Near Optimal A-B Testing*

Nikhil Bhat
Graduate School of Business
Columbia University
email: nbhat15@gsb.columbia.edu

Vivek F. Farias
Sloan School of Management
Massachusetts Institute of Technology
email: vivekf@mit.edu

Ciamac C. Moallemi
Graduate School of Business
Columbia University
email: ciamac@gsb.columbia.edu

Deeksha Sinha
Operations Research Center
Massachusetts Institute of Technology
email: deeksha@mit.edu

Initial Version: October 2015
Current Revision: September 2018

Abstract

We consider the problem of A-B testing when the impact of the treatment is marred by a large number of covariates. Randomization can be highly inefficient in such settings, and thus we consider the problem of optimally allocating test subjects to either treatment with a view to maximizing the precision of our estimate of the treatment effect. Our main contribution is a tractable algorithm for this problem in the online setting, where subjects arrive, and must be assigned, sequentially, with covariates drawn from an elliptical distribution with finite second moment. We further characterize the gain in precision afforded by optimized allocations relative to randomized allocations, and show that this gain grows large as the number of covariates grows. Our dynamic optimization framework admits several generalizations that incorporate important operational constraints such as the consideration of selection bias, budgets on allocations, and endogenous stopping times. In a set of numerical experiments, we demonstrate that our method simultaneously offers better statistical efficiency and less selection bias than state-of-the-art competing biased coin designs.

1. Introduction

The prototypical example of an 'A-B test' is the design of a clinical trial where one must judge the efficacy of a treatment or drug relative to some control. In a different realm, A-B testing today plays an increasingly pivotal role in e-commerce, ranging from the optimization of content and graphics for online advertising, to the design of optimal layouts and product assortments for webpages. E-commerce properties will even use A-B testing as a means of finding the best third party vendor for a specific service on their website (such as, say, recommendations or enterprise search).

*The authors wish to thank Anthony Atkinson, Steve Chick, Shane Hendersen, Nathan Kallus, Costis Maglaras, Assaf Zeevi, and Jose Zubizarreta for helpful discussions. The second author's work was supported in part by NSF CAREER Grant CMMI-1054034. The third author was supported by NSF Grant CMMI-1235023.

A natural approach to A-B testing is to independently, and with equal probability, assign each subject to either the treatment or control groups. Following such a randomized allocation, the benefit of the treatment relative to the control can be estimated from the outcomes of subjects in the two groups. The notion of a subject here can range from a patient in the clinical trial setting to a web-surfer or impression in the e-commerce setting. Similarly, the notion of a treatment can vary from an actual medical treatment in the clinical trial setting to the decision to show a specific ad in the e-commerce setting. While randomized allocation is simple and can easily be shown to yield unbiased estimates of the treatment effect under a minimal set of assumptions, the *efficiency* of this procedure (or, the sample size needed to get a statistically significant estimate of the treatment effect) can prove onerous in practice. To see, why consider the following challenges:

1. Limited Sample Size: In the clinical trial setting, the number of subjects is limited for several reasons. As an example, the cost of managing a single subject through a clinical trial is tens of thousands of dollars (see, e.g., Steensma and Kantarjian, 2014). In the e-commerce setting, one may need to conduct many thousands of A-B tests in an ongoing fashion. As an example, consider an advertising firm that uses A-B testing on live impressions (i.e., web-surfers) to mechanically decide the appropriate messaging, text size, font, color etc. for the creative content it generates for an online advertising campaign. In this domain, a reduction in the sample size needed to learn can, due to scale, result in dramatic, continual cost savings.

2. Confounding Effects: Running counter to the need for quick inference, the impact of a particular treatment (or design decision) may be marred by a potentially large number of covariates. The presence of these covariates makes the inference of the treatment effect more challenging, since the difference in outcome of the treatment and control groups might be due to a lack of 'balance' in the covariates in the two groups. While the law of large numbers assures us that a large enough sample size will 'wash out' the impact of this imbalance of covariates, the requisite sample size may grow exceedingly large when the number of covariates is large and/or the treatment effect is small.

3. 'Small' Treatment Effects: Similar to the covariate imbalance issue above, the incremental impact of the treatment under study may be relatively 'small'. This creates a challenge in the measurement of small treatment effects, which, despite their magnitude, many nevertheless be important in settings where the selected treatments will be applied on a sufficiently large scale. More precisely, if one imagined a model where the outcome is additively impacted by the treatment and exogenous noise, we expect the sample size required to discern the treatment from noise to grow quadratically with the ratio of the standard deviation of the exogenous noise to the treatment effect. To (heuristically) see why, observe that if S_n is the sum of n independent, zero mean random variables, each with standard deviation σ, $\theta > 0$ is some constant, and $\Phi(\cdot)$ is the cumulative distribution of the standard normal, then by the

central limit theorem, we expect

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq \theta\right) \sim 2\Phi\left(\frac{\theta\sqrt{n}}{\sigma}\right).$$

This suggests that, in order to differentiate a treatment effect with magnitude θ from exogenous noise with standard deviation σ, we need on the order of σ^2/θ^2 samples.

4. Operational Constraints: As already alluded to, A-B tests can be expensive, either because of an explicit cost related to managing test subjects or the implicit risk of testing a sub-optimal treatment. These issues clearly impact the choice of sample size and frequently imply a budget on the number of subjects allocated to the alternative treatment whose efficacy we seek to measure. It is also not unusual to dynamically 'stop' a trial based on ones confidence in the outcome. In clinical trials, one cares about 'selection bias' in addition to efficiency; measures such as selection bias speak to concerns of robustness (to modeling errors or manipulation), or even fairness. Taken together, these operational constraints further complicate an already challenging problem.

Addressing theses challenges motivates considering the careful design of such A-B tests. In particular, given a collection of subjects, some of whom must be chosen for treatment, and others assigned to a control, we would like an assignment that 'balances' the distribution of covariates across the two groups. This in turn could conceptually yield an efficient estimate of the treatment effect, the primary concern alluded to above.

Given the broad applicability of an efficient A-B test, it is perhaps not surprising that a large body of literature within the statistical theory of the design of experiments has considered this very problem, starting with the nearly century old work of Fisher (1935). While we defer a review of this substantial literature to Section 1.2, a very popular approach to dealing with the problem of achieving covariate balance is the use of 'stratification'. In this approach, the subjects are divided into a number of groups based on the covariates. In other words, the covariate space is divided into a number of regions and subjects whose covariates lie in a certain region are grouped together. Further, each of the groups is randomly split to be allocated to the treatment or the control. Unfortunately, stratification does not scale gracefully with the number of covariates since the number of groups required in stratification will grow exponentially with the dimension.[1] Another natural idea would be to 'match' subjects with similar covariates, followed by assigning one member of a match to the treatment and the other to the control. Such a design would try to mimic an idealistic scenario in which, for n subjects under the experiment, we have $n/2$ pairs of 'twins'. If the matched subjects are indeed close to each other in the space of covariates, we would have that the distribution of covariates in the treatment and control is close to each other, which would cancel out the effect of these covariates. While this latter approach does allow us to consider a large number of covariates, the literature only appears to present heuristics motivated by these ideas.

[1] Rule C in Table 1 of Atkinson (2002) illustrates that with as few as 10 covariates, methods based on stratification are hardly better than randomization.

To add a further challenge beyond those already discussed, an additional (and very important) requirement apparent from the applications above is that the process of allocating subjects (or impressions) to a particular treatment (or creative) must be made *sequentially, in an online or dynamic fashion.* Again, there is a literature on dynamic allocation starting with seminal work by Efron (1971) on 'biased coin designs' (BCDs). While a BCD seeks to balance the number of subjects in the treatment and control groups, there is by now a robust literature on so-called covariate adaptive BCDs. These schemes extend Efron's original proposal so that one cares about balance in not just the number of subjects across the two groups but also seeks balance in the covariate distribution. Viewed from the perspective of dynamic optimization, all of these heuristics can be seen as myopic schemes that in making an allocation at a given point in time fail to hedge against the future stream of arriving subjects. In fact, the literature surprisingly does not consider the design of an 'optimal' online allocation of subjects to treatments — or online A-B testing in our parlance — as a principled dynamic optimization problem where dynamic programming techniques for optimal sequential decision-making can be applied.

The present paper casts the problem of computing an efficient estimate of the treatment effect in an A-B test as a dynamic optimization problem. Despite this being a high-dimensional control problem, we show that one can efficiently compute near-optimal solutions to this problem when covariates are elliptically distributed. We show that our approach yields Pareto improvements over state of the art alternatives covariate adaptive BCD approaches. As a secondary contribution, we also show that that the important 'offline' variant of the problem also admits an efficient optimal algorithm and tightly characterize the value of optimization in that setting.

1.1. This Paper

Our approach, in a nutshell, is to formulate online A-B testing as a (computationally challenging) dynamic optimization problem and develop approximation and exact algorithms for the same. In particular, the present paper considers the setting where a subject's response is linear in the treatment and covariates; as we discuss later, this is a canonical model and is widely encountered in the literature on experiment design. We consider the problem of maximizing the precision of our estimate of the treatment effect by optimally allocating subjects to either the treatment or control group. We formulate this problem as a dynamic optimization problem and make the following contributions:

1. Offline Allocation: In the offline setting, i.e., where the allocation can be made after observing all subjects, we show that the problem can be solved efficiently by using as a subroutine a generalization of the MAX-CUT SDP relaxation of Goemans and Williamson (1995). While not our main result, this result shows that the problem of *offline* A-B testing (which is still valuable in some traditional applications) can surprisingly be solved efficiently. We also characterize the value of optimized allocations relative to randomization in this setting and show that this value grows large as the number of covariates grows.

2. Sequential Allocation: In the online setting — which is the algorithmic focal point of our work — our optimization problem is, not surprisingly, a high dimensional dynamic optimization problem with dimension that grows like the number of covariates. *We show how to break the curse of dimensionality here.* In particular, we show that the state space of this dynamic optimization problem collapses if covariates come from an elliptical family of distributions (a family that includes, for example, the multivariate Gaussian). This yields an *efficient* algorithm that is provably optimal in the elliptical distribution setting and that can nonetheless be employed when covariates are not from an elliptical family.

3. A General Framework: We show that our dynamic optimization formulation permits the consideration of criteria beyond just the variance of the treatment effect. Specifically, we extend our formulation to a framework that can accommodate the simultaneous minimization of selection bias; the minimization of general separable cost functions of the allocation; endogenous (optimal) stopping criteria (as opposed to a-priori fixed sample sizes); and budgets on the sample size for a given treatment, to name just a few applications of the framework.

4. Experimental Comparisons: We compare our approach to sequential allocation with a host of so-called covariate adaptive BCD approaches, several of which are considered state-of-the-art. It is typical to measure the performance of such approaches not just in terms of efficiency, but also with respect to the so-called selection bias they induce. Here we show that our approach yields a *Pareto improvement* over these alternatives. In addition to synthetic data, we run our experiment on real user impression data from Yahoo.com. We show similar Pareto gains despite the fact that the covariates in the real data are categorical.

Thus, our main contribution is providing an algorithm for the challenging problem of sequential A-B testing that can be shown to be near-optimal when covariates are drawn from an elliptical family. The algorithm is applicable to a canonical family of treatment models and also applies to the simultaneous optimization of several criteria. Given the vast extant literature on this problem, and the fact that it is nominally high-dimensional, it is a pleasant surprise that such an algorithm exists.

1.2. Related Literature

The theory of optimal experiment design (which, in a sense, subsumes the problems we consider here) starts with the seminal work of Fisher (1935). Important textbook expositions of this mature topic include that of Pukelsheim (2006) and Cook et al. (1979), the latter of which discusses the notion of covariate matching as it applies to practice. While not our primary focus, the 'offline' problem we discuss in this paper is of practical relevance in the social sciences; see Raudenbush et al. (2007), for an application and heuristics. Kallus (2013) studies an approach to this problem based on linear mixed integer optimization with an application to clinical trials. In a follow-up paper, Bertsimas et al. (2015) presents a robust optimization framework for the offline problem with an

emphasis on allocations of treatments that are robust to the specific form of the model of each subject's response as a function of the treatments and subject covariates (we merely consider linear functions here). The value of optimization has also recently received attention from the economics community; Kasy (2013) discusses several optimization formations that complement those proposed by Bertsimas et al. (2015). Unlike Kallus (2013); Bertsimas et al. (2015) however, Kasy (2013) offers no algorithmic approach to solve the problems he proposes (and unfortunately, his problem formulations appear largely intractable). In contrast, we focus on a class of models where the treatment effect is linear in the observed covariates and offer efficient approximation algorithms for the same. Our formulation is closely related to the case of squared loss with a non-informative prior in the verbiage of Kasy (2013). Our offline problem may be viewed as a special case of the problem of D_a-optimal experiment design and fortuitously coincides with an optimality criterion that already enjoys wide acceptance. By virtue of their computational efficiency, our techniques can be brought to bear in settings where the size of the problem can be very large rendering brute-force techniques for optimization (such as those suggested by Kasy (2013)) infeasible.

The problem that is of greatest algorithmic interest to us is the 'online' allocation problem, where treatments must be assigned to subjects as they arrive. With regard to this sequential problem, Efron (1971) proposed an allocation strategy, referred to as a 'biased coin design' (BCD), that sought to 'balance' the number of subjects in each trial while minimizing certain types of selection bias. Now whereas Efron's BCD seeks only to balance the number of subjects between test and control groups, there is by now a robust literature on so-called covariate adaptive BCDs (CA-BCDs). Such schemes seek balance not just in the number of subjects but also in the covariate distribution between groups. Perhaps the most widely used CA-BCD is the procedure proposed by Pocock and Simon (1975) wherein the authors recommend a 'bias' that depends on a generic cost function of the covariate imbalance between the two groups. Atkinson (1982, 1999) proposed the first CA-BCD whose design is rooted in theory, specifically to the notion of D_a optimality in experiment design; of course this approach comes at the cost of assuming a treatment effect model. A number of model-based CA-BCD proposals have followed, including Smith's rule (Smith, 1984b,a); the Bayesian procedure of Ball et al. (1993); and rule ABCD, proposed by Baldi Antognini and Zagoraiou (2011), to name a few. The so-called minimization approach of Pocock and Simon (1975) (which applies to generic cost functions of covariate imbalance) has also been recently analyzed by Hu and Hu (2012), who prescribe a more refined class of cost functions that lead to asymptotic balance. Alternatives to the CA-BCD procedure have also been proposed recently: Kapelner and Krieger (2014) presents an approach to achieving covariate balance based on ideas from the theory of online matching.

Viewed from the perspective of dynamic optimization, except for the heuristic proposed by Kapelner and Krieger (2014), all of the above approaches can be regarded as *myopic* policies. Such policies only consider the immediate impact of an allocation decision, and do not consider the impact on future decisions. In general, myopic policies will not be optimal. It is worth noting that for all of the aforementioned procedures, the theoretical analysis available, if any, is always

in a limiting regime where sample size grows large keeping the number of covariates fixed. Little is understood in finite samples. More generally, Rosenberger and Sverdlov (2008) note that "very little is known about the theoretical properties of covariate-adaptive designs". In contrast, we see that our approach yields provably optimal allocations in finite samples for a host of optimality criteria. As we see in our experimental work, this also translates to Pareto improvements over several of the schemes described above, even on real data. It is worth noting however, that such statements of optimality require restrictions on the types of treatment models one can consider, as well as distributional assumptions on the covariates.

Related but Distinct Problems.

It is important to distinguish the experiment design problems considered here from 'bandit' problems, particularly those with side information (e.g., Woodroofe, 1979; Langford and Zhang, 2007) as both classes of problems frequently find application in very related applications. In theory, the experimental design setting is appropriate when an irrevocable decision of what treatment is appropriate must be made (e.g., the number of ads to display with search results), whereas the bandit setting is appropriate in a setting where the decision can be changed over time to optimize the (say) long-run average value of some objective (e.g., maximizing revenues by finding the best audience for a specific campaign). In practice, the choice of which framework to use is frequently complicated by operational considerations. For instance consider the problem of deciding between two distinct creatives in an advertising campaign. The bandit formulation is elegant and quite natural for this setting (Hauser et al., 2009; Schwartz et al., 2017). Despite this, it is common industry practice to make such decisions using frequent A-B tests[2]. From a methodological perspective, an important difference is that solution methods for bandit problems need to address an 'exploitation-exploration' trade-off between learning the best alternative and collecting rewards to optimize the objective, while there is no such trade-off in our experimental design setting.

Other problems in marketing science are also close in spirit to the A-B testing problem we study. Adaptive conjoint analysis seeks to learn the tastes of an individual (or a group of individuals) by asking a sequence of questions (or presenting a sequence of choices). In an effort to learn accurately with as small a number of questions, Toubia et al. (2003, 2004) propose a dynamic optimization procedure that is in the spirit of the ellipsoid method in convex optimization.

Another closely related class of problems are ranking and selection problems where the task is to pick the best of a set of alternatives with a budget on samples (for an overview, see Kim and Nelson, 2006). In our lexicon, the emphasis in such problems is choosing from multiple (typically, greater than two) treatments in the *absence* of observable covariates on a sample. Interestingly, recent progress on this class of problems has also heavily employed dynamic optimization techniques (see, e.g., Chick and Gans, 2009; Chick and Frazier, 2012; Chick et al., 2017).

As a final note, the major emphasis in our work is on A-B testing with a fixed budget on

[2]For example, consider the following case study by one of the largest providers of commercial A-B testing infrastructure: `https://blog.optimizely.com/2014/02/03/case-study-sony-ab-tests-banner-ads/`.

samples. It is interesting to consider A-B tests that can be 'stopped' with continuous monitoring. Doing so can introduce a significant bias towards false discovery; Johari et al. (2017) have recently made exciting progress on this problem.

2. Model

In this section we describe the model. Given the model assumptions in Section 2.1, our problem is to maximize the precision of our estimate of the treatment effect. In Section 2.2 we pose the two optimization problems that are of interest. One of them is the offline problem where all subjects can be observed before making allocation decisions and the other is the sequential problem where subjects must be allocated without knowing the future arrivals. In Section 2.3 we present a simple upper bound on the precision of any estimate of the treatment effect given an allocation; this allows us to define the notions of efficiency and loss. Section 2.4 concludes with an intuitive interpretation of our optimization problems.

2.1. Setup

We must learn the efficacy of a treatment by observing its effect on n subjects. The kth subject is assigned a treatment $x_k \in \{\pm 1\}$. The kth subject is associated with a covariate vector (i.e., side information or context) $Z_k \in \mathbb{R}^p$. We assume that impact of the treatment on the kth subject is given by:

$$y_k = x_k \theta + Z_k^\top \kappa + \epsilon_k.$$

This assumes a linear dependence of the covariates and treatment decision on the outcome. The treatment effect $\theta \in \mathbb{R}$ and the weights on the covariates $\kappa \in \mathbb{R}^p$ are unknown. Our aim is to estimate θ. The $\{\epsilon_k\}$ are i.i.d. zero mean random variables with variance σ^2. The key restriction imposed by this model is that the impact of treatment is additive, an assumption that is ubiquitous in all of the related literature on the topic. Further, we assume that there is no endogeneity, i.e. the idiosyncratic noise in the model, ϵ_k, is uncorrelated with any of the covariates in Z_k.[3]

Letting $Z \in \mathbb{R}^{n \times p}$ be the matrix whose kth row is $Z_k^\top$, throughout this paper, we will assume that:

Assumption 1. *The first column of Z is a vector of all ones. Further, Z is full rank and $p \leq n-1$.*

The requirement that one of the covariates be a constant ensures that θ is interpreted as a treatment effect, otherwise it could be learned from the assignment of a single treatment. The crucial assumption is that $p \leq n-1$, which nonetheless allows for a large number of covariates.[4] In fact the scenario where $p \sim n$ is particularly relevant. Our problem formulation does not apply

[3]The assumption of no endogeneity is required for the least square estimate of θ under a given allocation to be unbiased. It is also required for our performance analysis. In general, it appears difficult to overcome bias in the face of the risk of model mis-specification while using a covariate dependent treatment assignment scheme.

[4]We will informally refer to p as the number of covariates even though, strictly speaking, it is the dimension of the linear model and could include second order terms, interaction terms between covariates, etc.

to the regime where $p > n$; indeed a formulation that is relevant to that regime is unclear to us since treatments must be assigned prior to having observed outcomes. For a particular allocation of treatments, x, let us denote by $\hat{\theta}_x$ the least squares estimator for θ.

2.2. Optimization Problem

We are interested in finding an experiment design with minimal variance or, equivalently, maximal precision. A standard calculation yields that the estimator $\hat{\theta}_x$ has precision

$$\mathrm{Prec}(\hat{\theta}_x) \triangleq \frac{1}{\mathrm{Var}(\hat{\theta}_x)} = \frac{x^\top P_{Z^\perp} x}{\sigma^2}, \tag{1}$$

where $P_{Z^\perp} \triangleq I - Z(Z^\top Z)^{-1} Z^\top$. Details are presented in the Electronic Companion to this paper.

We can now immediately state the *offline experiment design problem*:

$$\begin{aligned} \text{(P1)} \triangleq \text{maximize} \quad & x^\top P_{Z^\perp} x \\ \text{subject to} \quad & x \in \{\pm 1\}^n. \end{aligned}$$

Here, given the collection of covariates Z, we seek to find the allocation x which yields the least squares estimate with maximal precision.

In many real world applications the assignments need to be made in a sequential fashion. Subjects arrive one at a time and the assignment must be made without the knowledge of subjects in the future. We formulate this as a dynamic optimization problem. To this end we must now assume the existence of a measure on the covariate process $\{Z_k\}$. We define a filtration $\{\mathcal{F}_k\}$ by setting, for each time k, $\mathcal{F}_k$ to be the sigma algebra generated by the first k covariates $(Z_1, \ldots, Z_k)$ and the first $k-1$ allocations $(x_1, \ldots, x_{k-1})$. The *online experiment design problem* is then given by:

$$\begin{aligned} \text{(P2)} \triangleq \text{maximize} \quad & \mathsf{E}\left[x^\top P_{Z^\perp} x\right] \\ \text{subject to} \quad & x \in \{\pm 1\}^n, \\ & x_k \text{ is } \mathcal{F}_k\text{-measurable}, \quad \forall\, 1 \le k \le n, \end{aligned}$$

where the expectation is over the distribution of the covariate process. Here, the objective is to maximized the expected *ex post* precision.[5]

2.3. Upper Bound, Efficiency, and Loss

The following upper bound on the precision of any unbiased estimator that is a straightforward consequence of the Cramér-Rao bound:

Proposition 1. *If $\epsilon \sim N(0, \sigma^2 I)$, then for any covariate matrix Z and any unbiased estimator $(\hat{\theta}, \hat{\kappa})$,*

[5]Note that, in the online case, because of Jensen's inequality, maximizing precision and minimizing variance are no longer equivalent objectives.

including non-least squares estimators, we have:

$$\mathrm{Prec}(\hat{\theta}_x) \leq \frac{n}{\sigma^2},$$

an upper bound on the optimal value of both problems (P1) and (P2). For non-Gaussian noise ϵ, this upper bound still holds for all least squares estimators.

This proposition, whose proof is provided for completeness in the Electronic Companion to this paper, shows that the precision of the optimal estimator[6] is $O(n)$. Consider the case when subjects are identical, i.e., $p = 1$ and $Z_k = 1$ for all k. It is easy to note that, in this case assuming n is even, the optimal design allocates half of the subjects to either treatment. Further, the precision of such a design is n/σ^2, the optimal achievable precision. For $p > 1$ this precision is less than this value. Thus the presence of covariates only makes the inference challenging.

Motivated by Proposition 1, we define *efficiency* as the the precision of an estimator normalized by the Cramér-Rao upper bound, i.e.,

$$\mathrm{Eff}(\hat{\theta}_x) \triangleq \frac{\mathrm{Prec}(\hat{\theta}_x)}{n/\sigma^2} \leq 1,$$

Loss is defined as the sub-optimality of an estimator relative to the upper bound measured additively in sample units:

$$\mathrm{Loss}(\hat{\theta}_x) \triangleq n - \sigma^2 \mathrm{Prec}(\hat{\theta}_x) \geq 0,$$

so that

$$\mathrm{Prec}(\hat{\theta}_x) = \frac{n - \mathrm{Loss}(\hat{\theta}_x)}{\sigma^2}.$$

We consequently see that loss can intuitively be thought of as "the effective number of subjects on whom information is lost due to the imbalance of the design" (Atkinson, 2014).

2.4. Problem Interpretation

Before moving on to algorithm design, we pause to interpret the offline and online problems presented above. First we begin with an intuitive interpretation of the objective. Define the imbalance vector in covariate values between the test and control groups, $\overline{\Delta}_n \in \mathbb{R}^p$, according to $\overline{\Delta}_n \triangleq \sum_{k=1}^n x_k Z_k = Z^\top x$. Notice that the empirical second moment matrix for the covariates is given by $\Gamma_n \triangleq Z^\top Z/n$. Then, it is easy to see that the objective of the offline problem (P1) reduces to

$$x^\top P_{Z^\perp} x = x^\top \left(I - Z(Z^\top Z)^{-1} Z^\top\right) x = n\left(1 - \overline{\Delta}_n^\top \Gamma_n^{-1} \overline{\Delta}_n\right).$$

[6] In what follows, given a function $f(\cdot)$ and a positive function $g(\cdot)$, as $n \to \infty$ we say $f(n) = O(g(n))$ if $\limsup_{n\to\infty} |f(n)|/g(n) < \infty$, we say $f(n) = o(g(n))$ if $\lim_{n\to\infty} |f(n)|/g(n) = 0$, we say $f(n) = \Omega(g(n))$ if $\limsup_{n\to\infty} |f(n)/g(n)| > 0$, and finally we say $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$.

Therefore, the offline problem (P1) is equivalent to minimizing the square of the weighted euclidean norm of $\overline{\Delta}_n$,

$$\left\|\overline{\Delta}_n\right\|^2_{\Gamma_n^{-1}} \triangleq \overline{\Delta}_n^\top \Gamma_n^{-1} \overline{\Delta}_n,$$

while (P2) seeks to minimize the expected value of this quantity where the expectation is over the covariate process and our allocations. Put simply, both problems seek to minimize the aggregate imbalance of covariates between the treatment and control groups, measured according to this norm.

As a final point, we note that the measure of 'imbalance' minimized in problems (P1) and (P2) was derived assuming a least squares estimator, and it is worth noting that this choice is not arbitrary. Specifically, note that the Cramér-Rao bound dictates that, provided x and Z are independent of ϵ, and further if ϵ is normally distributed, then for *any* unbiased estimator of the treatment effect $\tilde{\theta}_x$, we have that

$$\text{Eff}(\tilde{\theta}_x) \leq \text{Eff}(\hat{\theta}_x)$$

where the right hand side quantity is the efficiency of the least square estimator. Now both problems (P1) and (P2) seek to find an allocation x to maximize the latter quantity, or its expected value, respectively. Consequently, both problems may be interpreted as seeking an allocation of samples to the test and control group with a view to maximizing the efficiency of our estimate of the treatment effect among *all* unbiased estimators of the treatment effect.

3. The Offine Optimization Problem

In this section, we consider the offline optimization problem (P1). We show that this combinatorial problem permits a tractable, constant factor approximation using an SDP-based randomized rounding algorithm. Moreover, in this setting, we can analyze the effect optimization has on the precision of the estimator of the treatment effect, as compared to randomization. To this end, we first obtain the mean precision of the randomized design. Surprisingly, precision is a simple function of n and p and does not depend on the data matrix Z. We show that when $p \sim n$, the randomization is rather inefficient and the precision is $O(1)$. This can be contrasted with the upper bound on precision given by Proposition 1 which is $\Omega(n)$. To conclude the section, we analyze the performance of the optimal allocation assuming a distribution on Z. We show that for any p, the precision of optimal allocation is $\Omega(n)$. Thus concluding that when $p \sim n$, randomization can be arbitrarily bad as compared to the optimal design.

3.1. Approximation Algorithm for (P1)

First, we observe that there is a tractable approximation algorithm to solve the combinatorial optimization problem (P1). In particular, consider the semidefinite program (SDP) over symmetric

positive semidefinite matrices $Y \in \mathbb{R}^{n\times n}$ given by[7]

$$
\begin{aligned}
\text{(P1-SDP)} \triangleq \quad & \text{maximize} && \operatorname{tr}\left(P_{Z^\perp} Y\right) \\
& \text{subject to} && Y_{kk} = 1, \qquad \forall\, 1 \le k \le n, \\
& && Y \succeq 0, \\
& && Y \in \mathbb{R}^{n\times n}.
\end{aligned}
$$

It is straight forward to see that (P1-SDP) is a relaxation of (P1) in the sense that it achieves higher objective value: given an optimal solution $\hat{x} \in \{\pm 1\}^n$ for (P1), define the symmetric positive definite matrix $\hat{Y} \triangleq \hat{x}\hat{x}^\top \in \mathbb{R}^{n\times n}$. Then, clearly $\hat{Y}$ satisfies the constraints of (P1-SDP). Also, $\operatorname{tr}\left(P_{Z^\perp}\hat{Y}\right) = \hat{x}^\top P_{Z^\perp}\hat{x}$, so the objective values for (P1) and (P1-SDP) coincide. Therefore, the optimal objective value of (P1-SDP) must be larger than that of (P1). Moreover, because it is an SDP, (P1-SDP) can be efficiently solved in polynomial time.

Based upon prior work on the MAX-CUT problem (Goemans and Williamson, 1995), the following result, due to Nesterov (1997), establishes that (P1-SDP) can be used as the basis of a randomized algorithm to solve (P1) with a constant factor guarantee with respect to the optimal design. The corresponding (randomized) allocation procedure is described in Algorithm 1.

procedure SDPAllocation(Z) ▷ Compute an allocation $\tilde{x}$
  Set $Y^* \succeq 0$ to be an optimal solution of the program (P1-SDP) given the data matrix Z.
  Set the matrix $V \in \mathbb{R}^{n\times n}$ with columns $v_1, \ldots, v_n \in \mathbb{R}^n$ so that the matrix decomposition $Y^* = V^\top V$ holds.
  Let $u \in \mathbb{R}^n$ be a vector chosen at random uniformly over the unit sphere.
  for $k \leftarrow 1, n$ **do**

$$\tilde{x}_k \leftarrow \begin{cases} +1 & \text{if } u^\top v_k \ge 0, \\ -1 & \text{if } u^\top v_k < 0. \end{cases}$$

  end for
  return $\tilde{x}$
end procedure

Algorithm 1: Randomized allocation algorithm based on (P1-SDP).

Theorem 1. *Given a data matrix $Z \in \mathbb{R}^{n\times p}$, set the allocation $\tilde{x} \in \mathbb{R}^n$ according to Algorithm 1. Then,*

$$\mathsf{E}_u\left[\tilde{x}^\top P_{Z^\perp}\tilde{x}\right] \ge \frac{2}{\pi} \max_{x\in\{\pm 1\}^n} x^\top P_{Z^\perp} x,$$

where the expectation is taken over the choice of random vector u in Algorithm 1. In order words, the expected value achieved by the vector $\tilde{x}$ in the offline experiment design problem (P1) is within a constant factor $2/\pi$ of the best possible.

Proof. This theorem is a direct consequence of Theorem 3.4.2 of Ben-Tal and Nemirovski (2001). That result states that any quadratic integer optimization problem with objective $x^\top Q x$, such that

[7] Here, $Y \succeq 0$ denotes that Y is a symmetric and positive semidefinite matrix.

$x \in \{\pm 1\}^n$, can be approximated within a relative error of $\pi/2$ using the prescribed algorithm, provided Q is positive semidefinite. Since $P_{Z^\perp}$ is positive semidefinite (indeed, it is a projection matrix), the result follows. ■

3.2. Optimal Allocations vs. Randomized Allocations

Randomization is the most popular technique used for A-B testing. In what follows, we will compare the performance of randomization to what can be achieved by the optimal offline allocation of (P1).

In its most basic variation, simple randomization partitions the population into two equally sized groups, each assigned a different treatment, where the partition is chosen uniformly at random over all such partitions (for simplicity, we will assume that the population is of even size). Denote by $X_{\text{rand}} \in \{\pm 1\}^n$ the random allocation generated by simple randomization, and denote by $\hat{\theta}_{X_{\text{rand}}}$ the resulting unbiased least squares estimator for θ.

Theorem 2. *If n is even, given a covariate matrix Z, define the expected precision and loss of simple randomization*

$$\text{Prec}_{\text{rand}} \triangleq \mathsf{E}_{X_{\text{rand}}}\left[\text{Prec}\left(\hat{\theta}_{X_{\text{rand}}}\right)\right], \qquad \text{Loss}_{\text{rand}} \triangleq \mathsf{E}_{X_{\text{rand}}}\left[\text{Loss}\left(\hat{\theta}_{X_{\text{rand}}}\right)\right],$$

where the expectations are taken over the random allocation X_{rand}. Then,

$$\text{Prec}_{\text{rand}} = \frac{n}{\sigma^2}\left(1 - \frac{p-1}{n-1}\right), \quad \text{Loss}_{\text{rand}} = \frac{n}{n-1}(p-1).$$

The proof relies on simple probabilistic arguments and is presented in the Electronic Companion to this paper. Surprisingly the precision and loss of the randomized allocation *does not* depend on the data matrix Z at all, as long as it is full rank and has a constant column.

Comparing with the upper bound of Proposition 1, we notice that in the large sample size regime where $n \to \infty$, simple randomization is asymptotically order optimal in the sense that it achieves precision that grows with order n — the maximum permitted by the upper bound of Proposition 1 — when $p \ll n$. This may not be the case when p is close to n, however. For example, if $p = n-1$, which is the maximum value p can take under Assumption 1, then $\text{Prec}_{\text{rand}} \approx 1/\sigma^2$, which is of *constant order*. In such a case, the least squares estimator $\hat{\theta}_{X_{\text{rand}}}$ will not asymptotically converge to θ as $n \to \infty$. In general, simple randomization is asymptotically order optimal any time that $p_n = o(n)$ as $n \to \infty$.

Now we consider the performance of the least squares estimator under the optimal design that would be obtained by solving the offline experiment design problem (P1). By construction, the optimal design will clearly have precision that is at least that of the randomized procedure. We would like to understand the magnitude of the possible improvement, however, and to see if it is material. Unlike in the simple randomized case, however, the precision of the optimal design depends on the covariate matrix Z. Moreover, it is difficult to obtain a closed-form expression for this precision as a function of Z.

We can illustrate this with a simple example. Consider the case where $p = n-1$. The precision of the optimal design is given by

$$\sup_{x \in \{\pm 1\}^n} \frac{x^\top P_{Z^\perp} x}{\sigma^2}.$$

Since $p = n-1$, the null space of $Z^\top$ is a one dimensional subspace of $\mathbb{R}^n$. Let $y \in \mathbb{R}^n$ be a non-zero vector such that $Z^\top y = 0$ and $\|y\|_2^2 = 1$. That is, y is a unit vector in the null space of $Z^\top$. It is easy to see that $P_{Z^\perp} = yy^\top$. Thus, the precision of the optimal design is

$$\sup_{x \in \{\pm 1\}^n} \frac{x^\top y y^\top x}{\sigma^2} = \sup_{x \in \{\pm 1\}^n} \frac{\left(y^\top x\right)^2}{\sigma^2} = \frac{\|y\|_1^2}{\sigma^2}. \tag{2}$$

Now, consider the following two cases:

1. y has only two non-zero components given by $1/\sqrt{2}$ and $-1/\sqrt{2}$. In this case, the optimal precision is $2/\sigma^2$. Thus, in this case, randomization is within a constant factor of optimal.

2. y has entries such that $|y_i| = 1/\sqrt{n}$ and $\mathbf{1}^\top y = 0$. In this case, the precision is n/σ^2. Thus, in this case, the optimal design achieves the Cramér-Rao upper bound and the performance is a significant improvement over the randomized design.

The preceding two cases show, that depending on the covariate matrix Z (which determines the vector y in the discussion above), the performance of the optimal design may be a drastic improvement over that of the randomized design. In order to study the performance of the optimal design, we proceed by making a certain probabilistic assumption on Z. Under this assumption, we will then analyze the distribution of performance of the optimal design. For this purpose, we will assume a distribution on the covariate matrix Z as follows:

Assumption 2. *Given (n, p) with $1 \leq p < n$, assume that the covariate matrix $Z \in \mathbb{R}^{n \times p}$ has independent and identically distributed rows. Further, assume that for each $1 \leq k \leq n$, the kth row $Z_k \in \mathbb{R}^p$ satisfies $Z_{k,1} = 1$, and that the vector of all components except the first satisfies $Z_{k,2:p} \sim N(0, \Sigma)$, i.e., it is distributed according to a multivariate normal distribution with zero mean and covariance matrix $\Sigma \in \mathbb{R}^{p-1 \times p-1}$.*

It is easy to check that, under Assumption 2, the covariate matrix Z will satisfy the full rank condition of Assumption 1 almost surely. Consider a sequence of problems indexed by the sample size n, and where the dimension of the covariates is given by $1 \leq p_n < n$. For each n, let $Z^{n,p_n} \in \mathbb{R}^{n \times p_n}$ be the data matrix satisfying Assumption 2. We have that:

Theorem 3. *Suppose that Assumption 2 holds with $\Sigma = \rho^2 I$. Let x^* be an optimal design obtained by solving (P1) with covariate matrix $Z = Z^{n,p_n}$, and let $\hat{\theta}_{x^*, Z^{n,p_n}}$ be the corresponding least squares estimator of θ. Denote the precision of this estimator by*

$$\mathrm{Prec}_*^{n,p_n} \triangleq \mathrm{Prec}\left(\hat{\theta}_{x^*, Z^{n,p_n}}\right).$$

Then, we have that for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathsf{P}\left(\frac{\mathrm{Prec}_*^{n,p_n}}{n} < \frac{1}{8\pi\sigma^2} - \epsilon\right) = 0,$$

where the probability is measured over the distribution of the covariates.

Theorem 3 states that, with high probability, the optimal offline optimization-based design always yields $\Omega(n)$ precision under Assumption 2. Note that this is true for all possible values of $p_n < n$ with $p_n = n - 1$ being the worst case (the latter fact is established in the proof). In contrast, Theorem 2 establishes that when $p = n - 1$, the precision one expects under randomized allocation is $O(1)$, so that the relative improvement from optimization for this value of p is $\Theta(n)$. In other words, if the number of covariates is comparable to the sample size, we might expect dramatic improvements over simple randomization through optimization. Moreover, while the optimal design requires solution of (P1), which may not be tractable, Theorem 1 suggests a tractable approximation which is guaranteed to achieve the same precision as the optimal design up to a constant factor.

The proof of Theorem 3 is presented in Section C. Here we provide a proof sketch. Let $Z^{n,p} \in \mathbb{R}^{n\times p}$ and $Z^{n,n-1} \in \mathbb{R}^{n\times n-1}$ be two covariate matrices defined on the same probability space (under the Assumption 2 with $\Sigma = \rho^2 I$) such that they are identical on the first p columns. We show that $\mathrm{Prec}_*^{n,p} \geq \mathrm{Prec}_*^{n,n-1}$. This establishes that $p = n - 1$ corresponds to the worst case precision and allows us to focus on the sequence $\mathrm{Prec}_*^{n,n-1}$. We then analyze the distribution of $Z^{n,n-1}$. We show that $\mathrm{Prec}_*^{n,n-1}$ can be written down as a function of a unit vector in the null space of $(Z^{n,n-1})^\top$, say $y_n \in \mathbb{R}^n$. Further, y_n describes a random one-dimensional subspace of $\mathbb{R}^n$ that is invariant to orthonormal transformations that leave the constant vector unchanged. There is a unique distribution that has this property. We then identify the distribution and compute the precision in closed-form using this distribution. In particular, we show that, as $n \to \infty$,

$$\frac{\mathrm{Prec}_*^{n,n-1}}{n} \to \frac{1}{8\pi\sigma^2},$$

where the convergence is in distribution.

4. Sequential Problem

We now consider the online experiment design problem (P2). Here, decisions must be made sequentially. At each time k, an allocation $x_k \in \{\pm 1\}$ must be made based only on the first k covariates and any prior allocations. In other words, x_k is $\mathcal{F}_k$-measurable.

In this section we show that the optimization problem is tractable. First, we pose a surrogate problem in which the objective of (P2) is simplified. The details of this simplification are provided in Section 4.1. In Section 4.2, we show that the reduction in performance when the surrogate problem is used to device an assignment policy is negligible. Focusing on the surrogate problem, we show that the surrogate problem is a p-dimensional dynamic program in Section 4.3. Surprisingly, if we assume that the data generating distribution for the covariates comes from the so-called *elliptical*

family then the state space collapses to two dimensions, making the dynamic program tractable. This state space collapse is presented in Section 4.4.

4.1. Formulation and Surrogate Problem

In order to formulate the sequential problem with an expected value objective, a probabilistic model for covariates is necessary. We will start by making the following assumption:

Assumption 3. *Given (n, p) with $1 \leq p < n$, assume that the covariate matrix $Z \in \mathbb{R}^{n \times p}$ has independent and identically distributed rows. Further, assume that for each $1 \leq k \leq n$, the kth row $Z_k \in \mathbb{R}^p$ satisfies $Z_{k,1} = 1$, and that the vector $Z_{k,2:p} \in \mathbb{R}^{p-1}$ of all components except the first has zero mean and covariance matrix $\Sigma \in \mathbb{R}^{p-1 \times p-1}$.*

Assumption 3 requires that the sequentially arriving covariates are i.i.d. with first and second moments. Assumption 2, by comparison, in addition imposes a Gaussian distribution.

Problem (P2) can be viewed as maximizing the expectation of terminal reward that is given by

$$x^\top P_{Z^\perp} x = x^\top \left(I - Z(Z^\top Z)^{-1} Z^\top \right) x = n - \frac{1}{n} \left(\sum_{k=1}^{n} x_k Z_k \right)^\top \Gamma_n^{-1} \left(\sum_{k=1}^{n} x_k Z_k \right), \tag{3}$$

where the sample second moment of covariates is given by

$$\Gamma_n \triangleq \frac{1}{n} \sum_{k=1}^{n} Z_k Z_k^\top.$$

We write this matrix in block form as

$$\Gamma_n = \begin{bmatrix} 1 & M_n^\top \\ M_n & \Sigma_n \end{bmatrix},$$

where,

$$\Sigma_n \triangleq \frac{1}{n} \sum_{k=1}^{n} Z_{k,2:p} Z_{k,2:p}^\top, \quad M_n \triangleq \frac{1}{n} \sum_{k=1}^{n} Z_{k,2:p}.$$

Here, M_n and Σ_n correspond to sample estimates of the covariate mean and covariance structure, respectively.

We define, for each k, the scalar *sample count imbalance* $\delta_k \in \mathbb{R}$ and the *covariate imbalance vector* $\Delta_k \in \mathbb{R}^{p-1}$ by

$$\delta_k \triangleq \sum_{\ell=1}^{k} x_\ell, \quad \Delta_k \triangleq \sum_{\ell=1}^{k} x_\ell Z_{\ell,2:p}. \tag{4}$$

The terminal reward (3) is equal to

$$x^\top P_{Z^\perp} x = n - \frac{1}{n} \begin{bmatrix} \delta_n & \Delta_n^\top \end{bmatrix} \begin{bmatrix} 1 & M_n^\top \\ M_n & \Sigma_n \end{bmatrix}^{-1} \begin{bmatrix} \delta_n \\ \Delta_n \end{bmatrix}.$$

Problem (P2) is then equivalent to

$$\begin{aligned}
\text{(P3)} \triangleq \text{minimize} \quad & \mathsf{E}\left[\begin{bmatrix}\delta_n & \Delta_n^\top\end{bmatrix}\begin{bmatrix}1 & M_n^\top \\ M_n & \Sigma_n\end{bmatrix}^{-1}\begin{bmatrix}\delta_n \\ \Delta_n\end{bmatrix}\right] \\
\text{subject to} \quad & x \in \{\pm 1\}^n, \\
& x_k \text{ is } \mathcal{F}_k\text{-measurable}, \quad \forall\, 1 \le k \le n.
\end{aligned}$$

Observe that the objective of (P3) corresponds to n times the loss of the estimator.

As $n \to \infty$, by the strong law of large numbers (under mild additional technical assumptions), $\Sigma_n \to \Sigma$ and $M_n \to 0$ almost surely. Motivated by this fact, in developing an efficient algorithm for (P3), our first move will be to consider a surrogate problem that replaces the sample covariance matrix Σ_n with the exact covariance matrix Σ and sets the sample mean M_n to the exact mean 0:

$$\begin{aligned}
\text{(P3}'\text{)} \triangleq \text{minimize} \quad & \mathsf{E}\left[\delta_n^2 + \|\Delta_n\|_{\Sigma^{-1}}^2\right] \\
\text{subject to} \quad & x \in \{\pm 1\}^n, \\
& x_k \text{ is } \mathcal{F}_k\text{-measurable}, \quad \forall\, 1 \le k \le n.
\end{aligned}$$

Here, given an arbitrary covariance matrix $\hat{\Sigma} \in \mathbb{R}^{p-1 \times p-1}$, we find it convenient to introduce the norm $\|\cdot\|_{\hat{\Sigma}^{-1}}$ on $\mathbb{R}^{p-1}$ defined by $\|z\|_{\hat{\Sigma}^{-1}} \triangleq (z^\top \hat{\Sigma}^{-1} z)^{1/2}$. In the present context, this norm is typically referred to as a Mahalanobis distance.

The roles of the sample count imbalance δ_n and the covariate imbalance vector Δ_n in the surrogate problem (P3′) are intuitive: requiring δ_n to be small balances the number of assignments between the two treatments (the focus of the so-called biased-coin designs). Requiring the same of Δ_n will tend to 'balance' covariates — when Δ_n is small, the empirical moments of the covariates across the two treatments are close. As discussed in the introduction, heuristics developed in the literature on the design of optimal trials tend to be driven by precisely these two forces.

For the rest of this section we will focus on the surrogate problem. We want to first justify the use of the surrogate objective. We do this by providing an approximation guarantee in Section 4.2. We then turn our attention on how to solve the surrogate problem via dynamic programming in the subsequent sections.

4.2. Approximation Guarantee for the Surrogate Problem

First, we show that the policy obtained by solving (P3′) is near optimal. Denote by $\hat{\mu}$ the measure over the sequence x_k induced by an optimal solution for the surrogate control problem (P3′), and let μ^* denote the measure induced by an optimal policy for our original dynamic optimization problem (P3). Now, δ_n and Δ_n are random variables given an allocation policy. Given a allocation policy μ, define

$$D_\mu^{n,p} \triangleq \mathsf{E}_\mu\left[\begin{bmatrix}\delta_n & \Delta_n^\top\end{bmatrix}\Gamma_n^{-1}\begin{bmatrix}\delta_n \\ \Delta_n\end{bmatrix}\right]$$

to be the objective value of (P3) under the allocation policy μ with sample size n and covariate dimension p. The following result is demonstrated, without loss of generality, under the assumption that Σ is the identity (otherwise, we simply consider setting $Z_{k,2:p}$ to $\Sigma^{-1/2}Z_{k,2:p}$):

Theorem 4. *Suppose that Assumption 2 holds with $\Sigma = I$ and let $\epsilon > 0$ be any positive real number. Consider a sequence of problems indexed by the sample size n, where the dimension of the covariates is given by $1 \leq p_n < n$ and $\gamma_n > 0$ are real numbers such that, for n sufficiently large, $n \geq L \max(p_n, l \log 2/\gamma_n)/\epsilon^2$. Then, as $n \to \infty$*

$$D_{\hat{\mu}}^{n,p_n} \leq \left(\frac{1+\epsilon}{1-\epsilon}\right)^2 D_{\mu^*}^{n,p_n} + \gamma_n n^2 + \gamma_n n^2 p_n + O\left(\sqrt{\frac{n}{p_n - 1}}\right).$$

Here, L and l are universal constants. In particular, selecting $\gamma_n \propto 1/n^4$ yields

$$D_{\hat{\mu}}^{n,p_n} \leq \left(\frac{1+\epsilon}{1-\epsilon}\right)^2 D_{\mu^*}^{n,p_n} + O\left(\sqrt{\frac{n}{p_n - 1}}\right). \tag{5}$$

The result above relies on the use of non-asymptotic guarantees on the spectra of random matrices with sub-Gaussian entries and can be found in the Electronic Companion to this paper.

The preceding result bounds the objective of the problem (P3) when (P3′) is used to devise an allocation policy. However, we are interested in the objective of the problem problem (P2), which is the precision or inverse variance of the design corresponding to the policy used. In particular, denote by $\mathrm{Prec}_{\mu}^{n,p}$ the expected precision of the estimator when allocations are made with a policy μ, for a problem with sample size n and covariate dimension p, i.e.,

$$\mathrm{Prec}_{\mu}^{n,p} = \frac{\mathsf{E}_{\mu}\left[x^\top P_{Z^\perp} x\right]}{\sigma^2} = \frac{n - D_{\mu}^{n,p}/n}{\sigma^2}. \tag{6}$$

Then, we have the following:

Corollary 1. *Suppose that Assumption 2 holds with $\Sigma = I$. Consider a sequence of problems indexed by the sample size n, where the dimension of the covariates is given by $1 \leq p_n < n$, and a fixed positive real number $\epsilon > 0$ such that*

$$\epsilon > \sqrt{L \limsup_{n\to\infty} p_n/n},$$

for a universal constant L. Then, as $n \to \infty$,

$$\frac{\mathrm{Prec}_{\hat{\mu}}^{n,p_n}}{\mathrm{Prec}_{\mu^*}^{n,p_n}} \geq 1 - \frac{4\epsilon^3}{(L-\epsilon^2)(1-\epsilon^2)} + o(1).$$

Corollary 1 gives the multiplicative loss in the precision by using an allocation derived from the surrogate problem (P3′). The multiplicative loss depends on the ratio p/n, which is captured in the choice of ϵ. For small values of ϵ the ratio of precision obtained by solving (P3′) and (P2) approaches 1. Note that this result holds in an asymptotic regime where p and n both increase to

infinity, as long as p/n remains small.

Proof of Corollary 1. Consider (5) in Theorem 4. This holds when

$$n \geq \frac{L \max(p_n, l \log 2/\gamma_n)}{\epsilon^2}$$

with $\gamma_n = b/n^4$ for some constant b. Equivalently,

$$n \geq \frac{L \max(p_n, 4l \log n + 2l \log b)}{\epsilon^2}.$$

For n sufficiently large, clearly the constraint that $n \geq L(4l \log n + 2l \log b)/\epsilon^2$ will be satisfied. Therefore, combined with the lower bound hypothesized for ϵ, (5) holds as $n \to \infty$.

Using (6),

$$\begin{aligned}
\mathrm{Prec}_{\mu^*}^{n,p_n} - \mathrm{Prec}_{\hat{\mu}}^{n,p_n} &= \frac{D_{\hat{\mu}}^{n,p_n} - D_{\mu^*}^{n,p_n}}{n\sigma^2} \\
&\leq \frac{\frac{(1+\epsilon)^2}{(1-\epsilon)^2} D_{\mu^*}^{n,p_n} - D_{\mu^*} + O\left(\sqrt{\frac{n}{p_n - 1}}\right)}{n\sigma^2} \\
&= \frac{4\epsilon D_{\mu^*}^{n,p_n}}{n\sigma^2(1-\epsilon)^2} + o(1) \\
&= \frac{4\epsilon}{(1-\epsilon)^2}\left(\frac{n}{\sigma^2} - \mathrm{Prec}_{\mu^*}^{n,p_n}\right) + o(1).
\end{aligned} \tag{7}$$

The first inequality follows from Theorem 4 and the last equality from (6).

Let $\mathrm{Prec}_{\mathrm{rand}}^{n,p_n}$ denote precision of the randomized policy. Using Theorem 2 and the optimality of μ^*, we have that

$$\frac{n}{\sigma^2} - \mathrm{Prec}_{\mu^*}^{n,p_n} \leq \frac{n}{\sigma^2} - \mathrm{Prec}_{\mathrm{rand}}^{n,p_n} = \frac{n}{\sigma^2}\frac{p_n - 1}{n-1} \leq \frac{n}{\sigma^2}\frac{p_n}{n} \leq \frac{\epsilon^2 n}{L\sigma^2}, \tag{8}$$

where the last inequality uses the fact that, by hypothesis, $p_n/n \leq \epsilon^2/L$. Substituting this into (7) we get that

$$\mathrm{Prec}_{\mu^*}^{n,p_n} - \mathrm{Prec}_{\hat{\mu}}^{n,p_n} \leq \frac{4\epsilon^3 n}{(1-\epsilon)^2 L\sigma^2} + o(1).$$

Now, using (8) we get that,

$$\mathrm{Prec}_{\mu^*}^{n,p_n} \geq \frac{n}{\sigma^2}\left(1 - \frac{\epsilon^2}{L}\right).$$

Thus, we have that,

$$\begin{aligned}
1 - \frac{\mathrm{Prec}_{\hat{\mu}}^{n,p_n}}{\mathrm{Prec}_{\mu^*}^{n,p_n}} &\leq \frac{4\epsilon n}{\mathrm{Prec}_{\mu^*}^{n,p_n}(1-\epsilon)^2 L\sigma^2} + o(1) \\
&\leq \frac{4\epsilon^3}{(L-\epsilon^2)(1-\epsilon^2)} + o(1).
\end{aligned}$$

This yields the result. ■

4.3. Dynamic Programming Decomposition

It is not difficult to see that (P3′) is a terminal cost dynamic program with state $(\delta_{k-1}, \Delta_{k-1}) \in \mathbb{R}^p$ at each time k. The pair (δ_k, Δ_k) can be interpreted as the post-decision state of the dynamic decision problem immediately after the kth allocation. In other words, given the past arrival sequence and actions, (δ_k, Δ_k) summarizes the the impact of this 'past' on the future objective. This is formally stated in the following proposition:

Proposition 2. *Suppose that Assumption 3 holds. For each $1 \leq k \leq n$, define the function $Q_k \colon \mathbb{R} \times \mathbb{R}^{p-1} \to \mathbb{R}$ by the Bellman equation*

$$Q_k(\delta_k, \Delta_k) \triangleq \begin{cases} \delta_n^2 + \|\Delta_n\|_{\Sigma^{-1}}^2, & \text{if } k = n, \\ \mathsf{E}\left[\min\limits_{u \in \{\pm 1\}} \; Q_{k+1}(\delta_k + u, \Delta_k + uZ_{k+1,2:p})\right], & \text{if } 1 \leq k < n. \end{cases} \tag{9}$$

Then,

1. *At each time k, the optimal continuation cost for the dynamic program (P3′) is given by $Q_k(\delta_k, \Delta_k)$. In other words, this is the expected terminal cost, given then covariates observed and the allocations made up to and including time k, assuming optimal decisions are made at all future times.*

2. *Suppose the allocation x_k^* at each time k is made according to*

$$x_k^* \in \operatorname*{argmin}_{u \in \{\pm 1\}} \; Q_k\left(\delta_{k-1} + u, \Delta_{k-1} + uZ_{k,2:p}\right).$$

Then, the sequence of allocations x^* is optimal for the online experiment design problem (P3′).

Proposition 2, whose proof is presented in the Electronic Companion to this paper, suggests a standard dynamic programming line of attack for the surrogate problem (P3′): optimal continuation cost functions $\{Q_k\}_{1 \leq k \leq n}$ can be computed via backward induction, and these can then be applied to determine an optimal policy. However, the dimension of this dynamic program is given by the number of covariates p. In general, the computational effort required by this approach will be exponential in p — this is the so-called curse of dimensionality. Thus, outside of very small numbers of covariates, say, $p \leq 3$, the standard dynamic programming approach is intractable. However, as we will now see, that the surrogate problem surprisingly admits an alternative, low dimensional dynamic programming representation.

4.4. State Space Collapse

Proposition 2 yields a dynamic programming approach for the surrogate problem (P3′) that is intractable for all but very small values of p. What is remarkable, however, is that if the covariate

data is assumed to have an *elliptical distribution*, then (P3′) can be solved via a tractable *two*-dimensional dynamic program. We first present the technical definition.

Definition 1. *A random variable X taking values in $\mathbb{R}^m$ has an elliptical distribution if the characteristic function $\varphi\colon \mathbb{C}^m \to \mathbb{C}$ has the form*

$$\varphi(t) \triangleq \mathsf{E}\left[\exp(it^\top X)\right] = \exp(i\mu^\top t)\Psi(t^\top \Sigma t),$$

for all $t \in \mathbb{C}^m$, given some $\mu \in \mathbb{R}^m$, $\Sigma \in \mathbb{R}^{m\times m}$, and a characteristic function $\Psi\colon \mathbb{C} \to \mathbb{C}$.

Elliptical distributions, studied extensively, for example, by Cambanis et al. (1981), are a generalization of the multivariate Gaussian distribution. The name derives from the fact that if an elliptical distribution has a density, then the contours of the density are ellipsoids in $\mathbb{R}^m$ parameterized by μ and Σ. A useful standard result for us (see, e.g., Cambanis et al., 1981) is that these distributions can be generated by independently generating the direction and the length of the deviation (in $\|\cdot\|_{\Sigma^{-1}}$-norm) from the center μ:

Proposition 3. *If X has an elliptical distribution with parameters μ, Σ, and Ψ, then there exists a non-negative random variable R such that,*

$$X \overset{d}{=} \mu + R\Sigma^{1/2}U,$$

where U is distributed uniformly on the unit sphere $\{x \in \mathbb{R}^{p-1} \mid \|x\|_2^2 = 1\}$ and U and R are independent.

Thus, any elliptical distribution can be identified with a vector $\mu \in \mathbb{R}^m$, a positive semidefinite matrix $\Sigma \in \mathbb{R}^{m\times m}$, and random variable R taking values on the non-negative real line. We denote such a distribution by $\mathrm{Ell}(\mu, \Sigma, R)$. It can be shown that if $R^2 \sim \chi^2_m$ is a chi-squared distribution with m degrees of freedom, then $\mathrm{Ell}(\mu, \Sigma, R)$ is a Gaussian distribution with mean μ and covariance Σ. Well-known distributions such as the multivariate t-distribution, Cauchy distribution, and logistic distribution also fall in the elliptical family.

We state the assumption needed for the state space collapse.

Assumption 4. *Given (n, p) with $1 \le p < n$, assume that the covariate matrix $Z \in \mathbb{R}^{n\times p}$ has independent and identically distributed rows. Further, assume that for each $1 \le k \le n$, the kth row $Z_k \in \mathbb{R}^p$ satisfies $Z_{k,1} = 1$, and that the vector $Z_{k,2:p} \in \mathbb{R}^{p-1}$ of all components except the first is distributed according to $\mathrm{Ell}(0, \Sigma, R)$, where it is assumed that the random variable R has finite second moment, and further that, without loss of generality,*[8] $\mathsf{E}[R^2] = p - 1$.

The following theorem shows how the p-dimensional dynamic program is reduced to a 2-dimensional one with Assumption 4.

Theorem 5. *Suppose that Assumption 4 holds. For each $1 \le k \le n$, define the function $q_k\colon \mathbb{Z} \times$*

[8] Note that under our assumption, it is easy to verify that each covariate vector $Z_{k,2:p}$ is zero mean. Our choice of normalization $\mathsf{E}[R^2] = p - 1$ ensures that the covariance matrix of $Z_{k,2:p}$ is given by Σ. This second moment requirement does exclude heavy-tailed elliptical distributions such as the Cauchy distribution. However, it is necessary so that our performance criteria (expected precision) is finite.

$\mathbb{R}_+ \to \mathbb{R}$ *according to*

$$q_k(m,\lambda) \triangleq \begin{cases} m^2+\lambda, & \text{if } k=n, \\ \mathsf{E}\left[\min\limits_{u\in\{\pm 1\}} \; q_{k+1}\left(m+u, \lambda + 2uRU_1\sqrt{\lambda}+R^2\right)\right], & \text{if } 1\le k<n. \end{cases} \tag{10}$$

Here, when $k<n$, the expectation is taken over independent random variables U and R that are the random variables in the stochastic decomposition of $Z_{1,2:p}$ from Assumption 4. Then,

1. *At each time k, the optimal continuation cost for the dynamic program (P3′) is given by*

$$Q_k(\delta_k,\Delta_k) = q_k\left(\delta_k, \|\Delta_k\|^2_{\Sigma^{-1}}\right).$$

   In other words, this is the expected terminal cost, given then covariates observed and the allocations made up to and including time k, assuming optimal decisions are made at all future times.

2. *Suppose the allocation x_k^* at each time k is made according to*

$$x_k^* \in \operatorname*{argmin}_{u\in\{\pm 1\}} \; q_k\left(\delta_{k-1}+u, \|\Delta_{k-1}+uZ_{k,2:p}\|^2_{\Sigma^{-1}}\right). \tag{11}$$

   Then, the sequence of allocations x^* is optimal for the online experiment design problem (P3′).

For the case of Gaussian distribution, the recursion (10) for solving the DP can be simplified according to the following corollary:

Corollary 2. *If Assumption 2 holds, then, for $1\le k\le n$, the functions $q_k^{\text{gauss}}\colon \mathbb{Z}\times\mathbb{R}_+\to\mathbb{R}$ are given by*

$$q_k^{\text{gauss}}(m,\lambda) \triangleq \begin{cases} m^2+\lambda, & \text{if } k=n, \\ \mathsf{E}\left[\min\limits_{u\in\{\pm 1\}} \; q_{k+1}^{\text{gauss}}\left(m+u, \left(\sqrt{\lambda}+u\eta\right)^2+\xi\right)\right], & \text{if } 1\le k<n. \end{cases} \tag{12}$$

Here, when $k<n$, the expectation is taken over independent random variables $(\eta,\xi)\in\mathbb{R}^2$, where $\eta\sim N(0,1)$ is a standard normal random variable, and $\xi\sim\chi^2_{p-2}$ is chi-squared random variable with $p-2$ degrees of freedom.[9]

We provide the proofs for Theorem 5 and Corollary 2 in the Electronic Companion to this paper. We make the following observations:

1. A key point is that, unlike the standard dynamic programming decomposition of Proposition 2, Theorem 5 provides a *tractable* way to solve the surrogate problem (P3′), independent of the covariate dimension p. This is because the recursion (10) yields a two-dimensional

[9] If $p=2$, we take $\xi\triangleq 0$.

dynamic program. One of the state variables of this program, m, is discrete, taking values on the integers from $-n$ to n. Further, one can show that, with high probability, the second state variable λ is $O(n^2)$ thereby allowing us to discretize the state-space on a two-dimensional mesh. The functions $\{q_k\}$ can be numerically evaluated on this grid via backward induction. Note that since the expectation in (10) is over a two-dimensional random variable, it can be computed via numerical integration. Further details of this procedure are given in Section 6.

2. Moreover, the functions $\{q_k\}$ do not directly depend on the matrix Σ at all and only indirectly depend on time horizon n through the remaining time $k-n$. In fact, they only depend on the covariate dimension p. For example, in the Gaussian case, this means that if these functions are computed offline, they can subsequently be applied to *all* p-dimensional problem with a Gaussian data distribution.

3. Finally, the algorithm assumes that the covariance matrix Σ is known. This is needed to compute the $\|\cdot\|_{\Sigma^{-1}}$-norm of Δ_k. In practice, Σ may not be known, and may need to be estimated from data. However, observe that Σ depends only on the distribution of covariates across the subject population, not on the outcome of experiments. In the applications we have in mind, there is typically a wealth of information about this population known in advance of the experimental trials. Hence, Σ can be estimated offline even if the number of covariates p is large and the number of experimental subjects n is small.

   For example, in an online advertising setting, and advertiser may want to compare two creatives using A-B testing with a limited number of experimental subjects. In advance of any experiments, the advertiser can use historical data from other trials or market surveys over the same population of subjects to estimate Σ.

5. Variations of the Sequential Problem: A Dynamic Programming Framework

The vanilla formulation of the sequential problem (P2) described in Section 2.2 solely optimizes statistical efficiency. In reality, a complete framework must allow the designer to model a number of additional constraints relevant to practical implementation, including budgets on allocations to the treatment arm; controlling selection bias in addition to maximizing efficiency; optimally stopping an experiment if efficiency objectives are met; and so forth. We will establish that the solution approach described in Section 4 applies to a substantially more general class of problem than the vanilla problem (P2).

To setup this dynamic programming framework, we introduce a few new concepts:

- We will think of the allocation at time $1 \leq k \leq n$ as a *bias* $v_k \in [0,1]$. Our optimization algorithm will yield the optimal bias at any given point in time, and then we pick an allocation

by flipping a coin with this bias, i.e., setting

$$x_k = \begin{cases} +1 & \text{with probability } v_k, \\ -1 & \text{with probability } 1 - v_k. \end{cases} \tag{13}$$

This is the same decision space as in a biased coin design.

- We are given convex stage wise costs, $c\colon [0,1] \to \mathbb{R}$, that are a function of bias. This can capture for instance, the 'cost' of a sample unit; the extent of 'non-randomness' in a given choice of bias, etc.
- The set of permitted bias v_k at any stage $1 \le k \le n$ can be constrained to an arbitrary convex set that is itself a function of the state at that time, $\mathcal{V}_k\left(\delta_{k-1}, \|\Delta_{k-1}\|^2_{\Sigma^{-1}}\right) \subset [0,1]$.
- Instead of a fixed time horizon n, we allow the experiment to be stopped early according to a stopping time $1 \le \tau \le n$. As we discuss below this allows us to model optimal early stopping based, for instance, on estimating the treatment effect with a desired precision.

Given these concepts, and an arbitrary parameter $\gamma \ge 0$, consider the following generalization of the problem (P3′):

$$\begin{aligned} (\text{P3}'') \triangleq \text{minimize} \quad & \mathsf{E}\left[\delta_\tau^2 + \|\Delta_\tau\|^2_{\Sigma^{-1}} + \gamma \sum_{k=1}^{\tau} c(v_k)\right] \\ \text{subject to} \quad & v_k \in \mathcal{V}_k\left(\delta_{k-1}, \|\Delta_{k-1}\|^2_{\Sigma^{-1}}\right), && \forall\, 1 \le k \le n, \\ & v_k \text{ is } \mathcal{F}_k\text{-measurable}, && \forall\, 1 \le k \le n. \end{aligned}$$

Following the same arguments as in Section 4.4, (P3″) can be solved according to optimal continuation costs given by the two-dimensional Bellman recursion[10]

$$q_k(m,\lambda) \triangleq \begin{cases} m^2 + \lambda, & \text{if } k = \tau, \\ \mathsf{E}\Big[\min\limits_{v \in \mathcal{V}_{k+1}(m,\lambda)} \gamma c(v) & \\ \qquad + v q_{k+1}\left(m+1, \lambda + 2RU_1\sqrt{\lambda} + R^2\right) & \text{if } 1 \le k < \tau, \\ \qquad + (1-v) q_{k+1}\left(m-1, \lambda - 2RU_1\sqrt{\lambda} + R^2\right)\Big], & \end{cases} \tag{14}$$

for each time k. Given the optimal continuation costs, an optimal decision v_k at each time k can

[10] In order for the decomposition (14) to apply, an additional technical assumption is needed on the stopping time τ: we assume that, for each $1 \le k < \tau$, the distribution of the random variable corresponding the to future stopped payoff $\delta_\tau^2 + \|\Delta_\tau\|^2_{\Sigma^{-1}}$ is conditionally independent of the history given the current state (δ_k, Δ_k).

be computed according to

$$
\begin{aligned}
v_k^* \in \underset{v \in \mathcal{V}_k\left(\delta_{k-1}, \|\Delta_{k-1}\|^2_{\Sigma^{-1}}\right)}{\operatorname{argmin}} \quad & \gamma c(v) \\
& + v q_k\left(\delta_{k-1}+1, \|\Delta_{k-1}+Z_{k,2:p}\|^2_{\Sigma^{-1}}\right) \\
& + (1-v) q_k\left(\delta_{k-1}-1, \|\Delta_{k-1}-Z_{k,2:p}\|^2_{\Sigma^{-1}}\right),
\end{aligned}
\tag{15}
$$

In the following, we illustrate how (P3″) addresses several practical variations of the sequential allocation problem:

Selection Bias. An important consideration that has emerged in the literature on A-B testing is managing so-called 'selection bias'. Following Blackwell and Hodges (1957), one commonly defines the selection bias of an allocation over n time steps as $\frac{2}{n}\sum_{k=1}^{n}|v_k - 1/2|$. Notice that perfect randomization has zero selection bias, whereas a fully deterministic procedure (where v_k is either 0 or 1) has the highest bias possible, one.

It is frequently important to balance this bias against efficiency (or, equivalently, loss). In particular, we want a Pareto optimal solution across the two criteria. Atkinson (2014) compares a multitude of state-of-the-art biased coin design (BCD) procedures and calls a procedure 'admissible' if it is not Pareto dominated by some other procedure. He finds that none of the heuristics he examines can be ruled out implying that *none* of these heuristics are Pareto optimal. But by varying $\gamma \geq 0$ in (P3″), we can generate a Pareto optimal solution at any point on the trade-off curve. Specifically, to incorporate selection bias into our framework, we simply define

$$
c(v) \triangleq |v - 1/2|, \quad \tau \triangleq n, \quad \mathcal{V}_k \triangleq [0,1]. \tag{16}
$$

Our approach can consequently produce any design on the Pareto frontier, and thus Pareto dominate state-of-the-art BCD designs. We will see this numerically in Section 6.

Notice that the optimal policy equation (15) in the setting of (16) is a linear program. Direct examination of this program yields an interesting insight: at every time k, the optimal action for (P3″) is restricted to $v_k \in \{0, 1/2, 1\}$. In other words, an optimal policy will only either take a deterministic action or fully randomize. This is in contrast to the main BCD heuristics developed in the literature (some of which we will describe shortly in Section 6.3), which tend to vary probabilities over the entire interval $[0,1]$.

Allocation Budget. Assuming a test with a total sample size of n, the designer may be happy to assign these samples to the control arm (the 'status quo') but may want to limit exposure to the test. Formally, we may want to have a budget B on the number of $+1$ allocations in the trial. As it turns out BCD does not naturally extend to this setting (Han et al., 2009; Kuznetsova and

Tymofyeyev, 2012). (P3″) can trivially incorporate a budget constraint, we simply define

$$c(v) \triangleq 0, \quad \tau \triangleq n, \quad \mathcal{V}_k\left(\delta_{k-1}, \|\Delta_{k-1}\|^2_{\Sigma^{-1}}\right) \triangleq \begin{cases} [0,1] & \text{if } k + \delta_{k-1} < 2B, \\ \{0\} & \text{otherwise.} \end{cases}$$

Endogenous Stopping. Consider the (not uncommon) scenario where there is an economic cost associated with every incremental sampling unit in a sequential trial, and all we care about is estimating the treatment effect up to a desired level of precision; see Johari et al. (2017) for a broader discussion of related problems. In such a scenario, we may opportunistically want to stop early so that the sample size is in fact picked endogenously. For concreteness, let us suppose that the unit cost per sample is a constant r. Assume further that it suffices to estimate the treatment effect with precision κ, unless the trial has run up to a sample size of n in which case we must stop. One can think of n here as an upper bound on sample size imposed by the trial designer. The objective is simply to minimize the expected cost of the trial. This problem is easily modeled in our framework. Specifically, (P3″) can capture this problem by defining

$$c(v) \triangleq r, \quad \tau \triangleq \min\left\{k \geq 1 \ : \ k - \frac{1}{k}\left(\delta_k^2 + \|\Delta_k\|^2_{\Sigma^{-1}}\right) \geq \kappa\sigma^2\right\} \wedge n, \quad \mathcal{V}_k \triangleq [0,1].$$

6. Experiments

This section focuses on numerical experiments with data. We will attempt to highlight the relative merits of our approach vis-à-vis simple randomization, as well as biased coin designs (BCDs). As discussed in the literature review, BCDs are an approach to minimizing loss (or equivalently, maximizing efficiency) by dynamically adjusting for covariate imbalances.

Our goal will be to show that for a given level of *selection bias*, our approach provides an improvement in efficiency (or a reduction in loss) over competing BCDs. Equivalently, our approach can achieve a given level of efficiency with a smaller level of selection bias. We will study these relative merits for varying values of sample size n, and the number of covariates p. Finally, while our analysis in Section 4 required the covariates to follow an elliptical distribution, such a requirement may not hold in real applications. As such we conduct experiments using click log data from Yahoo! wherein the covariates are categorical; we show that our approach enjoys similar relative merits in this setting.

6.1. BCDs, Loss, and Selection Bias

Let $v_k \in [0,1]$ denote the probability that the kth allocation is set to $x_k = +1$ under a given allocation rule $\mathcal{A}$. Recall from Section 5 that a measure of selection bias under $\mathcal{A}$ is defined according to

$$\text{Bias}_{\mathcal{A}} \triangleq \mathsf{E}\left[\frac{2}{n}\sum_{k=1}^{n}|v_k - 1/2|\right] \in [0,1].$$

(Here, we have normalized the bias to be contained in the unit interval.) This measure captures the extent of randomness (or, equivalently, how predictable any given allocation is) under $\mathcal{A}$ (Blackwell and Hodges, 1957). Also, recall our definition of loss,

$$\text{Loss}_{\mathcal{A}} \triangleq n - \mathsf{E}\left[x^\top P_{Z^\perp} x\right] = \mathsf{E}\left[x^\top Z \left(Z^\top Z\right)^{-1} Z^\top x\right] \geq 0.$$

The loss under $\mathcal{A}$ is interpreted as the effective number of samples on which information is lost due to an imbalance in covariates. It is well known that any allocation rule engenders a trade-off between loss and selection bias, so that a comparison between rules ideally compares the entire trade-off curve attained by the two rules (Atkinson, 2002). We will do precisely this in the experiments that follow.

Observe that the expressions for bias and loss do not depend on the experimental outcomes $\{y_k\}$. From an empirical perspective, this is helpful: we can assess any rule $\mathcal{A}$, given only access to the covariate distribution. The conclusions we draw on the relative merits of one approach with respect to another hold across any linear model for the given covariate structure.

6.2. Data

We run our experiments on two different data distributions for the covariates. Assumption 3 holds in both cases. Thus, $\{Z_k\}$ are i.i.d. and $Z_{k,1}$ is assumed to be 1. We run our experiments with the following sampling distributions for $Z_{2:p}$:

Synthetic Gaussian Data. In our synthetic experiments, we assume that $Z_{2:p}$ follows multivariate normal distribution. This is, of course, an elliptical distribution, so that Assumption 2 is satisfied. For the covariance matrix Σ, we set $\Sigma_{ii} = 1.0$ and $\Sigma_{ij} = 0.1$ for any $j \neq i$.

Yahoo! User Data. To experiment on data from a more realistic setting, we use a dataset of user click log data from the Yahoo! front page.[11] The users here are visitors to 'Featured Tab of the Today Module' on the Yahoo! front page. In the dataset, each user has 136 associated features, such as age and gender. Each feature is binary, taking values in $\{0, 1\}$. Some of these features were constant throughout the dataset, and these were discarded. Duplicate and co-linear features were discarded as well. Features were selected at random until up to $p = 40$ features were collected. Feature selection was repeated independently in each simulation trial.

Our algorithm requires the covariance matrix of the data as an input. For this purpose, we estimate the covariance matrix from a portion of the dataset. This estimate is obtained by simply taking a sample average across 1 million data points kept aside from the rest of the experiments.

Finally, for evaluation purposes, we require a generative model for the data. To this end, from a set of 1 million data points we sample individual data points, with replacement. In other words, as the sampling distribution we use the empirical distribution of the 1 million data points used for

[11] This dataset is obtained from the Yahoo! Labs repository of datasets available for academic research, and can be downloaded as "R6B — Yahoo! Front Page Today Module User Click Log Dataset, version 2.0" at `http://webscope.sandbox.yahoo.com/catalog.php?datatype=r`.

testing. Such a sampling procedure is intended to mimic the arrival of users on the Yahoo! front page.

6.3. Algorithms

Dynamic Programming (Our Approach). The problem at hand is addressed by the dynamic programming formulation described in Sections 5. As such, we are required to compute the 2-dimensional value functions given by $\{q_k\}_{1\leq k\leq n}$. These functions are computed offline by backward induction following (14). Here, we provide the computational details for this operation. In particular, given $q_{k+1}(\cdot,\cdot)$, we compute $q_k(\cdot,\cdot)$ as follows:

1. Discretization: The first state variable m is discrete and can take values from $-n$ to n. We discretize values for the second state variable λ on a geometric mesh taking values λ_0^i for $\lambda_0 \triangleq 1.5$ and $0 \leq i \leq 26$. The maximum value value of λ was chosen so that $\|\Delta_k\|^2_{\Sigma^{-1}}$ has a low probability of exceeding it.

2. Sampling: For each discretized pair (m,λ) we estimate $q_k(m,\lambda)$ via Monte Carlo simulation. In particular, $N = 10{,}000$ pairs[12] $(\xi,\eta) \in \mathbb{R}^2$ are sampled from the appropriate distributions and $q_k(m,\lambda)$ is estimated according to (14) using the corresponding empirical measure. We use the same sample set of (ξ,η) for all (m,λ) at which this is evaluated.

3. Interpolation: Given an (m,λ) such that λ is not a discretized mesh point, we estimate $q_{k+1}(m,\lambda)$ in the Bellman recursion (14) by linear interpolation between the closest points in the discretized mesh.

Biased Coin Designs. In addition to our own dynamic programming algorithm, we will consider several other rules proposed in the literature. These include: Rule ABCD (Baldi Antognini and Zagoraiou, 2011), which following Atkinson (2014), we refer to as Rule J; Smith's rule (Rule S) (Smith, 1984b,a); Atkinson's rule (Rule A) (Atkinson, 1982), and the Bayesian procedure of Ball et al. (1993) (Rule B). Rules J, S, and B are all parameterized by a scalar parameter, which we denote ρ, that may take values in $(0,\infty)$. Rule A is a special case of Rule S taking $\rho = 1$. As $\rho \to 0$, these rules become equivalent to randomization. On the other hand, as $\rho \to \infty$, these rules become entirely deterministic in nature. As such, for values of ρ close to zero, one expects low selection bias whereas as $\rho \to \infty$ one expects to see a reduction in loss at the expense of selection bias; a deterministic rule has the largest possible selection bias of 1. In order to precisely specify each of these rules, define

$$d_k\left(u_{k+1}, Z_{k+1,2:p}\right) \triangleq \left(1 - u_{k+1}\delta_k/k - u_{k+1}Z_{k+1,2:p}^\top \Sigma^{-1}\Delta_k/k\right)^2$$

[12] In all examples, our algorithm assumes that the covariate data is generated from a multivariate normal, even when this was not true (Yahoo! dataset). In this case, when $Z_{2:p} \sim N(I,\Sigma)$ is multivariate normal, $\lambda + 2uRU_1\sqrt{\lambda} + R^2$ has the same distribution as $\left(\sqrt{\lambda} + u\eta\right)^2 + \xi$ where η is a standard normal and ξ is a chi-squared random variable with $p-2$ degrees of freedom. See also Corollary 2.

where $u_{k+1} \in \{\pm 1\}$, $Z_{k+1,2:p} \in \mathbb{R}^{p-1}$, and δ_k and Δ_k have the usual definitions (4). For background on the function $d_k(\cdot,\cdot)$, see Atkinson (1982); this quantity arises naturally in the sequential design of D_A-optimal experiments. The rules described above then take the following form:

1. **Rules S/A**: Assign $x_{k+1} = +1$ with probability

$$v_{k+1} \triangleq \frac{d_k(+1, Z_{k+1,2:p})^\rho}{d_k(+1, Z_{k+1,2:p})^\rho + d_k(-1, Z_{k+1,2:p})^\rho}.$$

   The parameter ρ can take values in $(0, \infty)$. Rule A corresponds to the special case where $\rho = 1$.

2. **Rule B**: Assign $x_{k+1} = +1$ with probability

$$v_{k+1} \triangleq \frac{(1 + d_k(+1, Z_{k+1,2:p}))^\rho}{(1 + d_k(+1, Z_{k+1,2:p}))^\rho + (1 + d_k(-1, Z_{k+1,2:p}))^\rho}.$$

   The parameter ρ can again take values in $(0, \infty)$. This rule is very similar to Rule S, but permits a Bayesian interpretation (Ball et al., 1993).

3. **Rule D**: Assign $x_{k+1} = +1$ deterministically if $d_k(+1, Z_{k+1,2:p}) > d_k(-1, Z_{k+1,2:p})$, set $x_{k+1} = -1$ otherwise. This rule is obtained in the limit as $\rho \to \infty$ for rules A, S, and B. Note that this deterministic rule is equivalent to a *myopic* policy that seeks to optimize the objective of (P3′) assuming that x_{k+1} is the final allocation to be made, and ignoring the impact of this allocation on future decision making.

4. **Rule J**: Define the 'discrepancy' after k allocations, $D_k(Z_{k+1,2:p})$ according to

$$D_k(Z_{k+1,2:p}) \triangleq \frac{2 - k\left(d_k(+1, Z_{k+1,2:p}) + d_k(-1, Z_{k+1,2:p})\right)}{d_k(+1, Z_{k+1,2:p}) - d_k(-1, Z_{k+1,2:p})},$$

   assuming $d_k(+1, Z_{k+1,2:p}) \neq d_k(-1, Z_{k+1,2:p})$. If $D_k(Z_{k+1,2:p}) < 0$, we assign $x_{k+1} = +1$ with probability

$$v_{k+1} \triangleq \frac{|D_k(Z_{k+1,2:p})|^\rho}{1 + |D_k(Z_{k+1,2:p})|^\rho}.$$

   If, on the other hand $D_k(Z_{k+1,2:p}) > 0$, we assign $x_{k+1} = +1$ with probability

$$v_{k+1} \triangleq \frac{1}{1 + |D_k(Z_{k+1,2:p})|^\rho}.$$

   Finally, if $D_k(Z_{k+1,2:p}) = 0$ or $d_k(+1, Z_k) = d_k(-1, Z_k)$, we simply randomize ($v_{k+1} = 1/2$). The parameter ρ can again take values in $(0, \infty)$.

6.4. Results

Our goal is to compare the statistical efficiency of our dynamic programming-based sequential algorithm to the various competing BCDs discussed above while controlling for selection bias.

In order to do this, we run each BCD procedure for an increasing sequence of value of ρ. The smallest value used, $\rho = 0$, is simply equivalent to randomized allocation. The largest value of ρ we considered for each scheme was chosen so that the rule was effectively deterministic. We implemented our sequential DPalgorithm for an increasing sequence of values of γ, tracing out a similar trade-off curve.

Results are reported in Figures 1, 2, and 3. Of these, Figures 1 and 2 show results on synthetic Gaussian data while Figure 3 shows results on the Yahoo! dataset. Each data point in these figures is the average of 10,000 independent Monte Carlo trials with shared randomness across all BCD rules and our own rule; and different data points were generated for each rule by varying their respective configurations of ρ and γ.

These figures reveal that:

1. For any target level of selection bias, our dynamic programming algorithm has the smallest loss among all of the alternatives implemented. In this way, the DP approach Pareto dominates all alternatives. The relative improvement in loss can be non-trivial: the loss incurred under our approach can be up to five times smaller for moderate budgets on selection bias. Put a different way the effective number of samples 'lost' due to covariate imbalance can be substantially smaller for a given budget on selection bias.

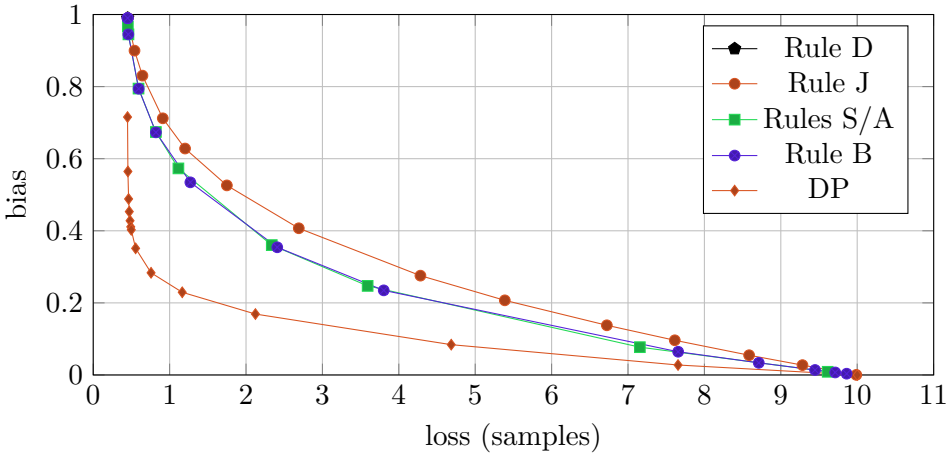
2. The relative improvement alluded to above is particularly pronounced for smaller values of p/n. Our intuition here is as follows: keeping n fixed one expects to require fewer non-random allocations for small p. As such, the importance of strategizing on *when* to employ a non-random allocation has greater impact in such a setting.

3. The relative merits of our sequential approach appear more pronounced in the setting where n is larger.

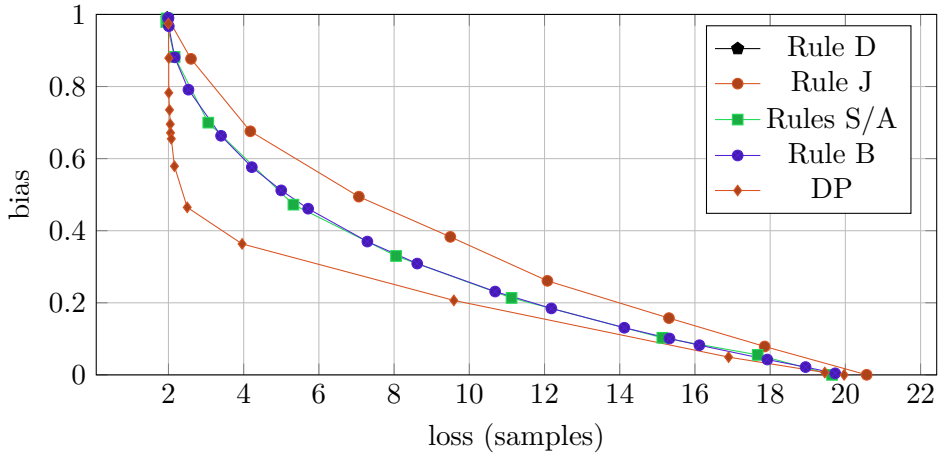
4. Finally, observe that Figure 3 shows results on the Yahoo! dataset, and that the covariates in this experiment are in fact categorical. Despite this we see that our approach exhibits similar improvements relative to the competing BCD schemes.

7. Conclusion

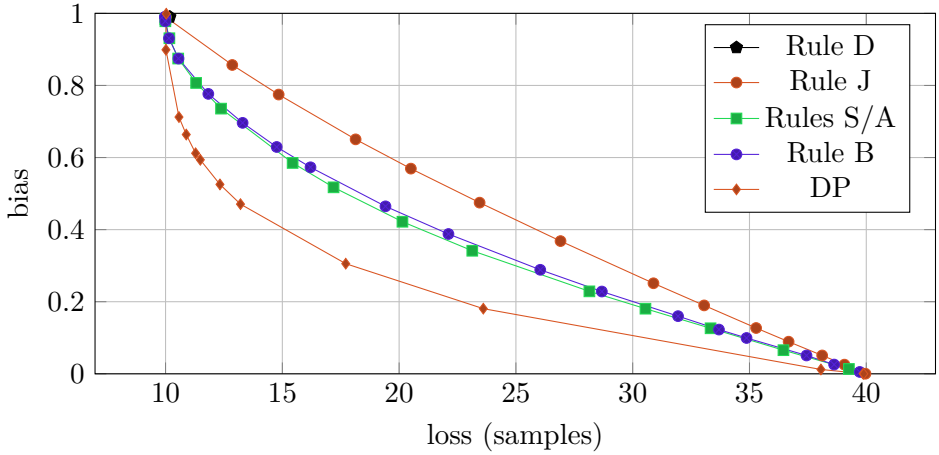
We conclude with a summary of what we have accomplished and what we view as key directions for further research. At a conceptual level, this paper illustrates the power of the 'optimization' viewpoint in what are inherently statistical problems: we have presented a provably near optimal solution to a problem for which a plethora of heuristics were available. In addition to establishing the appropriate approach to this problem, the algorithms we have developed are eminently practical and easy to implement — a property that is crucial for the sorts of applications that motivated this work. On a more pragmatic note, we have quantified the *value* of these sorts of optimization approaches establishing precise estimates of the benefits optimization approaches provide over straightforward



(a) $p = 10$

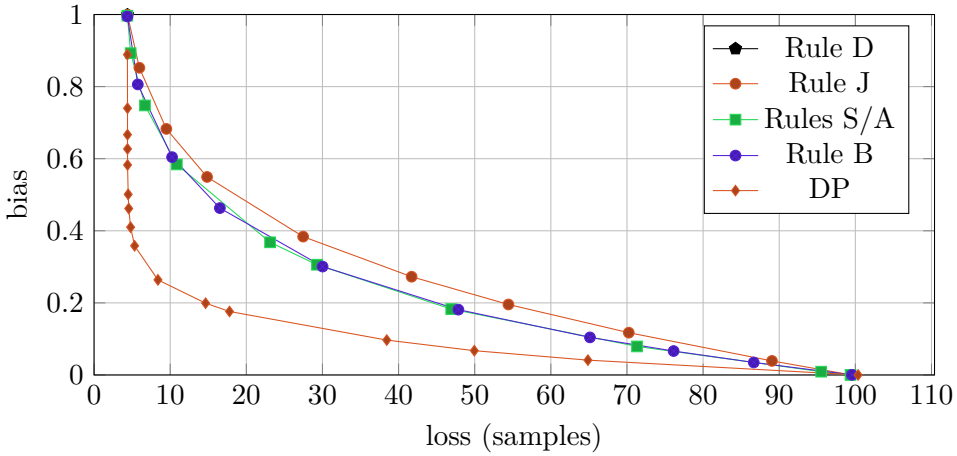


(b) $p = 20$

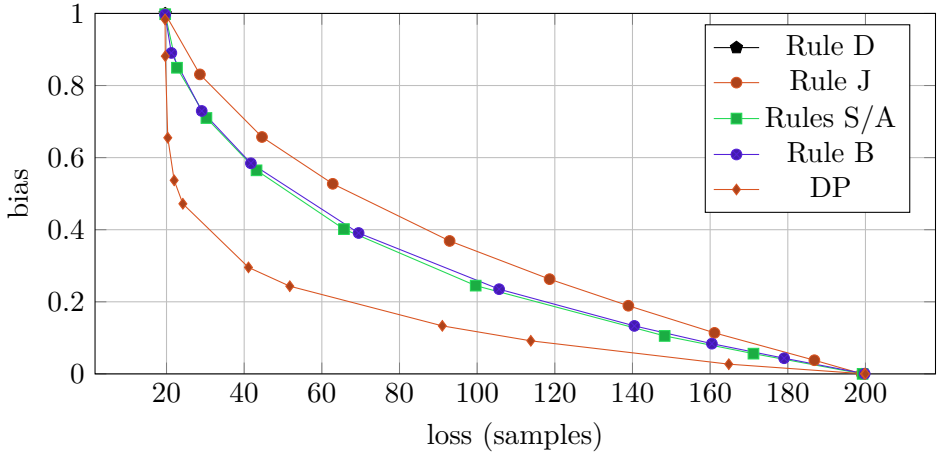


(c) $p = 40$

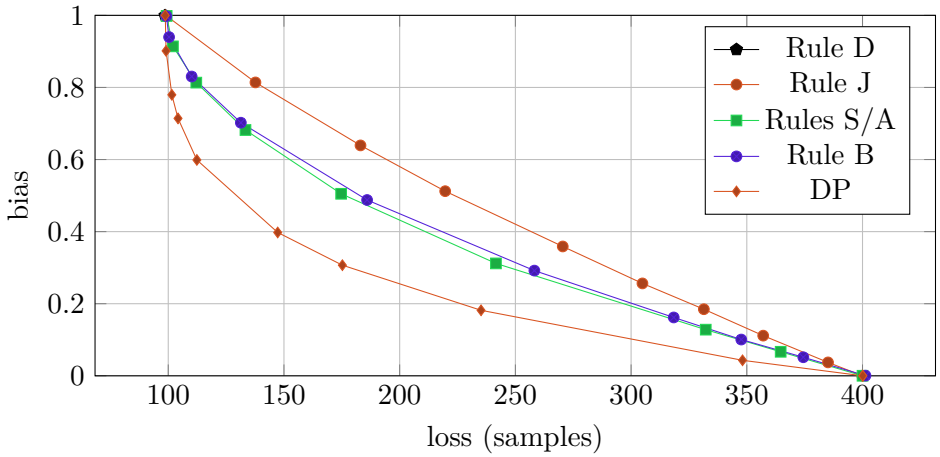
Figure 1: Bias-loss trade-off on synthetic Gaussian data for $n = 100$ and varying values of p.



(a) $p = 100$

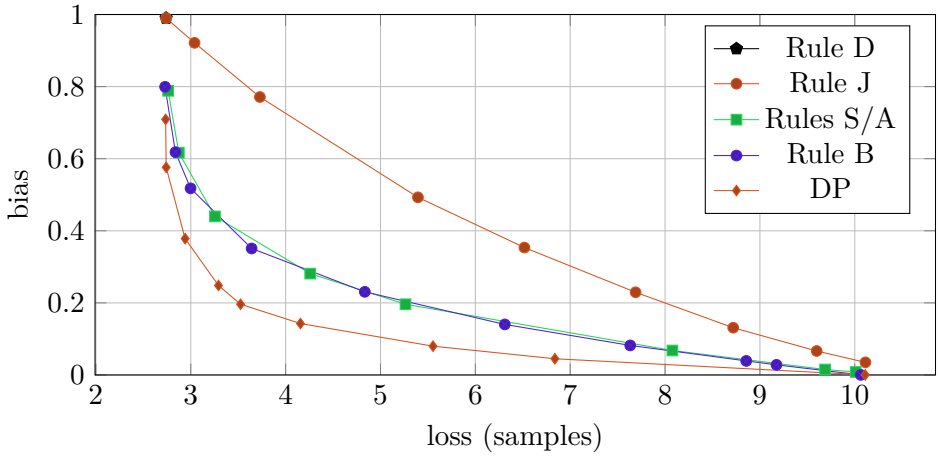


(b) $p = 200$

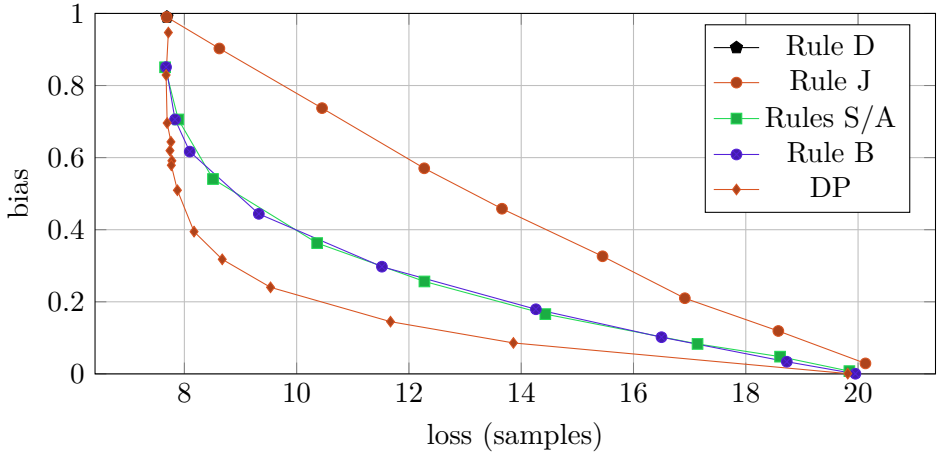


(c) $p = 400$

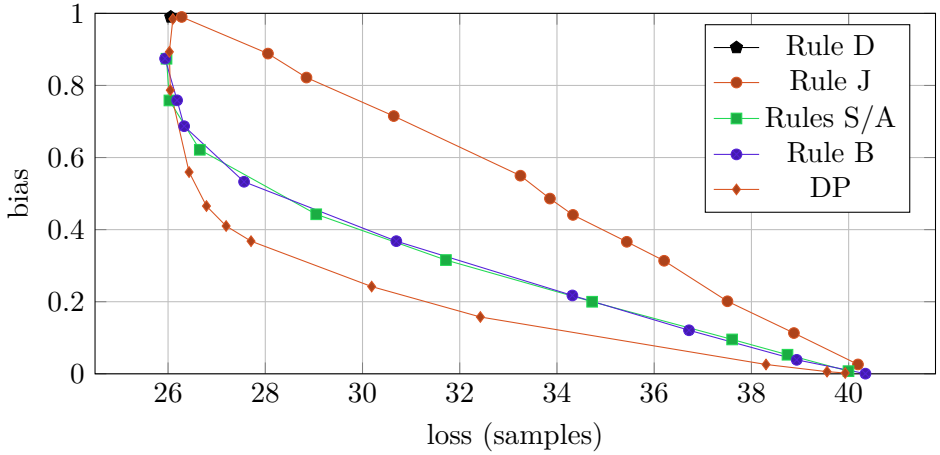
Figure 2: Bias-loss trade-off on synthetic Gaussian data for $n = 1000$ and varying values of p.



(a) $p = 10$



(b) $p = 20$



(c) $p = 40$

Figure 3: Bias-loss trade-off on the Yahoo! dataset for $n = 100$ and varying values of p.

randomization. These estimates illustrate that in so-called high dimensional setting — i.e., in settings where the number of covariates is large, such approaches can provide order of magnitude improvements in sampling efficiency.

Our progress does come at the expense of structural assumptions on the relationship between the observed effect and observable covariates. In particular, we assumed a linear model with exogenous noise. Any such structural assumption is restrictive. In the event that these assumptions fail, they could result in biased estimates of the treatment effect. With that said, it appears difficult to overcome the risk of such a bias while using a covariate dependent treatment assignment scheme. In addition to these structural assumptions, we also required that the experiment designer have some knowledge on the distribution of the covariates (their covariance matrix). Our theoretical results made further distributional assumptions on these covariates. Much remains to be done to mitigate the impact of these limiting assumptions, and as such a number of directions remain for future research. We highlight several here in parting:

1. Normality: To what extent can our assumption on the normality of covariates be relaxed? Can we develop approximation guarantees for the situation when covariates are not normally distributed?

2. Non-linear models: Can we allow for a nonlinear dependence on covariates? One direction to accomplish this is perhaps a reliance of some manner of non-parametric 'kernel' approach. The good news here is that the value of optimization is likely to be even higher in such an infinite-dimensional setting.

3. More than two alternatives: The present paper considers only the two alternative setting, an important direction for future work would be to consider settings where there is a larger number of choices.

References

A. C. Atkinson. Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika*, 69(1):61–67, 1982.

A. C. Atkinson. Optimum biased-coin designs for sequential treatment allocation with covariate information. *Statistics in Medicine*, 18(14):1741–1752, 1999.

A. C. Atkinson. The comparison of designs for sequential clinical trials with covariate information. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 165(2):349–373, 2002.

A. C. Atkinson. Selecting a biased-coin design. *Statistical Science*, 29(1):144–163, 2014.

A. Baldi Antognini and M. Zagoraiou. The covariate-adaptive biased coin design for balancing clinical trials in the presence of prognostic factors. *Biometrika*, 98(3):519–535, 2011.

F. G. Ball, A. F. M. Smith, and I. Verdinelli. Biased coin designs with a Bayesian bias. *Journal of Statistical Planning and Inference*, 34(3):403–421, 1993.

A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. Society for Industrial and Applied Mathematics, 2001.

D. P. Bertsekas. *Abstract Dynamic Programming*. Athena Scientific, Belmont, MA, 2013.

D. Bertsimas, M. Johnson, and N. Kallus. The power of optimization over randomization in designing experiments involving small samples. *Operations Research*, 63(4):868–876, 2015.

D. Blackwell and J. L. Hodges. Design for the control of selection bias. *The Annals of Mathematical Statistics*, 28(2):449–460, 1957.

S. Cambanis, S. Huang, and G. Simons. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(3):368–385, 1981.

S. Chick, M. Forster, and P. Pertile. A Bayesian decision theoretic model of sequential experimentation with delayed response. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5): 1439–1462, 2017.

S. E. Chick and P. Frazier. Sequential sampling with economics of selection procedures. *Management Science*, 58(3):550–569, 2012.

S. E. Chick and N. Gans. Economic analysis of simulation selection problems. *Management Science*, 55(3): 421–437, 2009.

T. D. Cook, D. T. Campbell, and A. Day. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Houghton Mifflin Boston, 1979.

B. Efron. Forcing a sequential experiment to be balanced. *Biometrika*, 58(3):403–417, 1971.

R. A. Fisher. *The Design of Experiments*. Oliver & Boyd, 1935.

M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42(6):1115–1145, November 1995.

B. Han, N. H. Enas, and D. McEntegart. Randomization by minimization for unbalanced treatment allocation. *Statistics in Medicine*, 28(27):3329–3346, 2009.

J. R. Hauser, G. L. Urban, G. Liberali, and M. Braun. Website morphing. *Marketing Science*, 28(2):202–223, 2009.

R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2012.

Y. Hu and F. Hu. Asymptotic properties of covariate-adaptive randomization. *The Annals of Statistics*, 40 (3):1794–1815, 2012.

A. T. James. Normal multivariate analysis and the orthogonal group. *The Annals of Mathematical Statistics*, pages 40–75, 1954.

R. Johari, L. Pekelis, and D. J. Walsh. Always valid inference: Bringing sequential analysis to A/B testing. Working paper, 2017.

N. Kallus. Regression-robust designs of controlled experiments. Working paper, 2013.

A. Kapelner and A. Krieger. Matching on-the-fly: Sequential allocation with higher power and efficiency. *Biometrics*, 70(2):378–388, 2014.

M. Kasy. Why experimenters should not randomize, and what they should do instead. *European Economic Association & Econometric Society*, 2013.

S.-H. Kim and B. L. Nelson. Selecting the best system. *Handbooks in operations research and management science*, 13:501–534, 2006.

O. M. Kuznetsova and Y. Tymofyeyev. Preserving the allocation ratio at every allocation with biased coin randomization and minimization in studies with unequal allocation. *Statistics in Medicine*, 31(8):701–723, 2012.

J. Langford and T. Zhang. The epoch-greedy algorithm for contextual multi-armed bandits. In *Advances in Neural Information Processing Systems*, pages 1096–1103, 2007.

Y. Nesterov. Semidefinite relaxation and nonconvex quadratic optimization. Technical report, Université Catholique de Louvain, Center for Operations Research and Econometrics, 1997.

S. J. Pocock and R. Simon. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, pages 103–115, 1975.

F. Pukelsheim. *Optimal Design of Experiments*. Society for Industrial and Applied Mathematics, 2006.

S. W. Raudenbush, A. Martinez, and J. Spybrook. Strategies for improving precision in group-randomized experiments. *Educational Evaluation and Policy Analysis*, 29(1):5–29, 2007.

W. F. Rosenberger and O. Sverdlov. Handling covariates in the design of clinical trials. *Statistical Science*, pages 404–419, 2008.

E. M. Schwartz, E. T. Bradlow, and P. S. Fader. Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science*, 2017.

R. L. Smith. Properties of biased coin designs in sequential clinical trials. *The Annals of Statistics*, pages 1018–1034, 1984a.

R. L. Smith. Sequential treatment allocation using biased coin designs. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 519–543, 1984b.

D. P. Steensma and H. M. Kantarjian. Impact of cancer research bureaucracy on innovation, costs, and patient care. *Journal of Clinical Oncology*, 32(5):376–378, 2014.

O. Toubia, D. I. Simester, J. R. Hauser, and E. Dahan. Fast polyhedral adaptive conjoint estimation. *Marketing Science*, 22(3):273–303, 2003.

O. Toubia, J. R. Hauser, and D. I. Simester. Polyhedral methods for adaptive choice-based conjoint analysis. *Journal of Marketing Research*, 41(1):116–131, 2004.

R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Y. Eldar and G. Kutyniok, editors, *Compressed Sensing, Theory and Applications*, pages 210–268. Cambridge University Press, 2012.

M. Woodroofe. A one-armed bandit problem with a concomitant variable. *Journal of the American Statistical Association*, 74(368):799–806, 1979.

Electronic Companion to "Near Optimal A-B Testing"

Nikhil Bhat
Graduate School of Business
Columbia University
email: nbhat15@gsb.columbia.edu

Vivek F. Farias
Sloan School of Management
Massachusetts Institute of Technology
email: vivekf@mit.edu

Ciamac C. Moallemi
Graduate School of Business
Columbia University
email: ciamac@gsb.columbia.edu

Deeksha Sinha
Operations Research Center
Massachusetts Institute of Technology
email: deeksha@mit.edu

Initial Version: October 2015
Current Revision: September 2018

A. Derivation of the Optimization Problem

Here derive the expression for precision used in Section 2. Denote the matrix $X \triangleq [x \; Z]$ and $\beta \triangleq [\theta \; \kappa^\top]^\top$. Thus our model is

$$y = X\beta + \epsilon.$$

The least squares estimate $\hat{\beta}$ of β is given by

$$\hat{\beta} = (X^\top X)^{-1} X^\top y = (X^\top X)^{-1} X^\top (X\beta + \epsilon) = \beta + (X^\top X)^{-1} X^\top \epsilon.$$

Then,

$$\mathrm{Var}(\hat{\beta}) = (X^\top X)^{-1} X^\top \mathrm{Var}(\epsilon\epsilon^\top) X (X^\top X)^{-1} = \sigma^2 (X^\top X)^{-1}.$$

Thus variance of $\hat{\theta} = \hat{\beta}_1$ is

$$\mathrm{Var}(\hat{\theta}) = \sigma^2 e_1^\top (X^\top X)^{-1} e_1 = \sigma^2 e_1^\top \begin{bmatrix} x^\top x & x^\top Z \\ Z^\top x & Z^\top Z \end{bmatrix}^{-1} e_1 = \frac{\sigma^2}{x^\top (I - Z(Z^\top Z)^{-1} Z^\top) x}.$$

Here, $e_1 \triangleq (1, 0, \ldots)$ is the first coordinate vector, and for the last equality we apply the block matrix inversion formula.

Proof of Proposition 1. By the Cramér-Rao bound we have that,

$$\mathrm{Cov}\left(\begin{bmatrix} \hat{\theta} \\ \hat{\kappa} \end{bmatrix} \right) \succeq I(\theta, \kappa)^{-1},$$

where $I(\theta, \kappa)$ is the Fisher information matrix. Under the Gaussian assumption for ϵ it is easy to see that,

$$I(\theta, \kappa)^{-1} = \sigma^2 \begin{bmatrix} x^\top x & x^\top Z \\ Z^\top x & Z^\top Z \end{bmatrix}^{-1}.$$

If e_1 is the unit vector along the first coordinate then,

$$\mathrm{Var}(\hat{\theta}) \geq e_1^\top I(\theta, \kappa)^{-1} e_1 = \frac{\sigma^2}{x^\top (I - Z(Z^\top Z)^{-1} Z^\top) x}.$$

Thus,

$$\mathrm{Prec}(\hat{\theta}) \leq \frac{x^\top \left(I - Z(Z^\top Z)^{-1} Z^\top\right) x}{\sigma^2} = \frac{n - x^\top Z(Z^\top Z)^{-1} Z^\top x}{\sigma^2} \leq \frac{n}{\sigma^2}.$$

The inequality follows since $Z(Z^\top Z)^{-1} Z^\top$ is positive semidefinite.

The last statement is consequence of the fact that $x^\top \left(I - Z(Z^\top Z)^{-1} Z^\top\right) x/\sigma^2$ is the precision of the optimal least squares estimator. ■

B. Performance of the Randomized Algorithm

We begin with a lemma that relies on some linear algebra arguments.

Lemma 1. *Consider a vector $a \in \mathbb{R}^{p-1}$ and an invertible $Q \in \mathbb{R}^{p-1 \times p-1}$ such that the matrix,*

$$\begin{bmatrix} 1 & a^\top \\ a & Q \end{bmatrix},$$

is invertible. Then,

$$\begin{bmatrix} 1 & a^\top \end{bmatrix} \begin{bmatrix} 1 & a^\top \\ a & Q \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ a \end{bmatrix} = 1.$$

Proof. By the block matrix inversion formula,

$$\begin{bmatrix} 1 & a^\top \\ a & Q \end{bmatrix}^{-1} = \begin{bmatrix} \rho^{-1} & -\rho^{-1} a^\top Q^{-1} \\ -\rho^{-1} Q^{-1} a & (Q - aa^\top)^{-1} \end{bmatrix} = \begin{bmatrix} \rho^{-1} & -\rho^{-1} a^\top Q^{-1} \\ -\rho^{-1} Q^{-1} a & Q^{-1} + \rho^{-1} Q^{-1} a a^\top Q^{-1} \end{bmatrix},$$

where $\rho \triangleq 1 - a^\top Q^{-1} a$. Thus,

$$\begin{aligned} \begin{bmatrix} 1 & a^\top \end{bmatrix} \begin{bmatrix} 1 & a^\top \\ a & Q \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ a \end{bmatrix} &= \rho^{-1} - 2\rho^{-1} a^\top Q^{-1} a + a^\top Q^{-1} a + \rho^{-1} \left(a^\top Q^{-1} a\right)^2 \\ &= \rho^{-1} - 2\rho^{-1}(1-\rho) + 1 - \rho + \rho^{-1} (1-\rho)^2 \\ &= \frac{1 - 2(1-\rho) + (1-\rho)\rho + 1 + \rho^2 - 2\rho}{\rho} = 1. \end{aligned}$$

■

Now we turn our attention to quantifying the performance of the randomized design.

Lemma 2. *Supposed the allocation x is chosen at random from the set $\{\pm 1\}^n$ independently of the covariate values Z, according to some distribution so that, for all $1 \le i < j \le n$,*

$$\mathsf{E}_x[x_i x_j] = \alpha,$$

for some constant α. Then,

$$\mathsf{E}_x\left[\mathrm{Loss}\left(\hat{\theta}_x\right)\right] = (1-\alpha)p + \alpha n,$$

where the expectation is taken over the distribution of x.

Proof. Define

$$\overline{Z} \triangleq \frac{1}{n}\sum_{k=1}^{n} Z_k, \qquad \Gamma_n \triangleq Z^\top Z/n,$$

so that (using Lemma 1)

$$Z^\top Z = n\Gamma_n, \qquad \overline{Z}^\top \Gamma_n^{-1}\overline{Z} = 1.$$

Then,

$$\begin{aligned}
\mathsf{E}_x\left[\mathrm{Loss}\left(\hat{\theta}_x\right)\right] &= \mathsf{E}_x\left[x^\top Z(Z^\top Z)^{-1} Z x\right] \\
&= \mathsf{E}_x\left[\left(\sum_{k=1}^{n} x_k Z_k\right)^\top (n\Gamma_n)^{-1}\left(\sum_{k=1}^{n} x_k Z_k\right)\right] \\
&= \frac{1}{n}\sum_{k=1}^{n}\sum_{\ell=1}^{n} \mathsf{E}_x\left[x_k x_\ell\right] Z_k^\top \Gamma_n^{-1} Z_\ell \\
&= \frac{1}{n}\left(\sum_{k=1}^{n} Z_k^\top \Gamma_n^{-1} Z_k + \alpha \sum_{k=1}^{n}\sum_{\ell\neq k} Z_k^\top \Gamma_n^{-1} Z_\ell\right) \\
&= \frac{1-\alpha}{n}\sum_{k=1}^{n} Z_k^\top \Gamma_n^{-1} Z_k + \frac{\alpha}{n}\left(\sum_{k=1}^{n} Z_k\right)^\top \Gamma_n^{-1}\left(\sum_{k=1}^{n} Z_k\right) \\
&= \frac{1-\alpha}{n}\sum_{k=1}^{n} \mathrm{tr}\left(\Gamma_n^{-1} Z_k Z_k^\top\right) + \alpha n \overline{Z}^\top \Gamma_n^{-1}\overline{Z} \\
&= (1-\alpha)\,\mathrm{tr}\left(\Gamma_n^{-1}\frac{1}{n}\sum_{k=1}^{n} Z_k Z_k^\top\right) + \alpha n \\
&= (1-\alpha)p + \alpha n.
\end{aligned}$$

■

Proof of Theorem 2. We can directly apply Lemma 2, with the observation that, under the pro-

posed randomized allocation, if $1 \leq i < j \leq n$,

$$\alpha \triangleq \mathsf{E}_x\left[x_i x_j\right] = \frac{n/2-1}{n-1} - \frac{n/2}{n-1} = -\frac{1}{n-1}.$$

■

C. Asymptotic Performance of the Optimal Design

In this section, we will prove Theorem 3. The theorem relies on Assumption 2 with $\Sigma = \rho^2 I$. In particular, we assume that $Z_{i,1} = 1$ and $Z_{i,j} \sim N(0, \rho^2)$ for $j > 1$. Further it is assumed that all entries of Z are independent.

We will place a sequence of problems of dimensions $1 \leq p < n$ on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$. To make the dependence on the dimension clear, we will denote the data matrix by $Z^{n,p}$. In this sequence of data matrices, $Z^{n,p}$ is formed by adding a column to $Z^{n,p}$. The additional column has the distribution $N(0, \rho^2 I_n)$. Let $\{Z^{n,n-1}\}_n$ be an independent sequence. Note that the sequence of matrices $\{Z^{n,p_n}\}_n$ defined using this generative model satisfy the assumptions laid out in Theorem 3.

Before we proceed let us set up some notation. Let $\mathrm{Gr}(k, \mathbb{R}^n)$ be the Grassmannian of dimension k in the vector space $\mathbb{R}^n$. In other words, it is the set of all subspaces of dimension k in $\mathbb{R}^n$. Let $\mathcal{S}^{n,p} \in \bigcup_{k=n-p}^{n} \mathrm{Gr}(k, \mathbb{R}^n)$ be the null space of $Z^{n,p\top}$. In other words, it the the orthogonal complement of the span of $Z^{n,p}$. In the following Lemma we show that the $Z^{n,p}$ is full rank.

Lemma 3. *The rank of $Z^{n,p}$ is p with probability 1. Thus, $\mathcal{S}^{n,p} \in Gr(n-p, \mathbb{R}^n)$ almost surely.*

Proof. We can prove this inductively. Since $Z^{n,1} = \mathbf{1}$, the statement is trivially true for $p = 1$. Assume that $Z^{n,p-1}$ is rank $p-1$. It implies that the span of $Z^{n,p-1}$ is a $p-1$ dimensional subspace, let us call it $\mathrm{span}(Z^{n,p-1})$. The pth column of $Z^{n,p}$ is non-degenerate Gaussian vector independent of $\mathrm{span}(Z^{n,p-1})$, call it $Z^{,p}$. $\mathsf{P}(Z^{,p} \in \mathrm{span}(Z^{n,p-1})) = 0$. Thus, almost surely, $Z^{n,p}$ is of rank p. ■

From the preceding lemma we can conclude that $\mathcal{S}^{n,n-1}$ is a 1 dimensional subspace, with probability 1. Now we derive an expression for the precision of the optimal estimator for $p = n-1$ in terms of $\mathcal{S}^{n,n-1}$. Let $A \triangleq \{\omega \in \Omega : \mathcal{S}^{n-1}(\omega) \in \mathrm{Gr}(1, \mathbb{R}^n)\}$. From now on, we assume $\Omega = A$ and all subsequent statements hold with probability one.

Consider a function $h : \mathrm{Gr}(1, \mathbb{R}^n) \to \mathbb{R}_+$, such that $h(\mathcal{S}) \triangleq \|y\|_1 / \|y\|_2$ for some non-zero $y \in \mathcal{S}$. It is trivial to check that this value is unique for any non-zero y in $\mathcal{S} \in \mathrm{Gr}(1, \mathbb{R}^n)$.

Lemma 4. *The precision of the optimal estimator for $p = n-1$ is given by $\sigma^{-2} h(\mathcal{S}^{n,n-1})^2$, almost surely.*

Proof. We know that the optimal precision for $n = p-1$ is given by $\sigma^{-2} x^{*\top} P_{Z^{n,n-1\perp}} x^*$, where x^* is the assignment that maximizes (P1). Now note that, $P_{Z^{n,n-1\perp}}$ can be given by $yy^\top / \|y\|_2^2$, for

any non-zero $y \in \mathcal{S}^{n,n-1}$. Thus the optimization problem (P1) is,

$$\begin{array}{ll} \text{maximize} & x^\top \dfrac{yy^\top}{\|y\|_2^2} x = \dfrac{(x^\top y)^2}{\|y\|_2^2} \\ \text{subject to} & x \in \{\pm 1\}^n. \end{array}$$

But the optimal x is such that $x_i = \text{sgn}(y_i^n)$. With this assignment, the optimal value is $\|y\|_1^2/\|y\|_2^2$. Thus the optimal precision for a given ω is given by $\|y\|_1^2/\sigma^2\|y\|_2^2 = h(\mathcal{S}^{n,n-1})^2/\sigma^2$. Thus,

$$\text{Prec}_*^{n,n-1} = \frac{h(\mathcal{S}^{n,n-1})^2}{\sigma^2},$$

almost surely. ■

Using the fact that we have all the $Z^{n,p}$s on the same probability space, it is easy to show that the precision monotonically decreases as p grows, for a fixed n.

Lemma 5. *For a fixed n, $\text{Prec}_*^{n,p}$ is a decreasing sequence in p. Thus,*

$$\inf_{1 \le p < n} \frac{\text{Prec}_*^{n,p}}{n} = \frac{\text{Prec}_*^{n,n-1}}{n}$$

Proof. We will prove that $\text{Prec}_*^{n,p}(\omega)$ is a decreasing sequence in p for a fixed n. By construction, $\mathcal{S}^{n,p}(\omega) \subset \mathcal{S}^{n,p-1}(\omega)$. Note that objective value of (P1) can be written as $x^\top P_{\mathcal{S}^{n,p}} x$, where $P_{\mathcal{S}^{n,p}}$ is the projection matrix for the subspace $\mathcal{S}^{n,p}$. For each $x \in \{\pm 1\}^n$ in the constraint set this value will monotonically decrease in p. Thus the optimal value will also decrease with p. This proves that $\text{Prec}_*^{n,p}$ is monotonically decreasing in p. ■

In the light of the preceding lemma we have that,

$$\inf_{1 \le p < n} \frac{\text{Prec}_*^{n,p}}{n} = \frac{\text{Prec}_*^{n,n-1}}{n} = \frac{h(\mathcal{S}^{n,n-1})^2}{n\sigma^2}. \tag{17}$$

In the last step we find the distribution of $\mathcal{S}^{n,n-1}$. For this purpose let us setup some more notation. Let $\mathcal{Q}^{\mathbf{1}} \subset \mathbb{R}^{n \times n}$ be the group of orthonormal matrices that leave the vector $\mathbf{1} \in \mathbb{R}^n$ invariant. In other words, it is a collection of matrices $Q \in \mathbb{R}^{n \times n}$ that satisfy,

$$QQ^\top = Q^\top Q = I,$$

and

$$Q\mathbf{1} = Q^\top \mathbf{1} = \mathbf{1}.$$

For any $\mathcal{S} \in \text{Gr}(k, \mathbb{R}^n)$, let $Q\mathcal{S} \triangleq \{Qx \mid x \in \mathcal{S}\}$. Let us also define $\mathcal{G}^{\mathbf{1}} \triangleq \{g \in \text{Gr}(1, \mathbb{R}^n) \mid \mathbf{1}^\top P_g \mathbf{1} = 0\}$.

Lemma 6. *$Q\mathcal{S}^{n,n-1}$ is distributed as $\mathcal{S}^{n,n-1}$, for any $Q \in \mathcal{Q}^{\mathbf{1}}$. There is a unique distribution on $\mathcal{G}^{\mathbf{1}}$ that has this invariance property. Further it has the same distribution as span$(\eta^n - \mathbf{1}\bar{\eta}^n)$*

with $\eta^n \sim N(0, I_n)$ *and* $\bar{\eta}^n = n^{-1}\mathbf{1}^\top \eta^n$. *Finally* $h(\mathcal{S}^{n,n-1})$ *has the same distribution as* $\|\eta^n - \mathbf{1}\bar{\eta^n}\|_1 / \|\eta^n - \mathbf{1}\bar{\eta^n}\|_2$

Proof. We first show that there is a unique probability distribution on $\mathcal{G}^{\mathbf{1}}$, say μ, such that $\mathcal{S}$ has the same distribution as $Q\mathcal{S}$ for any $Q \in \mathcal{Q}^{\mathbf{1}}$, if $\mathcal{S}$ is distributed as μ. For this purpose we use Theorem 4.1 of James (1954). $\mathcal{Q}^{\mathbf{1}}$ is a transitive compact topological group of transformations of $\mathcal{G}^{\mathbf{1}}$ to itself. Thus by the aforementioned theorem, there exists a unique measure that is invariant under transformations by $Q \in \mathcal{Q}^{\mathbf{1}}$.

Now we prove that $\text{span}(\eta^n - \mathbf{1}\bar{\eta^n})$ has the specified invariance property. First note that the covariance matrix of $\eta^n - \mathbf{1}\bar{\eta}^n$ is of the form $cI + d\mathbf{1}\mathbf{1}^\top$ for some $c, d \in \mathbb{R}$. Thus the covariance matrix of $Q(\eta^n - \frac{1}{n}\mathbf{1}^\top \eta^n)$ is $Q(cI + d\mathbf{1}\mathbf{1}^\top)Q^\top = cI + d\mathbf{1}\mathbf{1}^\top$. Since both of them are mean 0 and the same covariance matrix, $\text{span}(\eta^n - \mathbf{1}\bar{\eta}^n)$ and $\text{span}(Q(\eta^n - \mathbf{1}\bar{\eta}^n))$ have the same distribution.

By the uniqueness of this distribution μ, we have that $\text{span}(\eta^n - \mathbf{1}\bar{\eta^n})$ is indeed distributed as $\mathcal{S}^{n,n-1}$. ■

The previous lemma explicitly gives the distribution of $h(\mathcal{S}^{n,n-1})$. Using this, we prove an asymptotic property about $h(\mathcal{S}^{n,n-1})^2/n$.

Lemma 7.

$$\frac{h(\mathcal{S}^{n,n-1})^2}{n} \to \frac{1}{8\pi},$$

in distribution.

Proof. From Lemma 6 we have that, $h(\mathcal{S}^{n,n-1})^2$ has the same distribution as $\frac{\|\eta^n - \mathbf{1}\bar{\eta^n}\|_1^2}{\|\eta^n - \mathbf{1}\bar{\eta^n}\|_2^2}$, with $\eta^n \sim N(0, I_n)$. Further,

$$\begin{aligned}
\frac{\|\eta^n - \mathbf{1}\bar{\eta^n}\|_2^2}{n} &= \frac{1}{n}\sum_{i=1}^n (\eta_i^n - \bar{\eta}^n)^2 \\
&= \frac{1}{n}\sum_{i=1}^n ((\eta^n{}_i)^2 - 2\eta_i^n\bar{\eta^n} + \bar{\eta^n}^2) \\
&= \frac{1}{n}\sum_{i=1}^n (\eta^n{}_i)^2 - \frac{2}{n}\sum_{i=1}^n \eta_i^n\bar{\eta}^n + (\bar{\eta}^n)^2 \\
&= \frac{1}{n}\sum_{i=1}^n (\eta^n{}_i)^2 - (\bar{\eta}^n)^2
\end{aligned}$$

By strong law of large numbers we have that,

$$\frac{1}{n}\sum_{i=1}^n (\eta^n{}_i)^2 \to 1, \quad \text{almost surely,}$$

and,

$$(\bar{\eta}^n)^2 \to 0, \quad \text{almost surely.}$$

Thus,

$$\frac{1}{\sqrt{n}}\|\eta^n - \mathbf{1}\bar{\eta^n}\|_2 \to 1, \quad \text{a.s.} \tag{18}$$

Now we look at $\frac{1}{n}\|\eta^n - \mathbf{1}\bar{\eta^n}\|_1$. By triangle inequality,

$$\frac{1}{n}\|\eta^n\| + \frac{1}{n}\|\mathbf{1}\bar{\eta^n}\|_1 \geq \frac{1}{n}\|\eta^n - \mathbf{1}\bar{\eta^n}\|_1 \geq \frac{1}{n}\|\eta^n\|_1 - \frac{1}{n}\|\mathbf{1}\bar{\eta^n}\|_1$$

Now by the strong law of large numbers,

$$\frac{1}{n}\|\mathbf{1}\bar{\eta^n}\|_1 = |\bar{\eta^n}| \to 0, \quad \text{a.s.}$$

Thus, $\frac{1}{n}\|\eta^n - \mathbf{1}\bar{\eta^n}\|_1$ and $\frac{1}{n}\|\eta^n\|_1$ must have the same limit (if it exists). Again by, strong law of large numbers that,

$$\frac{1}{n}\sum_{i=1}^{n}|\eta_i^n| \to \mathsf{E}[|\xi|] = \frac{1}{2\sqrt{2\pi}},$$

where $\xi \sim N(0,1)$ is a standard normal. Thus,

$$\frac{1}{n}\|\eta^n - \mathbf{1}\bar{\eta^n}\|_1 \to \frac{1}{2\sqrt{2\pi}}. \tag{19}$$

From (18) and (19) and using Slutsky's lemma we have that,

$$\frac{\|\eta^n - \mathbf{1}\bar{\eta^n}\|_1}{\sqrt{n}\|\eta^n - \mathbf{1}\bar{\eta^n}\|_2} \to \frac{1}{2\sqrt{2\pi}}, \quad \text{almost surely.}$$

By continuity of $x \mapsto x^2$,

$$\frac{\|\eta^n - \mathbf{1}\bar{\eta^n}\|_1^2}{n\|\eta^n - \mathbf{1}\bar{\eta^n}\|_2^2} \to \frac{1}{8\pi}, \quad \text{almost surely.} \tag{20}$$

Finally by Equation (20) and the fact that $h(\mathcal{S}^{n,n-1})^2$ has the same distribution as $\frac{\|\eta^n - \mathbf{1}\bar{\eta^n}\|_1^2}{\|\eta^n - \mathbf{1}\bar{\eta^n}\|_2^2}$,

$$\frac{h(\mathcal{S}^{n,n-1})^2}{n} \to \frac{1}{8\pi},$$

in distribution. ■

Proof of Theorem 3. Using Lemmas 4 and 5, we have,

$$\frac{\mathrm{Prec}_*^{n,p_n}}{n} \geq \frac{\mathrm{Prec}_*^{n,n-1}}{n} = \frac{h(\mathcal{S}^{n-1})^2}{n\sigma^2}.$$

Finally using Lemma 7 we have,

$$\frac{h(\mathcal{S}^{n-1})^2}{n\sigma^2} \to \frac{1}{8\pi\sigma^2},$$

in distribution. Thus for any $\epsilon > 0$,

$$\mathsf{P}\left(\left|\frac{\mathrm{Prec}_*^{n,n-1}}{n} - \frac{1}{8\pi\sigma^2}\right| > \epsilon\right) \to 0.$$

Therefore,

$$\mathsf{P}\left(\frac{\mathrm{Prec}_*^{n,n-1}}{n} - \frac{1}{8\pi\sigma^2} < -\epsilon\right) \to 0.$$

Finally,

$$\mathsf{P}\left(\frac{\mathrm{Prec}_*^{n,p_n}}{n} - \frac{1}{8\pi\sigma^2} < -\epsilon\right) \to 0.$$

■

D. Approximation Guarantee for the Surrogate Problem

We assume without loss that $\Sigma = I$ and begin by establishing a corollary to a basic theorem from the non-asymptotic analysis of random matrices. Let us denote by Γ_n the matrix $\frac{1}{n}Z^\top Z$. Then we have the following approximation result:

Lemma 8. *Provided $n \geq \frac{L}{\epsilon^2}\max(p_n, l\log 2/\gamma_n)$, then with probability at least $1-\gamma_n$, we have*

$$\|\Gamma_n - I\| \leq \epsilon$$

where L and l are universal constants.

Proof. Let $Z_i^\top$ be a generic row of Z. We first observe that for any x satisfying $\|x\|_2^2 = 1$, we have

$$\mathsf{E}\left(x^\top Z_i\right)^2 = 1$$

so that the rows of Z are isotropic. Moreover, the sub-Gaussian norm of $x^\top Z_i$ is bounded, uniformly over all x of unit norm, by a universal constant (say, K). This fact follows from a calculation identical to that in equation 5.6 of Vershynin (2012). Consequently, we may apply Theorem 5.39 (specifically see equation 5.23) in Vershynin (2012), so that we have that with probability at least $1 - 2\exp(-c_K s^2)$,

$$\|\Gamma_n - I\| \leq C_K\sqrt{\frac{p_n}{n}} + \frac{s}{\sqrt{n}}$$

where $C_K(\triangleq C)$ and $c_K(\triangleq c)$ depend only on K, and can thus be taken as universal constants. Consequently, if $n \geq \max\left(\frac{4C^2 p_n}{\epsilon^2}, \frac{4\log 2/\gamma_n}{c\epsilon^2}\right)$, then we immediately have the result of the lemma by taking $s = \sqrt{\frac{\log 2/\gamma_n}{c}}$, $L = 4C^2$ and $l = \frac{1}{C^2 c}$. ■

Lemma 8 implies using Lemma 5.36 of Vershynin (2012) (or basic linear algebraic manipulations) that

$$1 - \epsilon \leq \sigma_{\min}\left(\frac{Z}{\sqrt{n}}\right) \leq \sigma_{\max}\left(\frac{Z}{\sqrt{n}}\right) \leq 1 + \epsilon \tag{21}$$

with probability at least $1-\gamma_n$. Here, $\sigma_{\max}$ and $\sigma_{\min}$ are, respectively, minimum and maximum singular values. Now, let us denote by $\hat{\mu}$ the measure over the sequence x_k induced by an optimal solution for the control problem (P3′) and let μ^* denote the measure induced by an optimal policy for our original dynamic optimization problem, (P3). We will demonstrate that an optimal solution to (P3′) is a near optimal solution to (P3). Before doing so, we establish some convenient notation: Denote

$$\bar{\Delta}_n = \begin{bmatrix} \delta_n \\ \Delta_n \end{bmatrix}$$

and recall

$$\Sigma_n \triangleq \frac{1}{n}\sum_{k=1}^{n} Z_{k,2:p_n} Z_{k,2:p_n}^\top$$

Proof of Theorem 4. Now, (21) is equivalently stated as:

$$1-\epsilon \leq \sqrt{\lambda_{\min}\left(\Gamma_n\right)} \leq \sqrt{\lambda_{\max}\left(\Gamma_n\right)} \leq 1+\epsilon,$$

with probability at least $1-\gamma_n$. This, in turn, implies that,

$$\frac{1}{1+\epsilon} \leq \sqrt{\lambda_{\min}\left(\Gamma_n^{-1}\right)} \leq \sqrt{\lambda_{\max}\left(\Gamma_n^{-1}\right)} \leq \frac{1}{1-\epsilon},$$

with probability at least $1-\gamma_n$. By the Courant-Fisher theorem (see, e.g., Horn and Johnson, 2012) we consequently have that,

$$\frac{\|\bar{\Delta}\|_2^2}{(1+\epsilon)^2} \leq \bar{\Delta}^\top \Gamma_n^{-1} \bar{\Delta} \leq \frac{\|\bar{\Delta}\|_2^2}{(1-\epsilon)^2}, \quad \forall\ \bar{\Delta} \in \mathbb{R}^{p_n}, \tag{22}$$

with probability at least $1-\gamma_n$.

Now note that

$$\bar{\Delta}^\top \Gamma_n^{-1} \bar{\Delta} = \|\bar{\Delta}\|_{\Gamma_n^{-1}}^2 \leq n^2, \tag{23}$$

for all feasible values of $\bar{\Delta} \in \mathbb{R}^{p_n}$. This follows from the non-negativity of the objective of (P2), which yields the inequality,

$$n - \frac{\|\bar{\Delta}\|_{\Gamma_n^{-1}}^2}{n} \geq 0.$$

Let A be the set of sample paths such that (22) holds. We have that,

$$\begin{aligned}
\mathsf{E}_{\hat{\mu}}\left[\left\|\bar{\Delta}_n\right\|^2_{\Gamma_n^{-1}}\right] &= \mathsf{E}_{\hat{\mu}}\left[\left\|\bar{\Delta}_n\right\|^2_{\Gamma_n^{-1}}\mathbb{I}_A + \left\|\bar{\Delta}_n\right\|^2_{\Gamma_n^{-1}}\mathbb{I}_{A^c}\right] \\
&\leq \frac{\mathsf{E}_{\hat{\mu}}\left[\left\|\bar{\Delta}_n\right\|^2_2\right]}{(1-\epsilon)^2} + n^2\mathsf{E}_{\hat{\mu}}\left[\mathbb{I}_{A^c}\right] \\
&\leq \frac{\mathsf{E}_{\hat{\mu}}\left[\left\|\bar{\Delta}_n\right\|^2_2\right]}{(1-\epsilon)^2} + \gamma_n n^2 \\
&\leq \frac{\mathsf{E}_{\mu^*}\left[\left\|\bar{\Delta}_n\right\|^2_2\right]}{(1-\epsilon)^2} + \gamma_n n^2
\end{aligned}$$

where the first inequality follows from the right hand side of (22) applied to each sample path in A and (23) applied to sample paths in A^c. The final inequality follows from the optimality of $\hat{\mu}$ for (P3′). We will now show that

$$\mathsf{E}_{\mu^*}\left[\left\|\bar{\Delta}_n\right\|^2_2\right] \leq (1+\epsilon)^2\mathsf{E}_{\mu^*}\left[\left\|\bar{\Delta}_n\right\|^2_{\Gamma_n^{-1}}\right] + n^2 p_n\gamma_n + O\left(\sqrt{\frac{n}{p_n-1}}\right)$$

together with the inequality above, this will yield the theorem. To prove this inequality, we first observe (as we did earlier) that on the set where (22) holds, i.e., the set A, $\left\|\bar{\Delta}_n\right\|^2_2 \leq (1+\epsilon)^2\left\|\bar{\Delta}_n\right\|^2_{\Gamma_n^{-1}}$. Thus,

$$\mathsf{E}_{\mu^*}\left[\left\|\bar{\Delta}_n\right\|^2_2\right] \leq (1+\epsilon)^2\mathsf{E}_{\mu^*}\left[\left\|\bar{\Delta}_n\right\|^2_{\Gamma_n^{-1}}\right] + \mathsf{E}_{\mu^*}\left[\left\|\bar{\Delta}_n\right\|^2_2\mathbb{I}_{A^c}.\right]$$

Now note that,

$$\left\|\bar{\Delta}_n\right\|^2_2 = \delta_n^2 + \|\Delta_n\|_2^2 \leq n^2 + \|\Delta_n\|_2^2\,.$$

The inequality follows since $|\delta_n| \leq n$. Thus,

$$\begin{aligned}
\mathsf{E}_{\mu^*}\left[\left\|\bar{\Delta}_n\right\|^2_2\mathbb{I}_{A^c}\right] &= n^2\gamma_n + \mathsf{E}_{\mu^*}\left[\|\Delta_n\|_2^2\,\mathbb{I}_{A^c}\right] \\
&\leq n^2\gamma_n + \mathsf{E}_{\mu^*}\left[\|\Delta_n\|_2^2\,\mathbb{I}_{\left\{\|\Delta_n\|_2^2 \geq \alpha_n(\gamma_n)\right\}}\right].
\end{aligned}$$

where $\alpha_n(\gamma_n)$ satisfies $\mathsf{P}_{\mu^*}\left(\|\Delta_n\|^2 \geq \alpha_n(\gamma_n)\right) = \gamma_n$. Applying Lemma 11 yields

$$\mathsf{E}_{\mu^*}\left[\|\Delta_n\|_2^2\,\mathbf{1}_{\|\Delta_n\|_2^2 \geq \alpha_n(\gamma_n)}\right] \leq n^2(p_n-1)\gamma_n + O\left(\sqrt{\frac{n}{p_n-1}}\right).$$

which yields the result. ■

To complete our proof of the theorem above, we must provide an upper bound on the quantity

$$\mathsf{E}_{\mu^*}\left[\|\Delta_n\|^2\,\mathbf{1}_{\|\Delta_n\|^2 \geq \alpha_n(\gamma_n)}\right]$$

where $\alpha_n(\gamma_n)$ satisfies $\mathsf{P}_{\mu^*}\left(\|\Delta_n\|^2 \geq \alpha_n(\gamma_n)\right) = \gamma_n$. In other words $\alpha_n(\gamma_n)$ is the γ_n percentile of $\|\Delta_n\|^2$. Let $\bar{Z}_n$ be a Gamma$(n(p_n-1)/2, 1)$ random variable, and let $\hat{\alpha}_n(\gamma_n)$ satisfy

$$\mathsf{P}\left(\bar{Z}_n \geq \hat{\alpha}_n(\gamma_n)\right) = \rho.$$

We have

Lemma 9.

$$\mathsf{E}_{\mu^*}\left[\|\Delta_n\|_2^2 \mathbf{1}_{\|\Delta_n\|_2^2 \geq \alpha_n(\gamma_n)}\right] \leq 2n\mathsf{E}\left[\bar{Z}_n \mathbf{1}_{\bar{Z}_n \geq \hat{\alpha}_n(\gamma_n)}\right]$$

Proof. Observe that

$$\begin{aligned}
\|\Delta_n\|_2^2 &= \left\|\sum_{k=1}^n x_k Z_{k,2:p_n}\right\|_2^2 \\
&\leq \left(\sum_{k=1}^n \|Z_{k,2:p_n}\|_2\right)^2 \\
&\leq n\sum_{k=1}^n \|Z_{k,2:p_n}\|_2^2 .
\end{aligned}$$

where the first inequality follows from the triangle inequality and the second from Cauchy-Schwartz. We then immediately have that

$$\mathsf{E}_{\mu^*}\left[\|\Delta_n\|_2^2 \mathbf{1}_{\|\Delta_n\|_2^2 \geq \alpha_n(\gamma_n)}\right] \leq n\mathsf{E}\left[\left(\sum_{k=1}^n \|Z_{k,2:p_n}\|_2^2\right) \mathbf{1}_{\sum_{k=1}^n \|Z_{k,2:p_n}\|_2^2 \geq \hat{\alpha}_n(\gamma_n)}\right].$$

But $\frac{1}{2}\sum_{k=1}^n \|Z_{k,2:p_n}\|_2^2 \triangleq \bar{Z}$ is distributed as a Gamma$(n(p_n-1)/2, 1)$ random variable and the claim follows. ■

Now Gamma random variables enjoy the following property on their tails:

Lemma 10. *If $\bar{Z} \sim$ Gamma$(k, 1)$ and $z(\gamma_k)$ is its ρth quantile (i.e., $z(\gamma_k)$ satisfies $\mathbb{P}(\bar{Z} \geq z(\gamma_k)) = \gamma_k$), then as $k \to \infty$,*

$$\mathsf{E}\left[\bar{Z}\mathbf{1}_{\bar{Z} \geq z(\gamma_k)}\right] \leq k\gamma_k + O\left(\frac{1}{\sqrt{k}}\right).$$

Proof. We have:

$$\begin{aligned}
\mathsf{E}\left[\bar{Z}\mathbf{1}_{\bar{Z}\geq z(\gamma_k)}\right] &= \int_{z(\gamma_k)}^{\infty} z\frac{z^{k-1}\exp(-z)}{\Gamma(k)}dz \\
&= \frac{\Gamma(k+1)}{\Gamma(k)}\int_{z(\gamma_k)}^{\infty}\frac{z^k\exp(-z)}{\Gamma(k+1)}dz \\
&= k\left[\frac{\Gamma(k+1,z(\gamma_k))}{k\Gamma(k)}\right] \\
&= k\left[\frac{k\Gamma(k,z(\gamma_k))+z(\gamma_k)^k\exp(-z(\gamma_k)}{k\Gamma(k)}\right] \\
&= k\left[\gamma_k+\frac{z(\gamma_k)^k\exp(-z(\gamma_k))}{k\Gamma(k)}\right]
\end{aligned}$$

where $\Gamma(\cdot,\cdot)$ is the right incomplete Gamma function. The final equality uses the fact that

$$\frac{\Gamma(k,z(\gamma_k))}{\Gamma(k)} = \gamma_k$$

by the definition of $z(\gamma_k)$. But $z^k\exp(-z)/(k\Gamma(k))$ is maximized at $z=k$, so that

$$\frac{z(\gamma_k)^k\exp(-z(\gamma_k))}{k\Gamma(k)} \leq \frac{k^k\exp(-k)}{k\Gamma(k)} = O\left(\frac{1}{k^{3/2}}\right)$$

where we have used Stirling's approximation for $\Gamma(k)$. The result follows. ■

We anticipate that tighter control on the big-oh error term is possible in the above proof, but this level of crudeness suffices. Using the preceding two lemmas now immediately yields:

Lemma 11.

$$\mathsf{E}_{\mu^*}\left[\|\Delta_n\|^2\mathbf{1}_{\|\Delta_n\|^2\geq\alpha_n(\gamma_n)}\right] \leq n^2(p_n-1)\rho + O\left(\sqrt{\frac{n}{p_n-1}}\right)$$

E. Dynamic Programming Formulation

Proof of Proposition 2. Consider the following n step Markov decision process (MDP):

1. The state at time k, $S_k=(\delta_{k-1},\Delta_{k-1},Z_k)$. The terminal state $S_n=(\delta_n,\Delta_n)$. The state space is $\mathcal{X}_k=\mathbb{R}^{2p}$ for non terminal time periods and is $\mathcal{X}_n=\mathbb{R}^p$ for the terminal time period.
2. The set of actions available to us is $\{\pm 1\}$.
3. At state S_k if action a_k is chosen, the state S_{k+1} is given by $(\delta_{k-1}+a_k,\Delta_{k-1}+a_kZ_{k,2:p},Z_{k+1})$. After n actions, the terminal state is $S_{n+1}=(\delta_n,\Delta_n)$.
4. There is no per step reward and the terminal reward is $S_{n+1}\mapsto\delta_n^2+\|\Delta_n\|^2_{\Sigma^{-1}}$.

Note that the MDP is finite horizon and the set of actions available at any point of time is finite, in particular 2. The problem (P3′) is just a terminal cost minimization MDP. It follows from

Proposition 4.2.1 in Bertsekas (2013) that a policy x^* that achieves the minimum expected cost. Further there exists a set of functions $J_k^* : \mathcal{X}_k \to \mathbb{R}$ such that $J_k^*(s_k)$ is the cost conditioned on $S_k = s_k$. Trivially,

$$J_{n+1}^*(\delta_n, \Delta_n) = \delta_n^2 + \|\Delta_n\|_{\Sigma^{-1}}^2.$$

These functions follow the recursion,

$$J_k^*(\delta_{k-1}, \Delta_{k-1}, Z_k) = \min_{u \in \{\pm 1\}} \mathsf{E}[J_{k+1}^*(\delta_{k-1} + u, \Delta_{k-1} + uZ_{k,2:p}, Z_{k+1})]. \tag{24}$$

Further x_k^*, the optimal policy, has the property that,

$$x_k^* \in \operatorname*{argmin}_{u \in \{\pm 1\}} \mathsf{E}[J_{k+1}^*(\delta_{k-1} + u, \Delta_{k-1} + uZ_{k,2:p}, Z_{k+1})]. \tag{25}$$

Let,

$$Q_k(\delta_k, \Delta_k) \triangleq \mathsf{E}\left[J_{k+1}^*(\delta_k, \Delta_k, Z_{k+1})\right]. \tag{26}$$

Using (24) and (26),

$$Q_k(\delta_k, \Delta_k) = \mathsf{E}\left[\min_{u \in \{\pm 1\}} Q_{k+1}(\delta_{k-1} + u, \Delta_{k-1} + uZ_{k,2:p})\right].$$

Further using (25) and (26),

$$x_k^* \in \operatorname*{argmin}_{u \in \{\pm 1\}} Q_k(\delta_{k-1} + u, \Delta_{k-1} + uZ_{k,2:p}).$$

This proves the dynamic programming proposition. ■

F. State Space Collapse

F.1. Proof of Theorem 5

In essence, the proof of Theorem 5 relies on the symmetry of the elliptical distribution for each covariate vector $Z_{k,2:p}$. In particular, for orthonormal matrix $Q \in \mathbb{R}^{p-1 \times p-1}$, $\Sigma^{-1/2} Z_{k,2:p}$ has the same distribution as $Q\Sigma^{-1/2} Z_{k,2:p}$. As a result of this spherical symmetry, under any non-anticipating policy, the distribution of the Mahalanobis distance $\|\Delta_{k+1}\|_{\Sigma^{-1}}$ at time $k+1$ is invariant across all Δ_k of a fixed Mahalanobis distance $\|\Delta_k\|_{\Sigma^{-1}}$ at time k. Thus, as opposed to having to maintain the p-dimensional state variable (δ_k, Δ_k), one merely needs to maintain the two-dimensional state variable $(\delta_k, \|\Delta_k\|_{\Sigma^{-1}})$.

To make this argument formal, we first define an inner product $\langle \cdot, \cdot \rangle_{\Sigma^{-1}}$ on $\mathbb{R}^{p-1}$ by

$$\langle \Delta, \Delta' \rangle_{\Sigma^{-1}} \triangleq \Delta^\top \Sigma^{-1} \Delta',$$

for $\Delta, \Delta' \in \mathbb{R}^{p-1}$. Using the symmetry of elliptical distribution, we can establish that:

Lemma 12. *Suppose $\Delta \in \mathbb{R}^{p-1}$ is a fixed $p-1$-dimensional vector and $X \sim \mathrm{Ell}(0, \Sigma, R)$ is an elliptically distributed $p-1$-dimensional random vector. Then,*

$$(\langle X, X\rangle_{\Sigma^{-1}}, \langle X, \Delta\rangle_{\Sigma^{-1}}) \stackrel{d}{=} \left(R^2, R\|\Delta\|_{\Sigma^{-1}} U_1\right).$$

In particular, when $X \sim N(0, \Sigma)$ has a Gaussian distribution, then,

$$(\langle X, X\rangle_{\Sigma^{-1}}, \langle X, \Delta\rangle_{\Sigma^{-1}}) \stackrel{d}{=} \left(\zeta^\top \zeta, \|\Delta\|_{\Sigma^{-1}} \zeta_1\right),$$

for an independent and normally distributed $p-1$-dimensional random vector $\zeta \sim N(0, I)$.

Proof. Since X follows the elliptical distribution,

$$X \stackrel{d}{=} R\Sigma^{1/2} U.$$

Thus,

$$\langle X, X\rangle_{\Sigma^{-1}} \stackrel{d}{=} R^2 U^\top \Sigma^{1/2} \Sigma^{-1} \Sigma^{1/2} U = R^2.$$

Also,

$$\langle X, \Delta\rangle_{\Sigma^{-1}} \stackrel{d}{=} R\Delta^\top \Sigma^{-1/2} U.$$

But, by the symmetry of the distribution of U, for any $h \in \mathbb{R}^{p-1}$, $h^\top U$ is has the same distribution as $\|h\|_2 U_1$. Due to independence of U and R, $(\langle X, X\rangle_{\Sigma^{-1}}, \langle X, \Delta\rangle_{\Sigma^{-1}})$ is distributed as $(R^2, R\|\Delta\|_{\Sigma^{-1}} U_1)$.

To prove the last statement, note that for the Gaussian case $(R, U) \sim (\|\zeta\|_2, \zeta/\|\zeta\|_2)$, if $\zeta \sim N(0, I)$. Thus,

$$(R^2, R\|\Delta\|_{\Sigma^{-1}} U_1) = \left(\|\zeta\|_2^2, \|\zeta\|_2 \|\Delta\|_{\Sigma^{-1}} e_1^\top \frac{\zeta}{\|\zeta\|_2}\right) = (\zeta^\top \zeta, \|\Delta\|_{\Sigma^{-1}} \zeta_1).$$

■

Now we are ready to prove the theorem.

Proof of Theorem 5. We will prove, by backward induction over $1 \le k \le n$, that

$$Q_k(\delta_k, \Delta_k) = q_k\left(\delta_k, \|\Delta_k\|_{\Sigma^{-1}}^2\right) \tag{27}$$

holds for all $\delta_k \in \mathbb{Z}$, $\Delta_k \in \mathbb{R}^{p-1}$. The result will then follow from Proposition 2.

Comparing (9) and (10), (27) clearly holds for $k = n$.

Now, assume that (27) holds for $k+1$. Then, from (9),

$$
\begin{aligned}
Q_k(\delta_k, \Delta_k) &= \mathsf{E}\left[\min_{u\in\{\pm 1\}} \; q_{k+1}\left(\delta_k + u, \|\Delta_k + uZ_{k+1,2:p}\|_{\Sigma^{-1}}^2\right)\right] \\
&= \mathsf{E}\left[\min_{u\in\{\pm 1\}} \; q_{k+1}\left(\delta_k + u, \|\Delta_k\|_{\Sigma^{-1}}^2 + \|Z_{k+1,2:p}\|_{\Sigma^{-1}}^2 + 2u\langle Z_{k+1,2:p}, \Delta_{k+1}\rangle_{\Sigma^{-1}}\right)\right] \\
&= \mathsf{E}\left[\min_{u\in\{\pm 1\}} \; q_{k+1}\left(\delta_k + u, \|\Delta_k\|_{\Sigma^{-1}}^2 + R^2 + 2uRe_1^\top U\|\Delta\|_{\Sigma^{-1}}\right)\right] \\
&\triangleq q_k\left(\delta_k, \|\Delta_k\|_{\Sigma^{-1}}^2\right).
\end{aligned}
\tag{28}
$$

The third equality follows from Lemma 12. ■

Finally, we prove Corollary 2.

Proof of Corollary 2. Following the proof of Theorem 5, we will simplify the expression for (28). In particular, using the final part of Lemma 12,

$$
\begin{aligned}
Q_k(\delta_k, \Delta_k) &= \mathsf{E}\left[\min_{u\in\{\pm 1\}} \; q_{k+1}^{\text{gauss}}\left(\delta_k + u, \|\Delta_k + uZ_{k+1,2:p}\|_{\Sigma^{-1}}^2\right)\right] \\
&= \mathsf{E}\left[\min_{u\in\{\pm 1\}} \; q_{k+1}^{\text{gauss}}\left(\delta_k + u, \|\Delta_k\|_{\Sigma^{-1}}^2 + R^2 + 2uRe_1^\top U\|\Delta\|_{\Sigma^{-1}}\right)\right] \\
&= \mathsf{E}\left[\min_{u\in\{\pm 1\}} \; q_{k+1}^{\text{gauss}}\left(\delta_k + u, \|\Delta_k\|_{\Sigma^{-1}}^2 + \zeta^\top\zeta + 2u\zeta_1\|\Delta\|_{\Sigma^{-1}}\right)\right] \\
&= \mathsf{E}\left[\min_{u\in\{\pm 1\}} \; q_{k+1}^{\text{gauss}}\left(\delta_k + u, \|\Delta_k\|_{\Sigma^{-1}}^2 + \xi + \eta^2 + 2u\eta\|\Delta\|_{\Sigma^{-1}}\right)\right] \\
&= \mathsf{E}\left[\min_{u\in\{\pm 1\}} \; q_{k+1}^{\text{gauss}}\left(\delta_k + u, (\|\Delta_k\|_{\Sigma^{-1}} + u\eta)^2 + \xi\right)\right].
\end{aligned}
$$

Here, $\xi \sim \chi_{p-2}^2$ if $p > 2$ and $\xi \triangleq 0$ if $p = 2$, and $\eta \sim N(0,1)$ are independent of each other. ■