Dynamic Data-Driven Estimation of Non-Parametric Choice Models

Nam Ho-Nguyen[1] and Fatma Kılınç-Karzan[1]

[1]Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, 15213, USA.

February 13, 2017; Revised May 19, 2017

Abstract

We study non-parametric estimation of consumer choice models. Non-parametric approaches were introduced to alleviate unreasonable assumptions and issues of suboptimal model fit/selection present in traditional parametric approaches, and are prevalent in several application areas. We present a generic convex optimization-based framework to efficiently learn a *simple* non-parametric choice model from data that is close to the best-fitting one, where simplicity is measured by sparsity of the underlying distribution. As opposed to the existing literature, we propose a solution method which enjoys provable convergence guarantees and extends naturally to the dynamic observation setting. Our computational study on the dynamic setting reveals the true impact of how much data are needed and at what rate to achieve the best trade-off in terms of estimation accuracy and model simplicity.

1 Introduction

A choice model specifies the probability distribution for rankings that consumers have over a set of items. Such models give choice probabilities, that is, the probability that a given consumer will choose an item from a given subset. Choice models are prevalent in several application areas such as revenue management, web page ranking, betting theory, social choice, marketing, and economics (see [7–9, 12, 19] and references therein). A good choice model aims to capture complex substitution behaviors of consumers in order to accurately predict demand from limited observations.

Choice model estimation has received quite a bit of interest. Traditional choice models often specify a *parametric* structure for the probability distribution (examples include the multinomial logit (MNL), nested logit, and mixed MNL models), see [19] and references therein. Imposing a parametric structure makes estimation of the necessary parameters a simpler task, but is often at the expense of overly facile assumptions on consumer behavior (such as independence of irrelevant alternatives in MNL models) preventing us from accurately capturing the substitution behavior. Because of this, the *non-parametric* approach of estimating the probability distribution for rankings directly have drawn growing interest in academia and practice [10, 17], and is shown through case studies [10] to lead to substantial improvement in prediction accuracy. However, learning (or even specifying) a full non-parametric model is intractable even for moderate-sized problems, since there are a factorial number of probabilities to estimate. Furthermore, today's technology enables us to collect data on a continuous basis. Unfortunately, existing methods have a notable deficiency: in order to update their estimates with the latest data, they must re-solve their choice models

entirely. In the case of non-parametric choice estimation, such dynamic considerations significantly compound the existing computational challenges.

Related literature. Earliest studies on non-parametric choice models appear in the economics and psychology literatures, e.g., Block and Marschak [4]. Mahajan and van Ryzin [14] showed that non-parametric models capture a number of parametric models as special cases. There are three main approaches employed for the *static* estimation of non-parametric choice models. Farias et al. [10] outline a constraint sampling based method to estimate the non-parametric model from observational data on consumer choices. Other methods based on maximum likelihood estimation (MLE) and norm minimization are presented by van Ryzin and Vulcano [20] and Bertsimas and Mišic [3] respectively. While [10] and [20] provide useful recovery results under some assumptions on the observational data, none of the methods in [3, 10, 20] come with provable convergence guarantees. Moreover, these prior methods are not equipped to deal with the more realistic *dynamic* data setting, where the firm continuously collect and wishes to utilize more consumer data even as we implement an estimation procedure.

Our contributions. We present a generic yet simple iterative framework for dynamically estimating non-parametric choice models and tackle the issues of the previous non-parametric approaches outlined above. We formulate the choice model estimation problem as the minimization of a generic distance metric between the theoretical distribution and an empirical one from the observations. Consequently, the generic nature of our distance metric allows us to present a unified view of the existing methods [3, 10, 20]. Unlike the existing literature, in this generic framework, using tools from convex duality and online convex optimization, we derive, for the first time, provably efficient convergence guarantees, i.e., we establish the number of iterations needed in order to estimate the choice model to within a given accuracy ϵ. Our analysis also upper-bounds the sparsity of our estimated model. As a result, our analysis exposes an explicit trade-off between the simplicity of the non-parametric choice model and its estimation accuracy. Furthermore, by employing a joint estimation-optimization (JEO) framework [2, 11], for the first time in the literature, we show that our method extends easily and efficiently to the *dynamic* setting, where new empirical observations are continuously collected and incorporated as new data into the choice model estimation process. On the algorithmic side, our general solution method enables us to recover primal solutions to convex optimization problems via a dual formulation. In particular, this technique can naturally be adapted to the dynamic case via the JEO framework, where the problem data are updated continuously. We also carry out a computational study, where we test the behavior of different distance metrics (ℓ_p-norms) in terms of estimation accuracy and sparsity of our learned models as well as the impact of number of observations and batch size used in the dynamic data setting on the same criteria. Our numerical results indicate that several choices in designing a generic dynamic setup mainly influence the simplicity of the estimated choice model as opposed to its accuracy.

Notation. For a positive integer $n \in \mathbb{N}$, we let $[n] = \{1, \ldots, n\}$, define $\Delta_n := \{x \in \mathbb{R}^n_+ : \sum_{i\in[n]} x_i = 1\}$ to be the standard simplex, and S_n to be the collection of rankings of the set $[n]$. We refer to a collection of objects b_j, $j \in J$ by the notation $\{b_j\}_{j\in J}$. Throughout the paper, the superscript, e.g., y^t, z^t, f^t, is used to attribute items to the t-th time period or iteration. The subscript is used to denote coordinates of a vector or matrix, e.g., β_{ij}. Given vectors x and y, $\langle x, y\rangle$ corresponds to the usual inner product of x and y. Given a norm $\|\cdot\|$ on a Euclidean space $\mathbb{E}$ and a real number $a > 0$, we denote its dual norm by $\|x\|_* = \min_y\{\langle x, y\rangle : \|x\| \leq 1\}$. We let $\partial f(x)$ be the subdifferential of f taken at x. We abuse notation slightly by denoting $\nabla f(x)$ for both the gradient of function f at x if f is differentiable and a subgradient of f at x, even if f

is not differentiable. If ϕ is of the form $\phi(x, y)$, then $\nabla_x \phi(x, y)$ denotes the subgradient of ϕ at x while keeping the other variables fixed at y. We denote the indicator function as $\mathbb{I}$, i.e., $\mathbb{I}(\mathcal{S}) = 1$ if statement $\mathcal{S}$ holds, and $\mathbb{I}(\mathcal{S}) = 0$ otherwise.

2 Data and Model

In the general non-parametric choice estimation framework, we consider a firm which has n products to sell. We assume that the firm chooses from a given set of m assortments $\mathcal{A}_1, \ldots, \mathcal{A}_m \subset [n]$, and the item 1 represents the 'no-buy' option and is present in all assortments $\mathcal{A}_j$, $j \in [m]$. Therefore, when presented with an assortment, the consumer will always choose an item from it (perhaps the no-buy option, i.e., item 1). We also denote $N := \sum_{j=1}^{m} |\mathcal{A}_j|$.

In this paper, we work with the following data collection process. When a consumer arrives, the firm displays an assortment $\mathcal{A}_j$ to the consumer. The consumer chooses a product $i_j \in \mathcal{A}_j$, and the firm observes this choice. A data set with K such observations will be a collection of pairs $\left\{i^k, \mathcal{A}^k\right\}_{k=1}^{K}$, where i^k denotes the item chosen and $\mathcal{A}^k$ denotes the assortment displayed for observation k. There are a number of useful statistics on this data set, which are defined as follows:

$$q_{ij} := \frac{1}{K} \sum_{k=1}^{K} \mathbb{I}(i^k = i, \mathcal{A}^k = \mathcal{A}_j) \quad \text{and} \quad q_j := \frac{1}{K} \sum_{k=1}^{K} \mathbb{I}(\mathcal{A}^k = \mathcal{A}_j) \tag{1}$$

$$p_{ij} := \frac{\sum_{k=1}^{K} \mathbb{I}(i^k = i, \mathcal{A}^k = \mathcal{A}_j)}{\sum_{k=1}^{K} \mathbb{I}(\mathcal{A}^k = \mathcal{A}_j)} = \frac{q_{ij}}{q_j}. \tag{2}$$

In words, q_{ij} is the proportion of observations where assortment $\mathcal{A}_j$ was displayed and item i was chosen, q_j is the proportion of observations where assortment $\mathcal{A}_j$ was displayed, and p_{ij} is the proportion of consumers who chose item i *given that* assortment $\mathcal{A}_j$ was displayed.

The non-parametric choice model is as follows. An incoming consumer will choose an item according to his/her ranking $\sigma \in S_n$ of the products in $[n]$, that is, when presented with an assortment $\mathcal{A}_j$, the consumer chooses the highest ranked product from $\mathcal{A}_j$, i.e., $\arg\min_{i \in \mathcal{A}_j} \sigma(i)$. The key assumption in this model is that the ranking σ of each incoming consumer is distributed i.i.d. according to some distribution λ on the set of all rankings S_n. Then λ represents the vector of probabilities of each $\sigma \in S_n$ being drawn, i.e., the probability that a consumer will have a ranking σ is $\lambda(\sigma)$.

For an item-assortment pair $i \in \mathcal{A}_j$, we define a binary vector $a_{ij} \in \{0, 1\}^{n!}$ to have entries $a_{ij}(\sigma) = \mathbb{I}\left(i = \arg\min_{i \in \mathcal{A}_j} \sigma(i)\right)$ for $\sigma \in S_n$. Thus, each entry of a_{ij} corresponds to a ranking σ, with $a_{ij}(\sigma) = 1$ if i is the highest ranked item in $\mathcal{A}_j$ according to σ, and $a_{ij}(\sigma) = 0$ otherwise. We define A to be the binary matrix of dimension $N \times n!$ with rows $a_{ij}^\top$. We denote the columns of A as the vector $a(\sigma) \in \{0, 1\}^N$, indexed by rankings $\sigma \in S_n$. Finally, for $j \in [m]$, we define A_j to be the submatrix of A with rows $a_{ij}^\top$ for $i \in \mathcal{A}_j$.

Then, based on our notation, given a distribution $\lambda \in \Delta_{n!}$ over rankings, the probability $\mathbb{P}_\lambda[i \mid \mathcal{A}_j]$ that a consumer chooses item i from $\mathcal{A}_j$ is simply represented as the inner product $\langle a_{ij}, \lambda \rangle$. Moreover, $A\lambda$ is nothing but the vector composed of the collection of probabilities $\{\mathbb{P}_\lambda[i \mid \mathcal{A}_j]\}_{i \in \mathcal{A}_j, j \in [m]}$. Finally, the probability distribution of a consumer choosing item $i \in \mathcal{A}_j$ given that they were offered assortment $\mathcal{A}_j$, i.e. $\{\mathbb{P}_\lambda[i \mid \mathcal{A}_j]\}_{i \in \mathcal{A}_j}$, is simply $A_j\lambda$. The statistics (1), (2) will be used to infer the best-fitting probability distribution λ, since p_{ij} is an empirical estimate of $\mathbb{P}[i \mid \mathcal{A}_j]$. We denote the collected vectors of $\{p_{ij}\}_{i \in \mathcal{A}_j, j \in [m]} = p \in \mathbb{R}^N$ and $\{p_{ij}\}_{i \in \mathcal{A}_j} = p_j \in \mathbb{R}^{|\mathcal{A}_j|}$ for $j \in [m]$.

3 Dynamic Learning of a Non-Parametric Choice Model

In Appendix A, we give a brief overview of three closely related approaches to learn the non-parametric choice model from data [3, 10, 20]. There, we also demonstrate that these seemingly disparate models used in [3, 10, 20] are special instantiations of a more general distance minimization framework, where λ is inferred by minimizing a distance measure $D(\cdot,\cdot)$ from the theoretical probabilities $A\lambda$ to the empirical observations p:

$$\min_{\lambda} \{D(A\lambda, p) : \ \lambda \in \Delta_{n!}\}. \tag{3}$$

In this paper, unlike prior literature, we provide a solution method equipped with *efficient convergence guarantees* for the *general* class of distance measures D. We assume that $D(\cdot, p)$ is convex and continuous on its domain; this is the case for the specific instantiations of D in [3, 10, 20] as well.

For a specified accuracy level $\epsilon > 0$, our goal is to obtain an (additive error) ϵ-approximate solution to this problem within a reasonable number of iterations. We also consider a dynamic variant of the problem where instead of fixed data p, we now have changing data p^t that converges to a limit p as we have more observations, i.e., $t \to \infty$. Under standard statistical assumptions, the estimates p^t obtained through (2) from a growing set of observations converge to the true distribution vector $p = \{\mathbb{P}[i, |\ \mathcal{A}_j]\}_{i\in\mathcal{A}_j, j\in[m]}$ (almost surely). In this setup we would still like to estimate λ, but we are only given access to the sequence $\{p^t\}_{t\geq 1}$. Also, note that (3) is intractable due to the $n!$ decision variables. To address these three challenges, we use three tools: convex duality, online convex optimization, and joint estimation-optimization.

3.1 Dual Formulation via Convex Conjugacy

Given that $D(\cdot, p)$ is convex and continuous on its domain, we can write $D(A\lambda, p)$ via its convex conjugate D^*:

$$D(A\lambda, p) = \sup_{y} \left\{ \langle A\lambda, y\rangle - D^*(y, p) : \ y \in \mathbb{R}^N \right\},$$
$$\text{where} \ \ D^*(y, p) := \sup_{z} \left\{ \langle y, z\rangle - D(z, p) : \ z \in \mathbb{R}^N \right\}.$$

Note that $D^*(y, p)$ is convex in y. In addition, if we assume that the (sub)gradients $\nabla_z D(z, p)$ are bounded $\|\nabla_z D(z, p)\| \leq R$ for some norm $\|\cdot\|$, then we can restrict the domain of y accordingly:

$$D(A\lambda, p) = \sup_{y:\|y\|\leq R} \left\{ \langle A\lambda, y\rangle - D^*(y, p) : \ y \in \mathbb{R}^N \right\}.$$

Henceforth, we denote $Y := \{y : \ \|y\| \leq R\}$ to be the corresponding norm ball. Based on these definitions, the problem (3) now admits a natural saddle point representation:

$$\mathrm{SV}(p) := \min_{\lambda} \{D(A\lambda, p) : \ \lambda \in \Delta_{n!}\} = \min_{\lambda\in\Delta_{n!}} \sup_{y\in Y} \{\langle A\lambda, y\rangle - D^*(y, p)\}. \tag{4}$$

In an ideal situation, we would solve (4) with a computationally efficient saddle point algorithm such as Mirror Prox [16], which leads to a convergence rate of $O\left(\log(N + n!)/T\right) \approx O\left((\log(N) + n\log(n))/T\right)$ after T iterations. However, each iteration of the Mirror Prox algorithm involves updating λ, i.e., $n!$ decision variables, which is intractable.

To remedy this, we transform the problem using standard convex conjugacy. The minimax theorem [18] allows us to write (4) as

$$\mathrm{SV}(p) = \sup_{y \in Y} \min_{\lambda \in \Delta_{n!}} \{\langle A\lambda, y\rangle - D^*(y,p)\}.$$

The second equality follows from the fact that $\langle A\lambda, y\rangle - D^*(y,p)$ is a linear function in λ, thus the minimum occurs at a vertex of $\Delta_{n!}$, i.e., some column of A. Thus, we arrive at the dual problem

$$\mathrm{SV}(p) = \sup_{y \in Y} f(y,p), \quad \text{where} \quad f(y,p) := \min_{\lambda \in \Delta_{n!}} \{\langle A\lambda, y\rangle - D^*(y,p)\}. \tag{5}$$

Notice that $f(y,p)$ is concave in y and has supergradients $\nabla_y f(y,p) = A\lambda - \nabla_y D^*(y,p)$, where $\lambda \in \arg\min_{\lambda' \in \Delta_{n!}} \langle A\lambda', y\rangle$. Thus, $\mathrm{SV}(p)$ is simply maximizing a concave function over a bounded convex domain Y. Moreover, the dimension of Y is N, which is much more manageable than $n!$.

Using online convex optimization, we show how to recover a primal solution λ for original problem (3) via solving the dual problem (5). We note that similar ideas were applied to online stochastic programming [1] and in an algorithmic approach to Approximate Caratheodory Theorem [15].

The dynamic variant of (5) can be described as follows: we would like to maximize $f(y,p)$ over Y, but we only have access to approximate data p^t, and hence we only see an approximate sequence of functions $f(y,p^t)$. This is now exactly a JEO problem [2, 11], for which methods exists to optimize $f(y,p)$ using a sequence $p^t \to p$ and employing efficient updates at each step t. In the next section, we show how to recover a primal solution λ for (3) with updating objective by solving a JEO version of the dual problem (5). Such a primal solution recovery from the dual problem was not examined in the previous literature including [2, 11], and thus it is a novel contribution.

3.2 Obtaining Primal Solutions via the Dual Problem

In Section 3.1, we established the equivalence between the primal problem (3) and its dual (5). We now show how to recover a solution for the primal problem via solving the dual problem in the dynamic data setting.

The key idea is as follows. For a given sequence of dual points $\{y^t\}_{t=1}^T$, we define concave functions f^t based on the current y^t and p^t by

$$f^t(y) := \langle A\lambda^t, y\rangle - D^*(y,p^t), \quad \text{where} \quad \lambda^t := \arg\min_{\lambda \in \Delta_{n!}} \langle A\lambda, y^t\rangle. \tag{6}$$

Notice that from (6), we also get a sequence of primal points $\{\lambda^t\}_{t=1}^T$. From these primal points, we build a candidate primal solution $\bar{\lambda}^T = \frac{1}{T}\sum_{t=1}^T \lambda^t$, which has the following optimality gap bound.

Theorem 3.1. *Given a sequence $\{y^t\}_{t=1}^T$, let λ^t be generated according to* (6)*. Then*

$$\begin{aligned} D\left(A\bar{\lambda}^T, p\right) - \min_{\lambda \in \Delta_{n!}} D(A\lambda, p) \le{}& \max_{y \in Y} \frac{1}{T}\sum_{t=1}^T f^t(y) - \frac{1}{T}\sum_{t=1}^T f^t(y^t) \\ &+ \max_{y \in Y} \frac{1}{T}\sum_{t=1}^T \left[D^*(y,p^t) - D^*(y,p)\right] \\ &+ \frac{1}{T}\sum_{t=1}^T \left[\min_{\lambda \in \Delta_{n!}} D(A\lambda, p^t) - \min_{\lambda \in \Delta_{n!}} D(A\lambda, p)\right]. \end{aligned}$$

Proof. Note that we have

$$D\left(A\bar{\lambda}^T, p\right) = D\left(\frac{1}{T}\sum_{t=1}^{T} A\lambda^t,\, p\right) = \max_{y\in Y}\left\{\left\langle \frac{1}{T}\sum_{t=1}^{T} A\lambda^t, y\right\rangle - D^*(y,p)\right\}$$
$$= \max_{y\in Y}\frac{1}{T}\sum_{t=1}^{T}\left[\langle A\lambda^t, y\rangle - D^*(y,p^t) + D^*(y,p^t) - D^*(y,p)\right]$$
$$= \max_{y\in Y}\frac{1}{T}\sum_{t=1}^{T}\left[f^t(y) + D^*(y,p^t) - D^*(y,p)\right]$$
$$\leq \max_{y\in Y}\frac{1}{T}\sum_{t=1}^{T} f^t(y) - \frac{1}{T}\sum_{t=1}^{T} f^t(y^t) + \frac{1}{T}\sum_{t=1}^{T} f^t(y^t) + \max_{y\in Y}\frac{1}{T}\sum_{t=1}^{T}\left[D^*(y,p^t) - D^*(y,p)\right],$$

where the inequality follows from decomposing the terms in the maximum. Furthermore, we have $f^t(y^t) = \langle A\lambda^t, y^t\rangle - D^*(y^t;p^t) = f(y^t;p^t)$. Hence, from $y^t \in Y$, we deduce

$$f^t(y^t) = f(y^t;p^t) \leq \max_{y\in Y} f(y,p^t) = \max_{y\in Y}\min_{\lambda\in\Delta_{n!}}\left\{\langle A\lambda, y\rangle - D^*(y,p^t)\right\} = \min_{\lambda\in\Delta_{n!}} D(A\lambda, p^t).$$

Substituting this bound into the above expression and rearranging terms gives us the desired result. □

Theorem 3.1 decomposes the optimality gap of our candidate point $\bar{\lambda}^T$ into three error terms. We now examine each of these terms. The latter two error terms

$$\max_{y\in Y}\frac{1}{T}\sum_{t=1}^{T}\left[D^*(y,p^t) - D^*(y,p)\right] \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T}\left[\min_{\lambda\in\Delta_{n!}} D(A\lambda,p^t) - \min_{\lambda\in\Delta_{n!}} D(A\lambda,p)\right]$$

are penalty terms that originate from the use of inexact data $p^t \approx p$, and they disappear in the static case when $p^t = p$. In the dynamic case, as $p^t \to p$, these two terms converge to zero if the distance function D and its conjugate D^* are sufficiently regular. For example, Lipschitz continuity of the functions $D^*(y,\cdot)$ (uniformly over $y \in Y$) and $\min_{\lambda\in\Delta_{n!}} D(A\lambda,\cdot)$ are sufficient for convergence of these two terms respectively. The first term, which is a regret term for the sequence $\{y^t\}_{t=1}^T$ on the functions $\{f^t\}_{t=1}^T$, can be shown to bound the regret of $\{y^t\}_{t=1}^T$ on the (approximate) dual objectives $\{f(y,p^t)\}_{t=1}^T$:

$$\max_{y\in Y}\frac{1}{T}\sum_{t=1}^{T} f^t(y) - \frac{1}{T}\sum_{t=1}^{T} f^t(y^t) \geq \max_{y\in Y}\frac{1}{T}\sum_{t=1}^{T} f(y,p^t) - \frac{1}{T}\sum_{t=1}^{T} f(y^t,p^t).$$

Thus, by choosing $\{y^t\}_{t=1}^T$ to have small regret on $\{f^t\}_{t=1}^T$, we are also near-optimal on the dual objectives. Online convex optimization (OCO) gives us several methods for minimizing regret. We state a typical bound, and defer a more complete overview of this to Appendix B.

Theorem 3.2 (see Appendix B)**.** *There exists a method to choose y^t using only y^{t-1}, f^{t-1} so as to guarantee*

$$\max_{y\in Y}\frac{1}{T}\sum_{t=1}^{T} f^t(y) - \frac{1}{T}\sum_{t=1}^{T} f^t(y^t) \leq O\left(\frac{1}{\sqrt{T}}\right).$$

Theorem 3.1 guarantees that, for T sufficiently large, an ϵ-approximate choice model $\bar{\lambda}^T$ will be returned in T iterations. Also, in each iteration, at most one new ranking will be introduced to $\bar{\lambda}^T$.

In our approach, the overall number of iterations T depends on two factors: the convergence rate of the regret bound, which for online Mirror Descent requires $T \geq O(1/\epsilon^2)$; and the convergence rate of the two error terms, which depends on the Lipschitz constants as well as the rate that $p^t \to p$. Therefore, T upper-bounds the sparsity of the learned model λ as well. As a result, our analysis in essence exposes an explicit trade-off between the sparsity of the non-parametric choice model λ and its estimation accuracy ϵ. To the best of our knowledge, this explicit connection has not been characterized in the literature before.

Let us summarize the assumptions we have made on the general distance function D: to obtain a bounded domain $Y = \{y : \|y\| \leq R\}$, we need bounded gradients $\|\nabla_z D(z,p)\| \leq R$; to obtain the Mirror Descent bound, we need $\|A\lambda^t - \nabla_y D^*(y,p^t)\|_* \leq G$ for all $y \in Y$, and since $A\lambda^t$ are always bounded, we need a bound on $\|\nabla_y D^*(y,p^t)\|_*$; finally, to bound the error terms, we require Lipschitz continuity of $D^*(y,p)$ and $\min_{\lambda\in\Delta_{n!}} D(A\lambda,p)$ in the p variable.

Example 3.3. *Suppose $D(A\lambda,p) = \|A\lambda - p\|$ is defined by a norm. Then standard convex analysis results imply that $\|\nabla_y D(y,p)\|_* \leq 1$, hence we can define $Y = \{y : \|y\|_* \leq 1\}$ in terms of the dual norm. Also, $D^*(y,p) = \langle y,p\rangle$ when $\|y\|_* \leq 1$ and ∞ otherwise, thus $\nabla_y D^*(y,p^t) = p^t$ for $y \in Y$, and is bounded for any norm, since the data vectors p^t are bounded. Therefore, the regret bound from Theorem 3.2 can be applied, and $\sqrt{2\Omega G^2/T} \to 0$ as $T \to \infty$. Furthermore, $D^*(y,p^t) - D^*(y,p) = \langle y, p^t - p\rangle \leq \|p^t - p\|$ for any $y \in Y$, and from standard analysis results we know that $\min_{\lambda\in\Delta_{n!}} D(A\lambda,p^t) - \min_{\lambda\in\Delta_{n!}} D(A\lambda,p) \leq \|p^t - p\|$, thus the two error terms are bounded by $\frac{2}{T}\sum_{t=1}^T \|p^t - p\|$, which converges to 0 since $p^t \to p$. This shows that when D is defined as any norm (in particular the ℓ_1-norm from Bertsimas and Mišic [3]), we can estimate a non-parametric choice model from a continuously updated sequence of data $p^t \to p$ efficiently.* ∎

Note that since the domain of λ is a simplex over the rankings, we have

$$A\lambda^t = a(\sigma^t), \quad \text{where} \quad \sigma^t = \arg\min_{\sigma\in S_n} \langle a(\sigma), y^t\rangle,$$

and hence the probability distribution $\bar{\lambda}^T$ assigns $1/T$ weight on each ranking σ^t, $t \in [T]$. Thus, each iteration, after computing y^t, we must solve the following combinatorial optimization problem over rankings:

$$\min_{\sigma} \left\{ \sum_{j\in[m]} \sum_{i\in\mathcal{A}_j} y_{ij}^t a_{ij}(\sigma) : \ \sigma \in S_n \right\}. \tag{7}$$

On the one hand, this problem is NP-hard, since we can show that it is a generalization of the linear ordering problem and the maximum weighted independent set problem, see e.g., [20, Proposition 3]. This is the main drawback of our approach. However, we note that the exact same combinatorial problem must be solved in all other approaches of learning a non-parametric choice model (see Appendix A and (10), (13), (16)). On the other hand, while we cannot avoid the NP-hardness in learning a non-parametric choice model from data, we note that (7) can be formulated as a (relatively) compact integer program with $O(n^2)$ variables and $O(n^3)$ constraints, and also it can be handled efficiently by off-the-shelf integer programming solvers, see [3]. Thus, by employing our suggested dual approach, we avoid the problem of having to handle $n!$ decision variables in the primal problem (3), and instead must solve a relatively compact integer program at each iteration.

Finally, we remark that our dual approach can apply to arbitrary domains for λ and A matrices besides the ones specified for learning non-parametric choice models in [3, 10, 20]. This flexibility can be attractive in utilizing additional a priori structural information on the choice model λ.

4 Computational Study

We carried out a computational study to compare the model fit of three different choices of distance functions $D(\cdot, \cdot)$ arising from three different ℓ_p-norms: $p \in \{1, 2, \infty\}$ on the same simulated data setup from [3, Section 5.3]. In particular, our ground truth choice model is a mixed MNL model generated by two meta-parameters: $K \in \mathbb{N}$, the number of MNL models in the mixture, and $L \in \mathbb{R}_+$ governing the intensity of the preferences. Given mixing probabilities $\{p_k\}_{k\in[K]}$ and K sets of utilities $\{u_{i,k}\}_{i\in\{0\}\cup[n]}$, $k \in [K]$, the mixed MNL model chooses an item $i \in \mathcal{A} \subseteq [n]$ with probability

$$\mathbb{P}[i \mid \mathcal{A}] = \sum_{k\in[K]} w_k \frac{u_{i,k}}{u_{0,k} + \sum_{i'\in\mathcal{A}} u_{i',k}}.$$

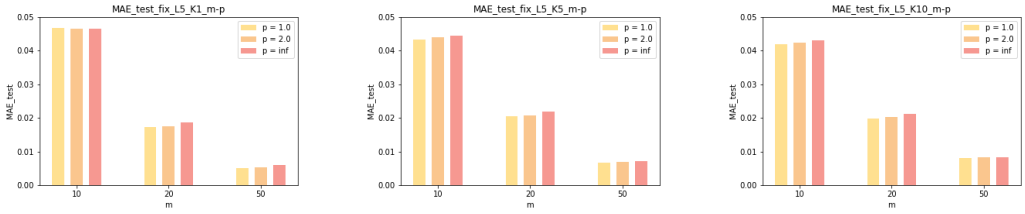
For each $k \in [K]$, we generate $n + 1$ parameters $q_{i,k} \sim U(0, 1)$, $i \in \{0\} \cup [n]$ (recall that 0 denotes the no-choice option present in each subset). The utilities $u_{i,k}$ are then set as follows: four randomly chosen $i \in \{0\} \cup [n]$ are set to $u_{i,k} = Lq_{i,k}$, while the rest are set to $u_{i,k} = q_{i,k}/10$. The mixing probabilities $\{w_k\}_{k\in[K]}$ are chosen randomly from the $(K - 1)$-dimensional simplex. We test the case $n = 10$ and parameter regimes $K \in \{1, 5, 10\}$ and $L \in \{5, 10, 100\}$.

For each combination of K and L, we generate 100 instances of the ground truth mixed MNL model. We then generate two sets of 100 subsets of maximum size $\lfloor n/2 \rfloor$ uniformly at random. We reserve one set for training, and the other set for testing. Our training regime is as follows: using the ground truth model, we compute the p_{train} vector, where $p_{\text{train},ij} = \mathbb{P}[i \mid \mathcal{A}_j]$, and $\mathcal{A}_j$ is are subsets from our training set. We vary the number of training subsets m we use by $m \in \{10, 20, 50\}$.

We use p_{train} to fit our model using the methods from Section 3. To evaluate model fit, we similarly compute p_{test} for our ground truth model using all 100 subsets in the test set, then compute $\hat{p}_{\text{test}}$ from our learned model, and examine the mean average error (MAE) between p_{test} and $\hat{p}_{\text{test}}$, defined as

$$\text{MAE}(p, \hat{p}) = \frac{1}{\text{length}(p)} \sum_{i,j} |p_{ij} - \hat{p}_{ij}|.$$

We terminate training when $\text{MAE}(p_{\text{train}}, \hat{p}_{\text{train}}) \leq 0.001$, and we evaluate performance by computing $\text{MAE}(p_{\text{test}}, \hat{p}_{\text{test}})$.

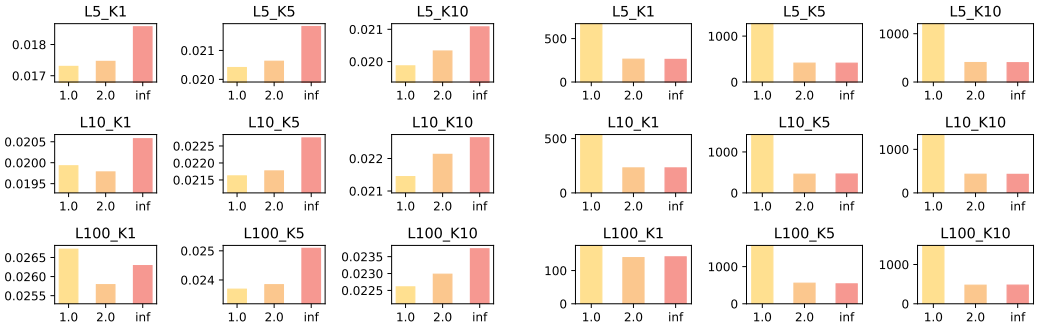


(a) $K = 1$ (b) $K = 5$ (c) $K = 10$

Figure 1: Test set MAE for $L = 5$.

Figure 1 displays the test set MAE for different K, m and p while $L = 5$ is fixed. As expected, we see that as m increases, the test MAE decreases. This trend holds for all combinations of L, K

and p, so henceforth, we examine results only for $m = 20$. Figure 1 also suggests a slight upward trend as we increase p. This can further be seen in Figure 2(a), where we examine the average

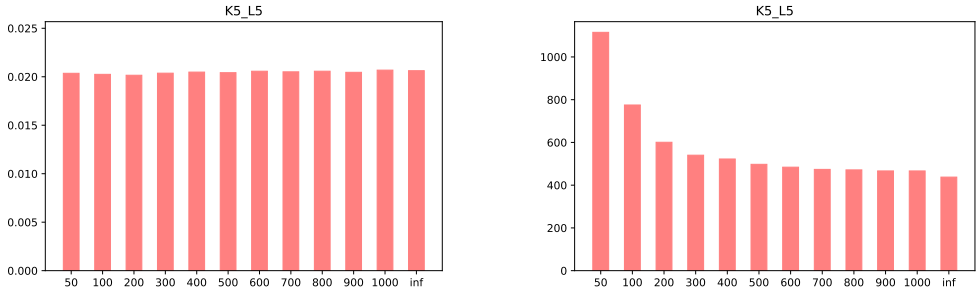


(a) Test set MAE ($m = 20$). (b) Average number of rankings learned ($m = 20$).

Figure 2: Test set MAE and # of ranking comparisons for $m = 20$

MAE across the 100 instances for each parameter combination of K and L. While the scales of the y-axes are quite small, and hence the differences quite minor, we notice that $p = 1$ often has the best test MAE. The two exceptions are when $L \in \{10, 100\}$ and $K = 1$, where $p = 2$ has the best fit. Also notice that in all cases but one, $p = \infty$ generally has worse fit than $p \in \{1, 2\}$. We plot the average number of rankings learned across the different parameter regimes in Figure 2(b), which is a measure of the sparsity of our model. While $p = 1$ often fits better than $p \in \{2, \infty\}$, it comes at the cost of a significantly denser model. The sparsity for $p \in \{2, \infty\}$ are generally quite similar. Thus, choosing $p = 1$ will give us the best fitting model on our test set, but comes at the cost of a significant loss of sparsity. Instead, $p = 2$ seems to lead to a good compromise.

We also test convergence of our method in the dynamic data setting when we only have access to p^t. In this setting, we fix the norm to be ℓ_2-norm, $L = 5$, $K = 5$, and $m = 20$. We generate p^1 by randomly generating 2000 observations of item-assortment choices from the ground truth model, and aggregating them according to (2). Each subsequent p^t is generated by adding $\kappa \in \mathbb{N}$ observations to the previous p^{t-1}. We stop when $\text{MAE}(\bar{p}^T, \hat{p}) \leq 0.001$, where $\bar{p}^T$ is the average of the p^t vectors seen. We test the effects of using different $\kappa \in \{50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000\}$.



(a) Effect of different κ on test MAE. (b) Effect of different κ on # of iterations.

Figure 3: Dynamically updating p^t for ℓ_2-norm, $L = 5$, $K = 5$, and $m = 20$

Figure 3(a) shows the average test MAE once the stopping criterion is reached. Here, $\kappa = \infty$

corresponds to using the true vector p. As we can see, the test MAE for different κ at the stopping criterion are virtually identical. Figure 3(b) plots the number of iterations for different κ, and shows that the learned model becomes more sparse when we increase κ. Note that the runs for $\kappa = 50$ take roughly twice as many iterations as $\kappa = 1000$, which means that runs for $\kappa = 1000$ use roughly 10 times as many observations to achieve effectively the same test MAE error. Surprisingly, using the *true* probabilities p to learn a model also achieves the same test MAE. We deduce that the main gains to additional observations are faster convergence and sparsity of our learned model, but test set accuracy remains unaffected. Also, observe that gains to sparsity diminish rapidly as κ increases.

All experiments were conducted on a server with 2.3 GHz processor and 64GB memory, and coded in Python 3.5. Gurobi 7.0 was used to solve the integer programming subproblems.

5 Future Directions

A number of research avenues are worthy of further investigation. Farias et al. [10] and van Ryzin and Vulcano [20] provide guarantees for the recovery of the correct choice model for their respective approaches. The question of whether similar correct recovery guarantees exist for our more general framework is appealing. In particular, [20, Corollary 1] states that as we get more accurate data $p^t \to p$, the MLE $\hat{\lambda}^t$, i.e., the solution to (12) with the data p^t, converges to the true non-parametric choice model λ with probability 1. However, $\hat{\lambda}^t$ is the full solution to (12). It would be nice to extend this to our more general framework, that is, obtain the same recovery guarantee $\hat{\lambda}^t \to \lambda$ with high probability when $p^t \to p$, with the crucial difference being that the $\hat{\lambda}^t$ are cheaply updated from previous time steps according to our simple online updates rather than being full solutions of (3) for each data point p^t. In addition, exploration of the combinatorial subproblem (7) to identify polynomial-time solvable cases or numerically efficient strategies is of further interest.

Acknowledgments

This research is supported in part by NSF grant CMMI 1454548.

References

[1] S. Agrawal and N.R. Devanur. Fast algorithms for online stochastic convex programming. In *Proceedings of the Twenty-sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1405–1424, Philadelphia, PA, USA, 2015. SIAM.

[2] H. Ahmadi and U. V. Shanbhag. Data-driven first-order methods for misspecified convex optimization problems: Global convergence and rate estimates. In *53rd IEEE Conference on Decision and Control*, pages 4228–4233, December 2014.

[3] D. Bertsimas and V.V. Mišic. Data-driven assortment optimization. Technical report, September 2015. `https://www.algomus.com/wp-content/uploads/2016/11/Data-Driven-Assortment-Optimization.pdf`.

[4] H.D. Block and J. Marschak. Random orderings and stochastic theories of responses. *Contributions to Probability and Statistics*, 2:97–132, 1960.

[5] G. Calafiore and M.C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.

[6] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, New York, NY, USA, 2006. ISBN 0521841089.

[7] A. Desir, V. Goyal, S. Jagabathula, and D. Segev. Assortment optimization under the mallows model. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 4700–4708. 2016.

[8] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th International Conference on World Wide Web*, WWW '01, pages 613–622, New York, NY, USA, 2001. ACM.

[9] V.F. Farias, S. Jagabathula, and D. Shah. A data-driven approach to modeling choice. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 504–512. 2009.

[10] V.F. Farias, S. Jagabathula, and D. Shah. A nonparametric approach to modeling choice with limited data. *Management Science*, 59(2):305–322, 2013.

[11] N. Ho-Nguyen and F. Kılınç-Karzan. Accelerating optimization under uncertainty via online convex optimization. Technical report, August 2016. `http://www.optimization-online.org/DB_HTML/2016/08/5571.html`.

[12] S. Jagabathula and D. Shah. Inferring rankings under constrained sensing. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 753–760. 2008.

[13] A. Juditsky and A. Nemirovski. First-order methods for nonsmooth convex large-scale optimization, I: General purpose methods. In S. Sra, S. Nowozin, and S.J. Wright, editors, *Optimization for Machine Learning*, Neural information processing series. MIT Press, 2012.

[14] S. Mahajan and G. van Ryzin. Stocking retail assortments under dynamic consumer substitution. *Operations Research*, 49(3):334–351, 2001.

[15] V. Mirrokni, R.P. Leme, A. Vladu, and S.C.-W. Wong. Tight bounds for approximate Carathéodory and beyond. Technical report, December 2015. `https://arxiv.org/abs/1512.08602`.

[16] A. Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15:229–251, 2004.

[17] P. Rusmevichientong, B. Van Roy, and P.W. Glynn. A nonparametric approach to multiproduct pricing. *Operations Research*, 54(1):82–98, 2006.

[18] M. Sion. On general minimax theorems. *Pacific J. Math.*, 8(1):171–176, 1958.

[19] K. Talluri and G. van Ryzin. *The theory and practice of revenue management*, volume 68 of *International Series in Operations Research & Management Science*. Springer US, 2004.

[20] G. van Ryzin and G. Vulcano. A market discovery algorithm to estimate a general class of nonparametric choice models. *Management Science*, 61(2):281–300, 2015.

A Existing Approaches to Non-Parametric Choice Estimation

In this section, we examine the existing approaches to learn the non-parametric choice model, i.e., infer an appropriate probability vector λ using the data collected via the process outlined in Section 2, and demonstrate how they are particular instantiations of our general model.

A.1 Revenue Prediction Approach

Let r_i be the revenue of item $i \in [n]$. Then the expected revenue of an assortment $\mathcal{A} \subset [n]$ under distribution λ is $\sum_{i \in \mathcal{A}} r_i \mathbb{P}_\lambda[i \mid \mathcal{A}]$. Farias et al. [10] seek to find the worst-case expected revenue from a distribution λ consistent with the given data in the sense that the theoretical probabilities $\mathbb{P}_\lambda[i \mid \mathcal{A}_j] = \langle a_{ij}, \lambda \rangle$ are precisely consistent with their empirical estimates p_{ij}. Since the probabilities $\mathbb{P}_\lambda[i \mid \mathcal{A}]$ are linear in λ, this can be formulated as a linear program (LP)

$$\min_{\lambda} \left\{ \sum_{i \in \mathcal{A}} r_i \mathbb{P}_\lambda[i \mid \mathcal{A}] : \ A\lambda = p, \ \lambda \in \Delta_{n!} \right\}.$$

We first make a few observations related to this model of Farias et al. [10]. In fact, when $\mathcal{A} = \mathcal{A}_j$ for some $j \in [m]$, we have $\mathbb{P}_\lambda[i \mid \mathcal{A}] = \langle a_{ij}, \lambda \rangle = p_{ij}$ due to the constraints $A\lambda = p$, hence the objective is constant. Thus the LP becomes a feasibility problem

$$\text{find} \quad \lambda \in \Delta_{n!} \quad \text{s.t.} \quad A\lambda = p. \tag{8}$$

That said, (8) is still computationally intractable even for moderate values of n because it involves $n!$ variables. Nonetheless, the dual of (8) admits the following robust LP interpretation:

$$\max_{\beta, \nu} \left\{ \langle \beta, p \rangle - \nu : \ \max_{\sigma \in S_n} \langle \beta, a(\sigma) \rangle \leq \nu \right\}. \tag{9}$$

Note that verifying the feasibility of a solution with respect to the robust constraint in (9), i.e.,

$$\max_{\sigma \in S_n} \langle \beta, a(\sigma) \rangle = \max_{\sigma} \left\{ \sum_{j \in [m]} \sum_{i \in \mathcal{A}_j} \beta_{ij} a_{ij}(\sigma) : \ \sigma \in S_n \right\} \leq \nu \tag{10}$$

is a combinatorial problem of the exact same form as (7). Farias et al. [10] suggests solving (9) using the constraint sampling technique [5] or by building an approximation to its robust counterpart obtained from approximating the uncertainty sets with an efficiently representable polyhedron.

In fact, (8) can be seen as choosing $\lambda \in \Delta_{n!}$ to minimize a (very harsh) distance measure:

$$\min_{\lambda \in \Delta_{n!}} D(A\lambda, p), \quad D(A\lambda, p) = \begin{cases} 0, & A\lambda = p \\ \infty, & \text{otherwise.} \end{cases} \tag{11}$$

In general, and specifically when the observations are noisy, there is no guarantee that there exists $\lambda \in \Delta_{n!}$ to fit the data p exactly, i.e., $A\lambda = p$. To remedy this, van Ryzin and Vulcano [20] and Bertsimas and Mišic [3] examine approaches that use less harsh distance measures $D(\cdot, \cdot)$.

A.2 Maximum Likelihood Estimation Approach

van Ryzin and Vulcano [20] propose the following method to learn λ via maximum likelihood estimation (MLE). We next describe their method and provide an alternative interpretation of their approach as the minimization of a particular distance measure, namely Kullback-Leibler (KL) divergence, between the true distributions $A_j\lambda$ and their empirical estimates p_j.

By (1), each item-assortment pair $i \in \mathcal{A}_j$ is seen Kq_{ij} times amongst the observations $\left\{i^k, \mathcal{A}^k\right\}_{k=1}^K$. Based on this, the log-likelihood of the observation set $\left\{i^k, \mathcal{A}^k\right\}_{k=1}^K$ is $\sum_{j\in[m]} \sum_{i\in\mathcal{A}_j} Kq_{ij} \log\left(\langle a_{ij}, \lambda\rangle\right)$. Thus, ignoring the constant K factor, the MLE problem is

$$\max_{\lambda} \left\{ \sum_{j\in[m]} \sum_{i\in\mathcal{A}_j} q_{ij} \log\left(\langle a_{ij}, \lambda\rangle\right) : \ \lambda \in \Delta_{n!} \right\}. \tag{12}$$

Throughout, we use the convention that when $q_{ij} = \langle a_{ij}, \lambda\rangle = 0$, we set $q_{ij}\log(\langle a_{ij}, \lambda\rangle) = 0$. This implies that if the optimal solution λ to (12) has $\mathbb{P}_\lambda[i \mid \mathcal{A}_j] = \langle a_{ij}, \lambda\rangle = 0$, then we must have $q_{ij} = 0$ also, i.e., we did not observe any choices of i from $\mathcal{A}_j$ in our data either.

Like (8), the problem (12) is very large, with $n!$ variables. A column generation technique is suggested in [20] to get around this, i.e., solve (12) on a subset of the variables, and use the optimality conditions to add variables as needed. The MLE column generating subproblem is constructed as

$$\max_{\sigma} \left\{ \sum_{j\in[m]} \sum_{i\in\mathcal{A}_j} \frac{q_{ij} a_{ij}(\sigma)}{\langle a_{ij}, \lambda(S)\rangle} : \ \sigma \in S_n \right\}. \tag{13}$$

The solution $\lambda(S)$ is optimal if (13) $\leq K$, otherwise the column σ^* maximizing (13) is added to the set S, and the process is repeated. Note that (13) has the same form as (7) and (10).

We next demonstrate that the MLE problem (12) admits a nice interpretation between the empirical estimates $\{p_j\}_{j\in[m]}$ and the distributions $\{A_j\lambda\}_{j\in[m]}$. To observe this, let us rewrite the objective in (12) as

$$\begin{aligned}
\sum_{j\in[m]} \sum_{i\in\mathcal{A}_j} q_{ij} \log\left(\langle a_{ij}, \lambda\rangle\right) &= \sum_{j\in[m]} q_j \sum_{i\in\mathcal{A}_j} p_{ij} \log\left(\langle a_{ij}, \lambda\rangle\right) \\
&= -\sum_{j\in[m]} q_j \underbrace{\sum_{i\in\mathcal{A}_j} p_{ij} \log\left(\frac{p_{ij}}{\langle a_{ij}, \lambda\rangle}\right)}_{=\mathrm{KL}(p_j, A_j\lambda)} + \underbrace{\sum_{j\in[m]} q_j \sum_{i\in\mathcal{A}_j} p_{ij} \log(p_{ij})}_{=\text{constant}}
\end{aligned}$$

where $\mathrm{KL}(a, b)$ is the KL divergence between two probability distributions a and b. Hence, (12) is equivalent to solving

$$\min_{\lambda} \left\{ \sum_{j\in[m]} q_j \,\mathrm{KL}(p_j, A_j\lambda) : \ \lambda \in \Delta_{n!} \right\}. \tag{14}$$

Thus, by defining $D(A\lambda, p) = \sum_{j\in[m]} q_j \,\mathrm{KL}(p_j, A_j\lambda)$, we see that the MLE is equivalent to (11) but with a different distance metric $D(\cdot, \cdot)$.

A.3 Norm-Minimization Approach

As opposed to the approaches outlined in Sections A.1 and A.2, in order to estimate a non-parametric choice model λ, Bertsimas and Mišic [3] suggest minimizing the ℓ_1-norm of $p - A\lambda$ by solving

$$\min_{\lambda} \{\|p - A\lambda\|_1 : \ \lambda \in \Delta_{n!}\}. \tag{15}$$

In fact, (15) can be cast as an LP, but it is still computationally intractable since the dimension of λ is $n!$. Similar to [20], [3] addresses this computational difficulty via a column generation approach. Again, (15) is of the same form as (11) where the distance metric $D(\cdot, \cdot)$ is selected to be $D(A\lambda, p) = \|p - A\lambda\|_1$. Furthermore, the column generating subproblem is of the form

$$\max_{\sigma} \left\{ \sum_{j \in [m]} \sum_{i \in \mathcal{A}_j} \beta_{ij}(S) a_{ij}(\sigma) - \nu(S) : \ \sigma \in S_n \right\}, \tag{16}$$

where $\beta(S), \nu(S)$ are from the dual solution to solving (15) on a subset of columns $\sigma \in S \subset S_n$. Again, this has the same form as (7), (10) and (13).

B Online Convex Optimization Based Framework

Online convex optimization (OCO) is commonly used to capture decision making in dynamic environments. Here we outline the basic OCO concepts; for further details and background, we refer the reader to Cesa-Bianchi and Lugosi [6].

In OCO, we are given a finite time horizon T, closed, bounded, and convex domain Z, and in each time period $t \in [T]$, a convex loss function $f^t : Z \to \mathbb{R}$ is revealed. At time periods $t \in [T]$ we must choose a decision $z^t \in Z$, and based on this we suffer a loss of $f^t(z^t)$ and receive some feedback typically in the form of first-order information on f^t. The main aim of OCO is to choose a sequence of points $\{z^t\}_{t=1}^T$ from the domain $\mathcal{Z}$ to bound the weighted regret

$$\sum_{t=1}^{T} \theta^t f^t(z^t) - \inf_{z \in \mathcal{Z}} \sum_{t=1}^{T} \theta^t f^t(z), \tag{17}$$

where $\theta = \{\theta^t\}_{t=1}^T \in \Delta_T$ is a collection of convex combination weights. The key restriction that separates OCO from standard optimization problems is that z^t must be chosen *before* observing f^t. The fact that there exist algorithms which bound (17) for *any* sequence $\{f^t\}_{t=1}^T$ is the crucial aspect of OCO which we exploit to solve the dynamic variant of (5).

A key class of algorithms which can be used for OCO (as well as standard offline convex optimization) are first-order methods (FOMs). Following the notation in the surveys [13], we outline the proximal setup for a general domain $\mathcal{Z}$. This setup forms the basis for several FOMs such as Mirror Descent and is used in their convergence analyses.

- *Norm*: $\|\cdot\|$ on the Euclidean space $\mathbb{E}$ where the domain $\mathcal{Z}$ lives, along with its dual norm $\|\zeta\|_* := \max_{\|z\| \le 1} \langle \zeta, z \rangle$.
- *Distance-Generating Function* (d.g.f.): A function $\omega(z) : \mathcal{Z} \to \mathbb{R}$, which is convex and continuous on $\mathcal{Z}$, and admits a selection of subgradients $\nabla\omega(z)$ that is continuous on the set $\mathcal{Z}^\circ := \{z \in \mathcal{Z} : \partial\omega(z) \neq \emptyset\}$ (here $\partial\omega(z)$ is a subdifferential of ω taken at z), and is strongly convex with modulus 1 with respect to $\|\cdot\|$:

$$\forall z', z'' \in \mathcal{Z}^\circ : \ \langle \nabla\omega(z') - \nabla\omega(z''), \ z' - z'' \rangle \ge \|z' - z''\|^2.$$

ALGORITHM 1: Generalized Mirror Descent

Input: time horizon T, positive step sizes $\{\gamma^t\}_{t=1}^T$, and a sequence of vectors $\{\xi^t\}_{t=1}^T$
Output: sequence $\{z^t\}_{t=1}^T$ from $\mathcal{Z}$.
$z^1 := \min_{z\in\mathcal{Z}} \omega(z)$;
for $t = 1, \ldots, T$ **do**
    $z^{t+1} = \mathrm{Prox}_{z^t}(\gamma^t \xi^t)$;
end

- *Prox-mapping*: Given a *prox center* $z \in \mathcal{Z}^\circ$,

$$\mathrm{Prox}_z(\xi) := \arg\min_{z'\in\mathcal{Z}} \{\langle \xi - \nabla\omega(z), z'\rangle + \omega(z')\} : \mathbb{E} \to \mathcal{Z}^\circ.$$

  When the d.g.f. is taken as the squared ℓ_2-norm, the prox mapping becomes the usual projection operation of the vector $z - \xi$ onto $\mathcal{Z}$.

- *Set width*: $\Omega = \Omega_z := \max_{z\in\mathcal{Z}} \omega(z) - \min_{z\in\mathcal{Z}} \omega(z)$.

For common domains $\mathcal{Z}$ such as simplex, Euclidean ball, and spectahedron, standard proximal setups, i.e., selection of norm $\|\cdot\|$, d.g.f. $\omega(\cdot)$, the resulting Prox computations and set widths Ω are discussed in [13, Section 1.7].

In the most basic setup, our functions f^t are convex and non-smooth. In this case, we utilize a generalization of Mirror Descent, outlined in Algorithm 1 for bounding the weighted regret (17).

We next state a bound on the weighted regret (17) in the most general case where our functions f^t need only satisfy convexity and Lipschitz continuity. More precisely, we will assume the following.

Assumption B.1. *A proximal setup of Section B exists for the domain $\mathcal{Z}$. Each function f^t is convex, and there exists $G \in (0, \infty)$ such that the subgradients of f^t are bounded, i.e., $\|\nabla f^t(z)\|_* \le G$ for all $z \in \mathcal{Z}$ and $t \in [T]$.*

Theorem B.1 ([11, Theorem 1])**.** *Suppose Assumption B.1 holds, and we are given weights $\theta \in \Delta_T$. Then Algorithm 1 with $\xi^t = \theta^t \nabla f^t(z^t)$, and step sizes $\gamma_t = \gamma := \sqrt{\frac{2\Omega}{\sup_{t\in[T]}(\theta^t)^2 G^2 T}}$ for all $t \in [T]$ results in*

$$\sum_{t=1}^T \theta^t f^t(z^t) - \inf_{z\in\mathcal{Z}} \sum_{t=1}^T \theta^t f^t(z) \le \sqrt{2\Omega \left(\sup_{t\in[T]} (\theta^t)^2\right) G^2 T}.$$

The bound on weighted regret in Theorem B.1 is optimized when the convex combination weights $\theta \in \Delta_T$ are set to be *uniform*, i.e., $\theta^t = 1/T$; in this case, the bound above becomes $O(1/\sqrt{T})$.