Journal of the Royal Statistical Society Series B: Statistical Methodology, 2026, **00**, 1–22
https://doi.org/10.1093/jrsssb/qkag067

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series B
Statistical Methodology

Original Article

Gaussianized design optimization for covariate balance in randomized experiments

Wenxuan Guo, Tengyuan Liang and Panos Toulis

Booth School of Business, The University of Chicago, Chicago, IL 60637, USA

Address for correspondence: Wenxuan Guo, Booth School of Business, The University of Chicago, Chicago, IL 60637, USA. Email: wxguo@chicagobooth.edu

Abstract

Achieving covariate balance in randomized experiments enhances the precision of treatment effect estimation. However, existing methods often require heuristic adjustments based on domain knowledge and are primarily developed for binary treatments. This paper presents Gaussianized Design Optimization, a novel framework for optimally balancing covariates in experimental design. The core idea is to Gaussianize the treatment assignments: we model treatments as transformations of random variables drawn from a multivariate Gaussian distribution and convert the design problem into a nonlinear continuous optimization over Gaussian covariance matrices. Compared to existing methods, our approach offers significant flexibility in optimizing covariate balance across a diverse range of designs and covariate types. Adapting the Burer–Monteiro approach for solving semidefinite programmes, we introduce first-order local algorithms for optimizing covariate balance, improving upon several widely used designs. Furthermore, we develop inferential procedures for constructing design-based confidence intervals under Gaussianization and extend the framework to accommodate continuous treatments. Simulations demonstrate the effectiveness of Gaussianization in multiple practical scenarios.

Keywords: continuous treatments, covariate balance, Mehler's formula, optimal experimental design

1 Introduction

Randomized experiments are considered the gold standard for causal inference in the study of treatment effects (Imbens & Rubin, 2015), because randomization controls potential confounders and leads to reliable treatment effect estimates. Nonetheless, under standard designs such as complete randomization and independent Bernoulli randomization (Neyman, 1923; Rosenberger & Lachin, 2015), covariate imbalance may arise by chance, reducing the estimation precision (Rosenberger & Sverdlov, 2008).

To mitigate covariate imbalance, a range of covariate-balancing designs has been proposed. Crucially, these approaches share a common design principle:

Assign similar units to different arms to balance covariates across groups.

Technically, this principle can be understood as shaping the covariance structure of the treatment assignments: for two units with similar covariates, negatively correlating their treatment assignments makes it likely that one receives treatment while the other serves as control, thereby improving covariate balance. As we discuss below, this principle is widely exploited by existing designs—in explicit or implicit ways—and is also the key motivation behind our work.

Matched-pair designs (Greevy et al., 2004) apply this principle directly by assigning opposite treatment arms within each pair of units, creating perfect negative correlations between their

Received: March 3, 2025. **Accepted:** February 5, 2026

assignment indicators. They are special cases of stratified randomization (R. A. Fisher, 1926), where units with similar covariates are partitioned into strata and a fixed proportion is treated within each stratum. This again induces negative correlations among assignment indicators within strata, thereby improving covariate balance. With additional information about the outcome model, one can construct an optimal matched-pair design that minimizes the estimation error among all stratified designs (Bai, 2022). With many covariates, however, constructing high-quality strata or pairs can be computationally intractable, and practical implementations often rely on heuristic or approximation algorithms (Higgins et al., 2016; Moore, 2012). In another line of work, rerandomization redraws assignments until certain balance criteria are met (Li et al., 2020; Morgan & Rubin, 2012). Such procedures implicitly select a covariance structure that leads to better balance. However, rerandomization can also be computationally challenging with high-dimensional covariates (Morgan & Rubin, 2012), and does not optimize any explicit covariate balance objectives.

To obtain optimal designs for covariate balance or estimation precision, much prior work adopts a model-based approach that minimizes a model-specific loss (Atkinson, 1982; G. W. Basse & Airoldi, 2018; Zhao, 2024). In classical regression designs, Cox and Reid (2000) discussed D-optimal and G-optimal designs, which minimize total prediction errors for a specified linear model. More recently, Bhat et al. (2020) employed D_A-optimality—a variant of D-optimality (Sibson, 1974)—to study optimal A/B tests in online experiments. Because these estimators and optimality criteria rely on a specific outcome model, the resulting designs are sensitive to model misspecification and can yield biased estimates (Box & Draper, 1959).

Recent algorithmic work has explored model-agnostic covariate-balancing designs closely related to the approach proposed in this paper. Harshaw et al. (2024) explicitly formulated a covariate balance objective as a function of the treatment covariance and proposed the Gram-Schmidt Walk (GSW) design to optimize it. Davezies et al. (2025) developed a cube method that achieves nearly exact balance on covariate moments. However, these methods focus on binary treatments and are tailored to particular objectives. To date, a general, model-agnostic optimality theory for covariate-balancing designs remains underexplored, especially one that provides algorithms for obtaining the optimal design beyond binary settings.

In our paper, we exploit the design principles outlined above and formalize the design problem as an optimization problem, named Gaussianized Design Optimization (GDO). The core idea is to model the treatment covariance by a Gaussian covariance matrix through Gaussianization. To fix ideas, consider n units in a binary treatment setting with assignment indicators $D_i \in \mathbb{D} = \{-1, +1\}$. We view $\mathbb{D}$ as the treatment space, which will be generalized to accommodate multiple treatments in Sections 2 and 3. Then, GDO models treatments $D = (D_1, \ldots, D_n)$ as random variables transformed from Gaussian vectors, which is referred to as Gaussianization:

$$D_i = \mathrm{sign}(T_i), \quad T := (T_1, \ldots, T_n) \sim \mathcal{N}(0, \Sigma), \quad \Sigma \in \mathcal{E}.$$

In words, the sign function assigns the treatment when $T_i > 0$ and the control when $T_i < 0$, with ties occurring with probability zero. With multiple treatments, we replace the sign function by a general discretization function $g(\cdot) : \mathbb{R} \to \mathbb{D}$ as defined in Section 3. We view the covariance matrix Σ as the design matrix, since it fully determines the joint distribution of treatments and therefore specifies the design. To ensure that the treatments $\{\mathrm{sign}(T_i)\}_{i=1}^n$ share the same marginal distribution, we restrict Σ to the feasible set $\mathcal{E} = \{\Sigma \mid \Sigma \succcurlyeq 0, \Sigma_{ii} = 1\}$, the correlation elliptope. For example, $\Sigma = I_n$ yields independent Bernoulli(1/2) randomization, while a block-diagonal Σ with 2×2 blocks $\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ induces a matched-pair design.

As a key contribution, we establish a Gaussianized representation of general designs via their covariance. Under Gaussianization, the covariance matrix Σ determines the design; hence there naturally exists a mapping f from $\mathcal{E}$ to $\mathbb{R}^{n \times n}$ such that

$$\mathrm{Cov}(D) = f(\Sigma).$$

Using Hermite polynomial expansions, we derive a closed-form expression for f and thereby fully parameterize the design via the Gaussian covariance Σ. This provides a computational formulation for design problems, as one can evaluate and improve a design by computing $f(\Sigma)$ and optimizing

Σ. Unlike prior work, this formulation is sufficiently general to accommodate diverse design objectives and extends beyond binary treatments to multiple treatment arms. Moreover, this approach relies crucially on Gaussianity, since a multivariate Gaussian distribution is uniquely determined by its mean and covariance Σ.

Given an observed covariate matrix $X \in \mathbb{R}^{n\times d}$, we adapt the optimality criteria in the literature and derive design objectives of the form:

$$\|X^\top \mathrm{Cov}(D)X\|_{\mathrm{norm}} \overset{\text{Gaussianization}}{\Rightarrow} \|X^\top f(\Sigma)X\|_{\mathrm{norm}}, \quad \mathrm{norm} \in \{\mathrm{op}, \mathrm{nuc}\}. \tag{1}$$

At a high level, the objective $\|X^\top \mathrm{Cov}(D)X\|_{\mathrm{norm}}$ characterizes the estimation error of a Horvitz-Thompson estimator that can be explained by covariates, and therefore measures the balance of a given design. It can be measured in either the operator norm (op) or the nuclear norm (nuc) as discussed in Section 3. Under Gaussianization, it reduces to a function on Σ through the analytical function f, which enables a direct optimization.

From an optimization perspective, directly optimizing over $\mathrm{Cov}(D)$ in (1) is computationally intractable. In Section 2, we show that the non-Gaussianized design optimization, analogous to the Max-Cut problem, amounts to solving a linear program over the cut polytope, which is an NP-hard task. Therefore, Gaussianization is a practical step to restrict the design space to the correlation elliptope and to enable feasible optimizations of (1) through gradient descent-type algorithms as described in Section 4. In summary, GDO transforms the design problem into a continuous optimization over Σ that is computationally tractable.

In certain experiments, the treatment variable is inherently continuous (e.g. a medication dosage), making it insufficient to confine the design to discrete treatment arms. To address this limitation, we further extend the GDO framework by allowing $\mathbb{D} = \mathbb{R}$ and propose Gaussian designs, which directly assign $T = (T_1, \ldots, T_n) \sim \mathcal{N}(0, \Sigma)$ as actual treatments. As shown in Section 5, this approach offers two main advantages. First, it allows testing structural properties of potential outcome functions, including monotonicity and convexity. Second, it enables covariate balance optimization akin to the discrete setup. Thus, Gaussian designs harness the flexibility of Gaussianization and contribute to the growing literature on continuous treatments (Colangelo & Lee, 2026; Hirano & Imbens, 2004; Imai & Van Dyk, 2004; Kennedy et al., 2017).

In Section 6, we investigate design-based inference under Gaussianization, where the outcome-generating model is fixed, and all randomness arises from the Gaussian treatments $\{T_i\}_{i=1}^n$. Under a local perturbation condition, we establish asymptotic normality for the treatment effect estimator and present valid inferential procedures. Together, these results establish a comprehensive framework that integrates design optimization, estimation, and inference under Gaussianization.

1.1 Related work

Randomized experiments, such as i.i.d. Bernoulli designs and complete randomization, are generally viewed as robust designs and are thus desirable in practice. R. A. Fisher (1925) and R. A. 1926. Fisher (1926) argued that randomization yields a valid estimate of experimental error, thereby enabling valid inference and significance testing. Wu (1981) framed robustness in terms of the worst-case mean squared error (MSE) by showing that complete randomization is a minimax design, meaning it minimizes the worst-case MSE. In time-series experiments such as N-of-1 trials, complete randomization has also been shown to be robust against both estimand choices and model misspecifications (Liang & Recht, 2025). See also Kallus (2018), Harshaw et al. (2024), G. W. Basse et al. (2023), Nordin and Schultzberg (2022), and Bai (2023) for related discussions.

In settings where covariate information is available, which is the focus of this paper, it is reasonable to adapt the experimental design so as to achieve covariate balance across treatment arms.

With few covariates, blocking (or stratified randomization) is the canonical way to reduce unwanted variation and increase precision (R. A. Fisher, 1926). Matched-pair designs (Greevy et al., 2004; Imbens & Rubin, 2015) are prime examples of blocking, where each block contains two units, and are optimal under certain conditions (Bai, 2022). However, blocking can be impractical with many covariates. This has motivated sampling-based techniques such as rerandomization (Morgan & Rubin, 2012), which has been shown to improve estimation precision in theory

(Li & Ding, 2020; Wang & Li, 2025). However, choosing the right trade-off between covariate balance criteria and the computational complexity of sampling can be challenging, especially with high-dimensional covariates. Moreover, blocking and rerandomization mainly focus on binary treatment settings, whereas our GDO framework applies to multiple treatments.

In a related line of work, covariate balancing is performed in a sequential manner, where the treatment is assigned to each unit based on prior treatments and covariates of other units gathered up to that point (Bugni et al., 2018; Ma et al. 2020, 2024; Ye et al., 2022). Such designs are collectively known as covariate-adaptive randomizations (CAR), and mainly target settings in clinical trials and online A/B tests. Our paper adopts imbalance measures akin to those in CAR procedures; for instance, the nuclear norm objective in (1) is equivalent in expectation to the imbalance of covariate means between treated and control units (Ma et al., 2024, e.g. Example 2.1). The key distinction between GDO and the CAR framework is that GDO is non-sequential and optimizes the joint distribution of treatment assignments, whereas CAR procedures optimize conditional assignment rules. Due to its sequential nature, the asymptotic theory for CAR (e.g. Theorem 3.1 of Ma et al., 2024) typically requires finite-moment conditions on the feature vector (see Assumption 2 of Ma et al., 2024). Such conditions are most natural for a fixed covariate dimension d, while modern experiments in personalized medicine and online A/B testing can involve many covariates (Bastani & Bayati, 2020; Zhang et al., 2022). In contrast, our framework accommodates high-dimensional settings where d can grow with n. In online supplementary material, Section S2.5 of the supplementary material, we compare the performance of the optimal Gaussianized design with CAR procedures and find that our method outperforms CAR in high-dimensional scenarios. In the related context of sequential estimation under adversarial covariate shifts, Liang (2024) established a dichotomy: a blessing in the regression setting, where the adversary shifts the covariate distribution toward that of an optimal experimental design; a curse in the classification setting, where the adversary shifts it toward the hardest design, thereby trapping subsequent estimation. This dichotomy further highlights how covariate distribution, experimental design, and estimation are deeply intertwined.

Directly relevant to our work, Harshaw et al. (2024) introduced the Gram-Schmidt Walk (GSW) design that formalizes the trade-off between covariate balance and robustness in binary treatment settings. Specifically, Harshaw et al. (2024) proposed $\|\mathrm{Cov}(D)\|_{\mathrm{op}}$ and $\|X^\top \mathrm{Cov}(D) X\|_{\mathrm{op}}$ as measures of robustness and covariate balance, respectively. The GSW design navigates the robustness-balance trade-off by proposing a weighted combination of the aforementioned measures, and employs a random walk to sequentially generate treatment assignments that optimize the weighted measure. Their measure of covariate balance (i.e. $\|X^\top \mathrm{Cov}(D) X\|_{\mathrm{op}}$) motivates us to study similar norm-based objectives.

Statistical inference under covariate-balancing designs—including CAR procedures—can be challenging due to the potentially complex covariate-treatment dependencies. See, for instance, Bugni et al. (2018, 2019), and Ma et al. (2020) for inference under stratified designs; Bai et al. (2022, 2024) for matching-based designs; and Li et al. (2018, 2020) for rerandomization. Our asymptotic results under Gaussianization follow this line of work with different proof techniques. In addition to asymptotic inference, Fisherian-style randomization inference provides a robust alternative to obtain finite-sample valid p-values for testing certain treatment effects (G. Basse et al., 2019; Ding et al., 2016; R. A. Fisher, 1935; Guo et al., 2025; Huang et al., 2025). These procedures can also be directly applied under Gaussianized designs.

Our paper also contributes to the growing literature on continuous treatments. Using techniques from the doubly robust methodology, Colangelo and Lee (2026) and Kennedy et al. (2017) studied the estimation of the average potential outcome function, while Hsu et al. (2024) tested functional properties such as monotonicity. Recently, Callaway et al. (2024) analyzed difference-in-differences setups with a continuous treatment, and discussed the identification of response functions and their derivatives. See Dong and Lee (2023), Schindl et al. (2024), and De Chaisemartin et al. (2022) for related studies. However, all these works consider i.i.d. data from observational studies, which is distinct from our experimental design setup.

2 An example: gaussianization with three treatment arms

To contextualize the idea, we first illustrate Gaussianized design optimization with a simple three-arm example, i.e. $\mathbb{D} = \{1, 2, 3\}$, supplemented with a simulation study. Following the

standard potential outcome framework (Neyman, 1923), we define $Y_i(k)$ as the potential outcome for unit i under treatment k for $k = 1, 2, 3$. The observed outcome for unit i is then defined as $Y_i = \sum_{k=1}^{3} \mathbb{I}\{D_i = k\} Y_i(k)$ and $D = (D_1, \ldots, D_n)$ is the treatment vector. Let $X \in \mathbb{R}^{n \times d}$ be the covariate matrix, and $X_i \in \mathbb{R}^d$ be unit i's covariates.

In this three-arm setup, the GDO framework reduces to the procedure below. Technical details are provided in Sections 3 and 4.

Procedure 1 (**High-Level Procedure of GDO**).

1. Estimand specification: For illustration, we focus on the average effect across all treatment arms

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

   This estimand serves as a simple representative example to demonstrate the multi-arm design procedure. In practice, researchers may target pairwise contrasts such as $\tau_1 - \tau_2$, which can be handled analogously within our framework by reweighting the covariate balance measure in Step 3. We use a Horvitz-Thompson estimator $\widehat{\tau}$ to obtain an unbiased estimate of this quantity.
2. Covariate balance measures: We propose the following covariate balance measures

$$\sum_{k=1}^{3} \|X^\top \mathrm{Cov}_k(D) X\|_{\mathrm{norm}}, \quad \mathrm{norm} \in \{\mathrm{op}, \mathrm{nuc}\},$$

   where $\mathrm{Cov}_k(D)$ is the covariance matrix of $(\mathbb{I}\{D_1 = k\}, \ldots, \mathbb{I}\{D_n = k\})$. The measures under the operator and nuclear norms capture the worst-case and average-case mean squared error (MSE) of $\widehat{\tau}$, respectively, as explained in Section 3.
3. Gaussianization: We model treatments by $D_i = g(T_i)$ and derive that $\mathrm{Cov}_k(D) = f_k(\Sigma)$, $k = 1, 2, 3$, with analytical expressions for f_k. Functions $\{f_k\}_{k=1}^3$ are explicitly given in Proposition 1. Gaussianization translates covariate balance measures into an analytical function on the Gaussian covariance Σ for $\mathrm{norm} \in \{\mathrm{op}, \mathrm{nuc}\}$: $\sum_{k=1}^{3} \|X^\top \mathrm{Cov}_k(D) X\|_{\mathrm{norm}} = \sum_{k=1}^{3} \|X^\top f_k(\Sigma) X\|_{\mathrm{norm}}$.
4. Design optimization: We apply a first-order algorithm (projected gradient descent on the Gaussianized space) in Section 4 to solve

$$\min_{\Sigma \in \mathcal{E}} \sum_{k=1}^{3} \|X^\top f_k(\Sigma) X\|_{\mathrm{norm}}. \tag{2}$$

   This returns a locally optimal Gaussian covariance matrix Σ^*.
5. Treatment assignment: Generate treatments through $D_i = g(T_i)$, $T \sim \mathcal{N}(0, \Sigma^*)$.

Procedure 1 applies to general experimental setups with $\mathbb{D} = \{1, \ldots, K\}$, where K is the total number of treatment arms (Section 3). Given Step 2, one naturally searches for a valid design that minimizes the covariate balance measure. However, optimization over the treatment covariance $\mathrm{Cov}_k(D)$ is computationally challenging: for binary treatment assignments ($K = 2$), we need to solve $\min_{\mathrm{Cov}_1(D)} \|X^\top \mathrm{Cov}_1(D) X\|_{\mathrm{nuc}} = \min_{\mathrm{Cov}_1(D)} \mathrm{tr}(XX^\top \mathrm{Cov}_1(D)) \Leftrightarrow \min_{C \in \mathcal{C}} \mathrm{tr}(XX^\top C)$. The feasible set of $\mathrm{Cov}_1(D)$ is affinely isomorphic to the cut polytope $\mathcal{C}$ (Huber & Maric, 2017), and the optimization problem is thus as hard as the Max-Cut problem (Barahona & Mahjoub, 1986),

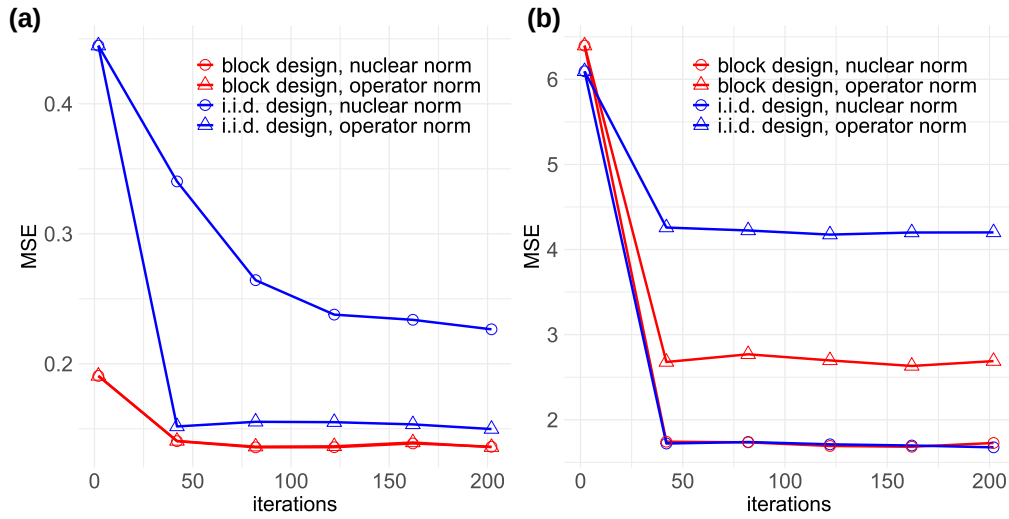

Figure 1. MSE of the Horvitz-Thompson estimators over iterations of design optimization. Panels (a) and (b) correspond to the single informative covariate setting and the uniform covariate setting, respectively.

which is NP-hard. In the optimization literature, Goemans and Williamson (1995) proposed an approximation algorithm for the Max-Cut problem, where the idea is to generate a cut vector by thresholding a correlated Gaussian vector, with the correlation matrix obtained as the solution to a semidefinite program. Our approach shares a similar procedure when $K = 2$: the Gaussianization step in our approach is precisely the Goemans-Williamson rounding, with the analytic function $f(x) = \arcsin(x)$ (derived from f_k in Proposition 1) shared in the analysis.

2.1 Simulation study

To demonstrate the benefits of design optimization, we conduct a simulation study that evaluates the MSE of $\widehat{\tau}$ under various designs. Two covariate structures are considered: (a) the first covariate has the largest scale and serves as the sole informative feature, and (b) all covariates are uniformly generated and are equally informative. We initialize our iterative algorithm using either an i.i.d. design ($\Sigma = I_n$) or a block design, where Σ is a block correlation matrix representing a classical block design constructed by sorting the first covariate. Details of the simulation are provided in online supplementary material, Section S2.

Figure 1 shows the MSE trajectories over iterations of Gaussianized design optimization. In setup (a), initializing with a block design yields a smaller MSE by leveraging the highly informative covariate. Furthermore, starting from the i.i.d. design and minimizing the covariate balance measure with the operator norm results in a lower final MSE. Different initial designs in setup (b) produce similar early-stage MSEs but diverge in their final performance. Notably, with suitable choices of the initial design and the norm, GDO reduces the MSE by more than 60% through iterations. Figure 2 provides heatmaps of the correlation matrices for different initializations and norms. Observe that Panels (b), (c), (g), and (h) preserve the block structure, which highlights how GDO makes local improvements.

3 A gaussianization framework

In this section, we formally introduce the Gaussianization framework, which includes both norm-based covariate balance measures and their Gaussianized representations. Our formulation accommodates general experimental setups with the discrete support $\mathbb{D} = \{1, \ldots, K\}$. We conclude this section by deriving a representation inspired by Liang and Tran-Bach (2022) using Mehler's formula (Mehler, 1866), a key technical insight that motivates our design optimization.

3.1 General covariate balance measures

We consider the potential outcome framework as in Section 2, and focus on uniform designs such that $\mathbb{P}(D_i = k) = 1/K$ for any $i = 1, \ldots, n$ and $k = 1, \ldots, K$. The Gaussianization framework can

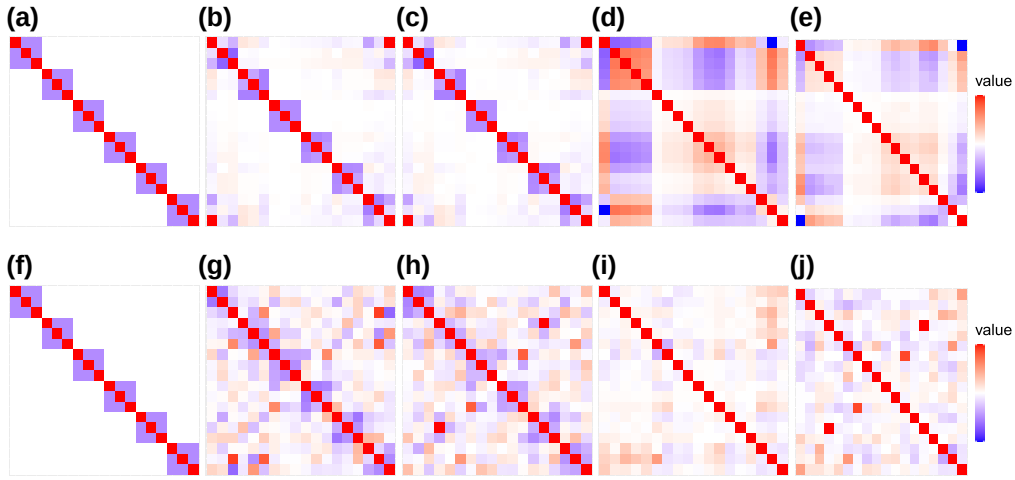

Figure 2. Heatmaps of covariance matrix Σ from Gaussianized design optimization. Panels (a)-(e) correspond to the single informative covariate setting, and panels (f)-(j) correspond to the uniform covariate setting. In each row, from left to right, we show the initial block design, the optimized block designs under the operator and nuclear norm, and the optimized i.i.d. designs under the operator and nuclear norm. The heatmap scale ranges from -1 to 1, with 0 at the midpoint.

also accommodate non-uniform designs where D_i follows non-uniform marginal treatment probabilities, by adjusting the discretization function g. The key requirement is that all treatment assignments share the same marginal distribution to enable effective design optimization.

We define our estimand as follows:

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

where $w = (w_1, \ldots, w_K)$ is a pre-specified vector. This can be a contrast vector, e.g. $w = (1, -1, 0, \ldots, 0)$, leading to the contrast $\tau_1 - \tau_2$. It can also be a weight vector, e.g. $w_k = 1/K$ and $\sum_{k=1}^{K} w_k = 1$, which reduces to the estimand in Section 2 given $K = 3$. These estimands encompass a rich class of causal quantities, and thus they are of primary interest in empirical research. To estimate τ_w, we use the Horvitz-Thompson estimator as mentioned in Section 2:

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

We focus on Horvitz-Thompson estimators, similar to previous works in the literature (Bai, 2022; Harshaw et al., 2024; Wang & Li, 2025). Notably, $\widehat{\tau}_w$ is the optimal linear unbiased sampling estimator of τ_w (Hege, 1967), and thus is desirable for design optimization. Alternatively, one could consider covariate-adjusted estimators (Chang, 2023; R. A. Fisher, 1935; List et al., 2024), but their performance is model-specific, potentially leading to biased estimation (Freedman, 2008). More discussion is provided in online supplementary material, Section S5.

While $\widehat{\tau}_w$ is unbiased, its mean squared error (MSE) depends on specific design structures through the covariance matrix of D.

Lemma 1 (Lemma 7 of Chang, 2023). Under uniform experimental designs, for $k = 1, \ldots, K$, we have

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

where $Y(k) = (Y_1(k), \ldots, Y_n(k))^\top$, and $\mathrm{Cov}_k(D)$ is defined in Section 2.

From Lemma 1, the MSE of the kth treatment effect is a quadratic form in the covariance matrix of the treatment assignment vector, $\mathrm{Cov}_k(D)$, evaluated at the (unknown) potential outcome vector $Y(k)$. Then, for a general estimator $\widehat{\tau}_w$, we utilize the AM-QM inequality to obtain

$$\mathbb{E}(\widehat{\tau}_w - \tau_w)^2 = \mathbb{E}\left(\sum_{k=1}^{K} w_k(\widehat{\tau}_k - \tau_k)\right)^2 \le K\sum_{k=1}^{K} w_k^2 \mathbb{E}(\widehat{\tau}_k - \tau_k)^2 = \frac{K^3}{n^2}\sum_{k=1}^{K} w_k^2 Y(k)^\top \mathrm{Cov}_k(D) Y(k).$$

The MSE bound on $\widehat{\tau}_w$ leads to measures of covariate balance. Specifically, following a similar idea as in Harshaw et al. (2024), let us assume for the moment that potential outcomes are perfectly linear in the covariates, i.e. $Y(k) = X\beta_k$, for some $\beta_k \in \mathbb{R}^d$. This reduces the MSE bound ('MB') to

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

In practice, even if we can somehow justify perfect linearity, the signals $\{\beta_k\}_{k=1}^K$ are in general unknown. Harshaw et al. (2024) formulated a worst-case MSE by assuming that the signal has a fixed norm with arbitrary directions. Following their idea, we consider a structural assumption that for $k = 1, \ldots, K$, $\|\beta_k\| \le M$, leading to a worst-case MSE measure:

$$\begin{aligned}\sup_{\|\beta_k\|\le M} \mathrm{MB} &\propto \sup_{\|\beta_k\|\le M} \sum_{k=1}^{K} w_k^2 \beta_k^\top X^\top \mathrm{Cov}_k(D) X\beta_k \propto \sum_{k=1}^{K} w_k^2 \sup_{\|\beta_k\|\le 1} \beta_k^\top X^\top \mathrm{Cov}_k(D) X\beta_k \\ &= \sum_{k=1}^{K} w_k^2 \|X^\top \mathrm{Cov}_k(D) X\|_{\mathrm{op}}.\end{aligned} \tag{3}$$

As an alternative, Isaki and Fuller (1982) and Chang (2023) have introduced the notion of 'anticipated variance' that measures an average MSE under a prior distribution on potential outcomes. Following their idea, we consider that $\{\beta_k\}_{k=1}^K$ are random signals with mean zero and identity covariance. This leads to an average-case MSE measure:

$$\begin{aligned}\mathbb{E}_{\beta_k}\mathrm{MB} &\propto \sum_{k=1}^{K} w_k^2 \mathbb{E}\beta_k^\top X^\top \mathrm{Cov}_k(D) X\beta_k = \sum_{k=1}^{K} w_k^2 \mathrm{tr}(X^\top \mathrm{Cov}_k(D) X \mathbb{E}\beta_k\beta_k^\top) \\ &= \sum_{k=1}^{K} w_k^2 \mathrm{tr}(X^\top \mathrm{Cov}_k(D) X) = \sum_{k=1}^{K} w_k^2 \|X^\top \mathrm{Cov}_k(D) X\|_{\mathrm{nuc}}.\end{aligned} \tag{4}$$

The derivation above leads to the formal definition of covariate balance measures.

Definition 1 For a uniform design with K treatments, we define the covariate balance measure as $\sum_{k=1}^K w_k^2 \|X^\top \mathrm{Cov}_k(D) X\|_{\mathrm{norm}}$, norm $\in$ {nuc, op}.

To summarize, by making two structural assumptions on the value of β_k, we derive covariate balance measures that only depend on X and the design D. Importantly, this motivates the study of objective (1) as the basis for optimal experimental design, as we show in the sequel. By setting $w_k = 1/K$ and $K = 3$, one recovers the measures exemplified in Procedure 1.

The definition of these covariate balance measures follows similar definitions in existing literature. Under the binary setting ($K = 2$), our nuclear-norm measure satisfies $\|X^\top \mathrm{Cov}(D) X\|_{\mathrm{nuc}} \propto \mathbb{E}\|\sum_{D_i=1} X_i - \sum_{D_i=2} X_i\|^2$. Notably, the squared norm inside the expectation is the imbalance measure of covariate means (Ma et al., 2024). The operator-norm measure is equivalent to the covariate balance measure in Harshaw et al. (2024), and it has been shown

that $\|X^\top \mathrm{Cov}(D)X\|_{\mathrm{op}} \propto \max_{\|\beta\|=1} \mathbb{E}((\sum_{D_i=1} X_i^\top \beta - \sum_{D_i=2} X_i^\top \beta)^2)$. The quantity above is referred to as a 'distributional extension of the squared Euclidean discrepancy' (Harshaw et al., 2024), and the squared term inside the expectation again captures an imbalance between two groups of covariates.

Notably, our measures are functions on the design distribution (by taking expectation over D) rather than treatment assignments, and this is deeply tied to our goal of design optimization. For an imbalance measure that directly depends on D (Ma et al., 2024; Morgan & Rubin, 2012), globally optimizing the measure involves searching over 2^n treatment assignments, which is generally intractable. As a result, covariate-adaptive randomization procedures optimize the treatment assignment in a sequential manner, conditionally on past observed treatments and covariates. Here, we take a different approach by focusing on the distribution-level covariate balance, which leads to efficient and meaningful optimization procedures over the entire treatment design distribution.

In practice, the choice between the nuclear and operator norm reflects an accuracy-robustness trade-off: the nuclear-norm objective pursues covariate balance more aggressively to reduce the estimation error on average, whereas the operator-norm objective offers moderate gains in balance while providing greater robustness. We demonstrate this point with simulations in online supplementary material, Section S2.4.

3.2 Gaussianized representation

We introduce a Gaussianized representation of the uniform design D through a map $g: \mathbb{R} \to \{1, \ldots, K\}$ as defined below:

$$g(t) = \begin{cases} i & \text{if } t \in \left(\Phi^{-1}\left(\frac{i-1}{K}\right), \Phi^{-1}\left(\frac{i}{K}\right)\right], \quad i = 1, \ldots, K-1 \\ K & \text{if } t \in (\Phi^{-1}\left(\frac{K-1}{K}\right), \infty) \end{cases},$$

where $\Phi(\cdot)$ is the standard normal CDF. In other words, we discretize the Gaussian treatments T according to the equally spaced quantiles. This recovers the uniform design since $g(T_i)$ is uniformly distributed on $\{1, \ldots, K\}$. In addition, one could adjust the quantiles in the function $g(\cdot)$ to represent non-uniform assignment probabilities.

When a uniform design is generated via a Gaussianized representation, i.e. $D_i = g(T_i)$ for $T \sim \mathcal{N}(0, \Sigma)$, the covariance structure of D is completely captured by Σ. Importantly, one can link the two covariance structures through analytical formulas.

Proposition 1 Under Gaussianization $D_i = g(T_i)$ for K treatment arms, we have $\mathrm{Cov}_k(D) = f_k(\Sigma)$, where $f_k: [-1, 1] \to \mathbb{R}$, $k = 1, \ldots, K$ are elementwise functions defined by

$$f_k(\rho) = \begin{cases} r_{1,1}(\rho) & \text{if } k = 1 \\ r_{K-1,K-1}(\rho) & \text{if } k = K \\ r_{k-1,k-1}(\rho) + r_{k,k}(\rho) - 2r_{k-1,k}(\rho) & \text{otherwise} \end{cases}.$$

For $i, j = 1, \ldots, K-1$ and $q_i = \Phi^{-1}(i/K)$, we have

$$\begin{aligned} r_{i,j}(\rho) &:= \mathrm{Cov}(\mathbb{I}\{X \le q_i\}, \mathbb{I}\{Y \le q_j\}) \\ &= \int_0^\rho \frac{1}{2\pi\sqrt{1-r^2}} \exp\left(-\frac{q_i^2 + q_j^2 - 2rq_iq_j}{2(1-r^2)}\right) \mathrm{d}r, \end{aligned} \tag{5}$$

where (X, Y) follows the bivariate normal distribution with mean zero, unit variance, and correlation ρ.

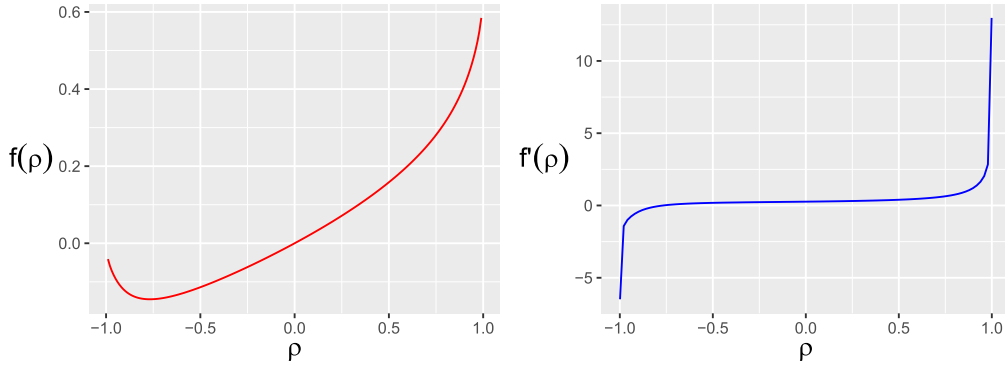

Figure 3. Function $f(\rho)$ and its derivative $f'(\rho)$ on $(-1, 1)$.

Proposition 1 provides a concrete procedure to compute the covariance matrix $\mathrm{Cov}_k(D)$. Importantly, it facilitates design optimization under Gaussianization, since one can formulate the covariate balance measures as objective functions on Σ:

$$\sum_{k=1}^{K} w_k^2 \|X^\top \mathrm{Cov}_k(D) X\|_{\mathrm{norm}} = \sum_{k=1}^{K} w_k^2 \|X^\top f_k(\Sigma) X\|_{\mathrm{norm}}. \tag{6}$$

In summary, we propose general covariate balance measures for uniform designs and derive their explicit Gaussianized representations. This Gaussianization enables feasible design optimization algorithms over the space of Gaussian covariance matrices, which will be the focus of Section 4.

Proposition 1 warrants more technical clarifications. First, its main benefit comes from (5), which provides analytical expressions for $\mathrm{Cov}_k(D)$. Alternatively, one may evaluate each covariance in (5) by Monte Carlo, but such simulation-based methods can be computationally challenging for large-scale randomized experiments. Second, we illustrate below the shape of f_k through the three-treatment example.

Remark 1 (Evaluation of f_k in the three-treatment example). Given $K = 3$, we evaluate $f(\rho) = \sum_{k=1}^{3} f_k(\rho)$, which maps Σ to the aggregated treatment covariance $\sum_{k=1}^{3} \mathrm{Cov}_k(D)$. This function represents the design optimization objective in Section 2, since for $w_k = 1/3$, the covariate balance measure in the nuclear norm reduces to $\sum_{k=1}^{3} w_k^2 \|X^\top \mathrm{Cov}_k(D) X\|_{\mathrm{nuc}} \propto \|X^\top \sum_{k=1}^{3} \mathrm{Cov}_k(D) X\|_{\mathrm{nuc}} = \|X^\top f(\Sigma) X\|_{\mathrm{nuc}}$. From Figure 3, positive Gaussian correlations often carry through discretization to positive treatment correlations, but strong negative Gaussian correlations can be diluted by discretization: $f(-1)$ and $f(0)$ induce similar correlations that are close to zero. The information loss from discretization is the cost of tractability, as directly optimizing the treatment covariance without Gaussianization is NP-hard. Finally, note that $f'(\rho)$ diverges at the endpoints $\rho = \pm 1$; this singular behavior of f' will guide us in developing optimization algorithms in Section 4.

3.3 Mehler's formula and proof sketch of Proposition 1

We prove Proposition 1 by leveraging a representation (Liang & Tran-Bach, 2022) of the bivariate normal distribution based on Mehler's formula and Hermite polynomials. To begin with, we define Hermite polynomials in the probabilists' convention.

Definition 2 For non-negative integers $m \geq 0$, define the mth order Hermite polynomial

$$\mathrm{He}_m(x) = \frac{(-1)^m}{\phi(x)} \frac{\mathrm{d}^m}{\mathrm{d}x^m} \phi(x).$$

with the standard normal density ϕ. Define normalized Hermite polynomials as $h_m(x) := \mathrm{He}_m(x)/\sqrt{m!}$.

Let L^2_ϕ be the class of square-integrable functions with respect to the standard normal distribution. Then, the set $\{h_m\}_{m=0}^{\infty}$ forms an orthonormal basis of L^2_ϕ as one can verify that $\mathbb{E}_{Z\sim\mathcal{N}(0,1)}[h_m(Z)h_{m'}(Z)] = \mathbb{I}\{m = m'\}$. For any function $g \in L^2_\phi$, the mth Hermite coefficient can be defined by $\alpha_m[g] := \mathop{\mathbb{E}}_{Z\sim\mathcal{N}(0,1)}[g(Z)h_m(Z)]$.

Let $p_\rho(x, y)$ be the density function of the bivariate normal distribution with unit variance and correlation ρ. Mehler's formula (Mehler, 1866) connects $p_\rho(x, y)$ to Hermite polynomials: $p_\rho(x, y) = \sum_{m=0}^{\infty} \rho^m h_m(x)h_m(y)\phi(x)\phi(y)$. That is, the density $p_\rho(x, y)$ can be decomposed into a sequence of products of Hermite polynomials and standard normal densities. Based on this result, we establish a representation for the covariance of functions on bivariate normal distributions.

Lemma 2 For $g, h \in L^2_\phi$, if $(X, Y) \in \mathbb{R}^2$ follow a bivariate normal distribution with mean zero, unit variance, and correlation ρ, we have

$$\mathop{\mathrm{Cov}}_{(X,Y)}[g(X), h(Y)] = \sum_{m=1}^{\infty} \alpha_m[g]\alpha_m[h]\rho^m.$$

Based on Mehler's formula and Lemma 2, we provide a proof sketch for Proposition 1. A complete proof can be found in online supplementary material, Section S3.1.

Proof Sketch. Here we focus on the proof of Equation (5), which is the key step in proving the result. Let $g(x) = \mathbb{I}\{x \le q_i\}$ and $h(x) = \mathbb{I}\{x \le q_j\}$. We leverage the derivative representation of Hermite polynomials (Definition 2) and Lemma 2 to obtain

$$\alpha_m[g] = -\frac{1}{\sqrt{m!}}\phi(q_i)\mathrm{He}_{m-1}(q_i), \quad \alpha_m[h] = -\frac{1}{\sqrt{m!}}\phi(q_j)\mathrm{He}_{m-1}(q_j),$$
$$\overset{\text{Lemma 2}}{\Longrightarrow} r_{ij}(\rho) = \sum_{m=1}^{\infty} \frac{1}{m!}\mathrm{He}_{m-1}(q_i)\mathrm{He}_{m-1}(q_j)\phi(q_i)\phi(q_j)\rho^m.$$

Notice that $r_{ij}(0) = 0$ and

$$r'_{ij}(\rho) = \sum_{m=1}^{\infty} \frac{1}{(m-1)!}\mathrm{He}_{m-1}(q_i)\mathrm{He}_{m-1}(q_j)\phi(q_i)\phi(q_j)\rho^{m-1} = p_\rho(q_i, q_j).$$

We obtain

$$r_{ij}(\rho) = \int_0^\rho p_r(q_i, q_j)\mathrm{d}r = \int_0^\rho \frac{1}{2\pi\sqrt{1-r^2}}\exp\left(-\frac{q_i^2 + q_j^2 - 2rq_iq_j}{2(1-r^2)}\right)\mathrm{d}r.$$

In summary, Proposition 1 can be proved by applying Mehler's formula to the covariance expression in (5). This technique will be used again in design optimization for the continuous setting (Section 5). Notably, this technical tool is designed for bivariate normal distributions, which further motivates the Gaussianization of treatments.

4 Gaussianized design optimization

In this section, we focus on solving the following optimization problem:

$$\min_{\Sigma\in\mathcal{E}} \|X^\top f(\Sigma)X\|_{\mathrm{norm}} =: l_{\mathrm{norm}}(\Sigma), \ \mathrm{norm} \in \{\mathrm{nuc}, \mathrm{op}\}, \tag{7}$$

Algorithm 1 Projected Gradient Descent for GDO (PGD-Gauss)

Data: $X \in \mathbb{R}^{n\times d}$, an evaluation function f, an initial design Σ^1, and the number of iterations T.

Result: Optimized covariance matrix Σ^*.

begin

Parametrize $\Sigma^1 = V^1(V^1)^\top$, where $V^1 \in \mathbb{R}^{n\times k}$, $\|v_i^1\| = 1$, and k equals the rank of Σ^1. Here v_i^1 is the i-th row of V^1.

for $t = 1, \ldots, T$ **do**

Compute $\nabla l_{\text{norm}}(\Sigma^t)$. $V^{t+1} = \left[I_n - \eta_t \nabla l_{\text{norm}}(\Sigma^t)\right] V^t$ for a suitable step size η_t.

$v_i^{t+1} \leftarrow v_i^{t+1}/\|v_i^{t+1}\|$.

$\Sigma^{t+1} \leftarrow V^{t+1}(V^{t+1})^\top$.

$\Sigma^* \leftarrow \Sigma^T$.

where f is a given elementwise function defined on $[-1, 1]$. Based on the linearity of the nuclear norm, the objective in (7) in $\|\cdot\|_{\text{nuc}}$ is equivalent to (6) by setting $f(\rho) = \sum_k w_k^2 f_k(\rho)$. Under the operator norm, the design optimization problem (6) is a weighted sum of objectives in the form of (7), and one can slightly modify the algorithm below to solve (6). Moreover, the general optimization problem (7) encompasses other covariate balance objectives in Section 5.

Formally, we propose Algorithm 1 to solve (7) above. This algorithm applies projected gradient descent (PGD-Gauss) on a factorized representation of Σ, similar to the Burer-Monteiro approach in semidefinite programming (Burer & Monteiro, 2003). The function f in design optimization may have an infinite derivative at ± 1 (Remark 1). Conceptually, this type of f will set ± 1 to be a barrier. Therefore, in Algorithm 1, we fix the diagonal values of Σ^t and only update the off-diagonal entries. That is, we consider

$$\nabla l_{\text{nuc}}(\Sigma^t) = (XX^\top - \text{diag}(XX^\top)) \circ f'(V^t V^{t\top}),$$
$$\nabla l_{\text{op}}(\Sigma^t) = (Xu_1u_1^\top X^\top - \text{diag}(Xu_1u_1^\top X^\top)) \circ f'(V^t V^{t\top}),$$

where $\circ$ is the Hadamard product, $u_1 \in \mathbb{R}^d$ is the leading eigenvector of $X^\top f(\Sigma^t)X$. For diagonal elements in the gradient, we adopt the convention $0 \times f'(\pm 1) = 0$. Notably, f' can be obtained by directly differentiating the analytic functions f_k defined in Proposition 1.

Since the objective function is non-convex in Σ in general, the PGD-Gauss only yields a local minimizer near the initial covariance matrix Σ^1. As explained in Section 2, GDO is not tailored to identify the global solution that perfectly balances the covariates, but rather serves as a tool for locally improving a given input design. The computational cost of the procedure is modest, as detailed in online supplementary material, Section S2 of the supplementary material.

By default, we initialize the design optimization by setting $\Sigma^1 = I_n$, which results in i.i.d. treatments. We view this as the baseline Gaussianized design, as it does not incorporate covariate information, and i.i.d. designs have a robust performance against unknown outcome-generating models (Harshaw et al., 2024; Wu, 1981). Therefore, the number of steps we run PGD-Gauss is an explicit tradeoff between robustness and covariate balance. In the simulations of Section 7, the i.i.d. initialization leads to satisfactory performance compared to state-of-the-art designs. From Figure 1, it is possible that alternative initializations can further improve the MSE objective, but doing so requires a justified choice of initial design and depends on the experimenter's preferences for covariate balance and design robustness.

Lastly, we demonstrate the theoretical benefits of the optimal Gaussianized design by analyzing the nuclear-norm objective under the binary treatment setup. In this setting, i.i.d. Bernoulli randomization is a common design that does not enforce covariate balance, and can be realized by setting $\Sigma = I_n$ under Gaussianization. Let Obj_{Bern} be the value of the nuclear-norm objective under Bernoulli randomization. Meanwhile, let Obj_* be the theoretically optimal objective value over all possible uniform designs. It holds that, for a constant C_0, $\frac{\text{Obj}_{\text{Bern}}}{\text{Obj}_*} \geq C_0 \frac{\sum_{i=1}^n \|X_i\|^2}{d \max_i \|X_i\|^2}$. The lower bound above typically diverges when the covariate dimension d grows at a slower rate compared to

$n\,(d/n \to 0)$, implying that Bernoulli randomization is substantially less balanced than the optimal design.

As we explained in Section 2, computing the theoretically optimal solution is generally intractable, and thus we study the objective value Obj_G achieved from Gaussianized design optimization. Under a regularity assumption on the solution of Obj_*, we show that $\frac{\mathrm{Obj}_G}{\mathrm{Obj}_*} \le (1-\frac{2}{\pi})\frac{\mathrm{Obj}_{\mathrm{Bern}}}{\mathrm{Obj}_*}+\frac{2}{\pi}$. Hence, Gaussianization interpolates between Bernoulli randomization and the optimal design.

Related analyses have appeared in prior work but under different setups. For example, Harshaw et al. (2024) studied a covariate balance objective measured by the operator norm. In particular, Sections 4.2 and 6.2 of Harshaw et al. (2024) reveal a substantial covariate balance gap between Bernoulli randomization and their design when $d/n \to 0$, which is consistent with our findings.

In our design optimization, we introduce Gaussianization to obtain an approximate solution to the intractable design problem. A similar idea was employed by Bhat et al. (2020) for near-optimal A/B testing, and our approximation ratio is analogous to their results in Section 3. Nevertheless, the approaches differ in important ways. First, Bhat et al. (2020) adopted a linear model and targeted the control of the OLS estimation error, whereas our analysis is model-agnostic and focuses on the estimation error of the Horvitz–Thompson estimator. Second, Bhat et al. (2020) focused on the binary case, whereas our framework and approximation analysis can accommodate multiple treatment arms. Complete proofs of the above results and extensions to multiple treatment arms can be found in online supplementary material, Section S3.4 of the supplementary material.

5 Gaussian design with continuous treatments

In this section, we extend Gaussianization to settings with continuous treatments. Specifically, we introduce a new experimental design, called Gaussian design, to assign continuous treatments based on a multivariate Gaussian distribution.

Definition 3 (Gaussian Design). A Gaussian design allocates treatment T_i to unit i, where $T = (T_1, \ldots, T_n) \sim \mathcal{N}(0, \Sigma)$ for some $\Sigma \in \mathcal{E}$.

When the actual treatment is restricted to a bounded interval $[a, b]$, one may compute a rescaled treatment assignment $(a+b)/2 + T_i(b-a)/(2t_\alpha)$, where $t_\alpha := \sqrt{2\log(2n/\alpha)}$ satisfies $\mathbb{P}(\max_i |T_i| \le t_\alpha) \ge 1-\alpha$. This ensures that all the rescaled treatments fall within $[a, b]$ with probability at least $1-\alpha$. Choosing α is analogous to setting an acceptance threshold in rerandomization, as both parameters control a family-wise tail probability. Hence we follow the rerandomization literature to use $\alpha = 0.001$ (Li et al., 2018). Gaussian designs directly allocate continuous treatments as above, and they are thus mechanically different from the Gaussianization perspective, where we focus on discrete treatments but model them with latent Gaussian variables. Compared to Gaussianization, Gaussian designs capture structural properties of potential outcome functions as discussed below.

5.1 Causal estimands

We denote $Y_i(t)$ as the response function for unit i and $t \in \mathbb{R}$, which generalizes the potential outcomes to continuous treatments. By an abuse of notation, we use $Y_i = Y_i(T_i)$ to denote the observed outcome for unit i. Given continuous treatments and response functions, we work with a class of causal effects of the form

$$\tau_w^c = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}} Y_i(t)w(t)\phi(t)\mathrm{d}t, \tag{8}$$

where $w(\cdot)$ is a pre-specified weight function on different treatment values. We use the superscript c in τ_w^c to distinguish it from τ_w under the discrete setting.

Similar to the discrete setup, we focus on Horvitz-Thompson-type estimators

$$\widehat{\tau}_w^c = \frac{1}{n}\sum_{i=1}^{n} Y_i(T_i)w(T_i) = \frac{1}{n}\sum Y_i W_i, \quad W_i := w(T_i).$$

Clearly, $\widehat{\tau}_w^c$ is an unbiased estimator of τ_w^c under Gaussian design. In the following, we provide several leading examples of the weight function $w(\cdot)$ in (8) to obtain meaningful causal estimands.

Example 1 (Average Treatment Effects on a Given Interval). Suppose we want to learn about the average treatment effect on a treatment interval $[l, r]$ (Fryges & Wagner, 2008). We may set $w(t) = \frac{\mathbb{I}\{t\in[l,\,r]\}}{(r-l)\phi(t)}$, which leads to

$$\tau_w^c = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{r-l}\int_l^r Y_i(x)\mathrm{d}x, \quad \widehat{\tau}_w^c = \frac{1}{n}\sum_{i=1}^{n} Y_i\frac{\mathbb{I}\{T_i\in[l,\,r]\}}{(r-l)\phi(T_i)}.$$

Example 2 (First derivative). Suppose $Y_i(t)$ is differentiable with $\mathbb{E}Y_i^2(T_i) < \infty$ and $\mathbb{E}|Y_i'(T_i)| < \infty$. To learn the first derivative of response functions, we consider $w(t) = t$ and obtain

$$\tau_w^c = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}} Y_i(t)t\phi(t)\mathrm{d}t \overset{(i)}{=} \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}} Y_i'(t)\phi(t)\mathrm{d}t, \quad \widehat{\tau}_w^c = \frac{1}{n}\sum_{i=1}^{n} Y_iT_i,$$

where (i) follows from Stein's Lemma. When the treatment is price, the first derivative corresponds to the price elasticity of demand (Mas-Colell et al., 1995). Notably, if we replace the base Gaussian density $\phi(t)$ with $\psi(t) = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ (a mixture of two Dirac measures), the causal estimand reduces to $\tau_w^c = \frac{1}{2n}\sum_{i=1}^{n}(Y_i(1) - Y_i(-1))$, corresponding to the average treatment effect under binary treatments.

Example 3 (Second derivative). Suppose $Y_i(t)$ is twice differentiable with $\mathbb{E}Y_i^2(T_i) < \infty$ and $\mathbb{E}|Y_i''(T_i)| < \infty$. To learn the second derivative, we consider $w(t) = t^2 - 1$ and obtain

$$\tau_w^c = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}} Y_i(t)(t^2-1)\phi(t)\mathrm{d}t \overset{(i)}{=} \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}} Y_i''(t)\phi(t)\mathrm{d}t,$$
$$\widehat{\tau}_w^c = \frac{1}{n}\sum_{i=1}^{n} Y_i(T_i^2-1),$$

where (i) follows by applying Stein's Lemma twice. $\widehat{\tau}_w^c$ serves as an unbiased estimator for the average second derivative of response functions. With expenditure-based treatments (e.g. advertising or incentives), the second derivative reflects diminishing marginal returns (Shephard & Färe, 1974).

5.2 Variance formula and measures of covariate balance

To analyze the variance of the estimators, we decompose $Y_i(t)$ as follows:

$$Y_i(t) = a_iY_0(t) + b_i, \quad a_i = X_i^\top\beta_1, \quad b_i = X_i^\top\beta_2. \tag{9}$$

In this decomposition, a_i and b_i control the scale and location of the ith response function, and they are perfectly linear in covariates. $Y_0(t)$ is a baseline response function, which is assumed to be known to the researcher. This assumption is justified as researchers often have prior knowledge of the shape of response functions, such as sigmoid dose-response curves in clinical trials (Meddings et al., 1989), and exponential utility functions in economics (Arrow, 1971).

Under (9), we analyze the variance of $\widehat{\tau}_w^c$. For two random vectors $X, Y \in \mathbb{R}^d$, we use the notation $\mathrm{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)^\top]$ and $\mathrm{Cov}(X) := \mathrm{Cov}(X, X)$. Then, under Equation (9), it holds that

$$\begin{aligned}\mathrm{Var}(\widehat{\tau}_w^c) &= \frac{1}{n^2}\left(\beta_1^\top X^\top \mathrm{Cov}(Y_0 \circ W) X\beta_1 + \beta_2^\top X^\top \mathrm{Cov}(W) X\beta_2 + 2\beta_1^\top X^\top \mathrm{Cov}(Y_0 \circ W,\, W) X\beta_2\right) \\ &\le \frac{2}{n^2}\left(\beta_1^\top X^\top \mathrm{Cov}(Y_0 \circ W) X\beta_1 + \beta_2^\top X^\top \mathrm{Cov}(W) X\beta_2\right).\end{aligned} \tag{10}$$

With a slight abuse of notation, we define $Y_0 = (Y_0(T_1), \ldots, Y_0(T_n))^\top$ and $W = (W_1, \ldots, W_n)^\top$, where $\circ$ denotes the Hadamard (elementwise) product, and the second line follows from the AM-GM inequality. From Equation (10), the estimation performance is characterized by quadratic forms analogous to those in the discrete setting (Lemma 1). In addition, the variance in inequality (10) depends on the coefficients β_1, β_2, which are unknown in general.

To make progress, we adopt a similar approach to that in Section 3. We first assume that β_1, β_2 are random signals with mean zero and identity covariance matrix, which leads to an average-case MSE measure:

$$\begin{aligned}\mathop{\mathbb{E}}_{\beta_1,\beta_2} \mathrm{Var}(\widehat{\tau}_w^c) &\le \frac{2}{n^2}\mathrm{tr}(X^\top(\mathrm{Cov}(Y_0 \circ W) + \mathrm{Cov}(W))X) \\ &\propto \|X^\top \mathrm{Cov}(Y_0 \circ W) X\|_{\mathrm{nuc}} + \|X^\top \mathrm{Cov}(W) X\|_{\mathrm{nuc}}.\end{aligned}$$

By assuming $\|\beta_1\| \le M$, $\|\beta_2\| \le M$, we obtain an upper bound on the worst-case MSE:

$$\begin{aligned}\sup_{\|\beta_1\|\le M, \|\beta_2\|\le M} \mathrm{Var}(\widehat{\tau}_w^c) &\le \sup_{\|\beta_1\|\le M, \|\beta_2\|\le M} \frac{2}{n^2}\left(\beta_1^\top X^\top \mathrm{Cov}(Y_0 \circ W) X\beta_1 + \beta_2^\top X^\top \mathrm{Cov}(W) X\beta_2\right) \\ &\propto \|X^\top \mathrm{Cov}(Y_0 \circ W) X\|_{\mathrm{op}} + \|X^\top \mathrm{Cov}(W) X\|_{\mathrm{op}}.\end{aligned}$$

These analytical steps lead to the formal definition of covariate balance measures under Gaussian designs.

Definition 4 For Gaussian designs with a baseline response function Y_0 and a weight function w, define the average and worst-case covariate balance measures as

$$\|X^\top \mathrm{Cov}(Y_0 \circ W) X\|_{\mathrm{norm}} + \|X^\top \mathrm{Cov}(W) X\|_{\mathrm{norm}},\ \mathrm{norm} \in \{\mathrm{nuc}, \mathrm{op}\}. \tag{11}$$

5.3 Gaussianized representation

Using Mehler's formula and Hermite coefficients in Section 3, we derive the following result, which is a direct application of Lemma 2.

Proposition 2 Suppose that $Y_0 w : t \mapsto Y_0(t)w(t) \in L^2_\phi$ and $w \in L^2_\phi$. Then we have $\mathrm{Cov}(Y_0 \circ W) = f_{Y_0,w}(\Sigma)$ and $\mathrm{Cov}(W) = f_w(\Sigma)$. Here, $f_{Y_0,w}$ and f_w are elementwise functions defined by

$$f_{Y_0,w}(\rho) = \sum_{m=1}^{\infty} \alpha_m[Y_0 w]^2 \rho^m, \quad f_w(\rho) = \sum_{m=1}^{\infty} \alpha_m[w]^2 \rho^m, \quad \rho \in [-1, 1],$$

where $\alpha_m[g]$ is the mth Hermite coefficient of the function g.

Proposition 2 demonstrates that the covariance matrices in covariate balance measures can be explicitly written as functions of Σ. This result facilitates optimization over Σ, similar to the role of Proposition 1 in the uniform design setup. Combining the results above, we formulate covariate balance measures $\|X^\top f_{Y_0,w}(\Sigma) X\|_{\mathrm{norm}} + \|X^\top f_w(\Sigma) X\|_{\mathrm{norm}}$, for $\mathrm{norm} \in \{\mathrm{nuc}, \mathrm{op}\}$. Consequently,

one may directly apply the algorithm proposed in Section 4 to Gaussian design and optimize the covariate balance. In online supplementary material, Section S2.3 of the supplementary material, we empirically evaluate the Gaussian design for continuous treatments using semi-synthetic experimental data. The results show that the optimized Gaussian design improves estimation precision and facilitates testing structural properties of the potential outcome functions, such as monotonicity and convexity.

6 Asymptotics and inference

In this section, we study asymptotic properties and inference under the Gaussianization $T \sim \mathcal{N}(0, \Sigma)$, where Σ is a solution obtained from the PGD-Gauss algorithm in Section 4, within the design-based framework. In design-based inference (Imbens & Rubin, 2015), we view the potential outcomes as fixed and the only randomness comes from the treatment assignment, i.e. the Gaussian treatment T. Here, we prove asymptotic normality under Gaussianization, and provide concrete procedures to compute confidence intervals. The key takeaway is that Gaussianization under the PGD-Gauss solution results in smaller variance compared to the i.i.d. Gaussianization ($\Sigma = I_n$), and thus improves estimation efficiency. Notably, our asymptotic theory allows high-dimensional covariates, where d can grow with, or even be larger than n.

For presentation purposes, we focus on the uniform design setup in Section 3 with treatments modeled by $D_i = g(T_i)$. Inference for continuous treatments is discussed in online supplementary material, Section S1. Throughout this section, we assume that the number of treatment arms K is fixed.

6.1 Asymptotic normality

Here, we focus on PGD-Gauss under the nuclear norm objective $\|X^\top f_k(\Sigma)X\|_{\text{nuc}}$. Recall that f_k defined in Proposition 1 is the covariance mapping with respect to treatment k, and thus the objective is a covariate balance measure for the kth average treatment effect. Similar normality results can be shown under the operator norm, but we focus on the nuclear norm for simplicity.

We study the asymptotic properties of $\widehat{\tau}_k$, the average treatment effect for arm k. We focus on implementing one step of PGD-Gauss with a step size η using the default initialization $\Sigma^1 = I_n$, and denote the resulting solution by Σ_η. We impose the following assumption on η and X.

Assumption 1 The covariates $X \in \mathbb{R}^{n\times d}$ satisfy $\|X_i\| = 1$, i.e. each row of X has unit norm. The step size in PGD-Gauss satisfies $\eta\|XX^\top - I_n\|_{\text{op}} = o(1)$.

Assumption 1 requires that Σ_η is a local perturbation of I_n by controlling the step size, which is the key condition to establish asymptotic normality. To better understand the step size condition, we may consider $X_i \overset{iid}{\sim} \mathcal{N}(0, \frac{1}{d} I_d)$, so that $\|X_i\| \approx 1$ in expectation. Then, results from random matrix theory suggest that $\|XX^\top\|_{\text{op}} \approx (1 + \sqrt{\frac{n}{d}})^2$ (Marčenko & Pastur, 1967; Vershynin, 2018). If, for intuition, we assume that $n > d$, the step size condition boils down to $\eta = o(d/n)$, the ratio of d to n.

To characterize the asymptotic distribution under the one-step PGD-Gauss, we define a sequence of ancillary potential outcomes. Using the f_k in Proposition 1, we define $\tilde{Y}(k) = f_k(I_n)^{-1/2} f_k(\Sigma_\eta)^{1/2} Y(k)$.

Theorem 1 Suppose Assumption 1 holds. Consider Gaussianization $T \sim \mathcal{N}(0, \Sigma_\eta)$, where Σ_η is the solution obtained from the one-step PGD-Gauss. If, as n goes to infinity, (a) $\max_{i=1,\dots,n} \tilde{Y}_i^2(k)/n \to 0$, (b) $n\text{Var}(\widehat{\tau}_k) = \frac{K-1}{n}\sum_{i=1}^n \tilde{Y}_i^2(k)$ has a positive limit, and (c) $\|Y(k)\|^2 \le nM$ for some constant $M > 0$, it holds that

$$\sqrt{n}\big(\widehat{\tau}_k - \tau_k\big) \xrightarrow{d} \mathcal{N}(0, \lim_{n\to\infty} n\text{Var}(\widehat{\tau}_k)) = \mathcal{N}\left(0, \lim_{n\to\infty} \frac{K-1}{n}\|\tilde{Y}(k)\|^2\right).$$

Theorem 1 establishes the asymptotic normality of $\widehat{\tau}_k$ under Gaussianization. Its proof relies on the asymptotic equivalence between $\widehat{\tau}_k$ under Σ_η and an ancillary estimator under i.i.d. Gaussianization. Due to the asymptotic equivalence, it suffices to prove the asymptotic normality

for the ancillary estimator using Lindeberg's central limit theorem. The proof and a generalization to multi-step PGD-Gauss can be found in online supplementary material, Section S3.2.

Notably, the variance term in Theorem 1 indicates the benefit of running PGD-Gauss for covariate balance. To see this, we may define $V(\Sigma) := \frac{K-1}{n}\|\tilde{Y}(k)\|^2$. From the proof of Theorem 1, $V(\Sigma)$ not only captures the variance limit, but also exactly matches the finite-sample variance of $\sqrt{n}(\widehat{\tau}_k - \tau_k)$, i.e. the MSE of $\widehat{\tau}_k$ after rescaling. The following proposition shows that $V(\Sigma_\eta)$ is strictly smaller than $V(I_n)$ on average. Denote by $\|A\|_F$ the Frobenius norm of a matrix A. Write $a_n = \Omega(b_n)$ if there exists a constant c such that $a_n \geq cb_n$ for n large enough.

Proposition 3 Suppose $Y(k) = X\beta_k$, where β_k is a random signal with zero mean and identity covariance matrix. In addition, suppose Assumption 1 holds and $\eta = o(1)$, $n\eta^3\|XX^\top - I_n\|_{\text{op}}^4 = o(\|XX^\top - I_n\|_F^2)$. If $f_k'(0) \neq 0$, we have $\mathbb{E}_{\beta_k} V(I_n) - \mathbb{E}_{\beta_k} V(\Sigma_\eta) = \Omega(\frac{\eta}{n}\|XX^\top - I_n\|_F^2) > 0$.

Proposition 3 suggests that for n large enough, there is a nonzero improvement in $V(\Sigma_\eta)$ in the average sense of random signals defined above. Therefore, $\widehat{\tau}_k$ has a smaller variance under Σ_η compared to the initial design I_n, which reveals the benefit of covariate balance. However, we clarify that the improvement in Proposition 3 is with respect to the non-asymptotic variance $V(\Sigma_\eta)$. Under Gaussian covariates $X_i \overset{iid}{\sim} \mathcal{N}(0, \frac{1}{d} I_d)$, the typical scales are $\eta = o(d/n)$, $\|XX^\top - I_n\|_F^2 \approx n^2/d$, such that the variance reduction is of order $o(1)$. This limitation reflects our focus on one-step updates. In practice, PGD-Gauss is run for many iterations proportional to n, and the per-step gains accumulate to a constant-level reduction, as shown in our numerical experiments across multiple setups. Theoretical conditions under which the one-step PGD-Gauss reduces the limiting variance remain an open and complex problem, which we consider as future work.

6.2 Inference

To make inference under Gaussianized designs, we need to estimate the variance of $\widehat{\tau}_k$. Under the one-step PGD-Gauss with a covariance matrix Σ_η, by Theorem 1 and Proposition 1, we write the asymptotic variance of $\widehat{\tau}_k$ as

$$V(\Sigma_\eta) = \frac{K-1}{n}\|\tilde{Y}(k)\|^2 = \frac{K^2}{n} Y(k)^\top f_k(\Sigma_\eta) Y(k) = \frac{K^2}{n}\sum_{i,j=1}^n Y_i(k) Y_j(k) f_k(\Sigma_{\eta,ij}).$$

We use a Horvitz-Thompson estimator to estimate the variance as below:

$$\widehat{V}_\eta = \frac{K^2}{n}\sum_{i,j=1}^n Y_i Y_j f_k(\Sigma_{\eta,ij}) \frac{\mathbb{I}\{g(T_i) = k, g(T_j) = k\}}{\mathbb{P}(g(T_i) = k, g(T_j) = k)}. \tag{12}$$

The joint treatment probabilities $\mathbb{P}(g(T_i) = k, g(T_j) = k)$ are determined by the design, and can be directly computed by Proposition 1, i.e. $\mathbb{P}(g(T_i) = k, g(T_j) = k) = f_k(\Sigma_{\eta,ij}) + 1/K^2$. The following result shows that $\widehat{V}_\eta$ is a consistent variance estimator.

Theorem 2 Suppose that $\max_i |Y_i(k)| = O(1)$ and Assumption 1 holds with η satisfying $n^2\eta^2\|XX^\top - I_n\|_{\text{op}}^2 = o(1)$. Then, $\widehat{V}_\eta$ is a well-defined variance estimator with $\mathbb{E}\widehat{V}_\eta = V(\Sigma_\eta)$ and $\text{Var}(\widehat{V}_\eta) = o(1)$.

Theorem 2 enables inference under the Gaussianized design Σ_η, as one can combine Theorems 1 and 2 to derive the design-based confidence interval $[\widehat{\tau}_k - z_{\alpha/2}\sqrt{\widehat{V}_\eta/n}, \widehat{\tau}_k + z_{\alpha/2}\sqrt{\widehat{V}_\eta/n}]$. Here, we set $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$ to obtain an asymptotic $(1-\alpha)$ confidence interval. Compared to Theorem 1, Theorem 2 requires a stronger condition $n^2\eta^2\|XX^\top - I_n\|_{\text{op}}^2 = o(1)$ on the step size η, as we need to bound higher moments of treatment assignments in the variance estimator.

In this section, we have focused on $\widehat{\tau}_k$ under the one-step PGD-Gauss. However, it is also desirable to construct confidence intervals for $\widehat{\tau}_w$ under a general Gaussian covariance matrix Σ, which

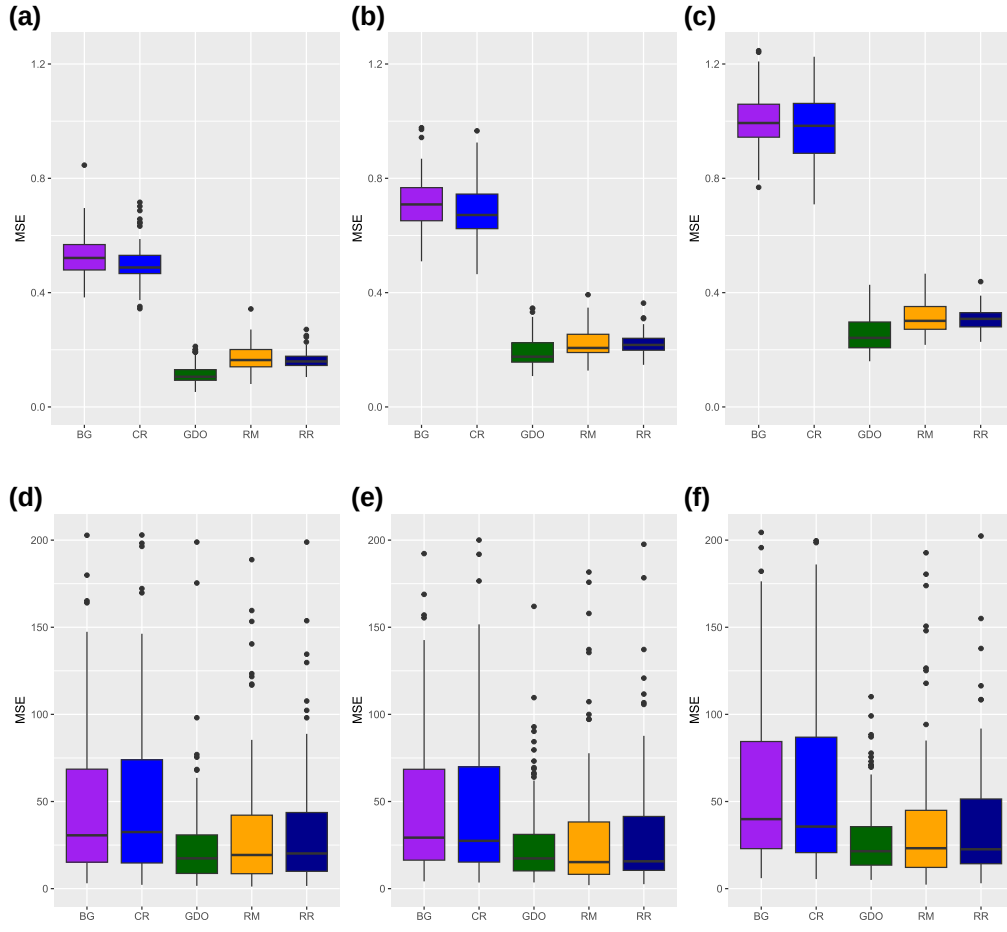

Figure 4. MSEs across different designs for estimating τ_1, τ_2, τ_{12} under linear and nonlinear models. (a) linear, τ_1. (b) linear, τ_2. (c) linear, τ_{12}. (d) nonlinear, τ_1. (e) nonlinear, τ_2 and (f) nonlinear, τ_{12}.

may be obtained by running PGD-Gauss until convergence. To this end, we discuss two general variance estimation approaches in online supplementary material, Section S1, including a conservative variance estimator and an alternative model-assisted procedure.

7 Simulations

In this section, we compare the estimation performance of different designs under a factorial setup. We set $n = 100$, $d = 5$, and $X_i \overset{iid}{\sim} \mathcal{N}(0, I_d)$. We consider a factorial design with two treatments $A_i \in \{0, 1\}$, $B_i \in \{0, 1\}$ with potential outcomes: $Y_i(A_i, B_i) = X_i^\top \beta_1 + A_i(X_i^\top \beta_2) + B_i(0.2 + X_i^\top \beta_3) + 0.5 A_i B_i + \varepsilon_i$, where $\beta_1 = (-1, -1, -2/3, -6/5, 0)$, $\beta_2 = (0, 0, -8/5, 8/5, 8/5)$, $\beta_3 = (2, 2, 2, 0, 0)^\top$, and $\{\varepsilon_i\}_{i=1}^n$ are correlated log-normal random variables with unit variance. To translate the factorial design to a standard uniform design, we encode the treatments by $D_i = 1 + 2A_i + B_i \in \{1, 2, 3, 4\}$. Then, we apply the Gaussianization techniques in Section 3 to model the treatments by $D_i = g(T_i)$ with the map g in Section 3, enabling Gaussianized design optimization.

In the factorial design under the potential outcome framework (Dasgupta et al., 2015), one is usually interested in the main effect of the first factor $\tau_1 := \frac{1}{2n}\sum_{i=1}^n (-Y_i(0, 0) - Y_i(0, 1) + Y_i(1, 0) + Y_i(1, 1)) = 0.25 + \frac{1}{n}\sum_{i=1}^n X_i^\top \beta_2$, the main effect of the second factor $\tau_2 := \frac{1}{2n}\sum_{i=1}^n (-Y_i(0, 0) + Y_i(0, 1) - Y_i(1, 0) + Y_i(1, 1)) = 0.45 + \frac{1}{n}\sum_{i=1}^n X_i^\top \beta_3$, and the interaction

effect $\tau_{12} := \frac{1}{2n}\sum_{i=1}^{n}(Y_i(0,0) - Y_i(0,1) - Y_i(1,0) + Y_i(1,1)) = 0.25$. We estimate them based on Horvitz-Thompson estimators.

We evaluate the MSE of Horvitz-Thompson estimators under different designs. We implement baseline Gaussianization (BG) with $\Sigma = I_n$, and Gaussianized design optimization (GDO) with Σ^*. The optimized covariance matrix Σ^* is obtained by running PGD-Gauss to solve the nuclear-norm objective with i.i.d. initialization and 200 iterations. For comparison purposes, we implement complete randomization (CR) (Dasgupta et al., 2015), recursive matching (RM) (Bai et al., 2024) and rerandomization (RR) (Li et al., 2020). RM and RR can be considered as state-of-the-art designs for covariate balance in the factorial setup.

The MSEs are presented in boxplots in Figure 4a–c, where we evaluate the MSEs based on 1,000 Monte Carlo runs, repeated over 100 independently generated datasets. We observe that, across all three estimation problems, GDO achieves the smallest MSE among five designs. In online supplementary material, Section S2 of the supplementary material, we further provide a scatter plot of the MSEs for τ_1 over the covariate balance objective in the nuclear norm, evaluated under all designs. We observe that (1) a smaller covariate balance measure indicates smaller MSE on average, and (2) GDO achieves the smallest covariate balance measure across all designs, trailed by RM and RR. In online supplementary material, Section S2, we present further simulation details on GDO loss curves and confidence intervals.

Next, we apply the same design procedure to balance the covariate matrix X, but generate the outcomes and evaluate the MSE based on nonlinear features of X_i, i.e. $\phi(X_i) := (X_{i1}, X_{i2}^3, X_{i3}X_{i4}, X_{i4}^2, X_{i5})$. Figure 4d–f visualizes the relative performance of different designs. Under the nonlinear setting, all designs suffer from an increase in MSE due to large nonlinear terms, but GDO still achieves the smallest MSE on average and maintains reliable performance.

8 Conclusion

In our paper, we develop a Gaussianization framework to optimize experimental designs for covariate balance. This approach accommodates general covariates and multiple treatment arms, which constitutes a key advantage over existing methods. Moreover, Gaussianization seamlessly extends to continuous treatments via the Gaussian design, which may be of independent interest in practical applications. As an extension, it would be interesting to consider more complex settings, such as those involving interference. Additionally, developing a general asymptotic theory for Gaussianized designs that extends beyond local perturbations remains an open problem. We consider these areas promising topics for future work.

Acknowledgments

We appreciate the constructive comments and feedback from the Editor, Associate Editor, and referees, which significantly improved the presentation of the manuscript.

Conflicts of interest: The authors declare that they have no conflicts of interest, financial or otherwise, regarding the publication of this article.

Funding

TL is supported by the National Science Foundation Career Award (DMS-2042473) and by the Wallman Society of Fellows at the University of Chicago. PT acknowledges support from the National Science Foundation award SES-2419009.

Data availability

The R code used to generate the numerical results in Section 7 and the Supplementary Material, and instructions for accessing the real data are available at https://github.com/kraneguo/gdo_paper.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series B*.

References

Arrow, K. (1971). The theory of risk aversion. In *Essays in the theory of risk-bearing* (pp. 90–120). North-Holland.

Atkinson A. C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika*, *69*(1), 61–67. https://doi.org/10.1093/biomet/69.1.61

Bai Y. (2022). Optimality of matched-pair designs in randomized controlled trials. *American Economic Review*, *112*(12), 3911–3940. https://doi.org/10.1257/aer.20201856

Bai Y. (2023). Why randomize? Minimax optimality under permutation invariance. *Journal of Econometrics*, *232*(2), 565–575. https://doi.org/10.1016/j.jeconom.2021.10.009

Bai Y., Liu J., & Tabord-Meehan M. (2024). Inference for matched tuples and fully blocked factorial designs. *Quantitative Economics*, *15*(2), 279–330. https://doi.org/10.3982/QE2354

Bai Y., Romano J. P., & Shaikh A. M. (2022). Inference in experiments with matched pairs. *Journal of the American Statistical Association*, *117*(540), 1726–1737. https://doi.org/10.1080/01621459.2021.1883437

Barahona F., & Mahjoub A. R. (1986). On the cut polytope. *Mathematical Programming*, *36*(2), 157–173. https://doi.org/10.1007/BF02592023

Basse G., Feller A., & Toulis P. (2019). Randomization tests of causal effects under interference. *Biometrika*, *106*(2), 487–494. https://doi.org/10.1093/biomet/asy072

Basse G. W., & Airoldi E. M. (2018). Model-assisted design of experiments in the presence of network-correlated outcomes. *Biometrika*, *105*(4), 849–858. https://doi.org/10.1093/biomet/asy036

Basse G. W., Ding Y., & Toulis P. (2023). Minimax designs for causal effects in temporal experiments with treatment habituation. *Biometrika*, *110*(1), 155–168. https://doi.org/10.1093/biomet/asac024

Bastani H., & Bayati M. (2020). Online decision making with high-dimensional covariates. *Operations Research*, *68*(1), 276–294. https://doi.org/10.1287/opre.2019.1902

Bhat N., Farias V. F., Moallemi C. C., & Sinha D. (2020). Near-optimal ab testing. *Management Science*, *66*(10), 4477–4495. https://doi.org/10.1287/mnsc.2019.3424

Box G. E., & Draper N. R. (1959). A basis for the selection of a response surface design. *Journal of the American Statistical Association*, *54*(287), 622–654. https://doi.org/10.1080/01621459.1959.10501525

Bugni F. A., Canay I. A., & Shaikh A. M. (2018). Inference under covariate-adaptive randomization. *Journal of the American Statistical Association*, *113*(524), 1784–1796. https://doi.org/10.1080/01621459.2017.1375934

Bugni F. A., Canay I. A., & Shaikh A. M. (2019). Inference under covariate-adaptive randomization with multiple treatments. *Quantitative Economics*, *10*(4), 1747–1785. https://doi.org/10.3982/QE1150

Burer S., & Monteiro R. D. (2003). A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, *95*(2), 329–357. https://doi.org/10.1007/s10107-002-0352-8

Callaway B., Goodman-Bacon A., & Sant'Anna P. H. (2024). *Difference-in-differences with a continuous treatment* (Technical report). National Bureau of Economic Research.

Chang H. (2023). 'Design-based estimation theory for complex experiments', arXiv, arXiv:2311.06891v2, preprint: not peer reviewed. .

Colangelo, K., & Lee, Y.-Y. (2026). Double debiased machine learning nonparametric inference with continuous treatments. *Journal of Business & Economic Statistics*, *44*(1), 67–79. https://doi.org/10.1080/07350015.2025.2505487

Cox D. R., & Reid N. (2000). *The theory of the design of experiments*. Chapman and Hall/CRC.

Dasgupta T., Pillai N. S., & Rubin D. B. (2015). Causal inference from 2k factorial designs by using potential outcomes. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *77*(4), 727–753. https://doi.org/10.1111/rssb.12085

Davezies L., Hollard G., & Merino P. V. (2025). 'Revisiting randomization with the cube method', arXiv, arXiv:2407.13613v3, preprint: not peer reviewed.

De Chaisemartin C., d'Haultfoeuille X., Pasquier F., & Vazquez-Bare G. (2022). 'Difference-in-differences estimators for treatments continuously distributed at every period', arXiv, arXiv:2201.06898, preprint: not peer reviewed.

Ding P., Feller A., & Miratrix L. (2016). Randomization inference for treatment effect variation. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *78*(3), 655–671. https://doi.org/10.1111/rssb.12124

Dong Y., & Lee Y.-Y. (2023). 'Nonparametric doubly robust identification of causal effects of a continuous treatment using discrete instruments', arXiv, arXiv:2310.18504, preprint: not peer reviewed.

Fisher R. A. (1925). *Statistical methods for research workers*. Edinburgh Oliver & Boyd.

Fisher R. A. (1926). The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain*, *33*, 503–513. https://doi.org/10.23637/rothamsted.8v61q
Fisher R. A. (1935). *The design of experiments*. Oliver and Boyd.
Freedman D. A. (2008). On regression adjustments to experimental data. *Advances in Applied Mathematics*, *40*(2), 180–193. https://doi.org/10.1016/j.aam.2006.12.003
Fryges H., & Wagner J. (2008). Exports and productivity growth: First evidence from a continuous treatment approach. *Weltwirtschaftliches Archiv*, *144*(4), 695–722. https://doi.org/10.1007/s10290-008-0166-8
Goemans M. X., & Williamson D. P. (1995). Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, *42*(6), 1115–1145. https://doi.org/10.1145/227683.227684
Greevy R., Lu B., Silber J. H., & Rosenbaum P. (2004). Optimal multivariate matching before randomization. *Biostatistics*, *5*(2), 263–275. https://doi.org/10.1093/biostatistics/5.2.263
Guo W., Lee J., & Toulis P. (2025). 'Ml-assisted randomization tests for detecting treatment effects in a/b experiments', arXiv, arXiv:2501.07722v1, preprint: not peer reviewed.
Harshaw C., Sävje F., Spielman D. A., & Zhang P. (2024). Balancing covariates in randomized experiments with the Gram–Schmidt walk design. *Journal of the American Statistical Association*, *119*(548), 2934–2946. https://doi.org/10.1080/01621459.2023.2285474
Hege V. S. (1967). An optimum property of the Horvitz–Thomson estimate. *Journal of the American Statistical Association*, *62*(319), 1013–1017. https://doi.org/10.1080/01621459.1967.10500912
Higgins M. J., Sävje F., & Sekhon J. S. (2016). Improving massive experiments with threshold blocking. *Proceedings of the National Academy of Sciences*, *113*(27), 7369–7376. https://doi.org/10.1073/pnas.1510504113
Hirano K., & Imbens G. (2004). The propensity score with continuous treatments. Applied Bayesian modeling and causal inference from incomplete-data perspectives.
Hsu Y.-C., Huber M., Lee Y.-Y., & Liu C.-A. (2024). Testing monotonicity of mean potential outcomes in a continuous treatment with high-dimensional data. *The Review of Economics and Statistics*, 1–44. https://doi.org/10.1162/rest_a_01416
Huang S., Li X., & Toulis P. (2025). 'Randomization tests for monotone spillover effects', arXiv, arXiv:2501.02454, preprint: not peer reviewed.
Huber M., & Maric N. (2017). 'Bernoulli correlations and cut polytopes', arXiv, arXiv:1706.06182v2, preprint: not peer reviewed.
Imai K., & Van Dyk D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, *99*(467), 854–866. https://doi.org/10.1198/016214504000001187
Imbens G. W., & Rubin D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
Isaki C. T., & Fuller W. A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, *77*(377), 89–96. https://doi.org/10.1080/01621459.1982.10477770
Kallus N. (2018). Optimal a priori balance in the design of controlled experiments. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *80*(1), 85–112. https://doi.org/10.1111/rssb.12240
Kennedy E. H., Ma Z., McHugh M. D., & Small D. S. (2017). Non-parametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *79*(4), 1229–1245. https://doi.org/10.1111/rssb.12212
Li X., & Ding P. (2020). Rerandomization and regression adjustment. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *82*(1), 241–268. https://doi.org/10.1111/rssb.12353
Li X., Ding P., & Rubin D. B. (2018). Asymptotic theory of rerandomization in treatment–control experiments. *Proceedings of the National Academy of Sciences*, *115*(37), 9157–9162. https://doi.org/10.1073/pnas.1808191115
Li X., Ding P., & Rubin D. B. (2020). Rerandomization in 2 k factorial experiments. *Annals of Statistics*, *48*(1), 43–63. https://doi.org/10.1214/18-AOS1790
Liang T. (2024). Blessings and curses of covariate shifts: Adversarial learning dynamics, directional convergence, and equilibria. *Journal of Machine Learning Research: JMLR*, *25*(140), 1–27. https://jmlr.org/papers/v25/23-0651.html
Liang T., & Recht B. (2025). Randomization inference when n equals one. *Biometrika*, *112*(2), Article asaf013. https://doi.org/10.1093/biomet/asaf013
Liang T., & Tran-Bach H. (2022). Mehler's formula, branching process, and compositional kernels of deep neural networks. *Journal of the American Statistical Association*, *117*(539), 1324–1337. https://doi.org/10.1080/01621459.2020.1853547
List J. A., Muir I., & Sun G. (2024). Using machine learning for efficient flexible regression adjustment in economic experiments. *Econometric Reviews*, *44*(1), 2–40. https://doi.org/10.1080/07474938.2024.2373446

Ma W., Li P., Zhang L.-X., & Hu F. (2024). A new and unified family of covariate adaptive randomization procedures and their properties. *Journal of the American Statistical Association*, *119*(545), 151–162. https://doi.org/10.1080/01621459.2022.2102986

Ma W., Qin Y., Li Y., & Hu F. (2020). Statistical inference for covariate-adaptive randomization procedures. *Journal of the American Statistical Association*, *115*(531), 1488–1497. https://doi.org/10.1080/01621459.2019.1635483

Marčenko V. A., & Pastur L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, *1*(4), 457. https://doi.org/10.1070/SM1967v001n04ABEH001994

Mas-Colell A., Whinston M. D., & Green J. R. (1995). *Microeconomic theory* (*Vol. 1*). Oxford University Press.

Meddings J., Scott R., & Fick G. (1989). Analysis and comparison of sigmoidal curves: Application to dose-response data. *American Journal of Physiology: Gastrointestinal and Liver Physiology*, *257*(6), G982–G989. https://doi.org/10.1152/ajpgi.1989.257.6.G982

Mehler F. (1866). Ueber die entwicklung einer function von beliebig vielen variablen nach laplaceschen functionen höherer ordnung. *Journal für die Reine und Angewandte Mathematik*, *66*, 161–176. https://doi.org/10.1515/crll.1866.66.161. In German.

Moore R. T. (2012). Multivariate continuous blocking to improve political science experiments. *Political Analysis: An Annual Publication of the Methodology Section of the American Political Science Association*, *20*(4), 460–479. https://doi.org/10.1093/pan/mps025

Morgan K. L., & Rubin D. B. (2012). Rerandomization to improve covariate balance in experiments. *Annals of Statistics*, *40*(2), 1263–1282. https://doi.org/10.1214/12-AOS1008

Neyman J. (1923). On the application of probability theory to agricultural experiments: Essay on principles (with discussion); section 9 (in polish) [Engl. transl. by D. M. Dabrowska and T. P. Speed (1990)]. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, *5*, 465–472. https://doi.org/10.1214/ss/1177012031

Nordin M., & Schultzberg M. (2022). Properties of restricted randomization with implications for experimental design. *Journal of Causal Inference*, *10*(1), 227–245. https://doi.org/10.1515/jci-2021-0057

Rosenberger W. F., & Lachin J. M. (2015). *Randomization in clinical trials: Theory and practice*. John Wiley & Sons.

Rosenberger W. F., & Sverdlov O. (2008). Handling covariates in the design of clinical trials. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, *23*(3), 404–419. https://doi.org/10.1214/08-STS269

Schindl K., Shen S., & Kennedy E. H. (2024). 'Incremental effects for continuous exposures', arXiv, arXiv:2409.11967v3, preprint: not peer reviewed.

Shephard R. W., & Färe R. (1974). *The law of diminishing returns*. In *Production theory: Proceedings of an international seminar held at the university at Karlsruhe May–July 1973* (pp. 287–318). Springer.

Sibson R. (1974). Da-optimality and duality. *Progress in Statistics*, *2*, 677–692.

Vershynin R. (2018). *High-dimensional probability: An introduction with applications in data science* (*Vol. 47*). Cambridge University Press.

Wang Y., & Li X. (2025). Asymptotic theory of the best-choice rerandomization using the mahalanobis distance. *Journal of Econometrics*, *251*, Article 106049. https://doi.org/10.1016/j.jeconom.2025.106049

Wu C.-F. (1981). On the robustness and efficiency of some randomized designs. *Annals of Statistics*, *9*(6), 1168–1177. https://doi.org/10.1214/aos/1176345634

Ye T., Yi Y., & Shao J. (2022). Inference on the average treatment effect under minimization and other covariate-adaptive randomization methods. *Biometrika*, *109*(1), 33–47. https://doi.org/10.1093/biomet/asab015

Zhang H., Hu F., & Yin J. (2022). Covariate-adaptive randomization with variable selection in clinical trials. *Stat*, *11*(1), Article e461. https://doi.org/10.1002/sta4.461

Zhao J. (2024). *Experimental design for causal inference through an optimization lens*. In *Tutorials in operations research: Smarter decisions for a better world* (pp. 146–188). INFORMS.