Sliced-Wasserstein Estimation with Spherical Harmonics as Control Variates

Rémi Leluc[1] Aymeric Dieuleveut[1] François Portier[2] Johan Segers[3] Aigerim Zhuman[3]

[1]CMAP, École Polytechnique, IPParis [2]CREST, ENSAI [3]LIDAM, ISBA, UCLouvain

Sliced-Wasserstein Distance (SW)

For probability measures $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$,

$$\mathrm{SW}_p^p(\mu,\nu,\mathrm{P}) = \int_{\mathbb{S}^{d-1}} \mathrm{W}_p^p(\theta_\sharp^\star \mu, \theta_\sharp^\star \nu)\, \mathrm{d}\,\mathrm{P}(\theta)$$

$\mathrm{P} \sim \mathcal{U}(\mathbb{S}^{d-1})$, integrand $f_{\mu,\nu}^{(p)} : \mathbb{S}^{d-1} \to \mathbb{R}$,

$$f_{\mu,\nu}^{(p)}(\theta) = \mathrm{W}_p^p(\theta_\sharp^\star \mu, \theta_\sharp^\star \nu)$$

Monte Carlo: Sample $\theta_i \sim \mathrm{P}$ and average $(f_{\mu,\nu}^{(p)}(\theta_i))_i$.

GOAL: Improve SW distance computation by improving the MC estimation using **Control Variates**.

Control Variates and OLSMC

Integral $\mathrm{I}(f)$ of square-integrable integrand $f \in L_2(\mathrm{P})$ on $(\Theta, \mathcal{F}, \mathrm{P})$ is approximated with $\theta_1, \ldots, \theta_n \sim \mathrm{P}$

$$\mathrm{I}(f) = \int_\Theta f(\theta)\mathrm{d}\,\mathrm{P}(\theta), \quad \mathrm{I}_n(f) = \frac{1}{n}\sum_{i=1}^n f(\theta_i).$$

$\varphi = (\varphi_1, \ldots, \varphi_s)^\top$ are **Control Variates**: $\mathrm{I}(\varphi_k) = 0$

For $\beta \in \mathbb{R}^s$, $\mathrm{I}(f - \beta^\top \varphi) = \mathrm{I}(f)$ yielding CV estimate

$$\mathrm{I}_n^{(\mathrm{cv})}(f,\beta) = \frac{1}{n}\sum_{i=1}^n (f(\theta_i) - \beta^\top \varphi(\theta_i)).$$

OLS framework: $\mathrm{I}(f)$ is the intercept of the LR model with features $\varphi_1, \ldots, \varphi_s$ and target response f,

$$(\mathrm{I}(f), \beta_\star(f)) \in \underset{(\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^s}{\arg\min}\ \mathrm{I}[(f - \alpha - \beta^\top\varphi)^2].$$

- **Ordinary Least Squares** Monte Carlo (OLSMC)

$$(\mathrm{I}_n^{\mathrm{ols}}(f), \beta_n(f)) \in \underset{(\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^s}{\arg\min}\ \|f_n - \alpha \mathbb{1}_n - \Phi\beta\|_2^2$$

$f_n = (f(\theta_1), \ldots, f(\theta_n))^\top \in \mathbb{R}^n$, $\mathbb{1}_n = (1, \ldots, 1)^\top \in \mathbb{R}^n$, $\Phi \in \mathbb{R}^{n\times s}$ is matrix of control variates $\Phi = (\varphi(\theta_i)^\top)_{i=1}^n$.

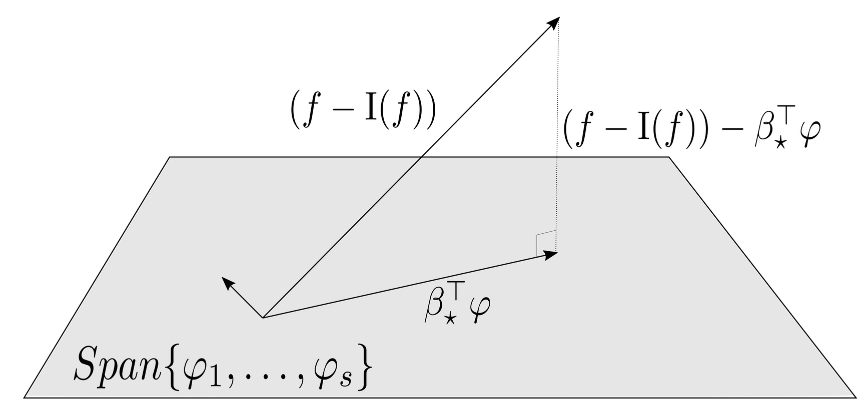

Figure 1: $L_2(\mathrm{P})$ projection of f onto $Span\{\varphi_1, \ldots, \varphi_s\}$.

Spherical Harmonics

The **Spherical Harmonics** $\{\varphi_{\ell,k} : \ell \geq 0, 1 \leq k \leq N_\ell^d\}$ form an orthonormal basis of the Hilbert space $L_2(\mathbb{S}^{d-1})$ so that for every $f \in L_2(\mathbb{S}^{d-1})$ we have

$$f = \sum_{\ell=0}^\infty \sum_{k=1}^{N_\ell^d} \hat{f}_{\ell,k}\varphi_{\ell,k} \quad \text{where} \quad \hat{f}_{\ell,k} = \int f \varphi_{\ell,k}\,\mathrm{d}\,\mathrm{P}\,.$$

$$\mathrm{I}(\varphi_{\ell,k}) = \int_{\mathbb{S}^{d-1}} \varphi_{\ell,k}(\theta)\,\mathrm{d}\,\mathrm{P}(\theta) = 0$$

SHCV Estimator

The SHCV estimate of maximum degree $2L$ is the OLSMC estimate with all spherical harmonics of even degree from 2 up to $2L$ as covariate matrix

$$\mathrm{SHCV}_{n,L}^p(\mu,\nu) = \mathrm{I}_n^{\mathrm{ols}}(f_{\mu,\nu}^{(p)})$$

(Linear rule) SHCV estimate can be represented as a linear rule $w^\top f_n$, where the weight vector $w \in \mathbb{R}^n$ **does not depend on the integrands**.

(Computing time) For K integrals, SHCV in $\mathcal{O}(Kn\omega_f + \omega(\Phi))$ compared to $\mathcal{O}(Kn\omega_f)$ for MC.

Sliced-Wasserstein Algorithms

Algorithm 1: Sliced Wasserstein Monte Carlo

Require: $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, number of random projections n

1: Sample random projections $\theta_1, \ldots, \theta_n \sim \mathrm{P}$
2: Compute $f_n = (f_{\mu,\nu}^{(p)}(\theta_i))_{i=1}^n$
3: Return average $\mathrm{MC}_n = (\mathbb{1}_n/n)^\top f_n$

Algorithm 2: Spherical Harmonics Control Variate

Require: $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, number of random projections n, spherical harmonics $\varphi = (\varphi_j)_{j=1}^s$

1: Sample random projections $\theta_1, \ldots, \theta_n \sim \mathrm{P}$
2: Compute $f_n = (f_{\mu,\nu}^{(p)}(\theta_i))_{i=1}^n$, $\Phi = (\varphi(\theta_i)^\top)_{i=1}^n$
3: Solve $(\mathrm{I}_n^{\mathrm{ols}}, \beta_n) \in \arg\min_{\alpha,\beta} \|f_n - \alpha\mathbb{1}_n - \Phi\beta\|_2^2$
4: Return $\mathrm{SHCV}_n = \mathrm{I}_n^{\mathrm{ols}}$

Theoretical Properties

For Gaussians $\mu = \mathcal{N}(a, \mathbf{A})$ and $\nu = \mathcal{N}(b, \mathbf{B})$

$$f_{\mu,\nu}^{(2)}(\theta) = |\theta^\top(a-b)|^2 + \left(\sqrt{\theta^\top \mathbf{A}\theta} - \sqrt{\theta^\top \mathbf{B}\theta}\right)^2$$

(Exact Rule) If $f_{\mu,\nu}^{(p)}$ is a polynomial of degree m, considering the SHCV estimate and control variates $\varphi = (\varphi_j)_{j=1}^{s_{L,d}}$, if $2L \geq m$ and $n > s_{L,d}$ then SHCV is exact: $\mathrm{SHCV}_{n,L}^p(\mu,\nu) = \mathrm{SW}_p^p(\mu,\nu)$.

(Affine transform) If $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ are related by $X \sim \mu$ and $\alpha X + b \sim \nu$ where $\alpha \in (0,\infty)$ and $b \in \mathbb{R}^d$ then the SHCV estimate is exact.

(Mean invariance) For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, the error of the SHCV method is (exactly) invariant under changes of the mean vectors m_μ and m_ν of μ and ν respectively.

Asymptotic Error Bound

Theorem 1 (Convergence rate). *Let $d \geq 2$, $p \in [1, \infty)$ and $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ be fixed. For any degree sequence $L = L_n$ such that $L = o(n^{1/(2(d-1))})$ as $n \to \infty$, the integration error satisfies*

$$\left|\mathrm{SHCV}_{n,L}^p(\mu,\nu) - \mathrm{SW}_p^p(\mu,\nu)\right| = \mathcal{O}_{\mathbb{P}}(L^{-1}n^{-1/2})$$

- For $d = 3$, with $L = n^{1/(2(d-1))}/\ell_n$ where $\ell_n \to \infty$ slowly, this yields the rate $n^{-3/4+o(1)}$ for the SHCV estimate, in comparison to the Monte Carlo rate $n^{-1/2}$.

Related Methods

- MC: standard MC estimate.
- CV_{low} and CV_{up}: the lower-CV and upper-CV estimates of [3] based on lower and upper bounds of a Gaussian approximation.
- CVNN: estimate of [1] based on nearest neighbors estimates acting as control variates.
- RQMC: (Randomized) Quasi Monte Carlo as in [2].
- SHCV: proposed estimate with Spherical Harmonics as Control Variates.

Numerical Experiments

(Gaussian) $\mathrm{SW}_2^2(\mu_m, \nu_m)$ with $\mu_m = m^{-1}\sum_{i=1}^m \delta_{x_i}$ and $\nu_m = m^{-1}\sum_{j=1}^m \delta_{y_j}$, $x_i \sim \mu = \mathcal{N}(a, \mathbf{A})$, $y_j \sim \nu = \mathcal{N}(b, \mathbf{B})$, $m = 1000$, means $a, b \sim \mathcal{N}_d(1_d, I_d)$ and covariance $\mathbf{A} = \Sigma_a\Sigma_a^\top$ and $\mathbf{B} = \Sigma_b\Sigma_b^\top$, entries of Σ_a, Σ_b drawn from $\mathcal{N}(0, 1)$.

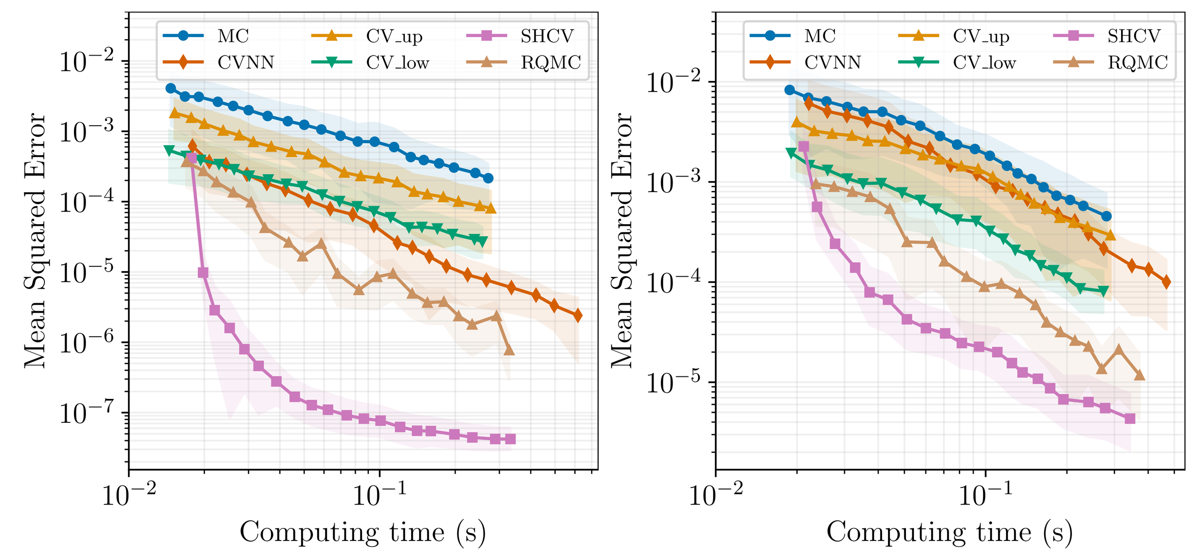

Figure 2: MSE for sampled Gaussian distributions supported on $m = 1000$ points, dimension $d \in \{3; 6\}$ (left/right).

(3D Point Clouds) dataset ShapeNetCore corresponding to the objects plane, lamp, and bed, each composed of $m = 2048$ points in $\mathbb{R}^3$.

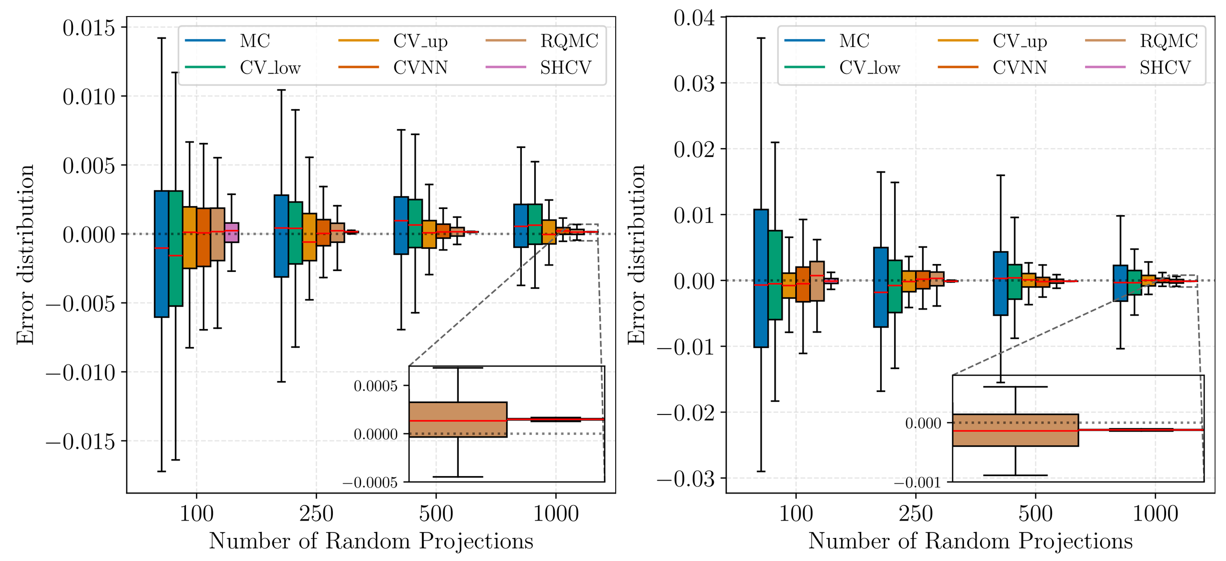

Figure 3: Boxplots of error $\widehat{\mathrm{SW}}_n(\mu_m, \nu_m) - \mathrm{SW}(\mu_m, \nu_m)$ for different SW estimates based on n random projections with $n \in \{100; 250; 500; 1000\}$ obtained over 100 independent runs.

References

[1] R. Leluc, F. Portier, J. Segers, and A. Zhuman. Speeding up Monte Carlo integration: Control neighbors for optimal convergence. *Bernoulli*, 2024.

[2] K. Nguyen, N. Bariletto, and N. Ho. Quasi-Monte Carlo for 3D Sliced Wasserstein. In *The Twelfth International Conference on Learning Representations*, 2024.

[3] K. Nguyen and N. Ho. Sliced Wasserstein Estimation with Control Variates. In *The Twelfth International Conference on Learning Representations*, 2024.