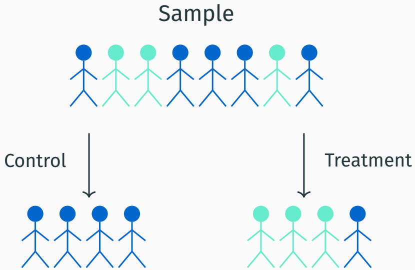
Gaussianized Design Optimization for Covariate Balance in Randomized Experiments

Wenxuan Guo, Fifth year Ph.D. student, University of Chicago
Tengyuan Liang, University of Chicago
Panos Toulis, University of Chicago

July 2025, The 26th ACM Conference on Economics and Computation (EC'25)

Randomized Experiments

- *Randomized experiments* are the gold standard in causal inference.
- Examples: i.i.d. Bernoulli randomization, complete randomization, and A/B tests.



- Blue and green denote old and new customers.
- Customer characteristics might be unbalanced.

Covariate Balance and Covariate-adaptive Designs

- Covariate balance: Different treatment groups share similar covariates.
- Covariate balance improves the estimation precision.
- Covariate-adaptive designs explicitly balance for the covariates.
- Widely used across different fields, including economics (Bruhn and McKenzie, 2009), clinical trials (Rosenberger and Lachin, 2015), and online business platforms (Kohavi and Longbotham, 2023).

Connection between Covariate Balance and Treatment Design

Objective: different treatment groups have similar covariates

$$\Updownarrow$$

Design: similar units receive distinct treatments

Two high-level questions.

- Design: How to define covariate similarity, and what is an ideal covariance structure?
- Sampling: How to sample treatments to achieve the desired covariance?

Existing Approaches

- Complete randomization (Neyman, 1923).
- Matched pair design (Fisher, 1926; Imbens and Rubin, 2015) to control pairwise distance of covariates.
- Rerandomization (Morgan and Rubin, 2012) based on imbalance measures.
- Gram-Schmidt Walk design (Harshaw et al., 2019).

However, existing designs have limitations.

1. Binary treatments.
2. Unclear if covariate balance is optimal.
3. Heuristic adjustments on the covariates, e.g., discretization, covariate selection, etc.

Gaussianized Design Optimization

We propose Gaussianized design optimization to optimally balance for covariates.

- Sampling: Model treatments D as transformations of multivariate Gaussian random variables.
- Design: Convert the design problem to an optimization problem over Gaussian covariance matrices.

Contributions.

1. Arbitrary number of treatment arms, even continuous.
2. Optimal covariate balance (local); Quantitative comparison of designs.
3. Different types of covariates, e.g., categorical/continuous, univariate/high-dimensional.

Problem Setup

- $\mathbb{D} = \{1, \ldots, K\}$: the treatment space.
- $D = (D_1, \ldots, D_n)$: the treatment vector with $D_i \in \mathbb{D}$.
- $Y_i(k)$: the potential outcome of unit i under treatment k.

Focus on a general causal effect with user-specified $(w_k)_{k=1}^K$:

$$\tau_w = \sum_{k=1}^K w_k \tau_k \ , \quad \tau_k = \frac{1}{n} \sum_{i=1}^n Y_i(k) \ .$$

Estimation:

- Observe $Y_i = \sum_{k=1}^K \mathbb{I}\{D_i = k\} Y_i(k)$.
- Use a Horvitz-Thompson estimator to estimate τ_w:

$$\widehat{\tau}_w = \sum_{k=1}^K w_k \widehat{\tau}_k \ , \quad \widehat{\tau}_k = \frac{1}{n} \sum_{i=1}^n \frac{\mathbb{I}\{D_i = k\}}{\mathbb{P}(D_i = k)} Y_i \ .$$

MSE Bound of Horvitz-Thompson Estimators

To derive the design objective, we analyze the Mean Squared Error (MSE) of the Horvitz-Thompson estimator.

- For $\widehat{\tau}_k$,

$$\mathbb{E}(\widehat{\tau}_k - \tau_k)^2 = \frac{K^2}{n^2} Y(k)^\top \mathrm{Cov}_k(D) Y(k) \ .$$

  $Y(k) = (Y_1(k), \ldots, Y_n(k))^\top$, $\mathrm{Cov}_k(D) = \mathrm{Cov}(\mathbb{I}\{D_1 = k\}, \ldots, \mathbb{I}\{D_n = k\})$.

- For a general estimator $\widehat{\tau}_w$,

$$\mathrm{MSE} = \mathbb{E}(\widehat{\tau}_w - \tau_w)^2 \leq \frac{K^3}{n^2} \sum_{k=1}^{K} w_k^2 Y(k)^\top \mathrm{Cov}_k(D) Y(k) =: \mathrm{MB} \ .$$

Goal of the experimental design is to minimize MSE through $\mathrm{Cov}_k(D)$.

Covariate Balance Measures

MB is unobservable as $Y(k)$ is unknown.

- Assume $Y(k) = X\beta_k$, such that balancing for covariates is beneficial.

$$\mathrm{MB} = \frac{K^3}{n^2} \sum_{k=1}^{K} w_k^2 \beta_k^\top X^\top \mathrm{Cov}_k(D) X \beta_k \ .$$

- Worst-case analysis (Harshaw et al., 2019): If $k = 1, \dots, K$, $\|\beta_k\| \le M$,

$$\sup_{\|\beta_k\| \le M} \mathrm{MB} \propto \sum_{k=1}^{K} w_k^2 \|X^\top \mathrm{Cov}_k(D) X\|_{\mathrm{op}} \ . \tag{1}$$

- Average-case analysis (Isaki and Fuller, 1982): If $\{\beta_k\}_{k=1}^K \sim (0, I_d)$,

$$\mathbb{E}_{\beta_k} \mathrm{MB} \propto \sum_{k=1}^{K} w_k^2 \|X^\top \mathrm{Cov}_k(D) X\|_{\mathrm{nuc}} \ . \tag{2}$$

(1) and (2) characterize the MSE under different structural assumptions.

Computational Challenges of Design Optimization

- Design optimization in the treatment space can be NP hard.
- Under $K = 2$ setting, the design optimization reduces to

$$\min_{\mathrm{Cov}(D)} \|X^\top \mathrm{Cov}(D) X\|_{\mathrm{nuc}} = \min_{\mathrm{Cov}(D)} \mathrm{tr}\big(XX^\top \mathrm{Cov}(D)\big) .$$

- Equivalent to the Max-Cut problem (NP-hard).
- Our solution is Gaussianization, motivated by Goemans and Williamson rounding of Max-Cut.

Gaussianization

- Gaussianization:

$$D_i = g(T_i)\ , \quad T \coloneqq (T_1, \ldots, T_n) \sim \mathcal{N}(0, \Sigma)\ .$$

- Σ is a design matrix from

$$\text{correlation elliptope:} \quad \mathcal{E} = \{\ \Sigma \in \mathbb{R}^{n \times n} \mid \Sigma \succeq 0, \Sigma_{ii} = 1\ \}\ .$$

- For example, to get a uniform design, set

$$g(t) = i\ , \quad \text{if } t \in \left(\Phi^{-1}\left(\frac{i-1}{K}\right), \Phi^{-1}\left(\frac{i}{K}\right)\right)\ , \quad i = 1, \ldots, K\ .$$

  $\Phi(\cdot)$ is the standard normal CDF.
- When $K = 2$ with $\mathbb{D} = \{\pm 1\}$,

$$g(T_i) = \mathrm{sign}(T_i)\ .$$

- How does this help with design optimization?

Gaussianized Representation

Proposition 1 (G., Liang, and Toulis)

Under the Gaussianization $D_i = g(T_i)$, we have

$$\mathrm{Cov}_k(D) = f_k(\Sigma) \ ,$$

where $f_k : [-1, 1] \to \mathbb{R}$ are elementwise functions with analytical expressions.

- Proposition 1 induces covariate balance measures on Σ:

$$\sum_{k=1}^{K} w_k^2 \|X^\top \mathrm{Cov}_k(D) X\|_{\mathrm{norm}} = \sum_{k=1}^{K} w_k^2 \|X^\top f_k(\Sigma) X\|_{\mathrm{norm}} \ .$$

- $f_k(\rho) = r_{k-1,k-1}(\rho) + r_{k,k}(\rho) - 2r_{k-1,k}(\rho)$ where

$$r_{i,j}(\rho) = \int_0^\rho p_r(q_i, q_j)\mathrm{d}r \ .$$

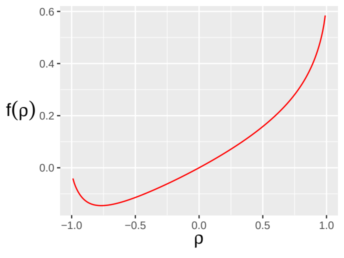
$q_i = \Phi^{-1}(i/K)$ and $p_\rho(x, y)$ is the density of $\mathrm{BN}(0, 0, 1, 1, \rho)$.

Gaussianized Design Optimization

Design of experiments now boils down to an optimization:

$$\min_{\Sigma \in \mathcal{E}} \sum_{k=1}^{K} w_k^2 \|X^\top f_k(\Sigma) X\|_{\mathrm{norm}} \ , \quad \mathrm{norm} \in \{\mathrm{nuc}, \mathrm{op}\} \ .$$

E.g., if $K = 3$, $w_k = 1/3$, and $\mathrm{norm} = \mathrm{nuc}$, the objective $= \left\|X^\top f(\Sigma) X\right\|_{\mathrm{nuc}}$.



- $\pm$ correlations in $\Sigma \Leftrightarrow \pm$ correlations in $f(\Sigma)$.
- Gaussianized design optimization:

$$\text{similar } (X_i, X_j) \Rightarrow \text{negative } f(\Sigma_{ij}) \Rightarrow \text{negative } \Sigma_{ij}.$$

Projected Gradient Descent

We propose a projected gradient descent algorithm on Σ to solve

$$\min_{\Sigma \in \mathcal{E}} \sum_{k=1}^{K} w_k^2 \|X^\top f_k(\Sigma) X\|_{\mathrm{norm}} =: l(\Sigma) \ .$$

At each iteration t:

1. Gradient descent: $\Sigma^t \leftarrow \Sigma^{t-1} - \eta \nabla l(\Sigma^{t-1})$.
2. Projection: $\Sigma^t \leftarrow \mathrm{Proj}_{\mathcal{E}}(\Sigma^t)$.

Remarks.

- Local optimality due to the non-convex objective $l(\cdot)$.
- Initialize design optimization from different designs.

Theoretical Guarantees

Consider design-based framework where the only randomness comes from the treatment assignment D.

Theorem 1 (G., Liang, and Toulis, Informal)

Let $T \sim \mathcal{N}(0, \Sigma_*)$ with Σ_* from Gaussianized design optimization.

Under regularity conditions on $\{Y_i(k)\}_{i=1}^n$ and conditions on the optimization algorithm, it holds that

$$\sqrt{n}\left(\widehat{\tau}_k - \tau_k\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2)) \ .$$

Let $\widehat{\tau}_k^{iid}$ be the estimator under i.i.d. uniform design.

$$\mathrm{Var}(\widehat{\tau}_k^{iid}) - \mathrm{Var}(\widehat{\tau}_k) > 0 \ .$$

- Design-based confidence intervals can be constructed.
- Nonzero variance reduction compared to the i.i.d. design.

Three-treatment Example: Gaussianized Design Optimization

- Suppose $\mathbb{D} = \{1, 2, 3\}$ and we want to estimate

$$\tau = \sum_{k=1}^{3} \frac{1}{3} \tau_k \ , \quad \tau_k = \frac{1}{n} \sum_{i=1}^{n} Y_i(k) \ .$$

- Formulate the Gaussianized design optimization

$$\min_{\Sigma \in \mathcal{E}} \sum_{k=1}^{3} \|X^\top f_k(\Sigma) X\|_{\mathrm{norm}} \ .$$

- Run projected gradient descent to get Σ^*.
- Generate $D_i = g(T_i)$, $T \sim \mathcal{N}(0, \Sigma^*)$.
- Perform estimation and inference for τ.

Three-treatment Example: MSE Trajectory

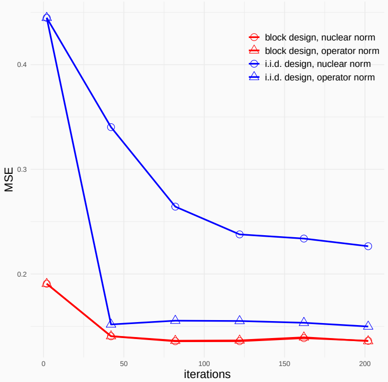
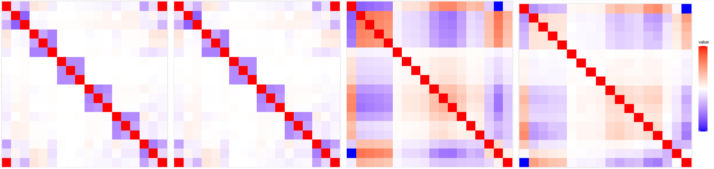


Figure 1: MSE of Horvitz-Thompson estimators over iterations of design optimization.

- Initialize Σ as i.i.d. and block designs.
- Gaussianized design optimization reduces the MSE by $> 60\%$.

Three-treatment Example: Optimized Covariance

(Initial Σ, design objective)



(a) block,op **(b)** block,nuc **(c)** iid,op **(d)** iid,nuc

Our design picks up cross-unit correlations in the Gaussian space.

- Different initial designs $\Rightarrow$ different covariance structure.
- Choices of norm affects the covariance in (c), (d).

Conclusion and Future Work

We develop a Gaussianization framework to optimize experimental designs for covariate balance.

- This approach allows general covariates and multiple treatment arms, offering great flexibility over existing methods.
- The paper contains design-based inference under Gaussianization, and extension to continuous treatments for studying dose-response relations.

Open problems:

- Optimal experimental design under interference.
- Practical confidence intervals by figuring out the best suitable variance bound.
- Randomization inference under Gaussianization.

Thank you!

- The complete version: Wenxuan Guo, Tengyuan Liang, and Panos Toulis “Gaussianized Design Optimization for Covariate Balance in Randomized Experiments,” `https://arxiv.org/abs/2502.16042`, 2025.