Strong Antithetic Variates: Theory and Applications

Zhankun Luo

November 18, 2025

Machine Intelligence and Networked Data Science Group (MINDS)
Elmore Family School of Electrical and Computer Engineering
Purdue University

Table of Contents

Motivation: Applications of Antithetic Variates

Monte Carlo Integration

- Estimate $\int_0^1 g(t)\mathrm{d}t$ using uniform samples
- Standard MC: $\frac{1}{n}\sum_{i=1}^{n} g(U_i)$ with $U_i \sim \mathrm{Unif}(0,1)$
- Antithetic MC: $\frac{1}{2n}\sum_{i=1}^{n}(g(U_i) + g(1-U_i))$
- **Key benefit:** Variance reduction through negative correlation

Function Approximation

- Approximate $g(\boldsymbol{x}) = \mathbb{E}_{\boldsymbol{Z}}[\phi(\boldsymbol{x}, \boldsymbol{Z})]$
- Antithetic estimator uses pairs $(\boldsymbol{Z}_i, \boldsymbol{a}(\boldsymbol{Z}_i))$
- **Key benefit:** Reduced variance for strongly isotonic functions

Motivation: Applications of Antithetic Variates (Continued)

Concentration Inequalities

- Improved Bernstein's inequality for antithetic samples
- Tighter concentration bounds for Lipschitz functions
- **Key benefit:** Better tail bounds with same sample size

Stochastic Optimization

- Antithetic SGD: $\boldsymbol{x}_{t+1} = \boldsymbol{x}_t - \eta \frac{\boldsymbol{g}(\boldsymbol{x}_t, \boldsymbol{Z}_t) + \boldsymbol{g}(\boldsymbol{x}_t, \boldsymbol{a}(\boldsymbol{Z}_t))}{2}$
- Reduced variance in gradient estimates
- **Key benefit:** Faster convergence, lower noise floor

Motivation: The Challenge

Pratical Applications

- Develop **antithetic maps** and **antithetic indices** as fundamental concepts
- Establish **strong variance reduction inequalities** with explicit bounds
- Connect to optimal transport theory for geometric insights
- Provide theoretical guarantees for applications in integration, approximation, concentration, and optimization

Existing Limitations

- Analyssis of Antithetic variates is limited to one-dimensional random variables,
- Or only limited to a special class of strongly isotonic functions.
- No theoretical guarantees for general multivariate functions and random vectors.

Optimality of Antithetic Map [2, 3]

Optimality Property

The antithetic map a_Z is the **unique function** that minimizes covariance (or equivalently, maximizes negative correlation) while preserving the distribution:

$$a_Z = \arg \min_{f : f(Z) \overset{d}{=} Z} \mathrm{Cov}(Z, f(Z)) = \arg \max_{f : f(Z) \overset{d}{=} Z} (-\mathrm{Cov}(Z, f(Z)))$$

where $f(Z) \overset{d}{=} Z$ means $f(Z)$ and Z are identically distributed.

Implications

- Among all functions f such that $f(Z)$ and Z are identically distributed, a_Z achieves the **minimum covariance**
- This optimality ensures **maximum variance reduction** for antithetic variates
- The antithetic index $\tau(Z) = -\mathrm{Cov}(Z, a_Z(Z))$ quantifies this optimal negative correlation

Antithetic Map and Index

Antithetic Map

For random variable Z with CDF F, the **antithetic map** is:

$$a_Z(z) = F^{-1}(1 - F(z))$$

The random variable $a_Z(Z)$ is the **antithetic pair** of Z.

Antithetic Index

For Z with finite second moment, the **antithetic index** is:

$$\tau(Z) = -\mathrm{Cov}(Z, a_Z(Z)) = \left(\int_0^1 F^{-1}(u)\mathrm{d}u\right)^2 - \int_0^1 F^{-1}(u)F^{-1}(1-u)\mathrm{d}u$$

Key property: $0 \leq \tau(Z) \leq \mathrm{Var}(Z)$, with equality when Z is symmetric.

Key Assumptions

Distribution Assumptions

- Z is **non-atomic** (continuous CDF)
- Z has **finite second moment**

Function Assumptions

- **Univariate:** $g : \mathbb{R} \to \mathbb{R}$ is monotone increasing and c-anti-Lipschitz:

$$g(x) - g(y) \geq c(x - y) \quad \text{for } x > y, \quad c > 0$$

- **Multivariate:** $\boldsymbol{g} : \mathbb{R}^d \to \mathbb{R}^p$ is c-strongly isotonic:

$$\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{y}) \geq c\|\boldsymbol{x} - \boldsymbol{y}\|_\infty \quad \text{for } \boldsymbol{x} \geq \boldsymbol{y} \text{ (i.e., } x_i \geq y_i \text{ for all } i\text{)}$$

Properties of Antithetic Index

Theorem 1: Key Properties

For non-atomic Z with finite second moment:

1. $a_Z(Z)$ and Z are **identically distributed**
2. a_Z is **monotone non-increasing**
3. **Non-negativity:** $\tau(Z) \geq 0$
4. **Upper bound:** $\tau(Z) \leq \mathrm{Var}(Z)$, with equality iff Z is symmetric
5. **Scale invariance:** $\tau(\alpha Z + \beta) = \alpha^2 \tau(Z)$

Examples

- Uniform $U \sim \mathrm{Unif}(0,1)$: $\tau(U) = \frac{1}{12} = \mathrm{Var}(U)$
- Gaussian $Z \sim \mathcal{N}(0,1)$: $\tau(Z) = 1 = \mathrm{Var}(Z)$
- Exponential $Z \sim \mathrm{Exp}(1)$: $\tau(Z) \approx 0.645 < 1 = \mathrm{Var}(Z)$

Strong Antithetic Variance Reduction

Theorem 2: Univariate Case

For random variable Z and monotone increasing c-anti-Lipschitz functions f, g:

$$\mathrm{Cov}(f(Z), g(a_Z(Z))) \leq -c^2\tau(Z) \leq 0$$

Theorem 3: Multivariate Case

For random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ with i.i.d. components (antithetic index $\tau > 0$) and c-strongly isotonic function $\boldsymbol{g} : \mathbb{R}^d \to \mathbb{R}^p$:

$$\mathrm{Var}\left(\frac{\boldsymbol{g}(\boldsymbol{X}) + \boldsymbol{g}(\boldsymbol{a}(\boldsymbol{X}))}{2}\right) \leq \frac{\mathrm{Var}(\boldsymbol{g}(\boldsymbol{X})) - c^2\tau pd}{2}$$

where $\boldsymbol{a}(\boldsymbol{X}) = (a(X_1), \ldots, a(X_d))$.

Optimal Transport Connection

Theorem 4: Geometric Interpretation

- The antithetic map $-a_Z$ is the **maximum correlation coupling** between distributions of Z and $-Z$
- For Q being the reflection of P about its mean:

$$W_2(P, Q) = \sqrt{2\mathrm{Var}(Z) - 2\tau(Z)}$$

  where W_2 is the 2-Wasserstein distance
- The antithetic index directly relates to optimal transport geometry

Significance

This connection provides a geometric understanding of why antithetic variates achieve optimal variance reduction.

Application 1: Monte Carlo Integration

Proposition: Antithetic Monte Carlo

For monotone increasing c-anti-Lipschitz function $g : \mathbb{R} \to \mathbb{R}$:

$$\mathrm{Var}\left(\frac{1}{2n}\sum_{i=1}^{n}\left(g\left(U_i\right) + g\left(1 - U_i\right)\right)\right) \leq \frac{\mathrm{Var}\left(g\left(U_1\right)\right)}{2n} - \frac{c^2}{24n}$$

where $U_1, \ldots, U_n \sim \text{Unif}(0, 1)$ are i.i.d.

Improvement

- Standard MC variance: $\frac{\mathrm{Var}(g(U_1))}{n}$
- Antithetic MC variance: $\leq \frac{\mathrm{Var}(g(U_1))}{2n} - \frac{c^2}{24n}$
- **Reduction:** At least $\frac{c^2}{24n}$ variance reduction

Application 2: Function Approximation

Proposition: Antithetic Function Approximation

For $g(\boldsymbol{x}) = \mathbb{E}_{\boldsymbol{Z}}[\phi(\boldsymbol{x}, \boldsymbol{Z})]$ where $\phi(\boldsymbol{x}, \cdot)$ is c-strongly isotonic:

$$\mathbb{E}\left[(\hat{g}(\boldsymbol{x}) - g(\boldsymbol{x}))^2\right] \leq \frac{\mathrm{Var}_{\boldsymbol{Z}}(\phi(\boldsymbol{x}, \boldsymbol{Z})) - c^2 \tau d}{2n}$$

where $\hat{g}(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\phi(\boldsymbol{x}, \boldsymbol{Z}_i) + \phi(\boldsymbol{x}, \boldsymbol{a}(\boldsymbol{Z}_i))}{2}$.

Extension: Lipschitz Functions

If ϕ is also L-Lipschitz (with $L > c$):

$$\mathbb{E}\left[(\hat{g}(\boldsymbol{x}) - g(\boldsymbol{x}))^2\right] \leq \frac{d}{2n} \left(L^2 \mathrm{Var}(Z) - c^2 \tau\right)$$

Application 3: Concentration Inequalities

Proposition: Improved Bernstein's Inequality

For i.i.d. zero-mean $X_1, \ldots, X_n$ with variance σ^2, $|X_i| \leq K$ almost surely, and antithetic index τ:

$$\mathbb{P}\left(\left|\sum_{i=1}^{2n} X_i\right| \geq t\right) \leq 2\exp\left(-\frac{t^2/2}{2n\left(\sigma^2 - \tau\right) + 2Kt/3}\right)$$

where $X_{n+1}, \ldots, X_{2n}$ are antithetic counterparts.

Improvement

Standard Bernstein bound has variance term $2n\sigma^2$, while antithetic version has $2n(\sigma^2 - \tau)$, providing tighter concentration.

Application 4: Stochastic Optimization

Theorem: Antithetic SGD (Non-convex)

For c-strongly isotonic gradient estimates, after T iterations:

$$\mathbb{E}\left[\|\nabla f(\hat{\boldsymbol{x}})\|_2^2\right] \leq \frac{2\Delta_1}{T\eta} + \frac{L\eta\left(\sigma^2 - c^2\tau pd\right)}{2}$$

where $\hat{\boldsymbol{x}}$ is uniformly chosen from $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T\}$ and $\Delta_1 = f(\boldsymbol{x}_1) - f(\boldsymbol{x}^*)$ with $\boldsymbol{x}^*$ being the optimal solution.

Theorem: Antithetic SGD (Strongly Convex)

For μ-strongly convex f:

$$\mathbb{E}\left[\|\boldsymbol{x}_{T+1} - \boldsymbol{x}^*\|_2^2\right] \leq (1 - \eta\mu)^T \delta_1 + \eta\frac{\sigma^2 - c^2\tau pd}{2\mu}$$

where $\delta_1 = \|\boldsymbol{x}_1 - \boldsymbol{x}^*\|_2^2$.

Summary

Key Contributions

- Introduced **antithetic maps** and **antithetic indices** as fundamental concepts
- Established **strong variance reduction inequalities** with explicit quantitative bounds
- Connected to **optimal transport theory** for geometric insights
- Provided theoretical guarantees for multiple applications

Technical Insights

- Variance reduction proportional to $c^2 \tau p d$: anti-Lipschitz constant, antithetic index, dimensions
- Antithetic map is the unique function maximizing negative correlation
- Optimal transport connection provides geometric interpretation

Future Directions

Existing Limitations

- Analyssis of Antithetic variates is limited to one-dimensional random variables,
- Or only limited to a special class of strongly isotonic functions.
- No theoretical guarantees for general multivariate functions and random vectors.
- Cannot be applied to finite sample data, without knowledge of CDF of the distribution.

Potential Extensions

- Formulation of Optimal Transport Problem [4] for Antithetic Variates
- Bounding the Statistic Accuracy of the Antithetic Variates with Finite Sample Data
- Constrution of Distribution based on Schrodinger Bridge Method [1] using Sinkhorn Algorithm [4]

General Definition: Multivariate Antithetic Map

General Definition in $\mathbb{R}^d$

For a random vector $\boldsymbol{z} \in \mathbb{R}^d$ with probability density p, the **antithetic map** $\alpha(\boldsymbol{z})$ is defined as:

$$\alpha = \arg\min_{f \in \mathcal{F}} \mathbb{E}[\langle \boldsymbol{z}, f(\boldsymbol{z}) \rangle]$$

where $\mathcal{F} = \{f : \mathbb{R}^d \to \mathbb{R}^d \mid f_{\#}p = p\}$ is the set of **measure-preserving maps under reflection**.

Brenier's Theorem: Structure

The antithetic map $\alpha(\boldsymbol{z})$ is, almost surely, the **negative gradient of a convex function**:

$$\alpha(\boldsymbol{z}) = -\nabla \phi(\boldsymbol{z})$$

for some convex function $\phi : \mathbb{R}^d \to \mathbb{R}$.

General Definition: Multivariate Antithetic Map

Key Properties

- **Anti-monotonicity:** $\langle \alpha(\boldsymbol{x}) - \alpha(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle \leq 0$ for all $\boldsymbol{x}, \boldsymbol{y}$
- **Measure preservation:** If $\boldsymbol{z} \sim p$, then $\alpha(\boldsymbol{z}) \sim p$
- **Special cases:**
  - Scalar: $\alpha(z) = F^{-1}(1 - F(z))$ (recovers the 1D definition)
  - Elliptical: $\alpha(\boldsymbol{z}) = 2\boldsymbol{\mu} - \boldsymbol{z}$ (for symmetric elliptical distributions)

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet.
Diffusion schrödinger bridge with applications to score-based generative modeling.
Advances in neural information processing systems, 34:17695–17709, 2021.

Abolfazl Hashemi, Dongmin Lee, and Anuran Makur.
Strong antithetic variance reduction inequalities.
In *2025 IEEE International Symposium on Information Theory (ISIT)*, pages 1–6. IEEE, 2025.

Abolfazl Hashemi, Dongmin Lee, and Anuran Makur.
Strong antithetic variates: Theory and applications.
2025.

Gabriel Peyré, Marco Cuturi, et al.
Computational optimal transport: With applications to data science.
Foundations and Trends® in Machine Learning, 11(5-6):355–607, 2019.