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

October 11, 2015

PRELIMINARY DRAFT — DO NOT DISTRIBUTE

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 efficiency 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. We further characterize the gain in efficiency afforded by optimized allocations relative to randomized allocations, and show that this gain grows large as the number of covariates grow. Randomized trials are the mainstay of A-B testing in modern e-commerce applications, where the number of covariates is often comparable to the number of experimental trials; we believe that our results are of particular practical value in those settings.

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).

A natural approach to A-B testing is to independently, and with equal probability, assign each subject (or sample) to either the treatment or control groups. Following such a randomized

*The authors wish to thank Steve Chick, Shane Henderson, 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.
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 sample here can range from a patient in the clinical trial setting to a web-surfer or impression in the e-commerce setting. The notion of a treatment can range from an actual medical treatment in the clinical trial setting to an action such as showing 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 number of samples 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 Samples: In the clinical trial setting, the size of the sample set is limited for a number of 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 creatives it generates for an online advertising campaign. In this domain, a reduction in the number of ‘samples’ 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, or alternative explanatory variables that are unrelated to the choice of treatment. 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 on the account of 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, this number of samples 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’. More precisely, if one imagined a model where the outcome is additively impacted by the treatment and exogenous noise, we expect the number of samples 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, with standard deviation $\sigma$, and $\theta > 0$ is some constant, then by the CLT, we expect

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

suggesting that in order to differentiate a treatment effect $\theta$ from exogenous noise with standard deviation $\sigma$, we needs on the order of $\sigma^2/\theta^2$ samples.
Addressing these 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 ‘cancels’ the impact of covariates. This in turn would conceptually address the two challenges discussed above and yield an ‘efficient’ test.

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 balance to cancel the impact of covariates is by using ‘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. 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.

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 fashion. Again, there is a literature on dynamic allocation starting with seminal work by Efron (1971). This online variant of the A-B testing problem is also a well studied one and while the literature pertaining to the problem abounds with heuristics, an efficient to implement, optimal ‘solution’ remains unknown. 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.

The principle contribution of this paper is to present the first provably near-optimal algorithm for online A-B testing that applies to a canonical class of treatment-effect models. As a secondary contribution, we also show that the important ‘offline’ variant of the problem also admits an efficient optimal algorithm for the same canonical class of treatment-effect models 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 ubiquitous in the literature on experiment design. We consider the problem of minimizing the variance of our estimate of the treatment effect by optimally allocating subjects to either the treatment or control group, and designing an appropriate estimator. We formulate this problem as a dynamic optimization problem and make the following contributions:

1. We characterize the value of optimized allocations relative to randomized allocations and show that this value grows large as the number of covariates grows. In particular, we show that there is a lot to be gained from ‘optimizing’ the process of online A-B testing relative to the simple randomized trials that are the mainstay of modern A-B testing. We show that the gains achieved using optimization are more dramatic when the number of covariates is close to the number of trials, a scenario quite pertinent to the e-commerce setting, for example.

2. 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.

3. 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. Using real data on user impressions on Yahoo.com, we show that our algorithm does indeed yield impressive gains in efficiency even when covariates do not arise from an elliptical family.

Thus, our main contribution is providing an algorithm for the challenging problem of online 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. Given the vast extant literature on this problem, 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. Finally, Kallus (2012) 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 subjects response as a function of the treatments and subject covariates (we merely consider linear functions here). For the linear response functions we consider here, our offline problem may be viewed as a special case of the problem of \( D_{opt} \) optimal experiment design for which, unlike the general case, we can present an efficient algorithm.

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 problem, Efron (1971) proposed a seminal strategy that sought to ‘balance’ the number of subjects in each trial while minimizing certain types of selection bias. This approach was generalized to problems where a number of known covariates for each subject were available by a number of authors, of which Pocock and Simon (1975) present a widely cited and used heuristic. Kapelner and Krieger (2013) provide a sophisticated and recently proposed heuristic intended for the same setting. Atkinson (1982, 1999) made the first attempt to root the design of dynamic allocation methods in theory by connecting them with the notion of \( D_{opt} \) optimality in experiment design. Atkinson, however, made no attempt to formulate a dynamic optimization problem and the heuristics he proposed seem difficult to justify from a dynamic optimization perspective. Smith (1984a,b) considers a similar dynamic optimization formulation, which is also addressed via heuristics. Finally, Rosenberger and Sverdlov (2008) provides an excellent overview of the variety of heuristics that exist for the dynamic allocation problem. The model we consider is narrower than that required for some of the work above in its requirement that the dependence on covariates be linear; this distinction is not substantive as our work permits easy extensions to more complicated dependencies. A second distinction is that in order to formulate a rigorous optimization problem we focus on a single objective — the variance in our estimate of the treatment effect — whereas some of the aforementioned literature considers issues such as manipulation (selection bias) by experiment designers, etc.

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. Here we simply note that the experiment design setting is appropriate when an irrevocable decision of what treatment is appropriate must be made (e.g., use creative A as opposed to B for a display campaign),
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). From a methodology 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 experiment design setting.

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., 2015).

As a final note, the emphasis in our work is on A-B testing with a fixed budget on 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; Pekelis et al. (2015) have recently made exciting progress on this problem.

2. Model

In this section we describe the model in more detail. Given the model assumptions in Section 2.1, our problem is to find an estimator of the treatment effect that minimizes the variance of the estimate. In Section 2.2 we pose the two optimization problem that are of interest. One of them is the offline problem where all subjects can be observed before making the decisions and the other is the sequential problem where subjects must be allocated without knowing the future arrivals.

In Section 2.3 we give an interpretation to our objective and explain why the allocations prescribed by the optimization problems make intuitive sense. Finally, in Section 2.4 we give a simple upper bound to both the optimization problems. This shows that the efficiency (inverse variance) of the optimal allocation is $O(n)$.

2.1. Setup

We must learn the efficacy of a treatment by observing its effect on $n$ subjects. The $k$th subject is assigned a treatment $x_k \in \{\pm 1\}$. The $k$th 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 $k$th 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 $\theta$. The $\{\epsilon_k\}$ are i.i.d. zero mean random variables with variance $\sigma^2$. 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. In other words, there are no unobserved covariates.

Letting $Z \in \mathbb{R}^{n \times p}$ be the matrix whose $k$th 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 $\theta$ 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. In fact the scenario where $p \sim n$ is particularly relevant. For a particular allocation of treatments, $x$, let us denote by $\hat{\theta}_x$ the least squares estimator for $\theta$.

### 2.2. Optimization Problem

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

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

where $P_{Z^\perp} \triangleq I - Z(Z^\top Z)^{-1}Z^\top$. Details are presented in Section A of the appendix.

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

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

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

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:

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

where the expectation is over the distribution of the covariate process.
2.3. Problem Interpretation

Before moving on to algorithm design, we pause to interpret the offline and online problems present above. First we begin with an intuitive interpretation of the objective. Define the imbalance in covariate effects between the test and control groups, \( \Delta_n \in \mathbb{R}^p \), according to \( \Delta_n \triangleq \sum_{k=1}^{n} x_k Z_k = Z^\top x \). 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

\[
x^\top P_{Z^\perp} x = x^\top \left( I - Z (Z^\top Z)^{-1} Z^\top \right) x = n \left( 1 - \Delta_n^\top \Gamma_n^{-1} \Delta_n \right).
\]

Therefore, the offline problem (P1) is equivalent to minimizing the square of the weighted euclidean norm of \( \Delta_n \),

\[
\|\Delta_n\|_{\Gamma_n^{-1}}^2 \triangleq \Delta_n^\top \Gamma_n^{-1} \Delta_n
\]

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.

We next seek a statistical interpretation of the two problems. To this end, we note the Cramér-Rao bound dictates that, provided \( x \) and \( Z \) are independent of \( \epsilon \), and further if \( \epsilon \) 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 efficient 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.

2.4. Upper Bound on Efficiency

We end this section with an upper bound on the efficiency 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 \((\hat{\theta}, \hat{\kappa})\), including non-least squares estimators, we have:

\[
\text{Eff}(\hat{\theta}) \leq \frac{n}{\sigma^2},
\]

an upper bound on the optimal value of both problems (P1) and (P2). For non-Gaussian noise \( \epsilon \), this upper bound still holds for all least squares estimators.
This proposition shows that the efficiency of the optimal estimator 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 treatments. Further, the efficiency of such a design is $\sigma^2/n$, the optimal achievable efficiency. For $p > 1$ this efficiency is less than this value. Thus the presence of covariates only makes the inference challenging.

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

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

where $I(\theta, \kappa)$ is the Fisher information matrix. Under the Gaussian assumption for $\epsilon$ 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,

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

Thus,

$$\text{Eff}(\hat{\theta}) \leq \frac{x^\top (I - Z(Z^\top Z)^{-1}Z^\top)x}{\sigma^2} = \frac{n - x^\top Z(Z^\top Z)^{-1}Z^\top x}{\sigma^2} \leq \frac{n}{\sigma^2}.$$

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

The last statement is consequence of the fact that $x^\top (I - Z(Z^\top Z)^{-1}Z^\top)x / \sigma^2$ is the efficiency of the optimal least squares estimator (see Section A of the appendix).

3. The Offline 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 efficiency of the estimator of the treatment effect, as compared to randomization. To this end, we first obtain the mean efficiency of the randomized design. Surprisingly, efficiency 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 efficiency is $O(1)$. This can be contrasted with the upper bound on efficiency 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 efficiency 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

\[
\begin{align*}
(P1-SDP) \triangleq & \quad \text{maximize} \quad \text{tr} \left( PZ^\perp Y \right) \\
& \text{subject to} \quad Y_{kk} = 1, \quad \forall \ 1 \leq k \leq n, \\
& \quad Y \succeq 0, \\
& \quad Y \in \mathbb{R}^{n \times n}.
\end{align*}
\]

It is straightforward to see that (P1-SDP) is a relaxation of (P1): given \( x \in \{\pm 1\}^n \) feasible for (P1), define the symmetric positive definite matrix \( Y \triangleq xx^\top \in \mathbb{R}^{n \times n} \). Then, clearly \( Y \) is feasible for (P1-SDP), and the objective values for (P1) and (P1-SDP) coincide. Moreover, because it is an SDP, (P1-SDP) can be efficiently solved in polynomial time.

Building 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:

**Theorem 1.** Let \( Y \) be an optimal solution to (P1-SDP), with a matrix factorization \( Y = V^\top V \), for some matrix \( V \in \mathbb{R}^{n \times n} \) with columns \( v_1, \ldots, v_n \in \mathbb{R}^n \). Let \( u \in \mathbb{R}^n \) be a vector chosen at random uniformly over the unit sphere, and define the random vector \( \tilde{x} \in \{\pm 1\}^n \) by setting

\[
\tilde{x}_k \triangleq \begin{cases} 
+1 & \text{if } u^\top v_k \geq 0, \\
-1 & \text{if } u^\top v_k < 0,
\end{cases}
\]

for each \( 1 \leq k \leq n \). Then,

\[
E_u \left[ \tilde{x}^\top PZ^\perp \tilde{x} \right] \geq \frac{2}{\pi} \max_{x \in \{\pm 1\}^n} x^\top PZ^\perp x,
\]

where the expectation is taken over the choice of random vector \( u \). In other 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 \( x^\top Qx \), such that \( x \in \{\pm 1\}^n \), can be approximated within a relative error of \( \pi/2 \) using the prescribed algorithm, provided \( Q \) is positive semidefinite. Since \( PZ^\perp \) 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 \( \theta \).

**Theorem 2.** If \( n \) is even, given a covariate matrix \( Z \), define the expected efficiency of simple randomization

\[
\text{Eff}_{\text{rand}} \triangleq \mathbb{E}_{X_{\text{rand}}} \left[ \text{Var} \left( \hat{\theta}_{X_{\text{rand}}} \right)^{-1} \right],
\]

where the expectation is taken over the random allocation \( X_{\text{rand}} \). Then,

\[
\text{Eff}_{\text{rand}} = \frac{n}{\sigma^2} \left( 1 - \frac{p - 1}{n} \right).
\]

The proof relies on simple probabilistic arguments and is presented in Section B of the appendix. Surprisingly the efficiency 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 efficiency 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 \( \text{Eff}_{\text{rand}} \approx 1/\sigma^2 \), which is of constant order. In such a case, the least squares estimator \( \hat{\theta}_{X_{\text{rand}}} \) will not asymptotically converge to \( \theta \) as \( n \to \infty \).

Now we consider the performance of the optimal estimator that would be obtained by solving the offline experiment design problem (P1). By construction, the optimal estimator will clearly have efficiency 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 efficiency of the optimal estimator depends on the covariate matrix \( Z \). Moreover, it is difficult to obtain a closed-form expression for this efficiency as a function of \( Z \).

We can illustrate this with a simple example. Consider the case where \( p = n - 1 \). The efficiency of the optimal estimator 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 efficiency of the optimal estimator is

$$\sup_{x \in \{\pm 1\}^n} \frac{x^\top yy^\top x}{\sigma^2} = \sup_{x \in \{\pm 1\}^n} \frac{(y^\top x)^2}{\sigma^2} = \frac{||y||^2}{\sigma^2}. \quad (2)$$

Now, consider the following two cases:

1. $y$ has only two non-zero components given by $1/\sqrt{n}$ and $-1/\sqrt{n}$. In this case, the optimal efficiency 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 $1^\top y = 0$. In this case, the efficiency is $n/\sigma^2$. 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 $k$th 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 covariance matrix $Z = Z^{n,p_n}$, and let $\hat{\theta}_{x^*,Z^{n,p_n}}$ be the corresponding least squares estimator of $\theta$. Define the efficiency of this estimator by

$$\text{Eff}_{x^*,Z^{n,p_n}} \triangleq \text{Var}(\hat{\theta}_{x^*,Z^{n,p_n}})^{-1}.$$ 

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

$$\lim_{n \to \infty} \mathbb{P}\left( \frac{\text{Eff}_{x^*,Z^{n,p_n}}}{n} < \frac{1}{8\pi \sigma^2} - \epsilon \right) = 0,$$

where the probability is taken over the distribution of the covariance matrix.
Theorem 3 states that, with high probability, the optimal offline optimization-based design always yields $\Omega(n)$ efficiently under Assumption 2. Comparing to Theorem 2, the optimal efficiency is a $\Theta(n)$ relative improvement over that of simple randomization when $p$ is large (i.e., $p = \Omega(n)$). In other words, if the number of covariates is comparable to the number of samples, 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 efficiency 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 $Z^{n,n-1} \in \mathbb{R}^{n \times p}$ 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 $\text{Eff}^{n,p}_* \geq \text{Eff}^{n,n-1}_*$. This establishes that $p = n - 1$ corresponds to the worst case efficiency and allows us to focus on the sequence $\text{Eff}^{n,n-1}_*$. We then analyze the distribution of $Z^{n,n-1}$. We show that $\text{Eff}^{n,n-1}_*$ 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 efficiency in the case in closed-form using this distribution. In particular, we show that,

$$\frac{\text{Eff}^{n,n-1}_*}{n} \rightarrow \frac{1}{8\pi \sigma^2},$$

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 loss in performance when the surrogate problem is used to devise 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 \(k\)th 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

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

where

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

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 \(\Sigma_n\) correspond to sample estimates of the covariate mean and covariance structure, respectively.

We define, for each \(k\), the scalar \(\delta_k \in \mathbb{R}\) and the vector \(\Delta_k \in \mathbb{R}^{p-1}\) by

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

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

\[
(P3) \triangleq \minimize \quad \mathbb{E} \left[ \begin{bmatrix} \delta_n & \Delta_n^\top \end{bmatrix} \begin{bmatrix} 1 & M_n^\top \\ M_n & \Sigma_n \end{bmatrix}^{-1} \begin{bmatrix} \delta_n \\ \Delta_n \end{bmatrix} \right]
\]
subject to \( x \in \{\pm 1\}^n \), \( x_k \) is \( F_k \)-measurable, \( \forall 1 \leq k \leq n \).

Observe that, 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 \( \Sigma_n \) with the exact covariance matrix \( \Sigma \) and sets the sample mean \( M_n \) to the exact mean \( 0 \):

\[
(P3') \triangleq \minimize \quad \mathbb{E} \left[ \delta_n^2 + \|\Delta_n\|^2_{\Sigma^{-1}} \right]
\]
subject to \( x \in \{\pm 1\}^n \), \( x_k \) is \( F_k \)-measurable, \( \forall 1 \leq k \leq n \).

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 \( \delta_n \) and \( \Delta_n \) in the surrogate problem (P3') are intuitive: requiring \( \delta_n \) 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 \( \Delta_n \) will tend to 'balance' covariates — when \( \Delta_n \) 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 \( \mu^* \) denote the measure induced by an optimal policy for our original dynamic optimization problem (P3). Now, \( \delta_n \) and \( \Delta_n \) are random variables given an allocation policy. Given a allocation policy \( \mu \), define

\[
D_{\mu}^{n,p} \triangleq \mathbb{E}_{\mu} \left[ \begin{bmatrix} \delta_n & \Delta_n^\top \end{bmatrix} \Gamma_n^{-1} \begin{bmatrix} \delta_n \\ \Delta_n \end{bmatrix} \right]
\]
to be the objective value of (P3) under the allocation policy $\mu$ with sample size $n$ and covariate dimension $p$. The following result is demonstrated, without loss of generality, under the assumption that $\Sigma$ 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 $\rho > 0$ are real numbers such that, for $n$ sufficiently large, $n \geq L \max(p_n, \log 2/\rho_n)/\epsilon^2$. Then, as $n \to \infty$

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

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

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

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 Section E of the appendix.

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, which is the efficiency or inverse variance of the design corresponding to the policy used. In particular, denote by $\text{Eff}^{n,p}_\mu$ the expected efficiency of the estimator when allocations are made with a policy $\mu$, for a problem with sample size $n$ and covariate dimension $p$, i.e.,

$$\text{Eff}^{n,p}_\mu = \frac{\mathbb{E}_{\mu} \left[ x^T P_{Z', x} \right]}{\sigma^2} = \frac{n - D_{\mu, \Sigma}^{n,p_n}/n}{\sigma^2}.$$}

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

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

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

$$\frac{\text{Eff}^{n,p}_\mu}{\text{Eff}^{n,p}_{\mu^*}} \geq 1 - \frac{4\epsilon^3}{(L - \epsilon^2)(1 - \epsilon^2)} + o(1).$$

Corollary 1 gives the multiplicative loss in the efficiency 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 $\epsilon$. For small values of $\epsilon$ the ratio of efficiency 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 (4) in Theorem 4. This holds when

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

with $\rho_n = b/n^4$ 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 $\epsilon$, (4) holds as $n \to \infty$.

Using (5),

$$\text{Eff}^{n,p}_n - \text{Eff}^{\tilde{n},p}_n = \frac{D^{n,p}_\mu - D^{n,p}_\mu^*}{n\sigma^2} \leq \frac{(1+\epsilon)^2}{(1-\epsilon)^2} D^{n,p}_\mu^* - D^{n,p}_\mu + O\left(\sqrt{\frac{n}{p_n - 1}}\right)$$

$$= \frac{4\epsilon D^{n,p}_\mu}{n\sigma^2 (1-\epsilon)^2} + o(1)$$

$$= \frac{4\epsilon}{(1-\epsilon)^2} \left( \frac{n}{\sigma^2} - \text{Eff}^{n,p}_\mu^* \right) + o(1).$$

The first inequality follows from Theorem 4 and the last equality from (5).

Let \( \text{Eff}^{n,p}_\mu \) denote efficiency of the randomized policy. Using Theorem 2 and the optimality of \( \mu^* \), we have that

$$\frac{n}{\sigma^2} - \text{Eff}^{n,p}_\mu^* \leq \frac{n}{\sigma^2} - \text{Eff}^{n,p}_\text{rand} = \frac{n}{\sigma^2} \frac{p_n - 1}{n - 1} \leq \frac{n}{\sigma^2} \frac{p_n}{n} \leq \frac{\epsilon^2 n}{L\sigma^2},$$

where the last inequality uses the fact that, by hypothesis, $p_n/n \leq \epsilon^2/L$. Substituting this into (6), we get that

$$\text{Eff}^{n,p}_\mu - \text{Eff}^{n,p}_\tilde{\mu} \leq \frac{4\epsilon^3 n}{(1-\epsilon)^2 L\sigma^2} + o(1).$$

Now, using (7) we get that,

$$\text{Eff}^{n,p}_\mu^* \geq \frac{n}{\sigma^2} \left( 1 - \frac{\epsilon^2}{L} \right).$$
Thus, we have that,
\[
1 - \frac{\text{Eff}^{n-p}_n}{\text{Eff}^{n-p}_n} \leq \frac{4\epsilon n}{\text{Eff}^{n-p}_n (1 - \epsilon)^2 L \sigma^2} + o(1)
\]
\[
\leq \frac{4\epsilon^3}{(L - \epsilon^2)(1 - \epsilon^2)} + o(1).
\]
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 \((\delta_k, \Delta_k)\) can be interpreted as the state of the dynamic decision problem. In other words, given the past arrival sequence and actions, \((\delta_k, \Delta_k)\) 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 \(Q^k : \mathbb{R} \times \mathbb{R}^{p-1} \to \mathbb{R}\) by the Bellman equation

\[
Q^k(\delta_k, \Delta_k) \triangleq \begin{cases} 
\delta_n^2 + \|\Delta_n\|_2^2, & \text{if } k = n, \\
\mathbb{E}\left[ \min_{u \in \{\pm 1\}} Q^{k+1}(\delta_k + u, \Delta_k + Z_{k+1:2:p}) \right], & \text{if } 1 \leq k < n.
\end{cases}
\]

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

\[
x^*_k \in \arg\min_{u \in \{\pm 1\}} Q^k(\delta_{k-1} + u, \Delta_{k-1} + uZ_k).
\]

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

Proposition 2, whose proof is presented in Section D of the appendix, 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 \( \varphi: \mathbb{C}^m \rightarrow \mathbb{C} \) has the form

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

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 \( \Psi: \mathbb{C} \rightarrow \mathbb{C} \).

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 \( \mu \) and \( \Sigma \). 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 \( \mu \):

**Proposition 3.** If \( X \) has an elliptical distribution with parameters \( \mu \), \( \Sigma \), and \( \Psi \), 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 semi-definite 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 Ell(\( \mu \), \( \Sigma \), \( R \)). It can be shown that if \( R^2 \sim \chi_m^2 \) is a chi-squared distribution with \( m \) degrees of freedom, then Ell(\( \mu \), \( \Sigma \), \( R \)) is a Gaussian distribution with mean \( \mu \) and covariance \( \Sigma \). 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 \( 1 \leq p < n \), 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 \( k \)th 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 $\text{Ell}(0, \Sigma, R)$, where it is assumed that the random variable $R$ has finite second moment, and further that, without loss of generality, $\mathbb{E}[R^2] = p - 1$.

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 $\ell \geq 1$, define the function $q^{\ell,p}: \mathbb{Z} \times \mathbb{R}_+ \rightarrow \mathbb{R}$ according to

$$q^{\ell,p}(m, \lambda) \triangleq \begin{cases} m^2 + \lambda, & \text{if } \ell = 1, \\ \mathbb{E} \left[ \min_{u \in \{\pm 1\}} q^{\ell-1,p} \left( m + u, \lambda + 2RU_1\sqrt{\lambda} + R^2 \right) \right], & \text{if } \ell > 1. \end{cases}$$

Here, when $\ell > 1$, 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 $(P_3')$ is given by

$$Q^k(\delta_k, \Delta_k) = q^{n-k+1,p} \left( \delta_k, \|\Delta_k\|_{\Sigma^{-1}}^2 \right).$$

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

$$x^*_k \in \arg\min_{u \in \{\pm 1\}} q^{n-k+1,p} \left( \delta_{k-1} + u, \|\Delta_{k-1} + uZ_k\|_{\Sigma^{-1}}^2 \right).$$

Then, the sequence of allocations $x^*$ is optimal for the online experiment design problem $(P_3')$.

For the case of Gaussian distribution, the recursion (9) for solving the DP can be simplified according to the following corollary:

**Corollary 2.** If Assumption 4 holds, then, for $\ell \geq 1$ and $p \geq 2$, the functions $q^{\ell,p}_{\text{gauss}}: \mathbb{Z} \times \mathbb{R}_+ \rightarrow \mathbb{R}$ are given by

$$q^{\ell,p}_{\text{gauss}}(m, \lambda) \triangleq \begin{cases} m^2 + \lambda, & \text{if } \ell = 1, \\ \mathbb{E} \left[ \min_{u \in \{\pm 1\}} q^{\ell-1,p}_{\text{gauss}} \left( m + u, (\sqrt{\lambda} + u\eta)^2 + \xi \right) \right], & \text{if } \ell > 1. \end{cases}$$

Here, when $\ell > 1$, 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.

---

1. 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 $\mathbb{E}[R^2] = p - 1$ ensures that the covariance matrix of $Z_{k,2:p}$ is given by $\Sigma$.

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

---

20
We defer the proof of Theorem 5 and Corollary 2 until Section 4.5 in order to make several remarks:

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 (9) yields a two-dimensional 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 \(\lambda\) is \(O(n^2)\) thereby allowing us to discretize the state-space on a two-dimensional mesh. The functions \(\{q^{\ell,p}\}_{\ell \geq 1}\) can be numerically evaluated on this grid via backward induction. Note that since the expectation in (9) is over a two-dimensional random variable, it can be computed via numerical integration. Further details of this procedure are given in Section 5.

2. Moreover, the functions \(\{q^{\ell,p}\}_{\ell \geq 1}\) do not depend on the matrix \(\Sigma\) or the time horizon \(n\). 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 \(\Sigma\) is known. This is needed to compute the \(\| \cdot \|_{\Sigma^{-1}}\)-norm of \(\Delta_k\). In practice, \(\Sigma\) may not be known, and may need to be estimated from data. However, observe that \(\Sigma\) 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, \(\Sigma\) 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, an 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 \(\Sigma\).

4.5. Proof of Theorem 5

In essence, the proof of Theorem 5 relies on the symmetry of the elliptical distribution for each covariate vector \(Z_{k,2p}\). In particular, for orthonormal matrix \(Q \in \mathbb{R}^{p-1 \times p-1}\), \(\Sigma^{-1/2}Z_{k,2p}\) has the same distribution as \(Q\Sigma^{-1/2}Z_{k,2p}\). 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 \(\Delta_k\) of a fixed Mahalanobis distance \(\|\Delta_k\|_{\Sigma^{-1}}\) at time \(k\). Thus, as opposed to having to maintain the \(p\)-dimensional state variable \((\delta_k, \Delta_k)\), 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 1.** Suppose $\Delta \in \mathbb{R}^{p-1}$ is a fixed $p-1$-dimensional vector and $X \sim \text{Ell}(0, \Sigma, R)$ is an elliptically distributed $p-1$-dimensional random vector. Then,

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

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

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

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

**Proof.** Since $X$ follows the elliptical distribution,

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

Thus,

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

Also,

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

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 $\|h\|_2U_1$. 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 $(R, U)$ is distributed as $(\|\zeta\|_2, \zeta/\|\zeta\|_2)$, if $\zeta \sim \mathcal{N}(0, I)$. Thus,

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

\[\blacksquare\]

Now we are ready to prove the theorem.

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

$$Q^k(\delta_k, \Delta_k) = q^{n-k+1,p} \left( \delta_k, \|\Delta_k\|_{\Sigma^{-1}}^2 \right)$$

holds for all $\delta_k \in \mathbb{Z}, \Delta_k \in \mathbb{R}^{p-1}$. The result will then follow from Proposition 2.
Comparing (8) and (9), (12) clearly holds for $k = n$.

Now, assume that (12) holds for $k + 1$. Then, from (8),

$$Q^k(\delta_k, \Delta_k) = E \left[ \min_{u \in \{\pm 1\}} q^{n-k,p} \left( \delta_k + u, \|\Delta_k + uZ_{k+1,2p}\|_{\Sigma^{-1}}^2 \right) \right]$$

$$= E \left[ \min_{u \in \{\pm 1\}} q^{n-k,p} \left( \delta_k + u, \|\Delta_k\|_{\Sigma^{-1}}^2 + \|Z_{k+1,2p}\|_{\Sigma^{-1}}^2 + 2u(Z_{k+1,2p}, \Delta_{k+1})\Sigma^{-1} \right) \right]$$

$$= E \left[ \min_{u \in \{\pm 1\}} q^{n-k,p} \left( \delta_k + u, \|\Delta_k\|_{\Sigma^{-1}}^2 + R^2 + 2uRe_1^\top U\|\Delta\|_{\Sigma^{-1}} \right) \right]$$

$$\triangleq q^{n-k+1,p} \left( \delta_k, \|\Delta_k\|_{\Sigma^{-1}}^2 \right).$$

The third equality follows from Lemma 1. ■

Finally, we prove Corollary 2.

**Proof of Corollary 2.** Following the proof of Theorem 5, we will simplify the expression for (13). In particular, using the final part of Lemma 1,

$$Q^k(\delta_k, \Delta_k) = E \left[ \min_{u \in \{\pm 1\}} g_{gauss}^{n-k,p} \left( \delta_k + u, \|\Delta_k + uZ_{k+1,2p}\|_{\Sigma^{-1}}^2 \right) \right]$$

$$= E \left[ \min_{u \in \{\pm 1\}} g_{gauss}^{n-k,p} \left( \delta_k + u, \|\Delta_k\|_{\Sigma^{-1}}^2 + R^2 + 2uRe_1^\top U\|\Delta\|_{\Sigma^{-1}} \right) \right]$$

$$= E \left[ \min_{u \in \{\pm 1\}} g_{gauss}^{n-k,p} \left( \delta_k + u, \|\Delta_k\|_{\Sigma^{-1}}^2 + \xi^\top \xi + 2u\eta\|\Delta\|_{\Sigma^{-1}} \right) \right]$$

$$= E \left[ \min_{u \in \{\pm 1\}} g_{gauss}^{n-k,p} \left( \delta_k + u, (\|\Delta_k\|_{\Sigma^{-1}}^2 + u\eta)^2 + \xi \right) \right].$$

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

5. Experiments

This section focuses on numerical experiments with data. Our goal is to show the gain our algorithm for online allocation enjoys over randomization as one varies the two key parameters of interest — the number of covariates $p$ and the number of samples $n$; as such this complements our theoretical analysis on the value of optimization in the offline setting.

In Section 4 we showed that if the covariates follow a distribution that falls in the elliptical family then then sequential optimization problem can be solved using a 2-dimensional dynamic program. In real applications this might not be true. For example, some of the covariates might be discrete or binary which rules out elliptical distributions. In this section we show that policy
obtained by solving the 2-dimensional dynamic program gives a large improvement in efficiency over randomization in realistic scenarios.

5.1. Efficiency Gain

As will be discussed shortly in Section 5.3, we are principally be interested in comparing our dynamic programming allocation policy with randomization. Our performance metric of interest is the efficiency of these procedures. In the present setting, from (1), the expected efficiency of the dynamic programming policy is given by

$$\text{Eff}_{dp} = \mathbb{E} \left[ \frac{x_{dp}^\top P_z x_{dp}}{\sigma^2} \right],$$

where the allocations $x_{dp}$ are determined by the policy, and the expectation is taken over realizations of the covariate matrix $Z$ and the resulting allocations $x_{dp}$. On the other hand, the efficiency of randomization is known from Theorem 2 to be

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

The results in a relative efficiency of

$$\frac{\text{Eff}_{dp}}{\text{Eff}_{rand}} = \frac{\mathbb{E} \left[ x_{dp}^\top P_z x_{dp} \right]}{n \left( 1 - \frac{p - 1}{n - 1} \right)}.$$

We call this quantity the efficiency gain of the dynamic programming approach.

Observe that the efficiency gain depends only on the distribution of covariates. In particular, estimating the efficiency gain does not require observations of experimental outcomes $y_k$. From an empirical perspective, this is helpful since it means that we can assess the allocation policies given only access to covariate distributions, as opposed to requiring observations of outcomes which would necessitate running actual experimental trials under different allocation policies. Indeed, the efficiency gain provides a relative assessment that holds for any outcome which can be modelled via linear dependencies on the covariates and treatments.

5.2. 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 Data.** In this case we consider the ideal case scenario where $Z_{2:p}$ follow the Gaussian distribution which falls in the elliptical. In other words Assumption 2 is satisfied. For the covariance matrix $\Sigma$, we set $\Sigma_{ii} = 1.0$ and $\Sigma_{ij} = 0.1$ for $j \neq i$.

**Yahoo User Data.** To experiment on data from a more realistic setting, we use a dataset on user
click log on the Yahoo! The users here are visitors to ‘Featured Tab of the Today Module’ on the Yahoo! front page. In the dataset, each user has 136 features associated with it such as age and gender. Each feature is binary, taking the values in $\{0, 1\}$. Some of these features are either 0 or 1 throughout the data set, and we discard these since they do not provide any additional information. The data also seems to have duplicate columns, which we again discard since these are redundant.

Our algorithm as an input requires the covariance matrix of the data. 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 roughly 1 million data points kept aside.

Finally, for evaluation purposes, we need a generative model for the data. To this end, from a set of 1 million data points we sample users, with replacement. In other words, as the sampling distribution we use the empirical distribution of the 1 million data points used for testing. Such a sampling procedure is intended to mimic arrival of users on Yahoo! front page.

5.3. Algorithms

Dynamic Programming (Our Approach). We approach this sequential problem using our dynamic programming based approach from Section 4. As the approximation for the $p$-dimensional value functions, we use the 2-dimensional value functions given by $\{q_{\text{gauss}}^{\ell,p}\}_{\ell \geq 1}$. These functions depend only on $p$ and can be computed offline numerically. Here we provide the details of how these are evaluated. $\{q_{\text{gauss}}^{\ell,p}\}_{\ell \geq 1}$ are computed using backward induction. In particular, given $q_{\text{gauss}}^{\ell-1,p}$ we compute $q_{\text{gauss}}^{\ell,p}$ as follows:

1. Discretization: The first state variable $m$ is discrete and can take values from $-n$ to $n$. We consider values for the second state variable $\lambda$ on a mesh of of the interval $[0, M]$. $M$ is conservatively set at a high value, i.e., a value such that $\|\Delta_k\|_2^{-1}$ has a low probability of exceeding $M$. In our case, we choose $M = 2 \times 10^3$ and a mesh step-size of 0.2. Using this procedure we have a 2-dimensional grid on the state space of the dynamic program.

2. Sampling: At each point $(m, \lambda)$ such that $-n \leq m \leq n$ and $\lambda$ is the mesh of $[0, M]$, we estimate $q_{\text{gauss}}^{\ell,p}(m, \lambda)$ via Monte Carlo sampling. In particular, $N$ pairs $(\xi, \eta)$ are sampled from the given distribution and $q_{\text{gauss}}^{\ell,p}(m, \lambda)$ is estimated according to (9) using the corresponding empirical measure. We use the same sample set for all $(m, \lambda)$ at which this is evaluated. The number of sample points ($N$) is chosen to be 10,000.

3. Interpolation: In the right side of (9), estimates of the value of the function $q_{\text{gauss}}^{\ell-1,p}(m', \lambda')$ for various values of $(m', \lambda')$ are obtained by linearly interpolating the closest values of the ‘quantized’ function described in the previous step; i.e., $q_{\text{gauss}}^{\ell-1,p}(m', \lambda')$ linearly interpolates $q_{\text{gauss}}^{\ell-1,p}(m', \lambda_l)$ and $q_{\text{gauss}}^{\ell-1,p}(m', \lambda_r)$ where $\lambda_l \leq \lambda' \leq \lambda_r$ are closest to $\lambda'$ on the mesh used for computing $q_{\text{gauss}}^{\ell-1,p}$.

---

Footnote: 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](http://webscope.sandbox.yahoo.com/catalog.php?datatype=r)
Randomization. This is the algorithm described in Section 3.2 where half of the subjects are assigned to the control group, uniformly at random. In randomization we can perform all allocations up front at time $t = 0$. This way it can be considered as a sequential procedure since we do not use information from future arrivals to make decisions.

5.4. Results

As an evaluation criteria, we use the relative efficiency gain, given by (14). Here, the expectation in the numerator of (14) is approximated through Monte Carlo simulation, which requires sampling covariates $Z$ from the aforementioned Gaussian and Yahoo! user data generative models, and computing the resulting dynamic programming allocations.

The aim of the experiments is to explore the gains of optimization for various values of the number of trials $n$ and covariate dimension $p$. To this end, we plot the efficiency gain for values of $n$ in the range $[p + 1, 100]$. We repeat this for various values $p$ between 10 and 40. We perform 10,000 Monte Carlo trials for each combination of $p$ and $n$ to make sure that confidence intervals are narrow (standard errors for the efficiency gain are uniformly below 0.01).

The results are plotted in Figure 1 and Figure 2. We see that there are significant gains to be obtained by optimization, both for the Gaussian model and Yahoo! user data model. Further, this improvement grows as the ratio of $p/n$ gets closer to 1. The efficiency ratio is close to 3 in some cases. This is consistent with the theory of the offline optimal design studied in Section 3 where we saw that randomization gives $O(1)$ performance when $p \sim n$ and optimized designs are $\Omega(n)$. As $n$ increases the efficiency gain decreases. However, even for the extreme case of $p = 10$ and
Figure 2: Efficiency gain (the ratio of the efficiencies of the optimal and the randomized design) as a function of the sample size \( n \) and the covariate dimension \( p \) for the Yahoo! dataset.

\( n = 100 \), there is close to 10% improvement in efficiency for both data sets.

6. 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 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 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.

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. Derivation of the Optimization Problem

Here derive the expression for efficiency used in Section 2. Denote the matrix
\[ X \triangleq [x \ Z] \]
and
\[ \beta \triangleq [\theta \ \kappa^\top] \].
Thus our model is
\[ y = X\beta + \epsilon. \]

The least squares estimate \( \hat{\beta} \) of \( \beta \) 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,
\[ \text{Var}(\hat{\beta}) = (X^\top X)^{-1}X^\top \text{Var}(\epsilon\epsilon^\top)X(X^\top X)^{-1} = \sigma^2 (X^\top X)^{-1}. \]

Thus variance of \( \hat{\theta} = \hat{\beta}_1 \) is
\[ \text{Var}(\hat{\theta}) = \sigma^2 e_1^\top (X^\top X)^{-1} e_1 = \sigma^2 e_1^\top \begin{bmatrix} x^\top x & x^\top Z \\ Z^\top x & Z^\top Z \end{bmatrix}^{-1} e_1 = \frac{\sigma^2}{x^\top (I - Z(Z^\top Z)^{-1}Z^\top)x}. \]

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

B. Performance of the Randomized Algorithm

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

Lemma 2. 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,
\[ \begin{bmatrix} 1 & a^\top \\ a & Q \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ a \end{bmatrix} = 1. \]

Proof. By matrix inversion lemma,
\[ \begin{bmatrix} 1 & a^\top \\ a & Q \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{\rho} & -\frac{a^\top Q^{-1}}{\rho} \\ -\frac{Q^{-1}a}{\rho} & Q^{-1} \end{bmatrix} \]
\[ = \begin{bmatrix} \frac{1}{\rho} & -\frac{a^\top Q^{-1}}{\rho} \\ -\frac{Q^{-1}a}{\rho} & Q^{-1} + \frac{\rho}{1+\kappa a^\top Q^{-1} a} \end{bmatrix}, \]
where,
\[ \rho \triangleq 1 - a^\top Q^{-1} a. \]
Thus,
\[
\begin{bmatrix}
1 & a^T \\
a & Q
\end{bmatrix}
\begin{bmatrix}
1 & a^T \\
a & Q
\end{bmatrix}^{-1}
\begin{bmatrix}
1 \\
a
\end{bmatrix}
= \begin{bmatrix}
1 & a^T \\
a & Q
\end{bmatrix}
\begin{bmatrix}
\frac{1}{\rho} - Q^{-1}a \\
-Q^{-1}a + \frac{(a^T Q^{-1}a)^2}{\rho}
\end{bmatrix}
\begin{bmatrix}
1 \\
a
\end{bmatrix}
= \frac{1}{\rho} - 2a^T Q^{-1}a + a^T Q^{-1}a + \frac{(a^T Q^{-1}a)^2}{\rho}
= \frac{1}{\rho} - 2(1 - \rho) + 1 - \rho + \frac{(1 - \rho)^2}{\rho}
= 1.
\]

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

**Proof of Theorem 2.** Let $X_{\text{rand}}$ be the allocation vector of the randomized allocation. Let $S \subset \{1, \ldots, n\}$ a random subset such that $S = \{i \leq n | X_{\text{rand},i} = 1\}$.

The efficiency of the estimator of $\theta$ is,
\[
\text{Var}\left(\hat{\theta}_{X_{\text{rand}}}\right)^{-1} \triangleq \sigma^{-2} X_{\text{rand}}^T \left( I - Z(Z^T Z)^{-1} Z^T \right) X_{\text{rand}}
= \sigma^{-2} \left( n - X_{\text{rand}}^T Z(Z^T Z)^{-1} Z^T X_{\text{rand}} \right).
\]

Let us define,
\[
\bar{Z}_S = \frac{2}{n} \sum_{k \in S} Z_k,
\]
and,
\[
\bar{Z}_{S^c} = \frac{2}{n} \sum_{k \notin S} Z_k,
\]
the mean covariates in $S$ and $S^c$ respectively. Also let,
\[
\bar{V} \triangleq \frac{1}{n} \sum_{1 \leq i \leq n} Z_i Z_i^T,
\]
and
\[
\bar{Z} \triangleq \frac{1}{n} \sum_{1 \leq i \leq n} Z_i.
\]
Thus,
\[
Z^T Z = n\bar{V}.
\]
We have that,
\[
\text{Var}(\hat{\theta}_{X_{\text{rand}}})^{-1} = \sigma^{-2} \left( n - X_{\text{rand}}^\top Z(Z^\top Z)^{-1} X_{\text{rand}} \right)
\]
\[
= \sigma^{-2} \left( n - \left( \sum_{i \in S} Z_i - \sum_{i \notin S} Z_i \right) \right) \left( n \bar{V} \right)^{-1} \left( \sum_{i \in S} Z_i - \sum_{i \notin S} Z_i \right)
\]
\[
= \sigma^{-2} \left( n - \left( \frac{n}{2} (\bar{Z}_S - \bar{Z}_{S^c}) \right)^\top \left( \frac{n}{2} (\bar{Z}_S - \bar{Z}_{S^c}) \right) \right)
\]
\[
= \frac{n}{\sigma^2} \left( 1 - \frac{1}{4} (\bar{Z}_S - \bar{Z}_{S^c})^\top \bar{V}^{-1} \left( \bar{Z}_S - \bar{Z}_{S^c} \right) \right)
\]

Now note that,
\[
\bar{Z}_S - \bar{Z}_{S^c} = 2(\bar{Z}_S - \bar{Z}),
\]
thus,
\[
\text{Var}(\hat{\theta}_{X_{\text{rand}}})^{-1} = \frac{n}{\sigma^2} \left( 1 - (\bar{Z}_S - \bar{Z})^\top \bar{V}^{-1} (\bar{Z}_S - \bar{Z}) \right)
\]
\[
= \frac{n}{\sigma^2} \left( 1 - \left( \frac{2}{n} \sum_{k \in S} (Z_k - \bar{Z}) \right) \right) \left( \bar{V}^{-1} \left( \frac{2}{n} \sum_{k \in S} (Z_k - \bar{Z}) \right) \right)
\]
\[
= \frac{n}{\sigma^2} \left( 1 - \frac{4}{n^2} \sum_{k,l \in S} (Z_k - \bar{Z})^\top \bar{V}^{-1} (Z_l - \bar{Z}) \right)
\]
\[
= \frac{n}{\sigma^2} \left( 1 - \frac{4}{n^2} \sum_{k,l=1} \sum_{k \neq l} (Z_k - \bar{Z})^\top \bar{V}^{-1} (Z_l - \bar{Z}) \right)
\]

Thus,
\[
E \left[ \text{Var}(\hat{\theta}_{X_{\text{rand}}})^{-1} \right] = \frac{n}{\sigma^2} \left( 1 - \frac{4}{n^2} E \left[ \sum_{k,l=1} \sum_{k \neq l} (Z_k - \bar{Z})^\top \bar{V}^{-1} (Z_l - \bar{Z}) \right] \right)
\]
\[
= \frac{n}{\sigma^2} \left( 1 - \frac{4}{n^2} \sum_{k=1}^n (Z_k - \bar{Z})^\top \bar{V}^{-1} (Z_k - \bar{Z}) \mathbb{P}(k \in S)
\right.
\]
\[
+ \sum_{k=1}^n \sum_{l=1}^n \sum_{l \neq k} (Z_k - \bar{Z})^\top \bar{V}^{-1} (Z_l - \bar{Z}) \mathbb{P}(k,l \in S) \right)
\]

But we observe that,
\[
\sum_{k=1}^n \sum_{l=1}^n \sum_{l \neq k} (Z_k - \bar{Z})^\top \bar{V}^{-1} (Z_l - \bar{Z}) = -(Z_k - \bar{Z})^\top \bar{V}^{-1} (Z_k - \bar{Z}).
\]

Also,
\[
\mathbb{P}(k \in S) = \frac{1}{2},
\]
and,
\[
\mathbb{P}(k,l \in S) = \frac{n/2 - 1}{2(n - 1)} = \frac{n - 2}{4(n - 1)}.
\]
Let us define,
\[
\tilde{C} \triangleq \frac{1}{n} \sum_{i=1}^{n} (Z_i - \bar{Z})(Z_i - \bar{Z})^\top.
\]
Substituting in the earlier equation we get,
\[
\text{Var} \left( \hat{\theta}_{\text{rand}} \right)^{-1} = \frac{n}{\sigma^2} \left( 1 - \frac{4}{n^2} \left[ \left( \frac{1}{2} - \frac{n-2}{4(n-1)} \right) \sum_{k=1}^{n} (Z_k - \bar{Z})^\top \bar{V}^{-1}(Z_k - \bar{Z}) \right] \right)
= \frac{n}{\sigma^2} \left( 1 - \frac{1}{n(n-1)} \sum_{k=1}^{n} (Z_k - \bar{Z})^\top \bar{V}^{-1}(Z_k - \bar{Z}) \right)
= \frac{n}{\sigma^2} \left( 1 - \frac{1}{n(n-1)} \sum_{k=1}^{n} \text{tr}(\bar{V}^{-1}(Z_k - \bar{Z})(Z_k - \bar{Z})^\top) \right)
= \frac{n}{\sigma^2} \left( 1 - \frac{\text{tr}(\bar{V}^{-1} \bar{C})}{n-1} \right)
= \frac{n}{\sigma^2} \left( 1 - \frac{\text{tr}(\bar{V}^{-1}(\bar{V} - \bar{Z}^\top \bar{Z}) \bar{Z})}{n-1} \right)
= \frac{n}{\sigma^2} \left( 1 - \frac{p-\bar{Z}^\top \bar{V}^{-1} \bar{Z}}{n-1} \right)
\]
Finally, using Lemma 2 we get that
\[
\bar{Z}^\top \bar{V}^{-1} \bar{Z} = 1.
\]
Thus we obtain the result. 

\[\square\]

**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 $(\Omega, \mathcal{F}, \mathbb{P})$. 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 $\text{Gr}(k, \mathbb{R}^n)$ be 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 $\mathcal{S}^{n,p} \subseteq \bigcup_{k=n-p}^{p} \text{Gr}(k, \mathbb{R}^n)$ 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, $\mathcal{S}^{n,p} \in \text{Gr}(n-p, \mathbb{R}^n)$ almost surely.

**Proof.** We can prove this inductively. Since $Z^{n,1} = 1$, 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 $\text{span}(Z^{n,p-1})$. The $p$th column of $Z^{n,p}$ is non-degenerate Gaussian vector independent of $\text{span}(Z^{n,p-1})$, call it $Z^p$. $P(Z^p \in \text{span}(Z^{n,p-1}) = 0$. Thus, almost surely, $Z^{n,p}$ is of rank $p$. 

\[\square\]
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 efficiency of the optimal estimator for $p = n - 1$ in terms of $\mathcal{S}_{n,n-1}$. Let $A = \{\omega \in \Omega : \mathcal{S}_{n-1}^{\perp}(\omega) \in \text{Gr}(1, \mathbb{R}^n)\}$. From now on, we assume $\Omega = A$ and all subsequent statements hold with probability one.

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

**Lemma 4.** Then efficiency 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 efficiency for $n = p - 1$ is given by $\sigma^{-2} x^* \top P_{\mathcal{S}_{n,n-1}^\perp} x^*$, where $x^*$ is the assignment that maximizes (P1). Now not that, $P_{\mathcal{S}_{n,n-1}^\perp}$ 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,

$$\text{maximize } x^\top yy^\top x = \frac{(x^\top y)^2}{\|y\|_2^2}$$
subject to $x \in \{-1,1\}^n$.

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

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

almost surely. ■

Using the fact that we have all the $\mathcal{S}_{n,p}$s on the same probability space, it is easy to show that the efficiency monotonically decreases as $p$ grows, for a fixed $n$.

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

$$\inf_{1 \leq p < n} \frac{\text{Eff}_{n,p}^*}{n} = \lim_{p \to n} \frac{\text{Eff}_{n,p}^*}{n} = \frac{\text{Eff}_{n,n-1}^*}{n} = \frac{h(\mathcal{S}_{n,n-1})^2}{n\sigma^2}.$$

**Proof.** We will prove that $\text{Eff}_{n,p}^*(\omega)$ 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 $x \in \{-1,1\}^n$ in the constraint set this value will monotonically decrease in $p$. Thus the optimal value will also decrease with $p$. This proves that $\text{Eff}_{n,p}^*$ is monotonically decreasing in $p$. ■

In the light of the preceding lemma we have that,

$$\inf_{1 \leq p < n} \frac{\text{Eff}_{n,p}^*}{n} = \lim_{p \to n} \frac{\text{Eff}_{n,p}^*}{n} = \frac{h(\mathcal{S}_{n,n-1})^2}{n\sigma^2}.$$
In the last step we find the distribution of $S_{n,n-1}$. For this purpose let us setup some more notation. Let $Q^1 \subset \mathbb{R}^{n \times n}$ be the group of orthonormal matrices that leaves $1$ 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

$$Q1 = Q^\top 1 = 1.$$

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

**Lemma 6.** $QS_{n,n-1}$ is distributed as $S_{n,n-1}$, for any $Q \in Q^1$. There is a unique distribution on $G^1$ that has this invariance property. Further it has the same distribution as $\text{span}(\eta^n - 1\bar{\eta}^n)$ with $\eta^n \sim N(0, I_n)$ and $\bar{\eta}^n = n^{-1}1^\top \eta^n$. Finally $h(S_{n,n-1})$ has the same distribution as $\|\eta^n - 1\bar{\eta}^n\|_1/\|\eta^n - 1\bar{\eta}^n\|_2$

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

Now we prove that $\text{span}(\eta^n - 1\bar{\eta}^n)$ has the specified invariance property. First note that the covariance matrix of $\eta^n - 1\bar{\eta}^n$ is of the form $cI + d11^\top$ for some $c,d \in \mathbb{R}$. Thus the covariance matrix of $Q(\eta^n - 1\bar{\eta}^n)$ is $Q(cI + d11^\top)Q^\top = cI + d11^\top$. Since both of them are mean 0 and the same covariance matrix, $\text{span}(\eta^n - 1\bar{\eta}^n)$ and $\text{span}(Q(\eta^n - 1\bar{\eta}^n))$ have the same distribution.

By the uniqueness of this distribution $\mu$, we have that $\text{span}(\eta^n - 1\eta^n)$ is indeed distributed as $S_{n,n-1}$.

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

**Lemma 7.**

$$\frac{h(S_{n,n-1})^2}{n} \rightarrow \frac{1}{8\pi},$$

in distribution.

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

\[
\frac{\|\eta^n - 1\eta^n\|_2^2}{n} = \frac{1}{n} \sum_{i=1}^{n} (\eta^n_i - \bar{\eta^n})^2 \\
= \frac{1}{n} \sum_{i=1}^{n} ((\eta^n_i)^2 - 2\eta^n_i \bar{\eta^n} + \bar{\eta^n}^2) \\
= \frac{1}{n} \sum_{i=1}^{n} (\eta^n_i)^2 - \frac{2}{n} \sum_{i=1}^{n} \eta^n_i \bar{\eta^n} + (\bar{\eta^n})^2 \\
= \frac{1}{n} \sum_{i=1}^{n} (\eta^n_i)^2 - (\bar{\eta^n})^2
\]

By strong law of large numbers we have that,

\[
\frac{1}{n} \sum_{i=1}^{n} (\eta^n_i)^2 \to 1, \text{ almost surely,}
\]

and,

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

Thus,

\[
(16) \quad \frac{1}{\sqrt{n}} \|\eta^n - 1\eta^n\|_2 \to 1, \text{ a.s.}
\]

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

\[
\frac{1}{n} \|\eta^n\| + \frac{1}{n} \|1\eta^n\|_1 \geq \frac{1}{n} \|\eta^n - 1\eta^n\|_1 \geq \frac{1}{n} \|\eta^n\|_1 - \frac{1}{n} \|1\eta^n\|_1
\]

Now by the strong law of large numbers,

\[
\frac{1}{n} \|1\eta^n\|_1 = |\bar{\eta^n}| \to 0, \text{ a.s.}
\]

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

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

where \( \xi \) is standard normal. Thus,

\[
(17) \quad \frac{1}{n} \|\eta^n - 1\eta^n\|_1 \to \frac{1}{2\sqrt{2\pi}}.
\]

From (16) and (17) and using Slutsky’s lemma we have that,

\[
\frac{\|\eta^n - 1\eta^n\|_1}{\sqrt{n} \|\eta^n - 1\eta^n\|_2} \to \frac{1}{2\sqrt{2\pi}}, \text{ almost surely.}
\]

36
By continuity of $x \mapsto x^2$, 

\begin{equation}
\frac{\|\eta^n - 1\eta^n\|^2}{n\|\eta^n - 1\eta^n\|^2} \rightarrow \frac{1}{8\pi}, \text{ almost surely.}
\end{equation}

Finally by Equation (18) and the fact that $h(S^{n,n-1})^2$ has the same distribution as $\frac{\|\eta^n - 1\eta^n\|^2}{\|\eta^n - 1\eta^n\|^2}$, 

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

in distribution. 

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

$$\frac{\text{Eff}^{n,p_n}_n}{n} \geq \frac{\text{Eff}^{n,n-1}_n}{n} = \frac{h(S^{n-1})^2}{n\sigma^2}.$$ 

Finally using Lemma 7 we have, 

$$\frac{h(S^{n-1})^2}{n\sigma^2} \rightarrow \mathcal{D} \frac{1}{8\pi\sigma^2}.$$ 

Thus for any $\epsilon > 0$ 

$$P \left( \left| \frac{\text{Eff}^{n,n-1}_n}{n} - \frac{1}{8\pi\sigma^2} \right| > \epsilon \right) \rightarrow 0.$$ 

Subsequently, 

$$P \left( \frac{\text{Eff}^{n,n-1}_n}{n} - \frac{1}{8\pi\sigma^2} < \epsilon \right) \rightarrow 0.$$ 

Finally, 

$$P \left( \frac{\text{Eff}^{n,p_n}_n}{n} - \frac{1}{8\pi\sigma^2} < \epsilon \right) \rightarrow 0.$$ 

D. 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 $(\delta_{k-1} + a_k, \Delta_{k-1} + a_k Z_{k,2p}, Z_{k+1})$. 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 $S_{n+1} \mapsto \delta_n^2 + \|\Delta_n\|^2_{\Sigma^{-1}}$. 

37
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 \(J^{*,k} : \mathcal{X}_k \rightarrow \mathbb{R}\) 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,
\[
J^{*,k}(\delta_{k-1}, \Delta_{k-1}, Z_k) = \min_{u \in \{\pm 1\}} \mathbb{E}[J^{*,k+1}(\delta_{k-1} + u, \Delta_{k-1} + uZ_{k,2p}, Z_{k+1})].
\]
Further \(x_k^*\), the optimal policy, has the property that,
\[
x^*_k \in \arg\min_{u \in \{\pm 1\}} \mathbb{E}[J^{*,k+1}(\delta_{k-1} + u, \Delta_{k-1} + uZ_{k,2p}, Z_{k+1})].
\]

Let,
\[
Q^k(\delta_k, \Delta_k) \triangleq \mathbb{E}\left[J^{*,k+1}(\delta_{k-1}, \Delta_{k-1}, Z_{k+1})\right].
\]
Using (19) and (21),
\[
Q^k(\delta_k, \Delta_k) = \mathbb{E}\left[\min_{u \in \{\pm 1\}} \mathbb{E}[Q^{k+1}(\delta_{k-1} + u, \Delta_{k-1} + uZ_{k,2p})]\right].
\]
Further using (20) and (21),
\[
x^*_k \in \arg\min_{u \in \{\pm 1\}} \mathbb{E}\left[Q^{k+1}(\delta_{k-1} + u, \Delta_{k-1} + uZ_{k,2p})\right].
\]
This proves the dynamic programming proposition.

\[\Box\]

\section*{E. 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 \(\Gamma_n\) the matrix \(\frac{1}{n}Z^\top Z\). Then we have the following approximation result:

\begin{lemma}
Provided \(n \geq \frac{L}{\epsilon^2} \max(p_n, l \log 2/\rho_n)\), then with probability at least \(1 - \rho_n\), we have
\[
\|\Gamma_n - I\| \leq \epsilon
\]
where \(L\) and \(l\) are universal constants.
\end{lemma}
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

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

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

■

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

$$(22) \quad 1 - \epsilon \leq \sigma_{\min}\left(\frac{Z}{\sqrt{n}}\right) \leq \sigma_{\max}\left(\frac{Z}{\sqrt{n}}\right) \leq 1 + \epsilon$$

with probability at least $1 - \rho_n$. 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 $\mu^*$ 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

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

and recall

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

Proof of Theorem 4. Now, $(22)$ is equivalently stated as:

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

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

$$\frac{1}{1 + \epsilon} \leq \sqrt{\lambda_{\min}(\Gamma_n^{-1})} \leq \sqrt{\lambda_{\max}(\Gamma_n^{-1})} \leq \frac{1}{1 - \epsilon},$$
with probability at least $1 - \rho_n$. By the Courant-Fisher theorem we consequently have that,

\begin{equation}
\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}^{pn},
\end{equation}

with probability at least $1 - \rho_n$.

Now note that

\begin{equation}
\bar{\Delta}^\top \Gamma_n^{-1} \bar{\Delta} = \|\bar{\Delta}\|_2^2 \leq n^2, \quad \forall \bar{\Delta} \in \mathbb{R}^{pn}.
\end{equation}

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

\[ n - \frac{\|\bar{\Delta}\|_2^2 \Gamma_n^{-1}}{n} \geq 0. \]

Let $A$ be the set of sample paths such that (23) holds. We have that,

\[
\mathbb{E}_\hat{\mu} \left[ \left\| \bar{\Delta}_n \right\|_2^2 \right] = \mathbb{E}_\hat{\mu} \left[ \left\| \bar{\Delta}_n \right\|_2^2 \mathbb{I}_A + \left\| \bar{\Delta}_n \right\|_2^2 \mathbb{I}_{A^c} \right]
\leq \frac{\mathbb{E}_\hat{\mu} \left[ \left\| \bar{\Delta}_n \right\|_2^2 \right]}{(1-\epsilon)^2} + n^2 \mathbb{E}_\hat{\mu} \left[ \mathbb{I}_{A^c} \right]
\leq \frac{\mathbb{E}_\hat{\mu} \left[ \left\| \bar{\Delta}_n \right\|_2^2 \right]}{(1-\epsilon)^2} + \rho_n n^2
\leq \frac{\mathbb{E}_{\mu^*} \left[ \left\| \bar{\Delta}_n \right\|_2^2 \right]}{(1-\epsilon)^2} + \rho_n n^2
\]

where the first inequality follows from the right hand side of (23) applied to each sample path in $A$ and (24) applied to sample paths in $A^c$. The final inequality follows from the optimality of $\hat{\mu}$ for (P3'). We will now show that

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

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 (23) holds, i.e., the set $A$, \( \left\| \bar{\Delta}_n \right\|_2^2 \leq (1+\epsilon) \left\| \bar{\Delta}_n \right\|_2^2 \). Thus,

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

Now note that,

\[
\left\| \bar{\Delta}_n \right\|_2^2 = \delta_n^2 + \left\| \Delta_n \right\|_2^2 \leq n^2 + \left\| \Delta_n \right\|_2^2.
\]
The inequality follows since $|\delta_n| \leq n$. Thus,

$$E_{\mu^*} \left[ \|\Delta_n\|_2^2 I_{A^c} \right] = n^2 \rho_n + E_{\mu^*} \left[ \|\Delta_n\|_2^2 I_{A^c} \right]$$

$$\leq n^2 \rho_n + E_{\mu^*} \left[ \|\Delta_n\|_2^2 I_{\{\|\Delta_n\|_2^2 \geq \alpha_n(\rho_n)\}} \right],$$

where $\alpha_n(\rho_n)$ satisfies $P_{\mu^*}(\|\Delta_n\|_2^2 \geq \alpha_n(\rho_n)) = \rho_n$. Applying Lemma 11 yields

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

which yields the result. ■

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

$$E_{\mu^*} \left[ \|\Delta_n\|_2^2 1_{\|\Delta_n\|_2^2 \geq \alpha_n(\rho_n)} \right]$$

where $\alpha_n(\rho_n)$ satisfies $P_{\mu^*}(\|\Delta_n\|_2^2 \geq \alpha_n(\rho_n)) = \rho_n$. In other words $\alpha_n(\rho_n)$ is the $\rho_n$ percentile of $\|\Delta_n\|^2$. Let $\bar{Z}_n$ be a Gamma$(n(p_n - 1)/2, 1)$ random variable, and let $\hat{\alpha}_n(\rho_n)$ satisfy

$$P (\bar{Z}_n \geq \hat{\alpha}_n(\rho_n)) = \rho.$$

We have

Lemma 9.

$$E_{\mu^*} \left[ \|\Delta_n\|_2^2 1_{\|\Delta_n\|_2^2 \geq \alpha_n(\rho_n)} \right] \leq 2nE \left[ \bar{Z}_n 1_{\bar{Z}_n \geq \hat{\alpha}_n(\rho_n)} \right]$$

Proof. Observe that

$$\|\Delta_n\|^2 = \left\| \sum_{k=1}^n x_k Z_{k,2:p_n} \right\|_2^2$$

$$\leq \left( \sum_{k=1}^n \|Z_{k,2:p_n}\|_2 \right)^2$$

$$\leq n \sum_{k=1}^n \|Z_{k,2:p_n}\|_2^2.$$

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

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

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 $\tilde{Z} \sim \text{Gamma}(k, 1)$ and $z(\rho_k)$ is its $\rho$th quantile (i.e., $z(\rho_k)$ satisfies $\mathbb{P}(\tilde{Z} \geq z(\rho_k)) = \rho_k$), then as $k \to \infty$,

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

**Proof.** We have:

$$
\mathbb{E} \left[ \tilde{Z} \mathbf{1}_{\tilde{Z} \geq z(\rho_k)} \right] = \int_{z(\rho_k)}^{\infty} z \frac{z^{k-1} \exp(-z)}{\Gamma(k)} \, dz
$$

$$
= \frac{\Gamma(k+1)}{\Gamma(k)} \int_{z(\rho_k)}^{\infty} \frac{z^k \exp(-z)}{\Gamma(k+1)} \, dz
$$

$$
= k \left[ \frac{\Gamma(k+1, z(\rho_k))}{k \Gamma(k)} \right]
$$

$$
= k \left[ \frac{k \Gamma(k, z(\rho_k)) + z(\rho_k)^k \exp(-z(\rho_k))}{k \Gamma(k)} \right]
$$

$$
= k \left[ \rho_k + \frac{z(\rho_k)^k \exp(-z(\rho_k))}{k \Gamma(k)} \right]
$$

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

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

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

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

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.**

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

42