Sliced-Wasserstein Estimation with Spherical Harmonics as Control Variates

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

[1] *Ecole Polytechnique, Institut Polytechnique de Paris, France*
[2] *CREST, ENSAI* [3] *LIDAM, ISBA, UCLouvain*

Paper Code

41^{st} ***I**nternational **C**onference on **M**achine **L**earning, 2024.*

Motivation: Comparing Measures and Spaces

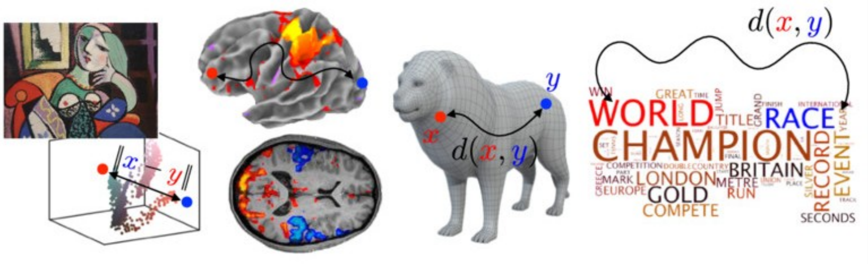
- **Probability distributions and histograms**

→ *images, vision, graphics, machine learning*

- **Optimal Transport**

(Monge, 1781; Kantorovich, 1942; Koopmans, 1949; Dantzig, 1951; Brenier, 1991; Otto, 2001; Villani et al., 2009; Figalli et al., 2010)

→ *takes into account metric d*



(Illustration from slides of Gabriel Peyré)

Motivation: Approximate Distance for OT

The Sliced-Wasserstein (SW) distance shares similar topological properties with the standard Wasserstein distance while having

better properties in terms of computational complexity

$\to \mathrm{W}_p(\mu_m, \nu_m)$ for discrete distributions μ_m and ν_m supported on m points, the worst-case computational complexity scales as $\mathcal{O}(m^3 \log m)$

$\to \mathrm{SW}_p(\mu_m, \nu_m)$ leverages projections and **fast 1d computations**.

Powerful framework for ML problems:

- Generative modeling (Deshpande et al., 2018, 2019; Liutkus et al., 2019)
- Autoencoders (Kolouri et al., 2018)
- Bayesian computation (Nadjahi et al., 2020)
- Image processing (Bonneel et al., 2015).

Ref: Rabin et al. (2012); Bonnotte (2013); Bayraktar and Guo (2021); Nadjahi et al. (2020)

Sliced-Wasserstein (SW) Distance

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

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

$\mathrm{P} \sim \mathcal{U}(\mathbb{S}^{d-1})$ and integrand $f^{(p)}_{\mu,\nu} : \mathbb{S}^{d-1} \to \mathbb{R}$, $f^{(p)}_{\mu,\nu}(\theta) = \mathrm{W}_p^p(\theta^\star_\sharp \mu, \theta^\star_\sharp \nu)$

Let $\theta_1, \ldots, \theta_n \overset{\text{i.i.d.}}{\sim} \mathrm{P}$, the naive MC estimator averages the values $(f^{(p)}_{\mu,\nu}(\theta_i))_i$.

$$\mathrm{I}_n^{\mathrm{mc}}(f) := \frac{1}{n} \sum_{i=1}^{n} f^{(p)}_{\mu,\nu}(\theta_i)$$

Research Goal

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

Ref: Rabin et al. (2012); Nguyen and Ho (2024); Nguyen et al. (2024); Glynn and Szechtman (2002); Oates et al. (2017); Portier and Segers (2019); Leluc et al. (2021); South et al. (2023)

Monte Carlo with Control Variates

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

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

Control Variates

Functions $\varphi_1, \dots, \varphi_s \in L_2(\mathrm{P})$ such that: $\forall 1 \leq j \leq s, \quad \mathrm{I}[\varphi_j] = 0.$

Let $\varphi = (\varphi_1, \dots, \varphi_s)^\top$, for any $\beta \in \mathbb{R}^s$, we have $\mathrm{I}[f - \beta^\top \varphi] = \mathrm{I}[f]$ leading to the CV estimate of $\mathrm{I}(f)$, parameterized by β

CV-Monte Carlo

$$\mathrm{I}_n^{(\mathrm{cv})}(f, \beta) = \frac{1}{n} \sum_{i=1}^n \left(f(X_i) - \beta^\top \varphi(X_i)\right), \quad \theta_1, \dots, \theta_n \sim \mathrm{P}\,.$$

$\rightarrow$ Optimal $\beta^\star$? Minimize the variance

Linear Regression Framework

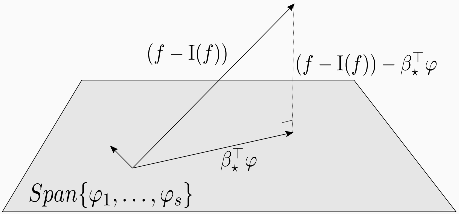
OLS framework: $\mathrm{I}(f)$ is the intercept of the LR model with features $\varphi_1, \ldots, \varphi_s$ and target response f,

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

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 $\Phi = (\varphi(\theta_i)^\top)_{i=1}^n$.



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

Spherical Harmonics

Polynomial spaces

Let $\mathscr{P}_\ell^d$ be the space of homogeneous polynomials of degree $\ell \geq 0$ on $\mathbb{R}^d$, i.e., $\mathscr{P}_\ell^d = \mathrm{Span}\{x_1^{a_1} \cdots x_d^{a_d} \mid a_k \in \mathbb{N}, \sum_{k=1}^d a_k = \ell\}$. Let $\mathscr{H}_\ell^d \subset \mathscr{P}_\ell^d$ be the space of harmonic polynomials: $\mathscr{H}_\ell^d = \{Q \in \mathscr{P}_\ell^d \mid \Delta Q = 0\}$.

Spherical Harmonics of degree $\ell \geq 0$

=

Restriction of elements in $\mathscr{H}_\ell^d$ to the sphere $\mathbb{S}^{d-1}$

Many applications in:

- Physics (*electromagnetic/gravitational fields, electron configurations*)
- Computer Graphics (*global illumination, radiance transfer*)
- Machine Learning (*spherical data representation*)

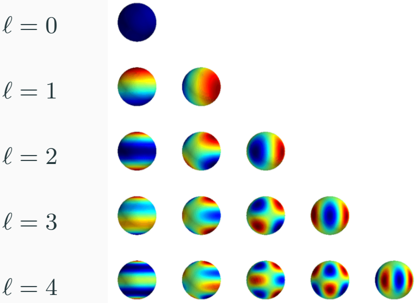
Ref: Atkinson and Han (2012); Dai (2013); Ramamoorthi and Hanrahan (2001); Basri and Jacobs (2003); Green (2003); Cohen et al. (2018); Dutordoir et al. (2020)

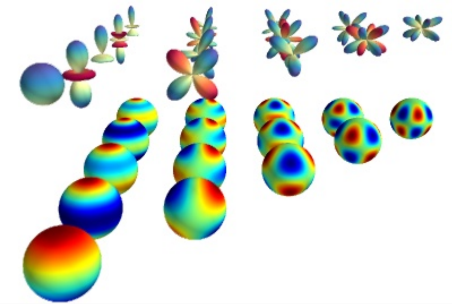
Spherical Harmonics are Control Variates

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

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

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





SHCV estimate

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

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

(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 and the additional cost $\omega(\Phi)$ of fitting the optimal control variates becomes negligible.

Theoretical Properties

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

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

(Exact Rule) If $f^{(p)}_{\mu,\nu}$ 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: $\mathrm{SHCV}^p_{n,L}(\mu,\nu) = \mathrm{SW}^p_p(\mu,\nu)$.

(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 Bound

Theorem (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

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

- 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 $n^{-1/2}$.

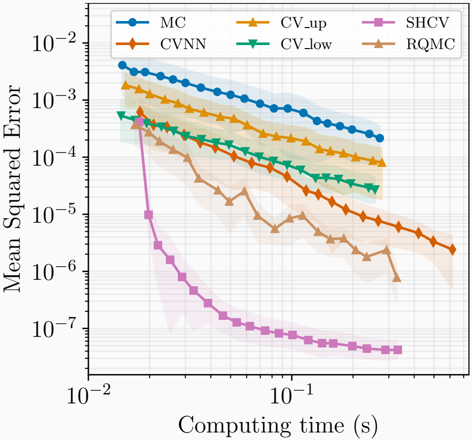
Numerical Experiments

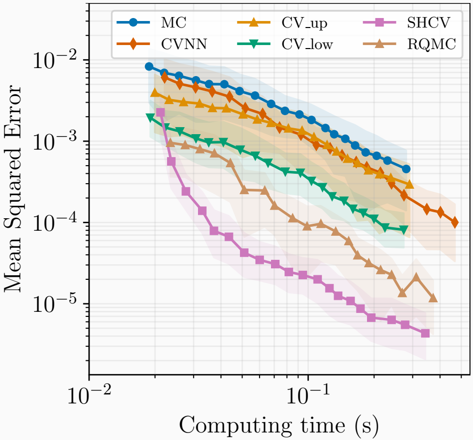
Methods in Competition:

- MC: standard MC estimate.
- CV$_{low}$ and CV$_{up}$: the lower-CV and upper-CV estimates of Nguyen and Ho (2024) based on lower and upper bounds of a Gaussian approximation.
- CVNN: estimate of Leluc et al. (2023) based on nearest neighbors estimates acting as control variates.
- RQMC: (Randomized) Quasi Monte Carlo as in Nguyen et al. (2024).
- SHCV: proposed estimate with **S**pherical **H**armonics as **C**ontrol **V**ariates.

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 $a, b \sim \mathcal{N}_d(\mathbb{1}_d, I_d)$ 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 $\mathcal{N}(0,1)$.





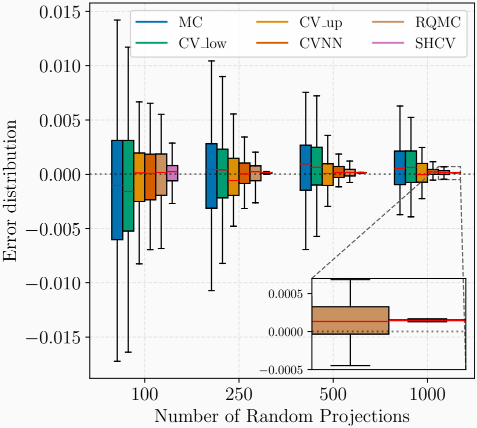
MSE for sampled Gaussian on $m = 1000$ points, dimension $d \in \{3; 6\}$ (left/right).

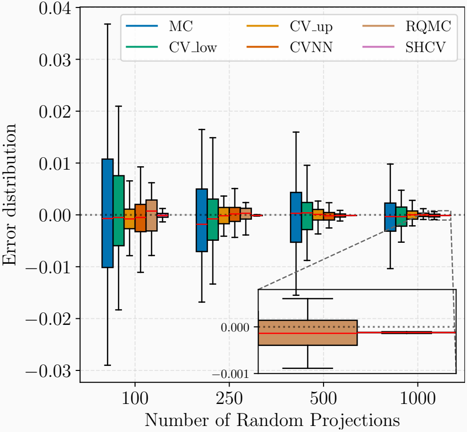
MSE and computing time (ms) for Gaussian distributions in dimension $d \in \{5; 10; 20\}$ based on $n = 500$ projections.

Method	$d = 5$		$d = 10$		$d = 20$	
	MSE	Time	MSE	Time	MSE	Time
MC	1.45e-3	81.1 ± 3.5	9.45e-4	80.7 ± 4.4	1.47e-3	81.1 ± 1.8
CV_{low}	2.67e-4	79.7 ± 1.1	3.45e-4	80.1 ± 1.4	3.82e-4	80.0 ± 1.0
CV_{up}	8.44e-4	83.0 ± 1.2	7.51e-4	83.0 ± 1.7	1.09e-3	83.1 ± 1.5
CVNN	4.29e-4	110 ± 2.2	1.12e-3	122 ± 1.6	2.14e-3	127 ± 1.4
QMC	2.91e-4	100 ± 1.2	2.37e-4	113 ± 1.4	6.60e-4	129 ± 1.4
RQMC	5.80e-5	96.3 ± 2.2	2.75e-4	113 ± 1.2	1.17e-3	130 ± 1.0
SHCV	**2.68e-6**	89.0 ± 6.3	**1.93e-4**	89.0 ± 4.5	**2.95e-4**	88.1 ± 2.8

Numerical Experiments

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





Conclusion and perspectives

Take-home messages

- We have developed a novel method for reducing the variance of MC estimation of the SW distance using spherical harmonics as control variates.
- The excellent practical performance of the SHCV estimate against state-of-the-art baselines is confirmed by theoretical properties and a convergence rate in probability for the integration error.

Perspectives

- In statistical inference with parametric probability measures, note that SHCV is compatible with the computation of gradient $\nabla_\phi \mathrm{SW}_p^p(\mu, \nu_\phi)$ and can be used for *generalized* SW flows (Kolouri et al., 2019).
- The proposed SHCV estimate focuses on the uniform distribution, it can be extended to more general probability distributions by combining control variates with importance sampling techniques as in Leluc et al. (2022).

References

Atkinson, K. and W. Han (2012). *Spherical Harmonics and Approximations on the Unit Sphere: An Introduction*. Lecture Notes in Mathematics. Springer.

Basri, R. and D. W. Jacobs (2003). Lambertian reflectance and linear subspaces. *IEEE transactions on pattern analysis and machine intelligence 25*(2), 218–233.

Bayraktar, E. and G. Guo (2021). Strong equivalence between metrics of Wasserstein type. *Electronic Communications in Probability 26*(none), 1 – 13.

Bonneel, N., J. Rabin, G. Peyré, and H. Pfister (2015). Sliced and radon Wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision 51*, 22–45.

Bonnotte, N. (2013). *Unidimensional and evolution methods for optimal transportation*. Ph. D. thesis, Université Paris Sud-Paris XI; Scuola normale superiore (Pise, Italie).

Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics 44*(4), 375–417.

Cohen, T. S., M. Geiger, J. Köhler, and M. Welling (2018). Spherical CNNs. In *International Conference on Learning Representations*.

Dai, F. (2013). *Approximation theory and harmonic analysis on spheres and balls*. Springer.

Dantzig, G. B. (1951). Application of the simplex method to a transportation problem. *Activity analysis and production and allocation*.

Deshpande, I., Y.-T. Hu, R. Sun, A. Pyrros, N. Siddiqui, S. Koyejo, Z. Zhao, D. Forsyth, and A. G. Schwing (2019). Max-sliced Wasserstein distance and its use for gans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10648–10656.

Deshpande, I., Z. Zhang, and A. G. Schwing (2018). Generative modeling using the sliced Wasserstein distance. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3483–3491.

Dutordoir, V., N. Durrande, and J. Hensman (2020). Sparse Gaussian processes with spherical harmonic features. In *International Conference on Machine Learning*, pp. 2793–2802. PMLR.

Figalli, A., F. Maggi, and A. Pratelli (2010). A mass transportation approach to quantitative isoperimetric inequalities. *Inventiones mathematicae 182*(1), 167–211.

Glynn, P. W. and R. Szechtman (2002). Some new perspectives on the method of control variates. In *Monte Carlo and Quasi-Monte Carlo Methods 2000: Proceedings of a Conference held at Hong Kong Baptist University, Hong Kong SAR, China, November 27–December 1, 2000*, pp. 27–49. Springer.

Green, R. (2003). Spherical harmonic lighting: The gritty details. In *Archives of the game developers conference*, Volume 56, pp. 4.

Kantorovich, L. (1942). On the transfer of masses (in russian). In *Doklady Akademii Nauk*, Volume 37, pp. 227.

Kolouri, S., K. Nadjahi, U. Simsekli, R. Badeau, and G. Rohde (2019). Generalized sliced Wasserstein distances. *Advances in neural information processing systems 32*.

Kolouri, S., P. E. Pope, C. E. Martin, and G. K. Rohde (2018). Sliced Wasserstein auto-encoders. In *International Conference on Learning Representations*.

Koopmans, T. C. (1949). Optimum utilization of the transportation system. *Econometrica: Journal of the Econometric Society*, 136–146.

Leluc, R., F. Portier, and J. Segers (2021). Control variate selection for Monte Carlo integration. *Statistics and Computing 31*(4), 50.

Leluc, R., F. Portier, J. Segers, and A. Zhuman (2022). A Quadrature Rule combining Control Variates and Adaptive Importance Sampling. In *Advances in Neural Information Processing Systems*, Volume 35, pp. 11842–11853. Curran Associates, Inc.

Leluc, R., F. Portier, J. Segers, and A. Zhuman (2023). Speeding up Monte Carlo integration: Control neighbors for optimal convergence. *arXiv preprint arXiv:2305.06151*.

Liutkus, A., U. Simsekli, S. Majewski, A. Durmus, and F.-R. Stöter (2019). Sliced-Wasserstein flows: Nonparametric generative modeling via optimal transport and diffusions. In *International Conference on Machine Learning*, pp. 4104–4113. PMLR.

Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, 666–704.

Nadjahi, K., V. De Bortoli, A. Durmus, R. Badeau, and U. Şimşekli (2020). Approximate Bayesian computation with the sliced-wasserstein distance. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5470–5474. IEEE.

Nadjahi, K., A. Durmus, L. Chizat, S. Kolouri, S. Shahrampour, and U. Simsekli (2020). Statistical and topological properties of sliced probability divergences. *Advances in Neural Information Processing Systems 33*, 20802–20812.

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

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

Oates, C. J., M. Girolami, and N. Chopin (2017). Control functionals for Monte Carlo integration. *Journal of the Royal Statistical Society Series B: Statistical Methodology 79*(3), 695–718.

Otto, F. (2001). The geometry of dissipative evolution equations: the porous medium equation.

Portier, F. and J. Segers (2019). Monte Carlo integration with a growing number of control variates. *Journal of Applied Probability 56*(4), 1168–1186.

Bibliography vi

Rabin, J., G. Peyré, J. Delon, and M. Bernot (2012). Wasserstein barycenter and its application to texture mixing. In *Scale Space and Variational Methods in Computer Vision: Third International Conference, SSVM 2011, Ein-Gedi, Israel, May 29–June 2, 2011, Revised Selected Papers 3*, pp. 435–446. Springer.

Ramamoorthi, R. and P. Hanrahan (2001). An efficient representation for irradiance environment maps. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pp. 497–500.

South, L. F., C. J. Oates, A. Mira, and C. Drovandi (2023). Regularized zero-variance control variates. *Bayesian Analysis 18*(3), 865–888.

Villani, C. et al. (2009). *Optimal transport: old and new*, Volume 338. Springer.