018].

- The gradient map $\nabla\phi$ induces a new probability measure $\mathrm{d}\varpi := e^{-W(\mathbf{y})}\mathrm{d}\mathbf{y}$ through $\varpi(E) = \pi(\nabla\phi^{-1}(E))$ for every Borel set E on $\mathbb{R}^d$. We say that ϖ is a push-forward measure of π under $\nabla\phi$.
- The potential functions have the following relation

$$e^{-V(\mathbf{x})} = e^{-W(\nabla\phi(\mathbf{x}))} \det(\nabla^2\phi(\mathbf{x})), \tag{1}$$

$$e^{-W(\mathbf{y})} = e^{-V(\nabla\phi^*(\mathbf{y}))} \det(\nabla^2\phi^*(\mathbf{y})). \tag{2}$$

2 Existing Algorithms

2.1 Langevin dynamics

$$\mathrm{d}\mathbf{x}_t = -\nabla_{\mathbf{x}} V(\mathbf{x}_t)\mathrm{d}t + \sqrt{2}\mathrm{d}\mathbf{B}_t. \tag{3}$$

Denote the evolutionary distribution as μ_t, the corresponding PDE is defined as

$$\partial_t\mu_t + \nabla \cdot (\mu_t \nabla \log(\frac{\pi}{\mu_t})) = 0. \tag{4}$$

2.2 Riemannian Langevin Dynamics

$$\mathrm{d}\mathbf{x}_t = \left(-G(\mathbf{x}_t)^{-1}\nabla_{\mathbf{x}} V(\mathbf{x}_t) + \nabla \cdot G(\mathbf{x}_t)^{-1}\right) \mathrm{d}t + \sqrt{2} G(\mathbf{x}_t)^{-1/2} \mathrm{d}\mathbf{B}_t, \tag{5}$$

where G is the Riemannian metric tensor. The corresponding PDE is defined as Ma et al. [2015]

$$\partial_t \mu_t + \nabla \cdot (\mu_t G^{-1} \nabla \log(\frac{\pi}{\mu_t})) = 0. \tag{6}$$

2.3 Mirror-Langevin Dynamics

Zhang et al. [2020], Ahn and Chewi [2020], Chewi et al. [2020], Jiang, Li et al. [2021]

$$\begin{cases} \mathrm{d}\mathbf{y}_t = & -\nabla_{\mathbf{x}} V(\mathbf{x}_t)\mathrm{d}t + \sqrt{2}\nabla^2 \phi(\mathbf{x})^{1/2} \mathrm{d}\mathbf{B}_t, \\ \mathbf{x}_t = & \nabla \phi^*(\mathbf{y}_t). \end{cases} \tag{7}$$

The Riemannian Langevin Dynamics (5) *yields the mirror-Langevin Dynamics* (7) *with* $G = \nabla^2 \phi$*.* The corresponding PDE has the following form Chewi et al. [2020]

$$\partial_t \mu_t + \nabla \cdot (\mu_t (\nabla^2 \phi)^{-1} \nabla \log(\frac{\pi}{\mu_t})) = 0. \tag{8}$$

2.4 Mirrored Langevin Dynamics

Hsieh et al. [2018]

$$\begin{cases} \mathrm{d}\mathbf{y}_t = & -\nabla_{\mathbf{y}} W(\mathbf{y}_t)\mathrm{d}t + \sqrt{2} \mathrm{d}\mathbf{B}_t, \\ \mathbf{x}_t = & \nabla \phi^*(\mathbf{y}_t). \end{cases} \tag{9}$$

Denote the evolutionary distribution in the dual space as ν_t, the corresponding PDE in the dual space is defined as

$$\partial_t \nu_t + \nabla \cdot (\nu_t \nabla \log(\frac{\varpi}{\nu_t})) = 0. \tag{10}$$

2.5 Mirrored Stein Operator

Shi et al. [2021] derives the following mirrored Stein operator from the generator of the mirror-Langevin dynamics

$$(\mathcal{M}_{\pi,\phi} g)(\mathbf{x}) = g(\mathbf{x})^T \nabla^2 \phi(\mathbf{x})^{-1} \nabla \log(\pi(\mathbf{x})) + \nabla \cdot (\nabla^2 \phi(\mathbf{x})^{-1} g(\mathbf{x})), \tag{11}$$

where g is a vector-valued function. The dissipation with mirrored Stein operator (11) (**NOTE: no direct proof**)

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbf{KL}(\mu_t || \pi) = -\mathbb{E}_{\mu_t} \left[(\mathcal{M}_{\pi,\phi} g_t)(\cdot) \right]. \tag{12}$$

The optimal g_t is

$$g_t = \mathbb{E}_{\mathbf{x}_t \sim \mu_t} \left[(\mathcal{M}_{\pi,\phi} K(\cdot, \mathbf{x}_t))(\cdot) \right], \tag{13}$$

where $\mathcal{M}_{\pi,\phi} K(\cdot, \mathbf{x}_t)$ applies $\mathcal{M}_{\pi,\phi}$ to each row of the matrix-valued function $K_{\mathbf{x}} = K(\cdot, \mathbf{x})$. The proposed dynamics is defined as

$$\begin{cases} \mathrm{d}\mathbf{y}_t = & g_t(\mathbf{x}_t)\mathrm{d}t, \\ \mathbf{x}_t = & \nabla \phi^*(\mathbf{y}_t) \end{cases} \tag{14}$$

2.6 Mirrored Stein Variational Gradient Descent

$$\begin{cases} \mathrm{d}\mathbf{y}_t = & \mathbb{E}_{\mathbf{x}' \sim \mu_t} \left[K(\mathbf{x}_t, \mathbf{x}') \nabla^2 \phi(\mathbf{x}')^{-1} \nabla \log(\pi(\mathbf{x}')) + \nabla_{\mathbf{x}'} \cdot (K(\mathbf{x}_t, \mathbf{x}') \nabla^2 \phi(\mathbf{x}')^{-1}) \right] \mathrm{d}t, \\ \mathbf{x}_t = & \nabla \phi^*(\mathbf{y}_t), \end{cases} \tag{15}$$

where $K(\mathbf{x}_t, \mathbf{x}') = k(\mathbf{x}_t, \mathbf{x}')\mathbf{I}$.

With kernel $k_\phi(\mathbf{y}_t, \mathbf{y}') = k(\nabla\phi^*(\mathbf{y}_t), \nabla\phi^*(\mathbf{y}'))$ (15) equals (MSVGD-with-primal-kernel)

$$\begin{cases} \mathrm{d}\mathbf{y}_t = \mathbb{E}_{\mathbf{y}'\sim\nu_t}\left[k_\phi(\mathbf{y}_t, \mathbf{y}')\nabla\log(\varpi(\mathbf{y}')) + \nabla_{\mathbf{y}'}k_\phi(\mathbf{y}', \mathbf{y}_t)\right]\mathrm{d}t, \\ \mathbf{x}_t = \nabla\phi^*(\mathbf{y}_t), \\ \nabla_{\mathbf{y}'}k_\phi(\mathbf{y}', \mathbf{y}_t) = \nabla^2\phi^*(\mathbf{y}')\nabla_{\mathbf{x}'}k(\mathbf{x}', \mathbf{x}_t) = \nabla^2\phi(\mathbf{x}')^{-1}\nabla_{\mathbf{x}'}k(\mathbf{x}', \mathbf{x}_t) \end{cases} \tag{16}$$

Replace the kernel $k_\phi(\mathbf{y}', \mathbf{y}_t)$ with $k(\mathbf{y}', \mathbf{y}_t)$, based on (16), we have (MSVGD-with-dual-kernel)

$$\begin{cases} \mathrm{d}\mathbf{y}_t = \mathbb{E}_{\mathbf{y}'\sim\nu_t}\left[k(\mathbf{y}_t, \mathbf{y}')\nabla\log(\varpi(\mathbf{y}')) + \nabla_{\mathbf{y}'}k(\mathbf{y}', \mathbf{y}_t)\right]\mathrm{d}t, \\ \mathbf{x}_t = \nabla\phi^*(\mathbf{y}_t), \end{cases} \tag{17}$$

which is related to the Mirrored Langevin Dynamics (9).

2.7 Stein Variational Mirror Descent

Matrix kernel

$$K_{\phi,t}(\mathbf{x}_1, \mathbf{x}_2) = \mathbb{E}_{\mathbf{x}_t\sim\mu_t}\left[k^{1/2}(\mathbf{x}_1, \mathbf{x}_t)\nabla^2\phi(\mathbf{x}_t)k^{1/2}(\mathbf{x}_t, \mathbf{x}_2)\right]. \tag{18}$$

3 Symmetric and Smoothing Vector Field

3.1 Riemannian LD in dual space

Recall that the SDE of Mirror-Langevin Dynamics is

$$\mathrm{d}\mathbf{y}_t = -\nabla_{\mathbf{x}}V(\mathbf{x}_t)\mathrm{d}t + \sqrt{2}\nabla^2\phi(\mathbf{x})^{1/2}\mathrm{d}\mathbf{B}_t, \tag{19}$$

which equals Riemannian Langevin dynamics in dual space

$$\mathrm{d}\mathbf{y}_t = \left(-\nabla^2\phi^*(\mathbf{y}_t)^{-1}\nabla_{\mathbf{y}}W(\mathbf{y}_t) + \nabla\cdot\nabla^2\phi^*(\mathbf{y}_t)^{-1}\right)\mathrm{d}t + \sqrt{2}\nabla^2\phi^*(\mathbf{y}_t)^{-1/2}\mathrm{d}\mathbf{B}_t. \tag{20}$$

And the corresponding PDE is

$$\partial_t\nu_t + \nabla\cdot(\nu_t(\nabla^2\phi^*)^{-1}\nabla\log(\frac{\varpi}{\nu_t})) = 0. \tag{21}$$

Proof.

$$\nabla^2\phi(\mathbf{x}_t)^{1/2} = \nabla^2\phi^*(\mathbf{y}_t)^{-1/2} \tag{22}$$

$$\begin{aligned} e^{-W(\mathbf{y})} =& e^{-V(\nabla\phi^*(\mathbf{y}))}\det(\nabla^2\phi^*(\mathbf{y})) \\ -\nabla_{\mathbf{y}}W(\mathbf{y}) =& -\nabla_{\mathbf{y}}V(\nabla\phi^*(\mathbf{y})) + \nabla_{\mathbf{y}}\log\det(\nabla^2\phi^*(\mathbf{y})) \\ =& -\nabla^2\phi^*(\mathbf{y})\nabla_{\mathbf{x}}V(\mathbf{x}) + \nabla_{\mathbf{y}}\log\det(\nabla^2\phi^*(\mathbf{y})) \\ -\nabla^2\phi^*(\mathbf{y})\nabla_{\mathbf{x}}V(\mathbf{x}) =& -\nabla_{\mathbf{y}}W(\mathbf{y}) - \nabla_{\mathbf{y}}\log\det(\nabla^2\phi^*(\mathbf{y})) \\ -\nabla_{\mathbf{x}}V(\mathbf{x}) =& -\nabla^2\phi^*(\mathbf{y})^{-1}\nabla_{\mathbf{y}}W(\mathbf{y}) - \nabla^2\phi^*(\mathbf{y})^{-1}\nabla_{\mathbf{y}}\log\det(\nabla^2\phi^*(\mathbf{y})) \end{aligned}$$

Now we need to proof $-\nabla^2\phi^*(\mathbf{y})^{-1}\nabla_{\mathbf{y}}\log\det(\nabla^2\phi^*(\mathbf{y})) = \nabla_{\mathbf{y}}\cdot\nabla^2\phi^*(\mathbf{y})^{-1}$. For the k-th dimensional, we have

$$\begin{aligned} &\left[\nabla_{\mathbf{y}}\log\det(\nabla^2\phi^*(\mathbf{y}))\right]_k + \left[\nabla^2\phi^*(\mathbf{y})\nabla_{\mathbf{y}}\cdot\nabla^2\phi^*(\mathbf{y})^{-1}\right]_k \\ =&\sum_{a=1}^d\sum_{b=1}^d\nabla_{y_a}\nabla^2\phi^*(\mathbf{y})_{kb}\left[\nabla^2\phi^*(\mathbf{y})^{-1}\right]_{ba} + \sum_{a=1}^d\sum_{b=1}^d\nabla^2\phi^*(\mathbf{y})_{kb}\nabla_{y_a}\left[\nabla^2\phi^*(\mathbf{y})^{-1}\right]_{ba} \\ =&\sum_{a=1}^d\nabla_{y_a}\left(\sum_{b=1}^d\nabla^2\phi^*(\mathbf{y})_{kb}\left[\nabla^2\phi^*(\mathbf{y})^{-1}\right]_{ba}\right) \\ =&\sum_{a=1}^d\nabla_{y_a}\mathbf{I}_{ka} = 0 \end{aligned}$$

□

3.2 Smoothing vector field with kernel function

Consider the following PDE

$$\partial_t \nu_t + \nabla \cdot (\nu_t \nabla \log(\frac{\varpi}{\nu_t})) = 0, \tag{23}$$

which can be solved via the following ODE

$$\begin{aligned}\mathrm{d}\mathbf{y}_t =& [\nabla_{\mathbf{y}} \log(\varpi(\mathbf{y}_t)) - \nabla_{\mathbf{y}} \log(\nu_t(\mathbf{y}_t))] \,\mathrm{d}t \\ =& \nabla^2\phi(\mathbf{x}_t)^{-1} [\nabla_{\mathbf{x}} \log(\pi(\mathbf{x}_t)) - \nabla_{\mathbf{x}} \log(\mu_t(\mathbf{x}_t))] \,\mathrm{d}t.\end{aligned}$$

Smoothing $\nabla^2\phi(\mathbf{x}_t)^{-1} [\nabla_{\mathbf{x}} \log(\pi(\mathbf{x}_t)) - \nabla_{\mathbf{x}} \log(\mu_t(\mathbf{x}_t))]$:

$$\begin{aligned}&\int k(\mathbf{x}', \mathbf{x}) \nabla^2\phi(\mathbf{x}')^{-1} [\nabla_{\mathbf{x}} \log(\pi(\mathbf{x}_t)) - \nabla_{\mathbf{x}} \log(\mu_t(\mathbf{x}_t))] \,\mathrm{d}\mu_t \\ =& \mathbb{E}_{\mathbf{x}'\sim\mu_t} \left[k(\mathbf{x}', \mathbf{x}) \nabla^2\phi(\mathbf{x}')^{-1} \nabla \log(\pi(\mathbf{x}'))\right] - \int k(\mathbf{x}', \mathbf{x}) \nabla^2\phi(\mathbf{x}')^{-1} \nabla \log(\mu_t(\mathbf{x}')) \mathrm{d}\mu_t \\ =& \mathbb{E}_{\mathbf{x}'\sim\mu_t} \left[k(\mathbf{x}', \mathbf{x}) \nabla^2\phi(\mathbf{x}')^{-1} \nabla \log(\pi(\mathbf{x}'))\right] + \int \nabla \cdot \left(k(\mathbf{x}', \mathbf{x}) \nabla^2\phi(\mathbf{x}')^{-1}\right) \mu_t(\mathbf{x}') \mathrm{d}\mathbf{x}' \\ =& \mathbb{E}_{\mathbf{x}'\sim\mu_t} \left[k(\mathbf{x}', \mathbf{x}) \nabla^2\phi(\mathbf{x}')^{-1} \nabla \log(\pi(\mathbf{x}')) + \nabla \cdot \left(k(\mathbf{x}', \mathbf{x}) \nabla^2\phi(\mathbf{x}')^{-1}\right)\right],\end{aligned}$$

which equals the update direction of Mirrored Stein Variational Gradient Descent.

4 Mirror ParVI with KL

Consider the following PDE

$$\partial_t \nu_t + \nabla \cdot (\nu_t (\nabla^2\phi^*)^{-1} \nabla \log(\frac{\varpi}{\nu_t})) = 0, \tag{24}$$

which can be solved via the following ODE

$$\mathrm{d}\mathbf{y}_t = \nabla^2\phi^*(\mathbf{y}_t)^{-1} [\nabla_{\mathbf{y}} \log(\varpi(\mathbf{y}_t)) - \nabla_{\mathbf{y}} \log(\nu_t(\mathbf{y}_t))] \,\mathrm{d}t. \tag{25}$$

We have

$$\begin{aligned}&\nabla^2\phi^*(\mathbf{y}_t)^{-1} [\nabla_{\mathbf{y}} \log(\varpi(\mathbf{y}_t)) - \nabla_{\mathbf{y}} \log(\nu_t(\mathbf{y}_t))] \\ =& \nabla^2\phi^*(\mathbf{y}_t)^{-1} \big[\nabla_{\mathbf{y}} \log(\pi(\nabla\phi^*(\mathbf{y}_t))) + \nabla_{\mathbf{y}} \log\det(\nabla^2\phi^*(\mathbf{y}_t)) - \nabla_{\mathbf{y}} \log(\mu_t(\nabla\phi^*(\mathbf{y}_t))) \\ &- \nabla_{\mathbf{y}} \log\det(\nabla^2\phi^*(\mathbf{y}_t))\big] \\ =& \nabla^2\phi^*(\mathbf{y}_t)^{-1} \nabla^2\phi^*(\mathbf{y}_t) [\nabla_{\mathbf{x}} \log(\pi(\mathbf{x}_t)) - \nabla_{\mathbf{x}} \log(\mu_t(\mathbf{x}_t))] \\ =& \nabla_{\mathbf{x}} \log(\pi(\mathbf{x}_t)) - \nabla_{\mathbf{x}} \log(\mu_t(\mathbf{x}_t))\end{aligned}$$

Thus, we have

$$\mathrm{d}\mathbf{y}_t = (\nabla_{\mathbf{x}} \log(\pi(\mathbf{x}_t)) - \nabla_{\mathbf{x}} \log(\mu_t(\mathbf{x}_t))) \,\mathrm{d}t \tag{26}$$

4.1 Mirror Functional Descent with Smoothing in Dual Space

We find that updating particles with smoothed (26) performs poorly when the mode of target distribution lies at the boundary, which be blamed for the contradictory between constrained support and the unconstrained kernel function in the primal space. Thus, we investigate kernel functions defined on the constrained domain (such as a simplex), or perform smoothing in the dual space.

We have the following results

$$\begin{aligned}\mathrm{d}\mathbf{y}_t &= \nabla^2\phi^*(\mathbf{y}_t)^{-1} [\nabla_{\mathbf{y}} \log(\varpi(\mathbf{y}_t)) - \nabla_{\mathbf{y}} \log(\nu_t(\mathbf{y}_t))] \,\mathrm{d}t, \\ &= \nabla^2\phi^*(\mathbf{y}_t)^{-1} \left[\nabla^2\phi(\mathbf{x}_t)^{-1} \left(\nabla_{\mathbf{x}} \log(\pi(\mathbf{x}_t)) - \nabla_{\mathbf{x}} \log\det(\nabla^2\phi(\mathbf{x}))\right) - \nabla_{\mathbf{y}} \log(\nu_t(\mathbf{y}_t))\right] \\ &= \left[\nabla_{\mathbf{x}} \log(\pi(\mathbf{x}_t)) - \nabla_{\mathbf{x}} \log\det(\nabla^2\phi(\mathbf{x})) - \nabla^2\phi^*(\mathbf{y}_t)^{-1} \nabla_{\mathbf{y}} \log(\nu_t(\mathbf{y}_t))\right] \mathrm{d}t,\end{aligned}$$

which can be directly smoothed with kernel functions in the dual space.

References

Kwangjun Ahn and Sinho Chewi. Efficient constrained sampling via the mirror-langevin algorithm. *arXiv preprint arXiv:2010.16212*, 2020.

Sinho Chewi, Thibaut Le Gouic, Chen Lu, Tyler Maunu, Philippe Rigollet, and Austin J Stromme. Exponential ergodicity of mirror-langevin diffusions. *arXiv preprint arXiv:2005.09669*, 2020.

Ya-Ping Hsieh, Ali Kavis, Paul Rolland, and Volkan Cevher. Mirrored langevin dynamics. *arXiv preprint arXiv:1802.10174*, 2018.

Qijia Jiang. Mirror langevin monte carlo: the case under isoperimetry.

Ruilin Li, Molei Tao, Santosh S Vempala, and Andre Wibisono. The mirror langevin algorithm converges with vanishing bias. *arXiv preprint arXiv:2109.12077*, 2021.

Yi-An Ma, Tianqi Chen, and Emily B Fox. A complete recipe for stochastic gradient mcmc. *arXiv preprint arXiv:1506.04696*, 2015.

Jiaxin Shi, Chang Liu, and Lester Mackey. Sampling with mirrored stein operators. *arXiv preprint arXiv:2106.12506*, 2021.

Kelvin Shuangjian Zhang, Gabriel Peyré, Jalal Fadili, and Marcelo Pereyra. Wasserstein control of mirror langevin monte carlo. In *Conference on Learning Theory*, pages 3814–3841. PMLR, 2020.