# Online A-Optimal Design and Active Linear Regression

**Xavier Fontaine** [1] **Pierre Perrault** [2] **Michal Valko** [3] **Vianney Perchet** [4 5]

## Abstract

We consider in this paper the problem of optimal experiment design where a decision maker can choose which points to sample to obtain an estimate $\hat{\beta}$ of the hidden parameter $\beta^\star$ of an underlying linear model. The key challenge of this work lies in the heteroscedasticity assumption that we make, meaning that each covariate has a different and unknown variance. The goal of the decision maker is then to figure out on the fly the optimal way to allocate the total budget of $T$ samples between covariates, as sampling several times a specific one will reduce the variance of the estimated model around it (but at the cost of a possible higher variance elsewhere). By trying to minimize the $\ell^2$-loss $\mathbb{E}[\|\hat{\beta} - \beta^\star\|^2]$ the decision maker is actually minimizing the trace of the covariance matrix of the problem, which corresponds then to online A-optimal design. Combining techniques from bandit and convex optimization we propose a new active sampling algorithm and we compare it with existing ones. We provide theoretical guarantees of this algorithm in different settings, including a $\mathcal{O}(T^{-2})$ regret bound in the case where the covariates form a basis of the feature space, generalizing and improving existing results. Numerical experiments validate our theoretical findings.

## 1. Introduction and related work

A classical problem in statistics consists in estimating an unknown quantity, for example the mean of a random variable, parameters of a model, poll results or the efficiency of a medical treatment. In order to do that, statisticians usually build estimators which are random variables based on the data, supposed to approximate the quantity to estimate. A way to construct an estimator is to make experiments and to gather data on the estimand. In the polling context an experiment consists for example in interviewing people in order to know their voting intentions. However if one wants to obtain a "good" estimator, typically an unbiased estimator with low variance, the choice of which experiment to run has to be done carefully. Interviewing similar people might indeed lead to a poor prediction. In this work we are interested in the problem of optimal design of experiments, which consists in choosing adequately the experiments to run in order to obtain an estimator with small variance. We focus here on the case of heteroscedastic linear models with the goal of actively constructing the design matrix. Linear models, though possibly sometimes too simple, have been indeed widely studied and used in practice due to their interpretability and can be a first good approximation model for a complex problem.

The original motivation of this problem comes from use cases where obtaining the label of a sample is costly, hence choosing carefully which points to sample in a regression task is crucial. Consider for example the problem of controlling the wear of manufacturing machines in a factory (Antos et al., 2010), which requires a long and manual process. The wear can be modeled as a linear function of some features of the machine (age, number of times it has been used, average temperature, ...) so that two machines with the same parameters will have similar wears. Since the inspection process is manual and complicated, results are noisy and this noise depends on the machine: a new machine, slightly worn, will often be in a good state, while the state of heavily worn machines can vary a lot. Thus evaluating the linear model for the wear requires additional examinations of some machines and less inspection of others. Another motivating example comes from econometrics, typically in income forecasting. It is usually assumed that the annual income is influenced by the individual's education level, age, gender, occupation, *etc.* through a linear model. Polling is also an issue in this context: what kind of individual to poll to gain as much information as possible about an explanatory variable?

The field of optimal experiment design (Pukelsheim, 2006) aims precisely at choosing which experiment to perform in order to minimize an objective function within a budget constraint. In experiment design, the distance of the produced hypothesis to the true one is measured by the covariance

[1]Centre Borelli, ENS Paris-Saclay, Palaiseau, France [2]Idemia, Courbevoie, France [3]Google DeepMind, Paris, France [4]CREST, ENSAE, Palaiseau, France [5]Criteo AI Lab, Paris, France. Correspondence to: Xavier Fontaine <xavier.fontaine@polytechnique.edu>.

*Proceedings of the $38^{th}$ International Conference on Machine Learning*, PMLR 139, 2021. 

matrix of the error (Boyd & Vandenberghe, 2004). There are several criteria that can be used to minimize a covariance matrix, the most popular being A, D and E-optimality. In this paper we focus on A-optimal design whose goal is to minimize the trace of the covariance matrix. Contrary to several existing works which solve the A-optimal design problem in an offline manner in the homoscedastic setting (Sagnol, 2010; Yang et al., 2013; Gao et al., 2014) we are interested here in proposing an algorithm which solves this problem sequentially, with the additional challenge that each experiment has an unknown and different variance.

Our problem is therefore close to "active learning" which is more and more popular nowadays because of the exponential growth of datasets and the cost of labeling data. Indeed, the latter may be tedious and require expert knowledge, as in the domain of medical imaging. It is therefore essential to choose wisely which data to collect and to label, based on the information gathered so far. Usually, machine learning agents are assumed to be passive in the sense that the data is seen as a fixed and given input that cannot be modified or optimized. However, in many cases, the agent can be able to appropriately select the data (Goos & Jones, 2011). Active learning specifically studies the optimal ways to perform data selection (Cohn et al., 1996) and this is crucial as one of the current limiting factors of machine learning algorithms are computing costs, that can be reduced since all examples in a dataset do not have equal importance (Freund et al., 1997). This approach has many practical applications: in online marketing where one wants to estimate the potential impact of new products on customers, or in online polling where the different options do not have the same variance (Atkeson & Alvarez, 2018).

In this paper we consider therefore a decision maker who has a limited experimental budget and aims at learning some latent linear model. The goal is to build a predictor $\hat{\beta}$ that estimates the unknown parameter of the linear model $\beta^\star$, and that minimizes $\mathbb{E}[\|\hat{\beta} - \beta^\star\|^2]$. The key point here is that the design matrix is constructed sequentially and actively by the agent: at each time step, the decision maker chooses a "covariate" $X_k \in \mathbb{R}^d$ and receives a noisy output $X_k^\top \beta^\star + \varepsilon$. The quality of the predictor is measured through its variance. The agent will repeatedly query the different available covariates in order to obtain more precise estimates of their values. Instinctively a covariate with small variance should not be sampled too often since its value is already quite precise. On the other hand, a noisy covariate will be sampled more often. The major issue lies in the heteroscedastic assumption: the unknown variances must be learned to wisely sample the points.

Antos et al. (2010) introduced a specific variant of our setting where the environment providing the data is assumed to be stochastic and i.i.d. across rounds. More precisely, they studied this problem using the framework of stochastic multi-armed bandits (MAB) by considering a set of $K$ probability distributions (or arms), associated with $K$ variances. Their objective is to define an allocation strategy over the arms to estimate their expected values uniformly well. Later, the analysis and results have been improved by Carpentier et al. (2011). However, this line of work is actually focusing on the case where the covariates are only vectors of the canonical basis of $\mathbb{R}^d$, which gives a simpler closed form linear regression problem.

There have been some recent works on MAB with heteroscedastic noise (Cowan et al., 2015; Kirschner & Krause, 2018) with natural connections to this paper. Indeed, covariates could somehow be interpreted as contexts in contextual bandits. The most related setting might be the one of Soare (2015). However, they are mostly concerned about best-arm identification while recovering the latent parameter $\beta^\star$ of the linear model is a more challenging task (as each decision has an impact on the loss). In that sense we improve the results of Soare (2015) by proving a bound on the regret of our algorithm. Other works as (Chen & Price, 2019) propose active learning algorithms aiming at finding a constant factor approximation of the classification loss while we are focusing on the statistical problem of recovering $\beta^\star$. Yet another similar setting has been introduced in (Riquelme et al., 2017a). In this setting the agent has to estimate several linear models in parallel and for each covariate (that appears randomly), the agent has to decide which model to estimate. Other works studied the problem of active linear regression, and for example Sugiyama & Rubens (2008) proposed an algorithm conducting active learning and model selection simultaneously but without any theoretical guarantees. More recently Riquelme et al. (2017b) have studied the setting of active linear regression with thresholding techniques in the homoscedastic case. An active line of research has also been conducted in the domain of random design linear regression (Hsu et al., 2011; Sabato & Munos, 2014; Dereziński et al., 2019). In these works the authors aim at controlling the mean-squared regression error $\mathbb{E}[(X^\top \beta - Y)^2]$ with a minimum number of random samples $X_k$. Except from the loss function that they considered, these works differ from ours in several points: they generally do not consider the heteroscedastic case and their goal is to minimize the number of samples to use to reach an $\varepsilon$-estimator while in our setting the total number of covariates $K$ is fixed. Allen-Zhu et al. (2020) provide a similar analysis but under the scope of optimal experiment design. Another setting similar to ours is introduced in (Hazan & Karnin, 2014), where active linear regression with a hard-margin criterion is studied. However, the minimization of the classical $\ell^2$-norm of the difference between the true parameter of the linear model and its estimator seems to be a more natural criterion, which justifies our investigations.

In this work we adopt a different point of view from the aforementioned existing works. We consider A-optimal design under the heteroscedasticity assumption and we generalize MAB results to the non-coordinate basis setting with two different algorithms taking inspiration from the convex optimization and bandit literature. We prove optimal $\widetilde{\mathcal{O}}(T^{-2})$ regret bounds for $d$ covariates and provide a weaker guarantee for more than $d$ covariates. Our work emphasizes the connection between MAB and optimal design, closing open questions in A-optimal design. Finally we corroborate our theoretical findings with numerical experiments.

## 2. Setting and description of the problem

### 2.1. Motivations and description of the setting

Let $X_1, \dots, X_K \in \mathbb{R}^d$ be $K$ covariates available to some agent who can successively sample each of them (several times if needed). Observations $Y$ are generated by a standard linear model, *i.e.*,

$$Y = X^\top \beta^\star + \varepsilon \quad \text{with } \beta^\star \in \mathbb{R}^d \, .$$

Each of these covariates correspond to an experiment that can be run by the decision maker to gain information about the unknown vector $\beta^\star$. The goal of optimal experiment design is to choose the experiments to perform from a pool of possible design points $\{X_1, \dots, X_K\}$ in order to obtain the best estimate $\hat{\beta}$ of $\beta^\star$ within a fixed budget of $T \in \mathbb{N}^*$ samples. In classical experiment design problems the variances of the different experiments are supposed to be equal. Here we consider the more challenging setting where each covariate has a specific and *unknown* variance $\sigma_k^2$, *i.e.*, we suppose that when $X_k$ is queried for the $i$-th time the decision maker observes

$$Y_k^{(i)} = X_k^\top \beta^\star + \varepsilon_k^{(i)} \, ,$$

where $\mathbb{E}[\varepsilon_k^{(i)}] = 0$, $\mathrm{Var}[\varepsilon_k^{(i)}] = \sigma_k^2 > 0$ and $\varepsilon_k^{(i)}$ is $\kappa^2$-subgaussian. We assume also that the $\varepsilon_k^{(i)}$ are independent from each other. This setting corresponds actually to online optimal experiment design since the decision maker has to design sequentially the sampling policy, in an adaptive manner.

A naive sampling strategy is to equally sample each covariate $X_k$. In our heteroscedastic setting, this will not produce the most precise estimate of $\beta^\star$ because of the different variances $\sigma_k^2$. Intuitively a point $X_k$ with a low variance will provide very precise information on the value $X_k^\top \beta^\star$ while a point with a high variance will not give much information (up to the converse effect of the norm $\|X_k\|$). This indicates that a point with high variance should be sampled more often than a point with low variance. Since the variances $\sigma_k^2$ are unknown, we need at the same time to estimate $\sigma_k^2$ (which might require lots of samples of $X_k$ to be precise) and to minimize the estimation error (which might require only a few examples of some covariate $X_k$). There is then a tradeoff between gathering information on the values of $\sigma_k^2$ and using it to optimize the loss; the fact that this loss is global, and not cumulative, makes this tradeoff "exploration *vs.* exploitation" much more intricate than in standard multi-armed bandits.

Usual algorithms handling global losses are rather slow (Agrawal & Devanur, 2014; Mannor et al., 2014) or dedicated to specific well-posed problems with closed form losses (Antos et al., 2010; Carpentier et al., 2011). Our setting can be seen as an extension of the two aforementioned works who aim at estimating the means of a set of $K$ distributions. Noting $\mu = (\mu_1, \dots, \mu_K)^\top$ the vector of the means of those distributions and $X_i = e_i$ the $i^{\text{th}}$ vector of the canonical basis of $\mathbb{R}^K$, we see (since $X_i^\top \mu = \mu_i$) that their objective is actually to estimate the parameter $\mu$ of a linear model. This setting is a particular case of ours since the vectors $X_i$ form the canonical basis of $\mathbb{R}^K$.

### 2.2. Definition of the loss function

As we mentioned it before, the decision maker can be led to sample several times the same design point $X_k$ in order to obtain a more precise estimate of its response $X_k^\top \beta^\star$. We denote therefore by $T_k \geq 0$ the number of samples of $X_k$, hence $T = \sum_{k=1}^K T_k$. For each $k \in [K]$[1], the linear model yields the following

$$T_k^{-1} \sum_{i=1}^{T_k} Y_k^{(i)} = X_k^T \beta^\star + T_k^{-1} \sum_{i=1}^{T_k} \varepsilon_k^{(i)} \, .$$

We define $\tilde{Y}_k = \sum_{i=1}^{T_k} Y_k^{(i)} / \sigma_k \sqrt{T_k}$, $\tilde{X}_k = \sqrt{T_k} X_k / \sigma_k$ and $\tilde{\varepsilon}_k = \sum_{i=1}^{T_k} \varepsilon_k^{(i)} / \sigma_k \sqrt{T_k}$ so that for all $k \in [K]$, $\tilde{Y}_k = \tilde{X}_k^T \beta^\star + \tilde{\varepsilon}_k$, where $\mathbb{E}[\tilde{\varepsilon}] = 0$ and $\mathrm{Var}[\tilde{\varepsilon}_k] = 1$. We denote by $\mathbb{X} = (\tilde{X}_1^\top, \cdots, \tilde{X}_K^\top)^\top \in \mathbb{R}^{K \times d}$ the induced design matrix of the policy. Under the assumption that $\mathbb{X}$ has full rank, the above Ordinary Least Squares (OLS) problem has an optimal unbiased estimator $\hat{\beta} = (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top \tilde{Y}$. The overarching objective is to upper-bound $\mathbb{E}\|\hat{\beta} - \beta^\star\|^2$, which can be easily rewritten as follows:

$$\mathbb{E}\left[\|\hat{\beta} - \beta^\star\|^2\right] = \mathrm{Tr}((\mathbb{X}^\top \mathbb{X})^{-1}) = \mathrm{Tr}\left(\sum_{k=1}^K \tilde{X}_k \tilde{X}_k^\top\right)^{-1}$$
$$= \frac{1}{T} \mathrm{Tr}\left(\sum_{k=1}^K p_k X_k X_k^\top / \sigma_k^2\right)^{-1} \, ,$$

where we have denoted for every $k \in [K]$, $p_k = T_k / T$ the proportion of times the covariate $X_k$ has been sampled. By definition, $p = (p_1, \dots, p_K) \in \Delta^K$, the simplex of

[1] $[K] = \{1, \dots, K\}$.

dimension $K-1$. We emphasize here that minimizing $\mathbb{E}\|\hat{\beta}-\beta^\star\|^2$ is equivalent to minimizing the trace of the inverse of the covariance matrix $\mathbb{X}^\top\mathbb{X}$, which corresponds actually to A-optimal design (Pukelsheim, 2006). Denote now by $\Omega(p)$ the following weighted covariance matrix

$$\Omega(p)=\sum_{k=1}^{K}\frac{p_k}{\sigma_k^2}X_kX_k^\top=\mathbb{X}^\top\mathbb{X}\,.$$

The objective is to minimize over $p\in\Delta^K$ the loss function $L(p)=\mathrm{Tr}\left(\Omega(p)^{-1}\right)$ with $L(p)=+\infty$ if $(p\mapsto\Omega(p))$ is not invertible, such that

$$\mathbb{E}\left[\|\hat{\beta}-\beta^\star\|^2\right]=\frac{1}{T}\mathrm{Tr}\left(\Omega(p)^{-1}\right)=\frac{1}{T}L(p)\,.$$

For the problem to be non-trivial, we require that the covariates span $\mathbb{R}^d$. If it is not the case then there exists a vector along which one cannot get information about the parameter $\beta^\star$. The best algorithm we can compare against can only estimate the projection of $\beta$ on the subspace spanned by the covariates, and we can work in this subspace.

The rest of this work is devoted to design an algorithm minimizing $\mathrm{Tr}\left(\Omega(p)^{-1}\right)$ with the difficulty that the variances $\sigma_k^2$ are unknown. In order to do that we will sequentially and adaptively choose which point to sample to minimize $\mathrm{Tr}\left(\Omega(p)^{-1}\right)$. This corresponds consequently to online A-optimal design. As developed above, the norms of the covariates have a scaling role and those can be renormalized to lie on the sphere at no cost, which is thus an assumption from now on: $\forall k\in[K]$, $\|X_k\|_2=1$. The following proposition shows that the problem we are considering is convex.

**Proposition 1.** *$L$ is strictly convex on $\Delta^d$ and continuous in its relative interior $\mathring{\Delta}^d$.*

The proof is deferred to Appendix C. Proposition 1 implies that $L$ has a unique minimum $p^\star$ in $\mathring{\Delta}^d$ and we note

$$p^\star=\operatorname*{arg\,min}_{p\in\Delta^d}L(p)\,.$$

Finally, we evaluate the performance of a sampling policy in term of "regret" *i.e.*, the difference in loss between the optimal sampling policy and the policy in question.

**Definition 1.** *Let $p_T$ denote the sampling proportions after $T$ samples of a policy. Its regret is then*

$$R(T)=\frac{1}{T}\left(L(p_T)-L(p^\star)\right)\,.$$

We will construct active sampling algorithms to minimize $R(T)$. A key step is the following computations of the gradient of $L$. Since $\nabla_k\Omega(p)=X_kX_k^T/\sigma_k^2$, it follows

$$\partial_{p_k}L(p)=-\frac{1}{\sigma_k^2}\mathrm{Tr}\left(\Omega(p)^{-2}X_kX_k^T\right)=-\frac{1}{\sigma_k^2}\left\|\Omega(p)^{-1}X_k\right\|_2^2.$$

As in several works (Hsu et al., 2011; Allen-Zhu et al., 2020) we will have to study different cases depending on the values of $K$ and $d$. The first one corresponds to the case $K\leq d$. As we explained it above, if $K<d$, the matrix $\Omega(p)$ is not invertible and it is impossible to obtain a sublinear regret, which makes us work in the subspace spanned by the covariates $X_k$. This corresponds to $K=d$. We will treat this case in Sections 3 and 4. The case $K>d$ is considered in Section 5.

## 3. A naive randomized algorithm

We begin by proposing an obvious baseline for the problem at hand. One naive algorithm would be to estimate the variances of each of the covariates by sampling them a fixed amount of time. Sampling each arm $cT$ times (with $c<1/K$) would give an approximation $\hat{\sigma}_k$ of $\sigma_k$ of order $1/\sqrt{T}$. Then we can use these values to construct $\hat{\Omega}(p)$ an approximation of $\Omega(p)$ and then derive the optimal proportions $\hat{p}_k$ to minimize $\mathrm{Tr}(\hat{\Omega}(p)^{-1})$. Finally the algorithm would consist in using the remainder of the budget to sample the arms according to those proportions. However, such a trivial algorithm would not provide good regret guarantees. Indeed the constant fraction $c$ of the samples used to estimate the variances has to be chosen carefully; it will lead to a $1/T$ regret if $c$ is too big (if $c>p_k^\star$ for some $k$). That is why we need to design an algorithm that will first roughly estimate the $p_k^\star$. In order to improve the algorithm it will also be useful to refine at each iteration the estimates $\hat{p}_k$. Following these ideas we propose Algorithm 1 which uses a pre-sampling phase (see Lemma 3 for further details) and which constructs at each iteration lower confidence estimates of the variances, providing an optimistic estimate $\tilde{L}$ of the objective function $L$. Then the algorithm minimizes this estimate (with an offline A-optimal design algorithm, see *e.g.*, (Gao et al., 2014)). Finally the covariate $X_k$ is sampled with probability $\hat{p}_{t,k}$. Then feedback is collected and estimates are updated.

**Proposition 2.** *For $T\geq 1$ samples, running Algorithm 1 with $N_i=p_i^oT/2$ (with $p^o$ defined by (2)) for all $i\in[K]$, gives final sampling proportions $p_T$ such that*

$$R(T)=\mathcal{O}_{\Gamma,\sigma_k}\left(\frac{\sqrt{\log T}}{T^{3/2}}\right)\,,$$

*where $\Gamma$ is the Gram matrix of $X_1,\dots,X_K$.*

The proof is postponed to Appendix D. Notice that we avoid the problem discussed by Erraqabi et al. (2017) (that is due to infinite gradient on the simplex boundary) thanks to presampling, allowing us to have positive empirical variance estimates with high probability.

**Algorithm 1** Naive randomized algorithm

**Require:** $d$, $T$, $\delta$ confidence parameter
**Require:** $N_1, \dots, N_d$ of sum $N$
Sample $N_k$ times each covariate $X_k$
$p_N \longleftarrow (N_1/N, \dots, N_d/N)$
Compute empirical variances $\hat{\sigma}_1^2, \dots, \hat{\sigma}_d^2$
**for** $N+1 \leq t \leq T$ **do**
  Compute $\hat{p}_t \in \arg\min \tilde{L}$, where $\tilde{L}$ is the same function as $L$, but with variances replaced by lower confidence estimates of the variances (from Theorem 1).
  Draw $\pi(t)$ randomly according to probabilities $\hat{p}_t$ and sample covariate $X_{\pi(t)}$
  Update $p_{t+1} = p_t + \frac{1}{t+1}(e_{\pi(t+1)} - p_t)$ and $\hat{\sigma}^2_{\pi(t)}$ where $(e_1, \dots, e_d)$ is the canonical basis of $\mathbb{R}^d$.
**end for**

## 4. A faster first-order algorithm

We now improve the relatively "slow" dependency in $T$ in the rates of Algorithm 1 – due to its naive reduction to a MAB problem, and because it does not use any estimates of the gradient of $L$ – with a different approach based on convex optimization techniques, that we can leverage to gain an order in the rates of convergence.

### 4.1. Description of the algorithm

The main algorithm is described in Algorithm 2 and is built following the work of Berthet & Perchet (2017). The idea is to sample the arm sampled which minimizes the norm of a proxy of the gradient of $L$, corrected by a positive error term, as in the UCB algorithm (Auer et al., 2002).

**Algorithm 2** Bandit algorithm

**Require:** $d$, $T$
**Require:** $N_1, \dots, N_d$ of sum $N$
Sample $N_k$ times each covariate $X_k$
$p_N \longleftarrow (N_1/N, \dots, N_d/N)$
Compute empirical variances $\hat{\sigma}_1^2, \dots, \hat{\sigma}_d^2$
**for** $N+1 \leq t \leq T$ **do**
  Compute $\nabla \hat{L}(p_t)$, where $\hat{L}$ is the same function as $L$, but with variances replaced by empirical variances.
  **for** $k \in [d]$ **do**
    $\hat{g}_k \longleftarrow \nabla_k \hat{L}(p_t) - 2\sqrt{\frac{3\log(t)}{T_k}}$
  **end for**
  $\pi(t) \longleftarrow \arg\min_{k \in [d]} \hat{g}_k$ and sample covariate $X_{\pi(t)}$
  Update $p_{t+1} = p_t + \frac{1}{t+1}(e_{\pi(t+1)} - p_t)$ and update $\hat{\sigma}^2_{\pi(t)}$
**end for**

$N_1, \dots, N_d$ are the number of times each covariate is sampled at the beginning of the algorithm. This stage is needed to ensure that $L$ is smooth. More details about that will be given with Lemma 3.

### 4.2. Concentration of the gradient of the loss

The cornerstone of the algorithm is to guarantee that the estimates of the gradients concentrate around their true value. To simplify notations, we denote by $G_k = \partial_{p_k} L(p)$ the true $\mathrm{k}^{\text{th}}$ derivative of $L$ and by $\hat{G}_k$ its estimate. More precisely, if we note $\hat{\Omega}(p) = \sum_{k=1}^K (p_k/\hat{\sigma}_k) X_k X_k^\top$, we have

$$G_k = -\sigma_k^{-2} \|\Omega(p)^{-1} X_k\|_2^2, \ \hat{G}_k \doteq -\hat{\sigma}_k^{-2} \|\hat{\Omega}(p)^{-1} X_k\|_2^2 \, .$$

Since $\hat{G}_k$ depends on the $\hat{\sigma}_k^2$, we need a concentration bound on the empirical variances $\hat{\sigma}_k^2$. As traditional results on the concentration of the variances (Maurer & Pontil, 2009; Carpentier et al., 2011) are generally obtained in the bounded setting, we prove in Appendix A the following bound in the case of subgaussian random variables.

**Theorem 1.** *Let $X$ be a centered and $\kappa^2$-sub-gaussian random variable sampled $n \geq 2$ times. Let $\delta \in (0,1)$. Let $c = (e-1)(2e(2e-1))^{-1} \approx 0.07$. With probability at least $1-\delta$, the following concentration bound on its empirical variance holds*

$$\left|\hat{\sigma}_n^2 - \sigma^2\right| \leq 3\kappa^2 \cdot \max\left(\frac{\log(4/\delta)}{cn}, \sqrt{\frac{\log(4/\delta)}{cn}}\right) .$$

Using Theorem 1 we claim the following concentration argument, which is the main ingredient of the analysis of Algorithm 2.

**Proposition 3.** *For every $k \in [K]$, after having gathered $T_k \leq T$ samples of covariates $X_k$, there exists a constant $\mathtt{C} > 0$ (explicit and given in the proof) such that, with probability at least $1-\delta$*

$$|G_k - \hat{G}_k| \leq \mathtt{C}\left(\sigma_k^{-1} \max_{i \in [K]} \frac{\sigma_i^2}{p_i}\right)^3 \cdot \max\left(\frac{\log(4TK/\delta)}{T_k}, \sqrt{\frac{\log(4TK/\delta)}{T_k}}\right) .$$

For clarity reasons we postpone the proof to Appendix B. Proving this proposition was one of the main technical challenges of our analysis. Now that we have it proven we can turn to the analysis of Algorithm 2.

### 4.3. Analysis of the convergence of the algorithm

In convex optimization several classical assumptions can be leveraged to derive fast convergence rates. Those assumptions are typically strong convexity, positive distance from

the boundary of the constraint set, and smoothness of the objective function, *i.e.*, that it has Lipschitz gradient. We prove in the following that the loss $L$ satisfies them, up to the smoothness because its gradient explodes on the boundary of $\Delta^d$. However, $L$ is smooth on the relative interior of the simplex. Consequently we will circumvent this smoothness issue by using a technique from (Fontaine et al., 2019) consisting in pre-sampling every arm a linear number of times in order to force $p$ to be far from the boundaries of $\Delta^d$.

Using the following notations $\mathbb{X}_0 \doteq (X_1^\top, \cdots, X_d^\top)^\top$ and $\Gamma \doteq \mathbb{X}_0\mathbb{X}_0^\top = \mathrm{Gram}(X_1, \ldots, X_d)$ we prove the following lemma in Appendix E.1.

**Lemma 1.** *The loss function $L$ verifies for all $p \in \Delta^d$,*

$$L(p) = \frac{1}{\det(\mathbb{X}_0^\top \mathbb{X}_0)} \sum_{k=1}^{d} \frac{\sigma_k^2}{p_k} \mathrm{Cof}(\mathbb{X}_0\mathbb{X}_0^\top)_{kk} \ .$$

With this expression, the optimal proportion $p^\star$ can be easily computed using the KKT theorem, with the following closed form:

$$p_k^\star = \sigma_k \sqrt{\mathrm{Cof}(\Gamma)_{kk}} / \sum_{i=1}^{d} \sigma_i \sqrt{\mathrm{Cof}(\Gamma)_{ii}} \ . \tag{1}$$

This yields that $L$ is $\mu$-**strongly convex** on $\Delta^d$, with $\mu = 2\det(\Gamma)^{-1} \min_i \mathrm{Cof}(\Gamma)_{ii}\sigma_i^2$. Moreover, this also implies that $p^\star$ is **far away from the boundary** of $\Delta^d$.

**Lemma 2.** *Let $\eta \doteq dist(p^\star, \partial\Delta^d)$ be the distance from $p^\star$ to the boundary of the simplex. We have*

$$\eta = \sqrt{\frac{K}{K-1}} \frac{\min_i \sigma_i \sqrt{\mathrm{Cof}(\Gamma)_{ii}}}{\sum_{k=1}^{d} \sigma_k \sqrt{\mathrm{Cof}(\Gamma)_{kk}}} \ .$$

*Proof.* This is immediate with (1) since $\eta = \sqrt{\frac{K}{K-1}} \min_i p_i^\star$. □

It remains to recover the smoothness of $L$. This is done using a pre-sampling phase.

**Lemma 3** (see (Fontaine et al., 2019))**.** *If there exists $\alpha \in (0, 1/2)$ and $p^o \in \Delta^d$ such that $p^\star \succcurlyeq \alpha p^o$ (component-wise) then sampling arm $i$ at most $\alpha p_i^o T$ times (for all $i \in [d]$) at the beginning of the algorithm and running Algorithm 2 is equivalent to running Algorithm 2 with budget $(1-\alpha)T$ on the smooth function $(p \mapsto L(\alpha p^o + (1-\alpha)p)$.*

We have proved that $p_k^\star$ is bounded away from 0 and thus a pre-sampling would be possible. However, this requires to have some estimate of each $\sigma_k^2$. The upside is that those estimates must be accurate up to some multiplicative factor (and not additive factor) so that a logarithmic number of samples of each arm is enough to get valid lower/upper bounds (see Corollary 1). Indeed, the estimate $\overline{\sigma}_k^2$ obtained satisfies, for each $k \in [d]$, that $\sigma_k^2 \in [\overline{\sigma}_k^2/2, 3\overline{\sigma}_k^2/2]$. Consequently we know that

$$\forall k \in [d], p_k^\star \geq \frac{1}{\sqrt{3}} \frac{\overline{\sigma}_k \sqrt{\mathrm{Cof}(\Gamma)_{kk}}}{\sum_{i=1}^{d} \overline{\sigma}_i \sqrt{\mathrm{Cof}(\Gamma)_{ii}}} \geq \frac{1}{2} p^o, \tag{2}$$

$$\text{where } p^o = \frac{\overline{\sigma}_k \sqrt{\mathrm{Cof}(\Gamma)_{kk}}}{\sum_{i=1}^{d} \overline{\sigma}_i \sqrt{\mathrm{Cof}(\Gamma)_{ii}}} \ .$$

This will let us use Lemma 3 and with a presampling stage as prescribed, $p$ is forced to remain far away from the boundaries of the simplex in the sense that $p_{t,i} \geq p_i^o/2$ at each stage $t$ subsequent to the pre-sampling, and for all $i \in [d]$. Consequently, this logarithmic phase of estimation plus the linear phase of pre-sampling ensures that in the remaining of the process, $L$ is actually smooth.

**Lemma 4.** *With the pre-sampling of Lemma 3, $L$ is smooth with constant $C_S$ where*

$$C_S \leq 432 \frac{\sigma_{\max}^2 \left(\sum_{k=1}^{d} \sigma_k \sqrt{\mathrm{Cof}(\Gamma)_{kk}}\right)^3}{\det(\Gamma)\sigma_{\min}^3 \sqrt{\min_k \mathrm{Cof}(\Gamma)_{kk}}} \ .$$

The proof is deferred to Appendix E.2. We can now state our main theorem that is proved in Appendix E.3.

**Theorem 2.** *Applying Algorithm 2 with $T \geq 1$ samples after having pre-sampled each arm $k \in [d]$ at most $p_k^o T/2$ times gives the following bound*[2]

$$R(T) = \mathcal{O}_{\Gamma,\sigma_k}\left(\frac{\log^2(T)}{T^2}\right) \ .$$

This theorem provides a fast convergence rate for the regret $R$ and emphasizes the importance of using the gradient information in Algorithm 2 compared to Algorithm 1.

## 5. Discussion and generalization to $K > d$

We discuss in this section the case where the number $K$ of covariate vectors is greater than $d$.

### 5.1. Discussion of the case $K > d$

In the case where $K > d$ it may be possible that the optimal $p^\star$ lies on the boundary of the simplex $\Delta^K$, meaning that some arms should not be sampled. This happens for instance as soon as there exist two covariate points that are exactly equal but with different variances. The point with the lowest variance should be sampled while the point with the highest

[2] The notation $\mathcal{O}_{\Gamma,\sigma_k}$ means that there is a hidden constant depending on $\Gamma$ and on the $\sigma_k$. The explicit dependency on these parameters is given in the proof.

one should not. All the difficulty of an algorithm for the case where $K > d$ is to be able to detect which covariate should be sampled and which one should not. In order to adopt another point of view on this problem it might be interesting to go back to the field of optimal design of experiments. Indeed by choosing $v_k = X_k/\sigma_k$, our problem consists exactly in the following constraint minimization problem given $v_1 \dots, v_K \in \mathbb{R}^d$:

$$\min \mathrm{Tr}\left(\sum_{j=1}^{K} p_j v_j v_j^\top\right)^{-1} \quad \text{under contraints } p \in \Delta^K \ . \quad \text{(P)}$$

It is known (Pukelsheim (2006)) that the dual problem of A-optimal design consists in finding the smallest ellipsoid, in some sense, containing all the points $v_j$:

$$\max \mathrm{Tr}(\sqrt{W})^2 \quad \text{(D)}$$
$$\text{u.c. } W \succ 0^3 \text{ and } v_j^\top W v_j \leq 1 \text{ for all } 1 \leq j \leq K \ .$$

In our case the role of the ellipsoid can be easily seen with the KKT conditions. We obtain the following proposition, proved in Appendix G.1.

**Proposition 4.** *The points $X_k/\sigma_k$ lie within the ellipsoid defined by the matrix $\Omega(p^\star)^{-2}$.*

This geometric interpretation shows that a point $X_k$ with high variance is likely to be in the interior of the ellipsoid (because $X_k/\sigma_k$ is close to the origin), meaning that $\mu_k > 0$ and therefore that $p_k^\star = 0$ *i.e.*, that $X_k$ should not be sampled. Nevertheless since the variances are unknown, one is not easily able to find which point has to be sampled. Figures illustrating the geometric interpretation can be found in Appendix G.2.

### 5.2. A theoretical upper-bound and a lower bound

We derive now a bound for the convergence rate of Algorithm 2 in the case where $K > d$.

**Theorem 3.** *Applying Algorithm 2 with $K > d$ covariate points gives the following bound on the regret:*

$$R(T) = \mathcal{O}\left(\log(T) T^{-5/4}\right) \ .$$

The proof is postponed to Appendix F.1.

One can ask whether this result is optimal, and if it is possible to reach the bound of Theorem 2. The following theorem provides a lower bound showing that it is impossible in the case where there are $d$ covariates. However the upper and lower bounds of Theorems 3 and 4 do not match. It is still an open question whether we can obtain better rates than $T^{-5/4}$.

**Theorem 4.** *In the case where $K > d$, for any algorithm on our problem, there exists a set of parameters such that $R(T) \gtrsim T^{-3/2}$.*

We prove Theorem 4 in Appendix F.2.

## 6. Numerical simulations

We now present numerical experiments to validate our results and claims. We compare several algorithms for active matrix design: a very naive algorithm that samples equally each covariate, Algorithm 1, Algorithm 2 and a Thompson Sampling (TS) algorithm (Thompson, 1933). We run our experiments on synthetic data with horizon time $T$ between $10^4$ and $10^6$, averaging the results over 25 rounds. We consider covariate vectors in $\mathbb{R}^K$ of unit norm for values of $K$ ranging from 3 to 100. All the experiments ran in less than 15 minutes on a standard laptop.

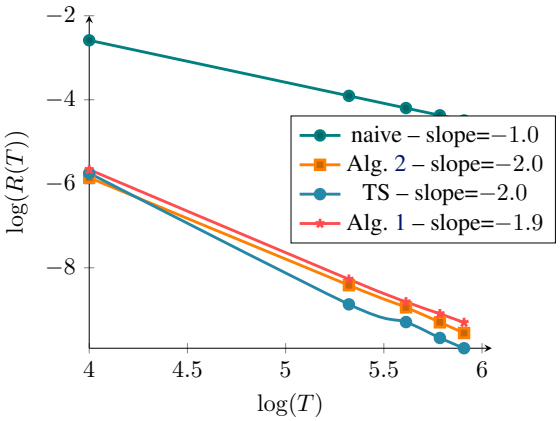


*Figure 1.* Regret as a function of $T$ in log–log scale in the case of $K = 3$ covariates in $\mathbb{R}^3$.

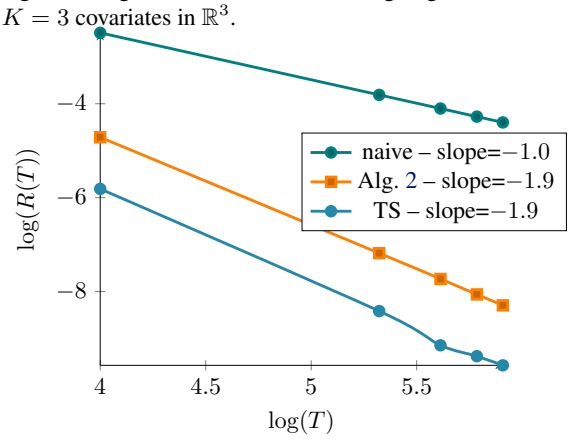


*Figure 2.* Regret as a function of $T$ in log–log scale in the case of $K = 4$ covariates in $\mathbb{R}^3$.

Let us quickly describe the Thompson Sampling algorithm. We choose Normal Inverse Gamma distributions for priors

for the mean and variance of each of the arms, as they are the conjugate priors for gaussian likelihood with unknown mean and variance. At each time step $t$, for each arm $k \in [K]$, a value of $\hat{\sigma}_k$ is sampled from the prior distribution. An approximate value of $\nabla_k L(p)$ is computed with the $\hat{\sigma}_k$ values. The arm with the lowest gradient value is chosen and sampled. The value of this arm updates the hyperparameters of the prior distribution.

In our first experiment we consider only 3 covariate vectors. We plot the results in log–log scale in order to see the convergence speed which is given by the slope of the plot. Results on Figure 1 show that both Algorithms 1 and 2, as well as Thompson sampling have regret $\mathcal{O}(1/T^2)$ as expected.

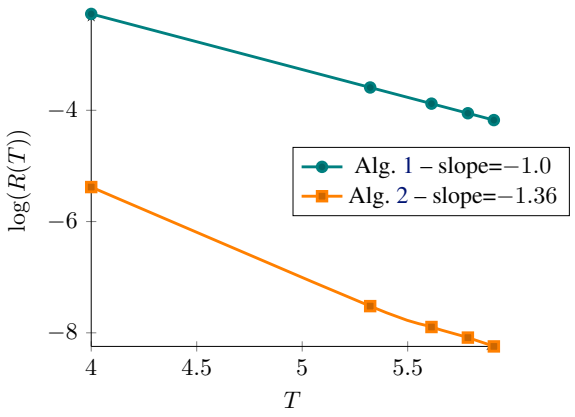


*Figure 3.* Regret as a function of $T$ in log–log scale in the case of $K = 4$ covariates in $\mathbb{R}^3$ in a challenging setting.

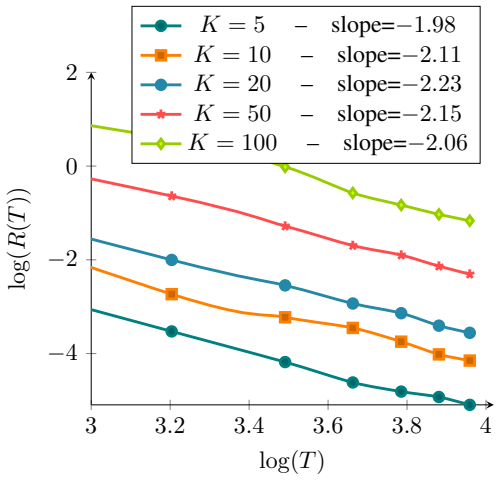


*Figure 4.* Regret as a function of $T$ for different values of $K$ in log–log scale.

We see that Thompson Sampling performs well on low-dimensional data. However it is approximately 200 times slower than Algorithm 2 – due to the sampling of complex Normal Inverse Gamma distributions – and therefore inefficient in practice. On the contrary, Algorithm 2 is very practical. Indeed its computational complexity is linear in time $T$ and its main computational cost is due to the computation of the gradient $\nabla \hat{L}$. This relies on inverting $\hat{\Omega} \in \mathbb{R}^{d \times d}$, whose complexity is $\mathcal{O}(d^3)$ (or even $\mathcal{O}(d^{2.807})$ with Strassen algorithm). Thus the overall complexity of Algorithm 2 is $\mathcal{O}(T(d^{2.8} + K))$ hence polynomial. This computational complexity advocates that Algorithm 2 is practical for moderate values of $d$, as in linear regression problems.

Figure 1 shows that Algorithm 1 performs nearly as well as Algorithm 2. However, the minimization step of $\hat{L}$ is time-consuming when $K > d$, since there is no close form for $p^\star$, which leads to approximate results. Therefore Algorithm 1 is not adapted to $K > d$. We also have conducted similar experiments in this case, with $K = d + 1$. The offline solution of the problem indicates that one covariate should not be sampled, *i.e.*, $p^\star \in \partial\Delta^K$. Results presented on Figure 2 prove the performances of Algorithm 2.

One might argue that the positive results of Figure 2 are due to the fact that it is "easy" for the algorithm to detect that one covariate should not be sampled, in the sense that this covariate clearly lies in the interior of the ellipsoids mentioned in Section 5.1. In the very challenging case where two covariates are equal but with variances separated by only $1/\sqrt{T}$, we obtain the results described on Figure 3. The observed experimental convergence rate is of the order of $T^{-1.36}$ which is much slower than the rates of Figure 2, and between the rates proved in Theorems 3 and Theorem 4.

Finally we run a last experiment with larger values of $K = d$. We plot the convergence rate of Algorithm 2 for values of $K$ ranging from 5 to 100 in $\log - \log$ scale on Figure 4. The slope is again approximately of $-2$, which is coherent with Theorem 2. We note furthermore that larger values of $d$ do not make Algorithm 2 impracticable, as inferred by its cubic complexity.

## 7. Conclusion

We have proposed an algorithm mixing bandit and convex optimization techniques to solve the problem of online A-optimal design, which is related to active linear regression with repeated queries. This algorithm has proven fast and optimal rates $\tilde{\mathcal{O}}(T^{-2})$ in the case of $d$ covariates that can be sampled in $\mathbb{R}^d$. One cannot obtain such fast rates in the more general case of $K > d$ covariates. We have therefore provided weaker results in this very challenging setting and conducted more experiments showing that the problem is indeed more difficult.

## Acknowledgements

Xavier Fontaine was supported by grants from Région Ile-de-France.

## References

Agrawal, S. and Devanur, N. R. Bandits with concave rewards and convex knapsacks. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pp. 989–1006, 2014.

Allen-Zhu, Z., Li, Y., Singh, A., and Wang, Y. Near-optimal discrete optimization for experimental design: A regret minimization approach. *Mathematical Programming*, pp. 1–40, 2020.

Antos, A., Grover, V., and Szepesvári, C. Active learning in heteroscedastic noise. *Theoretical Computer Science*, 411(29-30):2712–2728, 2010.

Atkeson, L. R. and Alvarez, R. M. *The Oxford handbook of polling and survey methods*. Oxford University Press, 2018.

Auer, P., Cesa-Bianchi, N., and Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.*, 47(2-3):235–256, May 2002. ISSN 0885-6125.

Berthet, Q. and Perchet, V. Fast rates for bandit optimization with upper-confidence frank-wolfe. In *Advances in Neural Information Processing Systems*, pp. 2225–2234, 2017.

Boyd, S. and Vandenberghe, L. *Convex optimization*. Cambridge university press, 2004.

Carpentier, A., Lazaric, A., Ghavamzadeh, M., Munos, R., and Auer, P. Upper-confidence-bound algorithms for active learning in multi-armed bandits. In *International Conference on Algorithmic Learning Theory*, pp. 189–203. Springer, 2011.

Chen, X. and Price, E. Active regression via linear-sample sparsification. In Beygelzimer, A. and Hsu, D. (eds.), *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pp. 663–695, Phoenix, USA, 25–28 Jun 2019. PMLR. URL http://proceedings.mlr.press/v99/chen19a.html.

Cohn, D. A., Ghahramani, Z., and Jordan, M. I. Active learning with statistical models. *Journal of artificial intelligence research*, 4:129–145, 1996.

Cowan, W., Honda, J., and Katehakis, M. N. Normal bandits of unknown means and variances: Asymptotic optimality, finite horizon regret bounds, and a solution to an open problem. *arXiv preprint arXiv:1504.05823*, 2015.

Dereziński, M., Warmuth, M. K., and Hsu, D. Unbiased estimators for random design regression. *arXiv preprint arXiv:1907.03411*, 2019.

Erraqabi, A., Lazaric, A., Valko, M., Brunskill, E., and Liu, Y.-E. Trading off Rewards and Errors in Multi-Armed Bandits. In Singh, A. and Zhu, J. (eds.), *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pp. 709–717, Fort Lauderdale, FL, USA, 20–22 Apr 2017. PMLR. URL http://proceedings.mlr.press/v54/erraqabi17a.html.

Fontaine, X., Berthet, Q., and Perchet, V. Regularized contextual bandits. In Chaudhuri, K. and Sugiyama, M. (eds.), *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pp. 2144–2153. PMLR, 16–18 Apr 2019. URL http://proceedings.mlr.press/v89/fontaine19a.html.

Freund, Y., Seung, H. S., Shamir, E., and Tishby, N. Selective sampling using the query by committee algorithm. *Machine learning*, 28(2-3):133–168, 1997.

Gao, W., Chan, P. S., Ng, H. K. T., and Lu, X. Efficient computational algorithm for optimal allocation in regression models. *Journal of Computational and Applied Mathematics*, 261:118–126, 2014.

Goos, P. and Jones, B. *Optimal design of experiments: a case study approach*. John Wiley & Sons, 2011.

Hazan, E. and Karnin, Z. Hard-margin active linear regression. In *International Conference on Machine Learning*, pp. 883–891, 2014.

Hsu, D., Kakade, S. M., and Zhang, T. An analysis of random design linear regression. *arXiv preprint arXiv:1106.2363*, 2011.

Kirschner, J. and Krause, A. Information directed sampling and bandits with heteroscedastic noise. In Bubeck, S., Perchet, V., and Rigollet, P. (eds.), *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pp. 358–384. PMLR, 06–09 Jul 2018. URL http://proceedings.mlr.press/v75/kirschner18a.html.

Mannor, S., Perchet, V., and Stoltz, G. Approachability in unknown games: Online learning meets multi-objective optimization. In *Conference on Learning Theory*, pp. 339–355, 2014.

Maurer, A. and Pontil, M. Empirical bernstein bounds and sample variance penalization. *arXiv preprint arXiv:0907.3740*, 2009.

Pukelsheim, F. *Optimal design of experiments*. SIAM, 2006.

Riquelme, C., Ghavamzadeh, M., and Lazaric, A. Active learning for accurate estimation of linear models. In Precup, D. and Teh, Y. W. (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 2931–2939, International Convention Centre, Sydney, Australia, 06–11 Aug 2017a. PMLR. URL http://proceedings.mlr.press/v70/riquelme17a.html.

Riquelme, C., Johari, R., and Zhang, B. Online active linear regression via thresholding. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017b.

Sabato, S. and Munos, R. Active regression by stratification. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems 27*, pp. 469–477. Curran Associates, Inc., 2014. URL http://papers.nips.cc/paper/5468-active-regression-by-stratification.pdf.

Sagnol, G. *Optimal design of experiments with application to the inference of traffic matrices in large networks: second order cone programming and submodularity*. PhD thesis, École Nationale Supérieure des Mines de Paris, 2010.

Soare, M. *Sequential Resource Allocation in Linear Stochastic Bandits* . Theses, Université Lille 1 - Sciences et Technologies, December 2015. URL https://hal.archives-ouvertes.fr/tel-01249224.

Sugiyama, M. and Rubens, N. Active learning with model selection in linear regression. In *Proceedings of the 2008 SIAM International Conference on Data Mining*, pp. 518–529. SIAM, 2008.

Thompson, W. R. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

Yang, M., Biedermann, S., and Tang, E. On optimal designs for nonlinear models: a general and efficient algorithm. *Journal of the American Statistical Association*, 108(504): 1411–1420, 2013.