Adaptive Multi-Objective Optimization for Primary and Guardrail Criteria

Blake Mason
bjmason@amazon.com

James McQueen
jmcq@amazon.com

1 Introduction

Organizations increasingly face the challenge of evaluating numerous potential user experiences, particularly in digital contexts where content can be easily modified through page layouts, personalized recommendations, and automated content generation. While creating these experiences has become increasingly straightforward, especially with the advent of generative AI, determining user preferences remains a critical bottleneck. A/B/n tests are the gold standard in determining whether a new treatment is an improvement over the business as usual. A/B tests are well understood, robust to a range of modeling assumptions, and allow the experimenter to measure a variety of different success criteria when determining whether to replace the control with the new treatment experience after the experiment concludes. A/B/n tests however, scale poorly to large numbers of treatments being tested simultaneously. As the number of treatments under consideration increases, the same pool of traffic arriving to the website is split into smaller fractions leading to noisier treatment effect estimates and more uncertain decision making.

Instead, organizations increasingly turn to *adaptive experimentation* to improve power in settings where the number of treatments is large. Generically, adaptive experimentation techniques such as multi-armed bandits compute performance statistics of different treatments as the experiment proceeds, and use this information to reallocate traffic towards the best performing treatments. In the extreme case, compared to a standard A/B/n test this can reduce the number of samples necessary to identify the best treatment by an $O(K)$ factor where K is the number of treatments [5]. A primary reason limiting the further adoption of adaptive experimentation techniques is that these algorithms most commonly optimize for a single success metric. By contrast, organizations frequently care about a range of different success criteria when evaluating whether to launch a treatment such as financial metrics or customer engagement. The core metric(s) a treatment seeks to improve are referred to as primary criteria and other metric(s) which the treatment seeks to not degrade are referred to as guardrail criteria. Since adaptive algorithms most often optimize based on a single metric, it is common to experiment in two phases. First, an adaptive algorithm selects a winning treatment according to a single primary launch criteria. Then the winning treatment enters a standard A/B test to measure the remaining guardrail criteria. This method is powerful but limited. If primary and guardrail criteria come at the expense of each other, then the organization may not be willing to launch the winning treatment and will need to start over. Instead, in this paper we study the problem of *adaptive multi-objective optimization with primary and guardrail criteria* where the experimenter wishes to use an adaptive algorithm to select a treatment that is likely to simultaneously meet both primary and guardrail criteria before proceeding to a standard A/B test to make a final launch decision.

1.1 Problem Setting

An organization is considering among a set $\mathcal{A} = \{0, \dots, K\}$ treatments to launch. Treatment 0 represents the control. In order to determine whether a treatment can be launched, the organization considers M different metric criteria representing different aspects of each treatment's performance. We consider the model of [11] but in the practical case that the businesses metrics are correlated with one another. That is, revenue and profit correlate to one another. Each time a customer is shown a treatment $a \in \mathcal{A}$, the organization observes a vector $\mathbf{x}_a \in \mathbb{R}^M$ of the observed outcomes for each of the M metrics. For all treatments a, we assume that $\mathbf{x}_a$ follows a sub Gaussian distribution with unknown mean $\boldsymbol{\mu_a}$ and known covariance proxy $\boldsymbol{\Sigma}_a \succ 0$. Without loss of generality, we assume that a treatment a improves over control if it is larger on average; that is $\mu_{a,i} \geq \mu_{0,m}$ for all metrics $m = 1, \dots, M$. In order to determine which treatment to launch into production, the organization first leverages an adaptive algorithm to select a treatment $\widehat{a} \in \mathcal{A}$ before performing an A/B test with equal allocations comparing $\widehat{a}$ to control to make a launch decision. At the end of the A/B test, for each metric m the experimenter performs a two sample z-test with significance level α_m to determine

if treatment $\widehat{a}$ meets its primary/guardrail criteria[1]. We assume that the organization has a known budget of N samples it can collect from users for to select $\widehat{a}$ and a budget N_V of samples it can collect to determine whether to launch $\widehat{a}$ into production.

As which treatment is best is in general ambiguous when there are multiple success criteria, in this work we seek to identify during the adaptive exploration phase the treatment that is most likely to satisfy both its primary and guardrail criteria during the subsequent A/B test to validate the winning treatment. Specifically, let $\mathcal{E}_{a,m}$ denote the probability that the z-score of a treatment $a \in \mathcal{A}$ exceeds its threshold for metric m with significance α_m after collecting $N_V/2$ samples. We seek to identify the unknown treatment $a^* = \arg\max_{a\in\mathcal{A}} \mathbb{P}\left(\bigcap_m \mathcal{E}_{a,m}\right)$ which maximizes the probability that it satisfies all of its primary and guardrail criteria. That is, the goal is to develop an algorithm that with high probability returns a treatment $\widehat{a} = a^*$.

2 Optimal Experimental Design to Identify a^*

Before developing an adaptive algorithm to identify a^*, we begin by deriving the optimal *static* allocation to identify a^* by leveraging G-Optimal experimental design [7, 9]. Let $\mathbf{e}_i \in \mathbb{R}^M$ denote the i^{th} standard basis vector. The following proposition helps characterize a^*:

Proposition 2.1. *Let $\boldsymbol{\xi}_a \in \mathbb{R}^M$ be a vector whose m^{th} entry is $\Phi^{-1}(\alpha_m)\sqrt{2/N_V}$ where $\Phi^{-1}(\cdot)$ is the inverse CDF of a standard normal distribution and $\boldsymbol{\Sigma}_a^{m,m}$ denotes the m^{th} diagonal entry of $\boldsymbol{\Sigma}_a$. In the setting of section 1.1, we have that*

$$a^* = \arg\max_{a\in\mathcal{A}\setminus\{0\}} \min_m \mathbf{e}_m^\top \left[\left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \left(\boldsymbol{\mu_a} - \boldsymbol{\mu_0}\right) + \boldsymbol{\xi} \right].$$

For brevity, let $\mathbf{z}_a := \left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \left(\boldsymbol{\mu_a} - \boldsymbol{\mu_0}\right) + \boldsymbol{\xi}_a$. Our goal therefore is to identify the treatment that solves the above max-min problem. Conceptually, the $\left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \left(\boldsymbol{\mu_a} - \boldsymbol{\mu_0}\right)$ term captures the difficulty in distinguishing arm a from control during the exploration phase whereas $\boldsymbol{\xi}_a$ captures the complexity of validating that a meets the launch and guardrail criteria. The best arm strikes a balance between these two objectives. That is, it is both sufficiently easy to find, but also sufficiently easy to verify. To see how to do this optimally, consider a set $\{\mathbf{x}_a^1, \mathbf{x}_a^2, \ldots \mathbf{x}_a^{N_a}\} \subset \mathbb{R}^M$ of N_a samples of arm a such as might be collected by showing treatment a to N_a users on the website. Define $\widehat{\boldsymbol{\mu}}_a := \frac{1}{n}\sum_{t=1}^{N_a} \mathbf{x}_a^t$. Hence, an unbiased estimate of $\mathbf{e}_m^\top \mathbf{z}_a$ is $\mathbf{e}_m^\top \widehat{\mathbf{z}}_a := \mathbf{e}_m^\top \left[\left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \left(\widehat{\boldsymbol{\mu}}_a - \widehat{\boldsymbol{\mu}}_0\right) + \boldsymbol{\xi}_a \right]$. Furthermore, as the above is a linear transformation of subgaussian random variables, by Theorem 2.7 of [8] we have that this estimate is also sub-Gaussian with covariance proxy $\text{Cov}(\mathbf{e}_m^\top \widehat{\mathbf{z}}_a) = \mathbf{e}_m^\top \left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \left(\frac{1}{N_a}\boldsymbol{\Sigma}_a + \frac{1}{N_0}\boldsymbol{\Sigma}_0\right) \left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \mathbf{e}_m$. As we know neither the treatment a nor the metric m which achieves the max-min given in Prop. 2.1, we instead (possibly naively) sample in order to minimize the maximum variance of all estimates $\mathbf{e}_m^\top \widehat{\mathbf{z}}_a$. In particular, as the above expression for the covariance does not depend on the unknown means, we can compute such a sampling allocation directly in a process referred to as G-Optimal design [7, 9]. Specifically, let $\triangle_{K+1}$ denote the probability simplex in $\mathbb{R}^{K+1}$. We seek to solve the following optimization

$$\lambda = \arg\min_{\lambda\in\triangle_{K+1}} \max_{a\in\mathcal{A}} \max_m \mathbf{e}_m^\top \left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \left(\frac{1}{\lambda_a}\boldsymbol{\Sigma}_a + \frac{1}{\lambda_0}\boldsymbol{\Sigma}_0\right) \left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \mathbf{e}_m. \tag{1}$$

By construction then, allocating samples to the different treatments according to λ minimizes the variance of $\mathbf{e}_m^\top \widehat{\mathbf{z}}_a^*$ and maximizes the probability of identifying a^*.

Proposition 2.2. *The optimization in Eq.* (1) *is convex.*

Proof. To show that the above can be solved via convex optimization, we first note that the argument of the above minimization over lambda can be written as a pointwise maximization over a and i for functions of λ. Furthermore, the constraint set $\lambda \in \triangle_{|\mathcal{A}_S|}$ is convex. Hence, by 3.2.3 of [1], to show that the above optimization is convex, it is sufficient to show that each function we are maximizing over is convex. Specifically, for any a and i define

$$f_{a,i}(\lambda) := \mathbf{e}_i^\top \left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \left(\frac{1}{\lambda_a}\boldsymbol{\Sigma}_a + \frac{1}{\lambda_0}\boldsymbol{\Sigma}_0\right) \left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \mathbf{e}_i.$$

[1]These ideas all directly extend to Bayesian criteria as well, but we omit that for brevity.

For any $a' \neq 0$ and $a' \neq a$, $f_{a,i}(\lambda)$ is constant. Thus any derivative or partial derivative for $\lambda_{a'}$ is 0. Next, we have that

$$\frac{d}{d\lambda_0} f_{a,i}(\lambda) = \frac{-1}{\lambda_0^2} \mathbf{e}_i^\top \left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \boldsymbol{\Sigma}_0 \left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \mathbf{e}_i$$

and likewise for $\frac{d}{d\lambda_a}$. Based on the above, we see that

$$\frac{\partial f}{\partial \lambda_0 \partial \lambda_a} = \frac{\partial f}{\partial \lambda_a \partial \lambda_0} = 0.$$

Thus all off-diagonal entries of the Hessian of $f_{a,i}(\lambda)$ are 0 for any a and i. Finally, note that

$$\frac{d}{d\lambda_0^2} = \frac{2}{\lambda_0^3} \mathbf{e}_i^\top \left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \boldsymbol{\Sigma}_0 \left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \mathbf{e}_i > 0$$

since $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_a$ are positive definite and $\lambda_0, \lambda_a \geq 0$ on the constraint set. Hence, for any a and i, the Hessian is a diagonal matrix with non-negative entries. Thus, we see that the Hessian of $f_{a,i}(\lambda)$ is positive semi-definite, implying the $f_{a,i}(\lambda)$ is convex. □

3 Sequential Halving for Multi-Objective Optimization

Taking the above idea a step further, we borrow ideas from [10] to perform sequential halving (c.f., [6]) via iterative rounds of optimal design to develop Algorithm 1. Much like standard sequential halving, Algorithm 1 proceeds in

Algorithm 1: G-optimal design for MOO via Sequential Halving

Input : Initialize $\mathcal{A}_1 \leftarrow \mathcal{A} \backslash \{0\}$. Rounding method Approx with parameter $\epsilon \in (0,1)$, budget N

Samples per round: $n \leftarrow \lfloor \frac{N}{\lceil \log_2(K) \rceil} \rfloor$;

for *Round* $r = 0, \ldots, \lceil \log_2(K) \rceil - 1$ **do**

  $\lambda_r^* \leftarrow \arg\min_{\lambda \in \triangle_{|\mathcal{A}_r|}} \max_{a \in \mathcal{A}_r} \max_m \mathbf{e}_m^\top \left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \left(\frac{1}{\lambda_a}\boldsymbol{\Sigma}_a + \frac{1}{\lambda_0}\boldsymbol{\Sigma}_0\right) \left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \mathbf{e}_m$

  $n_r^{(0)} n_r^1, \ldots, n_r^{(A)} \leftarrow \text{Approx}(\lambda_r^*, n, \epsilon)$

  for *Treatment* $a = 0, 1, \ldots, K$ **do**

    Sample treatment a n_r^a times and observe $\left\{\mathbf{x}_a^1, \ldots, \mathbf{x}_a^{n_r^a}\right\}$

    Compute $\widehat{\boldsymbol{\mu}}_a^r \leftarrow \frac{1}{n_r^a} \sum_{i=1}^{n_r^a} \mathbf{x}_a^i$

  $\widehat{z}_{a,m}^r = \mathbf{e}_m^\top \left[\left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \left(\widehat{\boldsymbol{\mu}}_a^r - \widehat{\boldsymbol{\mu}}_0^r\right) + \boldsymbol{\xi}_a\right]$

  Let $\mathcal{A}_{r+1}$ be the $\lceil |\mathcal{A}_r|/2 \rceil$ arms with maximum $\min_m \widehat{z}_{a,m}^r$

Return single treatment in $\widehat{a} \in \mathcal{A}_{\lceil \log_2(K) \rceil}$

rounds each of which are receive an identical budget of samples, and at the end of each round, the bottom half of treatments are eliminated. In this particular case, at the start of each round, the algorithm computes an optimal design λ_r^* over the remaining treatments in $\mathcal{A}_r$. It then samples according to this distribution to estimate $\min_m \widehat{\mathbf{z}}_{a,m}^r$ and eliminates the half of remaining treatments for which this is the smallest. Intuitively, this can be seen as progressively solving the max-min optimization in Prop. 2.1. As the rounds proceed, fewer treatments remain and those that do receive a greater number of samples giving a more precise estimate of the unknown $\boldsymbol{\mu_a} - \boldsymbol{\mu_0}$. Lastly, as a technical point the algorithm relies on a generic rounding procedure referred to as Approx$(\cdot)$ which given a budget of samples, a distribution λ, and a tolerance ϵ specifies an integer number of times to sample each treatment. This is done to decrease the variance of $\min_m \widehat{\mathbf{z}}_{a,m}^r$, and we direct the reader to Appendix B of [4] for examples of polynomial time algorithms to compute this. It is tempting to use ideas such as the RIPS estimator from [2] to avoid the need for rounding. However, as we have a fixed budget of N samples, this leads to a dependence on the unknown $\boldsymbol{\mu_a} - \boldsymbol{\mu_0}$ values. The following theorem bounds the probability that the treatment $\widehat{a}$ returned by Alg. 1 is not a^*.

Theorem 3.1. *In the setting of section 1.1, we have that*

$$\mathbb{P}(\widehat{a} \neq a^*) \leq 3M \log_2(K) \max_a \exp\left(\frac{-N}{4(1+\epsilon)} \cdot \frac{\min_m \max_{m'} \left(z_{a,m} - z_{a^*,m'}\right)^2}{\min_{\lambda \in \triangle_{|\mathcal{A}|}} \max_{m,a'} f(\lambda, a', m)}\right)$$

where

$$f(\lambda, a', m) = \mathbf{e}_m^\top \left(\boldsymbol{\Sigma}_{a'} + \boldsymbol{\Sigma}_0\right)^{-1/2} \left(\frac{1}{\lambda_{a'}}\boldsymbol{\Sigma}_{a'} + \frac{1}{\lambda_0}\boldsymbol{\Sigma}_0\right) \left(\boldsymbol{\Sigma}_{a'} + \boldsymbol{\Sigma}_0\right)^{-1/2} \mathbf{e}_m.$$

In particular, the above shows that the probability that Alg. 1 makes a mistake decays exponentially in the number of samples. The min-max in the numerator of characterizes how easy it is to distinguish the performance of any treatment a on any metric m relative to that of a^*. The denominator gives the variance of the estimate of $\min_m \widehat{\mathbf{z}}^r_{a,m}$. Intuitively, this is a signal to noise ratio. As either the treatments become easier to distinguish or the variance decreases, the probability that the algorithm makes a mistake also decreases. There is room for improvement in this bound however. We conjecture that it is possible to have the same treatment a appear in both the numerator and the denominator rather than a in the former and a' in the latter as this would match the information theoretic lower bound stated in [4] in the special case that there is a single metric.

4 Extension to Bayesian Models

While the above algorithm is designed with the frequentist setting in mind, the extension to a Bayesian observation model is straightforward. Note that in the Bayesian case, there is a better estimate of $\boldsymbol{\mu}_a - \boldsymbol{\mu}_0$ than $\widehat{\boldsymbol{\mu}}_a - \widehat{\boldsymbol{\mu}}_0$ as we have access to a prior over $\boldsymbol{\mu}_a$ and $\boldsymbol{\mu}_0$. Assume that $\boldsymbol{\mu}_0,\ \boldsymbol{\mu}_a \sim \mathcal{N}(\mathbf{m}, \frac{1}{2}\mathbf{C})$. Then $\boldsymbol{\mu}_a - \boldsymbol{\mu}_0 \sim \mathcal{N}(0, \mathbf{C})$. Hence,

$$\boldsymbol{\mu}_a - \boldsymbol{\mu}_0 | \widehat{\boldsymbol{\mu}_a}, \widehat{\boldsymbol{\mu}_0}, \boldsymbol{\Sigma_a}, \boldsymbol{\Sigma_0}, \mathbf{C} \sim \mathcal{N}\left(\tilde{\boldsymbol{\mu}}_{a0}, \tilde{\boldsymbol{\Sigma}}_{a0}\right).$$

for

$$\tilde{\boldsymbol{\mu}}_{a0} = \left(\mathbf{C}^{-1} + \left(\frac{1}{N_a}\boldsymbol{\Sigma_a} + \frac{1}{N_0}\boldsymbol{\Sigma_0}\right)^{-1}\right)^{-1} \left(\frac{1}{N_a}\boldsymbol{\Sigma_a} + \frac{1}{N_0}\boldsymbol{\Sigma_0}\right)^{-1} (\widehat{\boldsymbol{\mu}_a} - \widehat{\boldsymbol{\mu}_0})$$

and

$$\tilde{\boldsymbol{\Sigma}}_{a0} = \left(\mathbf{C}^{-1} + \left(\frac{1}{N_a}\boldsymbol{\Sigma_a} + \frac{1}{N_0}\boldsymbol{\Sigma_0}\right)^{-1}\right)^{-1}$$

In the Bayesian regime, rather than performing a 2-sample z-test to determine of a treatment meets its launch/guardrail criteria, for a candidate treatment a, it is more common to check if $\mathbb{P}(\mu_{a,m} \geq \mu_{0,m}) \geq q_m$ for all metrics m. That is, the experimenter will only accept treatment a if it improves over control with probability q_m for all metrics m. Following Proposition 1 of [11], define

$$\xi_{a,m} = \frac{\Phi^{-1}(1-q_m)}{\sqrt{N_v/2}} \sqrt{1 + 2\frac{\boldsymbol{\Sigma}_a^{m,m} + \boldsymbol{\Sigma}_0^{m,m}}{\boldsymbol{C}^{m,m} N_v}}.$$

where $\boldsymbol{C}^{m,m}$ denotes the $m^{\textbf{th}}$ diagonal entry of $\boldsymbol{C}$ and we recall that N_v denotes the budget of samples reserved for validation. Let $\boldsymbol{\xi}_a$ be the vector that is entry-wise equal to $\xi_{a,m}$. We estimate $\boldsymbol{\mu}_a - \boldsymbol{\mu}_0$ using its a posterior mean. Therefore we have that

$$\begin{aligned}
&\mathbf{e}_i^\top \left[\left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \left(\boldsymbol{\mu}_a - \boldsymbol{\mu}_0\right) + \boldsymbol{\xi}_a\right] \Big| \widehat{\boldsymbol{\mu}_a}, \widehat{\boldsymbol{\mu}_0}, \boldsymbol{\Sigma_a}, \boldsymbol{\Sigma_0}, \mathbf{C} \\
&\quad \sim \mathcal{N}\left(\mathbf{e}_i^\top \left[\left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \tilde{\boldsymbol{\mu}}_{a0} + \boldsymbol{\xi}_a\right], \mathbf{e}_i^\top \left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \tilde{\boldsymbol{\Sigma}}_{a0} \left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \mathbf{e}_i\right).
\end{aligned}$$

Following the same approach as the frequentist regime, we sample so as to minimize the maximum posterior variance over allocations λ. Specifically, we wish to solve

$$\min_{\lambda \in \triangle_{K+1}} \max_{a \in \mathcal{A}} \max_m \mathbf{e}_m^\top \left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \left(\mathbf{C}^{-1} + \left(\frac{1}{\lambda_a}\boldsymbol{\Sigma_a} + \frac{1}{\lambda_0}\boldsymbol{\Sigma_0}\right)^{-1}\right)^{-1} \left(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_0\right)^{-1/2} \mathbf{e}_m. \tag{2}$$

The remainder of the algorithm proceeds as before but altering the computation of λ_r^* to instead depend on the output of equation (2). We note however that it is not immediately obvious that the above optimization is convex due to the covariance term. It may be possible to leverage a result such as [3] showing that matrix inversion is a convex mapping to reduce this to a problem of compositions of convex operators, but we defer this to future work. From the standpoint of statistical guarantees of the algorithm, it is possible to prove a similar theoretical result as Theorem 3.1 by swapping a use of Hoeffding's inequality for a Gaussian tail bound and otherwise using the same mechanics.

References

[1] S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[2] R. Camilleri, K. Jamieson, and J. Katz-Samuels. High-dimensional experimental design and kernel bandits. In *International Conference on Machine Learning*, pages 1227–1237. PMLR, 2021.

[3] C. Davis. Notions generalizing convexity for functions defined on spaces of matrices. In *Proceedings of the Symposia in Pure Mathematics*, volume 7, pages 187–201. American Mathematical Society Providence, RI, USA, 1963.

[4] T. Fiez, L. Jain, K. G. Jamieson, and L. Ratliff. Sequential experimental design for transductive linear bandits. *Advances in neural information processing systems*, 32, 2019.

[5] K. Jamieson, M. Malloy, R. Nowak, and S. Bubeck. On finding the largest mean among many, 2013.

[6] Z. Karnin, T. Koren, and O. Somekh. Almost optimal exploration in multi-armed bandits. In *International conference on machine learning*, pages 1238–1246. PMLR, 2013.

[7] F. Pukelsheim. *Optimal design of experiments*. SIAM, 2006.

[8] O. Rivasplata. Subgaussian random variables: An expository note. *Internet publication, PDF*, 5, 2012.

[9] M. Soare, A. Lazaric, and R. Munos. Best-arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 27, 2014.

[10] Z. Xiong, R. Camilleri, M. Fazel, L. Jain, and K. Jamieson. A/b testing and best-arm identification for linear bandits with robustness to non-stationarity. In *International Conference on Artificial Intelligence and Statistics*, pages 1585–1593. PMLR, 2024.

[11] Q. Zhang, T. Fiez, Y. Liu, and W. Liu. Multi-metric adaptive experimental design under fixed budget with validation, 2025.