MICHIGAN
ROSS SCHOOL OF BUSINESS

Working Paper

Adaptive Parametric and Nonparametric Multi-product Pricing via Self-Adjusting Controls

Qi (George) Chen
Stephen M. Ross School of Business
University of Michigan

Stefanus Jasin
Stephen M. Ross School of Business
University of Michigan

Izak Duenyas
Stephen M. Ross School of Business
University of Michigan

Ross School of Business Working Paper
Working Paper No. 1258
December 2014

Adaptive Parametric and Nonparametric Multi-product Pricing via Self-Adjusting Controls

Qi (George) Chen, Stefanus Jasin, Izak Duenyas
Stephen M. Ross School of Business, University of Michigan, Ann Arbor, MI, USA
georgeqc, sjasin, duenyas@umich.edu

We study a multi-period network revenue management (RM) problem where a seller sells multiple products made from multiple resources with finite capacity in an environment where the demand function is unknown a priori. The objective of the seller is to jointly learn the demand and price the products to minimize his expected revenue loss. Both the parametric and the nonparametric cases are considered in this paper. It is widely known in the literature that the revenue loss of any pricing policy under either case is at least $\Omega(\sqrt{k})$. However, there is a considerable gap between this lower bound and the performance bound of the best known heuristic in the literature. To close the gap, we develop several self-adjusting heuristics with strong performance bound. For the general parametric case, our proposed Parametric Self-adjusting Control (PSC) attains a $\mathcal{O}(\sqrt{k})$ revenue loss, matching the theoretical lower bound. If the parametric demand function family further satisfies a well-separated condition, by taking advantage of passive learning, our proposed Accelerated Parametric Self-adjusting Control achieves a much sharper revenue loss of $\mathcal{O}(\log^2 k)$. For the nonparametric case, our proposed Nonparametric Self-adjusting Control (NSC) obtains a revenue loss of $\mathcal{O}(k^{1/2+\epsilon}\log k)$ for any arbitrarily small $\epsilon > 0$ if the demand function is sufficiently smooth. Our results suggest that in terms of performance, the nonparametric approach can be as robust as the parametric approach, at least asymptotically. All the proposed heuristics are computationally very efficient and can be used as a baseline for developing more sophisticated heuristics for large-scale problems.

Key words: Revenue management; learning; self-adjusting control; maximum likelihood estimation; spline approximation; asymptotic analysis

---

1. Introduction

Revenue management (RM), which was first implemented in the 1960s by legacy airline companies to maintain their edge in the competitive airline market, has recently become widespread in many industries such as hospitality, fashion goods, and car rentals (Talluri and van Ryzin 2005). The sellers in these industries face the common challenge of using a *fixed* capacity of perishable resources to satisfy *volatile* demand of products or services. If the seller fails to satisfy the demand appropriately, a considerable amount of profit[1] is at stake either due to the zero salvage value of unused capacity or the loss of potential revenue. Given the high stakes, RM is aimed at helping firms to make optimal decisions such that the right products are sold to the right customer at the right time and at the right price. One type of operational leverage often employed by the sellers is dynamic pricing: By adjusting the prices over time, the seller can effectively control the rate at which the demand arrives so that he can better match volatile demand with the available capacity.

Despite its potential benefits (Talluri and van Ryzin 2005), the efficacy of dynamic pricing hinges upon knowledge of how market demand responds to price adjustment, i.e., the knowledge of the underlying demand as a function of price; this is not always accessible to the sellers. Although many sellers have adopted sophisticated statistical methods, the estimated demand functions are inevitably subject to estimation error, which in turn affects the quality of the sellers' pricing decisions. The negative impact of inaccurate demand function estimation is further magnified in practice because typical RM industries tend to have an enormous sales volume; so, small error can potentially lead to a huge loss in revenue in absolute term. Given this limitation, the key issue faced by most RM practitioners is how to price dynamically when the demand function is either not perfectly known or completely unknown a priori.

This paper studies joint learning and pricing problem in a general network RM setting with multiple products and multiple capacitated resources for both the parametric and the nonparametric demand cases. For each case, we develop a heuristic that is not only easy to implement for large scale problems but also has a provable analytical performance bound. Our bounds significantly improve the performance bounds of existing heuristics in the literature.

Literature review. Our research draws on two streams of literature: the RM literature and the statistics literature. A large body of RM literature has investigated the traditional dynamic pricing problem when the seller knows the underlying demand function. The prevailing view is that, even in this simple case where learning is not in play, computing an optimal pricing policy is already computationally challenging. This is because the common technique for solving sequential decision problems, the so-called Dynamic Program (DP), suffers from the well-known curse of dimensionality. This curse of dimensionality is exacerbated in most RM industries because the sellers typically have to manage thousands of prices on a daily basis.[2] Due to this challenge, instead of finding the optimal pricing policy, a considerable body of existing literature has focused on developing computationally implementable heuristics with provably good performance. (See Bitran and Caldentey (2003) and Elmaghraby and Keskinocak (2003) for a comprehensive review of the literature.)

Within the RM literature, some papers develop heuristics based on solving a deterministic pricing problem, i.e., the deterministic counterpart of the original stochastic problem, which is computationally much easier to solve than the DP. This approach was first proposed by Gallego and van Ryzin (1994, 1997). They develop a static price control by first solving the deterministic pricing problem at the beginning of the selling season and then using its optimal solution throughout the selling season subject to the available capacity. Although the proposed heuristic is easy to implement, its drawback is also obvious: It does not utilize the progressively revealed demand realization, which leaves an open room for further improvement. Indeed, one intuitively appealing idea that has been studied in the literature involves *re-optimizing* the deterministic pricing problem in order to incorporate the progressively revealed demand realization. Maglaras and Meissner (2006) show that the re-optimized static control (RSC) cannot perform worse than static price control without re-optimization. However, it is not clear whether re-optimization guarantees a much better performance. A more recent study by Jasin (2014) shows that RSC actually performs much better than static price control. Despite this, there are still computational challenges to implement RSC in practice. Although solving the deterministic pricing problem is much easier than solving the DP, frequent re-optimizations of the deterministic pricing problem may not be practically feasible in some industries such as airlines and hotels. To address this concern, Jasin (2014) proposes a *self-adjusting* heuristic called *Linear Rate Correction* (LRC) that requires only a single optimization at the beginning of the selling season and autonomously updates the prices according to some pre-specified re-optimization-free update rules throughout the remaining selling season. Surprisingly, this simple heuristic guarantees the same performance as RSC in an asymptotic sense. Motivated by this result, in this paper, we develop self-adjusting heuristics akin to LRC when the demand function is unknown and show that the proposed self-adjusting heuristics achieve the best achievable performance bounds.

To develop a joint learning and pricing heuristic, we need to incorporate a demand learning mechanism. This requires us to use and generalize some of the standard results in the statistics literature. The statistics literature is replete with studies that attempt to estimate an unknown function from a family of candidate functions based on noisy observations. Depending on the assumptions being made about the candidate function family, this research area can be further categorized into two subfields, the parametric case and the nonparametric case, both of which have wide applications in practice. In the parametric case, researchers typically assume that the candidate function family can be fully characterized by a fixed, finite, number of parameters (i.e., a parameter vector). Popular examples include the linear, exponential, and logit function families with unknown parameters. Commonly used parametric estimation techniques include Least Squares, Generalized Least Squares, and Maximum Likelihood (ML). (See Borovkov (1999) for details.) Parametric models are widely used in industries where historical data is readily available to the sellers to infer the structural form of the demand function. Unlike the parametric case, in the nonparametric case, no information on the functional form is available. As one can imagine, the estimation problem becomes much harder because now the seller may need to estimate the function value at an infinite number of points (i.e., all points in the domain of the function) to fully characterize the underlying function. As the dimension of the domain increases, the estimation difficulty increases exponentially, which leads to another type of curse of dimensionality. Despite this technical challenge, there are applications where the nonparametric approach is more appropriate than the parametric approach, e.g., the case of new product introduction where market response to the innovative product cannot be easily inferred from historical data of similar products.

Recent works on joint learning and pricing for the capacitated RM have combined statistical learning method with dynamic pricing heuristic. In this stream of research, the central trade-off is between the cost of learning the demand function (exploration) and the reward of using the optimal price computed based on the estimated demand function (exploitation). The longer the time the seller spends on learning the demand function, the less opportunity there is for the seller to exploit the knowledge of the newly learned demand function. On the flip side, if the exploration time is too short, it will result in poor demand estimation, which yields highly sub-optimal prices. The important question is how to properly balance the exploration and exploitation to yield the maximum possible expected revenue. As mentioned earlier, even in the simpler setting with known demand function, determining the optimal policy is already difficult, let alone finding the optimal policy when the demand function is unknown. Hence, a more reasonable goal is to find heuristics that may not necessarily be optimal, but have provable good performance.

Following the standard convention in the literature, we use the revenue earned by a clairvoyant who knows the demand function and faces no variability in demand arrival as a benchmark. Since both the variability in demand realization and the informational uncertainty of the demand function are not present, we can easily imagine that this benchmark always serves as an upper bound for the expected revenue under any heuristic (e.g., Besbes and Zeevi (2012)). Indeed, it has been shown in the literature that the revenue difference between the benchmark and any feasible pricing heuristic is at least $\Omega(\sqrt{k})$ for both the parametric and the nonparametric cases, where k represents the size of the problem (see the last paragraph in §2 for more details). This result naturally raises the following questions: (1) Is the lower bound on revenue loss actually tight? (2) Does knowing the functional form of demand have a big impact on revenue performance (i.e., is there a performance difference between the parametric and the nonparametric approaches)? We want to highlight here that most existing literature on joint learning and pricing has focused primarily on the setting of a single-leg RM (single product and single resource). Besbes and Zeevi (2009) is among the first to investigate this problem under both the parametric and the nonparametric cases. Their heuristic for the parametric case yields a revenue loss of $\mathcal{O}(k^{2/3}\log^{0.5} k)$ whereas their heuristic for the nonparametric case guarantees a revenue loss of $\mathcal{O}(k^{3/4}\log^{0.5} k)$. This suggests that there is a considerable gap between the performance of the parametric approach and the nonparametric approach. Recent works by Wang et al. (2014) and Lei et al. (2014) have managed to significantly shrink this gap; they develop sophisticated nonparametric heuristics that guarantee a $\mathcal{O}(\sqrt{k}\log^{4.5} k)$ and $\mathcal{O}(\sqrt{k})$ revenue loss, respectively. Thus, for the setting of single-leg RM, existing works in the literature have not only managed to completely close the gap between the performance of the parametric and the nonparametric approaches, at least in the asymptotic sense, but also shown that the theoretical lower bound of $\Omega(\sqrt{k})$ is tight.

The network RM problem with multiple products and multiple resources is significantly more challenging than the single-leg RM. To the best of our knowledge, the only existing literature that addresses the joint learning and pricing problem in the setting of network RM is Besbes and Zeevi (2012). They consider the nonparametric case only and show that the performance bound of their proposed heuristic is $\mathcal{O}(k^{(n+2)/(n+3)}\log^{0.5} k)$, where n is the number of products. Observe that the fraction $(n+2)/(n+3)$ in the bound highlights the curse of dimensionality for network RM since the performance bound quickly deteriorates as the number of products n increases. If, however, the true demand function is sufficiently smooth (e.g., infinitely differentiable), they show that it is possible to construct a nonparametric heuristic that guarantees a $\mathcal{O}(k^{2/3+\epsilon}\log^{0.5} k)$ revenue loss for some $\epsilon > 0$ that can be arbitrarily small. Thus, the best known nonparametric heuristic for the general network RM setting in the literature has a performance guarantee no better than $\mathcal{O}(k^{2/3}\log^{0.5} k)$. As one can see, there is still a considerable gap between the lower bound of $\Omega(\sqrt{k})$ and the performance bound of $\mathcal{O}(k^{2/3}\log^{0.5} k)$. It is then not clear whether, in the general network RM, the lower bound can actually be attained by any heuristic (including the parametric approach), and whether there is an inevitable performance gap between the parametric approach and the nonparametric approach. We address these questions in our paper.

Proposed heuristics and our contributions. In this paper, we develop several heuristics for the capacitated joint learning and pricing problem for both the parametric and the nonparametric cases. Our heuristics combine statistical demand learning with a self-adjusting heuristic that is based on a heuristic in Jasin (2014) for the known demand setting. Our contributions are as follows:

1. For the parametric case, we develop a heuristic called *Parametric Self-adjusting Control* (PSC) that combines Maximum Likelihood (ML) estimation with self-adjusting price updates, and derive an analytical performance bound. To the best of our knowledge, this is the first paper that develops a joint learning and pricing heuristic in the network RM setting with parametric demand model. We show that PSC is rate-optimal. To be precise, the revenue loss of PSC is $\mathcal{O}(\sqrt{k})$ (Theorem 1), which matches the theoretical lower bound. In addition, we also show that if the parametric demand function family satisfies the so-called well-separated condition, then we can outperform the $\Omega(\sqrt{k})$ lower bound. We develop an *Accelerated Parametric Self-adjusting Control* (APSC), a variation of PSC, that attains a much sharper performance bound of $\mathcal{O}(\log^2 k)$ (Theorem 2).

2. For the nonparametric case, we develop a heuristic called *Nonparametric Self-adjusting Control* (NSC) that combines Spline Estimation with demand linearization and self-adjusting price

updates. We also provide an analytical performance bound. To the best of our knowledge, this is the first paper that introduces Spline Approximation Theory to the literature of joint learning and pricing. We show that if the underlying demand function is sufficiently smooth, the revenue loss of our heuristic is $\mathcal{O}(k^{1/2+\epsilon}\log k)$ for some $\epsilon > 0$ that can be arbitrarily small (Theorem 3). This is the tightest bound of its kind (i.e., it significantly improves the $\mathcal{O}(k^{2/3+\epsilon}\log^{0.5} k)$ bound of Besbes and Zeevi (2012)) and is only slightly worse than the theoretical lower bound of $\Omega(\sqrt{k})$.

3. From the operational perspective, our results indicate that, if demand is sufficiently smooth, not knowing the functional form of demand function should not hurt the performance by too much. Since the parametric approach is subject to model mis-specification, it can potentially hurt performance (see Figure 6 for an illustration). Thus, if the seller is not very confident about the functional form of the demand, using a nonparametric approach may yield a more robust revenue. In addition, we want to point out that our heuristics are computationally very easy to implement because they only require one (or two) deterministic optimization(s) throughout the selling season. Given the enormous complexity and scale of typical RM applications, this is an obviously appealing feature. Needless to say, if desirable, the firms can also incorporate occasional re-optimizations during the exploitation stage to further improve the performance of our heuristics.

4. On the technical side, aside from the analysis of self-adjusting heuristics mentioned above, our results also contribute to the broader literature in several ways. First, for the parametric estimation, we employ a geometric argument to derive a large deviation bound for multidimensional ML estimation with *non-i.i.d.* observations (Lemma 4). This expands our understanding on the behavior of ML estimator in non-i.i.d. observation framework. Second, for nonparametric estimation, we approximate the demand function using a linear combination of spline basis functions and derive a large deviation bound for this estimated demand function and its Jacobian matrix (Lemma 7). This result extends the application of Spline Approximation Theory to the case where observations are subject to stochastic errors. Finally, we derive a nonparametric Lipschitz-type stability result for a class of optimization problems (Lemma 8). The proof techniques used here are of independent interest for the perturbation analysis of potentially other classes of optimization problems.

The remainder of the paper is organized as follows. We first formulate the problem in §2. We then introduce our heuristics and evaluate their performances for the parametric and the nonparametric case in §3 and §4 respectively. Finally, we conclude the paper in §5. All the proofs of the results can be found in the online appendix.

2. Problem Formulation

Notation. The following notation will be used throughout the paper. (Other notation will be introduced when necessary.) We denote by $\mathbb{R}$, $\mathbb{R}_+$, and $\mathbb{R}_{++}$ the set of real, nonnegative real, and positive real numbers respectively. For column vectors $a = (a_1;\dots;a_n) \in \mathbb{R}^n, b = (b_1;\dots;b_n) \in \mathbb{R}^n$, we denote by $a \succeq b$ if $a_i \geq b_i$ for all i, and by $a \succ b$ if $a_i > b_i$ for all i. Similarly, we denote by $\mathbb{Z}$, $\mathbb{Z}_+$, and $\mathbb{Z}_{++}$ the set of integers, nonnegative integers, and positive integers respectively. We denote by $\cdot$ the inner product of two vectors and by $\otimes$ the tensor product of sets or linear spaces. We use a prime to denote the transpose of a vector or a matrix, an I to denote an identity matrix with a proper dimension, and an $\mathbf{e}$ to denote a vector of ones with a proper dimension. For any vector $v = [v_j] \in \mathbb{R}^n$, $||v||_p := (\sum_{j=1}^n |v_j|^p)^{1/p}$ is its p-norm ($1 \leq p \leq \infty$) and, for any real matrix $M = [M_{ij}] \in \mathbb{R}^{n\times n}$, $||M||_p := \sup_{||v||_p=1} ||Mv||_p$ is its induced p-norm. For example, $||M||_1 = \max_{1\leq j\leq n} \sum_{i=1}^n |M_{ij}|$, $||M||_2 =$ the largest eigenvalue of $M'M$, and $||M||_\infty = \max_{1\leq i\leq n} \sum_{j=1}^n |M_{ij}|$. (Note that $||M||_1 = ||M'||_\infty$.) For any function $f : X \to Y$, we denote by $||f(.)||_\infty := \sup_{x\in X} ||f(x)||_\infty$ the infinity-norm of f. We use ∇ to denote the usual derivative operator and use a subscript to indicate the variables with respect to which this operation is applied to. (No subscript ∇ means that the derivative is applied to all variables.) If $f : \mathbb{R}^n \to \mathbb{R}$, then $\nabla_x f = (\frac{\partial f}{\partial x_1};\dots;\frac{\partial f}{\partial x_n})$; if, on the other hand, $f = (f_1;\dots;f_n) : \mathbb{R}^n \to \mathbb{R}^n$, then

$$\nabla_x f = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}.$$

Finally, we introduce some commonly used functional spaces. We denote by $\mathsf{C}^s(\mathcal{S})$ the set of functions whose first sth order partial derivatives are continuous on its domain $\mathcal{S}$, by $\mathsf{P}^s([a,b])$ the set of single variate polynomial functions with degree s on an interval $[a,b] \subseteq \mathbb{R}$, e.g., $\mathsf{P}^1([0,1])$ is the set of all linear functions on the interval $[0,1]$.

The model. We consider the problem of a monopolist selling his products to incoming customers during a finite selling season and aiming to maximize his total expected revenue. There are n types of products, each of which is made up of a combination of a subset of m types of resources. For example, in the airline setting, a product refers to a multi-flight itinerary and a resource refers to a seat in a single-leg flight; in the hotel setting, a product refers to a multi-day stay and a resource refers to a one-night stay at a particular room. We denote by $A = [A_{ij}] \in \mathbb{R}^{m\times n}$ the *resource consumption matrix*, which characterizes the types and amounts of resources needed by each product. To be precise, a single unit of product j requires A_{ij} units of resource i. Without loss of generality, we assume that the matrix A has full row rank. (If this is not the case, then one can apply the standard row elimination procedure to delete the redundant rows. See Jasin (2014).) We denote by $C \in \mathbb{R}^m$ the vector of initial capacity levels of all resources at the beginning of the selling season. Since, in many industries (e.g., hotels and airlines), replenishment of resources during the selling season is either too costly or simply not feasible, following the standard model in the literature (Gallego and van Ryzin 1997), we will assume that the seller has no opportunity to procure additional units of resources during the selling season. In addition, we also assume without loss of generality that the remaining resources at the end of the selling season have zero salvage value.

We consider a discrete-time model with T *decision* periods, indexed by $t = 1, 2, ..., T$. At the beginning of period t, the seller first decides the price $p_t = (p_{t,1}; \ldots; p_{t,n})$ for his products, where p_t is chosen from a convex and compact set $\mathcal{P} = \otimes_{l=1}^{n} [\underline{p}_l, \bar{p}_l] \subseteq \mathbb{R}^n$ of feasible price vectors. The posted price p_t, in turn, induces a demand, or sale, for one of the products with a certain probability. Here, we implicitly assume that at most one sale for one product occurs in each period. This is without loss of generality since we can always slice the selling season fine enough to guarantee that at most one customer arrives in each period. Let $\Delta^{n-1} := \{(x_1; \ldots; x_n) \in \mathbb{R}^n | \sum_{i=1}^{n} x_i \leq 1$, and $x_i \geq 0$ for all $i\}$ denote the standard $(n-1)$-simplex. Let $\lambda^*(.) : \mathcal{P} \to \Delta^{n-1}$ denote the induced *demand rate* or *purchase probability* vector; we also call $\lambda^*(.)$ the underlying *demand function*[3]. Contrary to most existing RM literature where it is assumed that the seller knows $\lambda^*(.)$ a priori, in this paper, we simply assume that this function can be estimated using statistical learning approaches. (In §3, we consider the *parametric* case where the seller knows the functional form of $\lambda^*(.)$ but its parameters are unknown. In §4, we consider the *nonparametric* case where the functional form of $\lambda^*(.)$ is also not known to the seller.) Let $\Lambda_{\lambda^*} := \{\lambda^*(p) : p \in \mathcal{P}\}$ denote the convex set of feasible demand rates and let $D_t(p_t) = (D_{t,1}(p_t); \ldots; D_{t,n}(p_t))$ denote the vector of realized demand in period t under price p_t. It should be noted that, although demands for different products in the same period are not necessarily independent, demands over different periods are assumed to be independent (i.e., D_t only depends on the posted price p_t in period t). By definition, we have $D_t(p_t) \in \mathcal{D} := \{D \in \{0,1\}^n : \sum_{j=1}^{n} D_j \leq 1\}$ and $\mathbb{E}[D_t(p_t)] = \lambda^*(p_t)$. This allows us to write $D_t(p_t) = \lambda^*(p_t) + \Delta_t(p_t)$, where $\Delta_t(p_t)$ is a zero-mean random vector. For notational simplicity, whenever it is clear from the context which price p_t is being used, we will simply write $D_t(p_t)$ and $\Delta_t(p_t)$ as D_t and Δ_t respectively. The sequence $\{\Delta_t\}_{t=1}^{T}$ will play an important role in our analysis later. Define the revenue function $r^*(p) := p \cdot \lambda^*(p)$ to be the one-period expected revenue that the seller can earn under price p. It is typically assumed in the literature that $\lambda^*(.)$ is invertible (see the regularity assumptions below). By abuse of notation, we can then write $r^*(p) = p \cdot \lambda^*(p) = \lambda \cdot p^*(\lambda) = r^*(\lambda)$ to emphasize the dependency of revenue on demand rate instead of on price. We make the following regularity assumptions about $\lambda^*(.)$ and $r^*(.)$.

Regularity Assumptions. *There exists positive constants $\bar{r}$, $\underline{v}$, $\bar{v}$ such that:*

R1. $\lambda^*(.) : \mathcal{P} \to \Lambda_{\lambda^*}$ *is in* $\mathsf{C}^2(\mathcal{P})$ *and it has an inverse function* $p^*(.) : \Lambda_{\lambda^*} \to \mathcal{P}$ *that is in* $\mathsf{C}^2(\Lambda_{\lambda^*})$;

R2. *There exists a set of turn-off prices* $p_j^\infty \in \mathbb{R} \cap \{\infty\}$ *for* $j = 1, \ldots, n$ *such that for any* $p = (p_1; \ldots; p_n)$, $p_j = p_j^\infty$ *implies that* $\lambda_j^*(p) = 0$.

R3. $||r^*(.)||_\infty \le \bar{r}$ *and* $r^*(.)$ *is strongly concave in* λ, *i.e.,* $-\bar{v}I \preceq \nabla^2_{\lambda\lambda} r^*(\lambda) \preceq -\underline{v}I$ *for all* $\lambda \in \Lambda_{\lambda^*}$.

Assumption R1 is fairly natural and is easily satisfied by many demand functions, e.g., linear demand, logit demand, and exponential demand. Assumption R2 is common in the literature. (See Besbes and Zeevi (2009) and Wang et al. (2014).) In particular, the existence of turn-off prices p_j^∞ allows the seller to effectively shut down the demand for any product whenever needed, e.g., in the case of stock-out. As for Assumption R3, the boundedness of $r^*(.)$ follows from the compactness of Λ_{λ^*} and the smoothness of $r^*(.)$. The strong concavity of $r^*(.)$ as a function of λ is a standard assumption in the literature and is satisfied by many commonly used demand functions such as linear, exponential, and logit functions. It should be noted that although some of these functions, such as logit, do not naturally correspond to a concave revenue function when viewed as a function of p, they are nevertheless concave when viewed as a function of λ. This highlights the benefit of treating revenue as a function of demand rate instead of as a function of price. Additional regularity assumptions will be provided later.

Admissible controls and the induced probability measures. Let $D_{1:t} := (D_1, D_2, \ldots, D_t)$ denote the history of the demand realized up to and including period t. Let $\mathcal{H}_t$ denote the σ-field generated by $D_{1:t}$. We define a *control* π as a sequence of functions $\pi = (\pi_1, \pi_2, \ldots, \pi_T)$, where π_t is a $\mathcal{H}_{t-1}$-measurable real function that maps the history $D_{1:t-1}$ to $\otimes_{j=1}^n [\underline{p}_j, \bar{p}_j] \cup \{p_j^\infty\}$. This class of controls is often referred to as *non-anticipating controls* because the decision in each period depends only on the accumulated observations up to the beginning of the period. Under policy π, the seller sets the price in period t equal to $p_t^\pi = \pi_t(D_{1:t-1})$ almost surely (a.s.). Let Π denote the set of all *admissible controls.* That is,

$$\Pi := \left\{ \pi : \sum_{t=1}^T A D_t(p_t^\pi) \preceq C \;\; \text{and} \;\; p_t^\pi = \pi_t(\mathcal{H}_{t-1}) \;\; a.s. \right\}.$$

In this paper, we will often suppress the dependency of Π on λ^* for notational brevity. Note that even though the seller does not know the underlying demand function, the existence of the turn-off prices $p_1^\infty, \ldots, p_n^\infty$ guarantees that this constraint can be satisfied if the seller applies p_j^∞ for product j as soon as the remaining capacity at hand is not sufficient to produce one more unit of product j. Let $\mathbb{P}_t^\pi$ denote the induced probability measure of $D_{1:t} = d_{1:t}$ under an admissible control $\pi \in \Pi$, i.e.,

$$\mathbb{P}_t^\pi(d_{1:t}) = \prod_{s=1}^{t}\left[\left(1-\sum_{j=1}^{n}\lambda_j^*(p_s^\pi)\right)^{\left(1-\sum_{j=1}^{n} d_{s,j}\right)}\prod_{j=1}^{n}\lambda_j^*(p_s^\pi)^{d_{s,j}}\right],$$

where $p_s^\pi = \pi_s(d_{1:s-1})$ and $d_s = [d_{s,j}] \in \mathcal{D}$ for all $s = 1,\dots,t$. (By definition of $\lambda^*(.)$, the term $1-\sum_{j=1}^{n}\lambda_j^*(p_s^\pi)$ can be interpreted as the probability of no-purchase in period s under price p_s^π.) For notational simplicity, we will write $\mathbb{P}^\pi := \mathbb{P}_T^\pi$ and denote by $\mathbb{E}^\pi$ the expectation with respect to the probability measure $\mathbb{P}^\pi$. The total expected revenue under $\pi\in\Pi$ is then given by:

$$R^\pi = \mathbb{E}^\pi\left[\sum_{t=1}^{T} p_t^\pi \cdot D_t(p_t^\pi)\right].$$

The deterministic formulation and performance metric. It is common in the literature to consider the deterministic analog of the dynamic pricing problem as follows:

$$\text{(P)} \qquad J^D := \max_{p_t\in\mathcal{P}} \left\{\sum_{t=1}^{T} r^*(p_t):\ \sum_{t=1}^{T} A\lambda^*(p_t) \preceq C\right\},$$

$$\text{or equivalently,} \quad (\text{P}_\lambda) \qquad J^D := \max_{\lambda_t\in\Lambda_{\lambda^*}} \left\{\sum_{t=1}^{T} r^*(\lambda_t):\ \sum_{t=1}^{T} A\lambda_t \preceq C\right\}.$$

By assumption R3, P_λ is a convex program and is computationally easy to solve. (To avoid triviality, we assume that P_λ has a feasible solution.) It can be shown that J^D is in fact an upper bound for the total expected revenue under any admissible control. That is, $R^\pi \le J^D$ for all $\pi\in\Pi$. (See Besbes and Zeevi (2012) for more details.) This allows us to use J^D as a benchmark to quantify the performance of any admissible pricing control. In this paper, we follow the convention and define the expected revenue loss of an admissible control $\pi\in\Pi$ as $\rho^\pi := J^D - R^\pi$. Let λ^D denote the optimal solution of P_λ and let $p^D = p^*(\lambda^D)$ denote the corresponding optimal deterministic price. (Since $r^*(\lambda)$ is strongly concave with respect to λ, by Jensen's inequality, it can be proved that the optimal solution is static, i.e., $\lambda_t = \lambda^D$ for all t.) Also, let μ^D denote the optimal dual solution corresponding to the capacity constraints in P_λ. Let $\mathsf{Ball}(x,r)$ be a closed Euclidean ball centered at x with radius r. We state our fourth regularity assumption below:

R4. (Interior Assumption) *There exists* $\phi > 0$ *such that* $\mathsf{Ball}(p^D,\phi) \subseteq \mathcal{P}$.

Assumption R4 is sufficiently mild. Intuitively, it states that the static price should neither be too low that it attracts too much demand nor too high that it induces no demand. A similar interior assumption has also been made in Jasin (2014) and Chen et al. (2014).

Asymptotic setting. As discussed in §1, most RM applications can be categorized as either moderate or large size, i.e., the seller is selling a lot of products. Motivated by this, following the standard convention in the literature (e.g., Besbes and Zeevi (2009) and Wang et al. (2014)), in this paper, we will consider a sequence of increasing problems where the length of the selling season and the initial capacity levels are both scaled by a factor of $k > 0$. (One can interpret k as the *size* of the problem. For example, $k = 500$ could correspond to a flight with capacity 500 seats and $k = 5,000$ could correspond to a large hotel with capacity 5,000 rooms.) To be precise, in the k^{th} problem, the length of the selling season and the initial capacity are given by kT and kC, respectively. The optimal deterministic solution is still λ^D and the optimal dual solution is still μ^D. Let $\rho^\pi(k)$ denote the expected revenue loss under an admissible control $\pi \in \Pi$ for the problem with scaling factor k. We are primarily interested in identifying the order of $\rho^\pi(k)$ for large k. (Intuitively, one would expect that a better-performing control should have a revenue loss that grows relatively slowly with respect to k.) The following notation will be used throughout the remainder of the paper. For any two functions $f : \mathbb{Z}_{++} \to \mathbb{R}$ and $g : \mathbb{Z}_{++} \to \mathbb{R}_+$, we write $f(k) = \Omega(g(k))$ if there exists $M > 0$ independent of k such that $f(k) \geq Mg(k)$. Similarly, we also write $f(k) = \Theta(g(k))$ if there exists $M, K > 0$ independent of k such that $Mg(k) \leq f(k) \leq Kg(k)$, and write $f(k) = \mathcal{O}(g(k))$ if there exists $K > 0$ independent of k such that $f(k) \leq Kg(k)$.

3. Parametric Demand Case

In this section, we consider the parametric demand case and develop two heuristics: *Parametric Self-adjusting Control* (PSC) and *Accelerated Parametric Self-adjusting Control* (APSC). For the general family of parametric demand, we show that PSC is rate-optimal, i.e., it guarantees a $\mathcal{O}(\sqrt{k})$ revenue loss. Thus, we have completely closed the gap with the theoretical lower bound of $\Omega(\sqrt{k})$. If the parametric family of demand satisfies a so-called "well-separated" condition, we show that it is possible to further improve the $\mathcal{O}(\sqrt{k})$ bound via APSC. In what follows, we discuss the parametric function family and its estimation procedure first before describing the heuristics.

Parametric demand function family. Let Θ be a compact subset of $\mathbb{R}^q$ where $q \in \mathbb{Z}_{++}$ is the number of unknown parameters. Under the parametric demand case, the seller knows that the underlying demand function $\lambda^*(.)$ equals $\lambda(.;\theta)$ for some $\theta \in \Theta$. Although the function $\lambda(.;\theta)$ is known, the true parameter vector θ^* is unknown and needs to be estimated from the data. Let $\Lambda_\theta := \{\lambda(p;\theta) : p \in \mathcal{P}\}$ denote the set of feasible demand rates under some parameter vector $\theta \in \Theta$. We assume that Λ_θ is convex. (It can be shown that, under the most commonly used parametric function families such as linear, logit, and exponential demand, Λ_θ is convex for all $\theta \in \Theta$.) The one-period expected revenue function is given by $r(p;\theta) := p \cdot \lambda(p;\theta)$. We assume that R1 and R3 hold not only for θ^*, but also for all $\theta \in \Theta$. (See parametric family assumptions below.) This means that the demand function $\lambda(p;\theta)$ is invertible; so, by abuse of notation, we can write

$r(p;\theta) = p \cdot \lambda(p;\theta) = \lambda \cdot p(\lambda;\theta) = r(\lambda;\theta)$. In addition to the regularity assumptions R1-R4, we also need further assumptions on the parametric demand function family given below. These are all standard assumptions in the literature and are immediately satisfied by commonly used demand function families such as linear, logit and exponential.

PARAMETRIC FAMILY ASSUMPTIONS. *There exist positive constants $\omega, \underline{v}, \bar{v}$ such that for all $p \in \mathcal{P}$ and for all $\theta \in \Theta$:*

P1. $\lambda(p;.) : \Theta \to \Delta^{n-1}$ *is in* $\mathsf{C}^1(\Theta)$. *For all* $\lambda, \lambda' \in \Lambda_\theta$, $||p(\lambda;\theta) - p(\lambda';\theta)||_2 \le \omega||\lambda - \lambda'||_2$.

P2. *For all* $1 \le i, j \le n$, $||\lambda(p;\theta) - \lambda(p;\theta^*)||_2 \le \omega||\theta - \theta^*||_2$, $|\frac{\partial \lambda_j}{\partial p_i}(p;\theta) - \frac{\partial \lambda_j}{\partial p_i}(p;\theta^*)| \le \omega||\theta - \theta^*||_2$.

P3. *R1 and R3 hold for all* $\theta \in \Theta$.

Similar to P and P_λ defined in §2, we define a deterministic pricing problem for *any* $\theta \in \Theta$ as

$$(\mathrm{P}(\theta)) \qquad J^D_\theta := \max_{p \in \mathcal{P}} \left\{ \sum_{t=1}^T r(p_t;\theta) : \sum_{t=1}^T A\lambda(p_t;\theta) \preceq C \right\},$$

$$\text{or equivalently, } (\mathrm{P}_\lambda(\theta)) \qquad J^D_\theta := \max_{\lambda_t \in \Lambda_\theta} \left\{ \sum_{t=1}^T r(\lambda_t;\theta) : \sum_{t=1}^T A\lambda_t \preceq C \right\}.$$

We denote by $p^D(\theta)$ (resp. $\lambda^D(\theta)$) the optimal solution of $\mathrm{P}(\theta)$ (resp. $\mathrm{P}_\lambda(\theta)$). In addition, we also denote by $\mu^D(\theta)$ the optimal dual solution corresponding to the capacity constraints of $\mathrm{P}(\theta)$. (Note that $\mu^D(\theta)$ is also the optimal dual solution corresponding to the capacity constraints of $\mathrm{P}_\lambda(\theta)$.) Observe that $\mathrm{P}(\theta^*)$ is equivalent to P defined in §2 in the sense that $\lambda^D(\theta^*) = \lambda^D, p^D(\theta^*) = p^D, \mu^D(\theta^*) = \mu^D$, and $J^D_{\theta^*} = J^D$.

Maximum likelihood estimator. As noted earlier, the seller does not know the true parameter vector θ^*. But, he can estimate this parameter vector using statistical methods. In this paper, we will focus primarily on *Maximum Likelihood* (ML) estimation. (The analysis of other statistical methods is beyond the scope of this paper.) The behavior of ML estimator has been intensively studied in the statistics literature. It not only has certain desirable theoretical properties, but is also widely used in practice. To guarantee the regular behavior of ML estimator, certain statistical conditions need to be satisfied. To formalize these conditions, it is convenient to first consider the distribution of a sequence of demand realizations when a sequence of $\tilde{q} \in \mathbb{Z}_{++}$ fixed price vectors $\tilde{p} = (\tilde{p}^{(1)}, \tilde{p}^{(2)}, \ldots, \tilde{p}^{(\tilde{q})}) \in \mathcal{P}^{\tilde{q}}$ have been applied. For all $d_{1:\tilde{q}} \in \mathcal{D}^{\tilde{q}}$, we define the distribution $\mathbb{P}^{\tilde{p},\theta}$ as follows:

$$\mathbb{P}^{\tilde{p},\theta}(d_{1:\tilde{q}}) = \prod_{s=1}^{\tilde{q}} \left[\left(1 - \sum_{j=1}^n \lambda_j(\tilde{p}^{(s)};\theta)\right)^{\left(1 - \sum_{j=1}^n d_{s,j}\right)} \prod_{j=1}^n \lambda_j(\tilde{p}^{(s)};\theta)^{d_{s,j}} \right].$$

Let $\mathbb{E}_{\theta}^{\tilde{p}}$ denote the expectation with respect to $\mathbb{P}^{\tilde{p},\theta}$. The PSC and APSC that we will develop later use a set of "exploration prices" $\tilde{p}$ in the first L periods and then use maximum likelihood estimation to estimate the demand parameters. The exploration prices that we use need to satisfy the following conditions to guarantee the regular behavior of ML estimator:

Statistical Conditions on Exploration Prices. *There exist constants* $0 < \lambda_{\min} < \lambda_{\max} < 1$, $c_f > 0$, *and a sequence of prices* $\tilde{p} = (\tilde{p}^{(1)}, \ldots, \tilde{p}^{(\tilde{q})}) \in \mathcal{P}^{\tilde{q}}$ *such that:*

S1. $\mathbb{P}^{\tilde{p},\theta}(.) \neq \mathbb{P}^{\tilde{p},\theta'}(.)$ *whenever* $\theta \neq \theta'$;

S2. *For all* $\theta \in \Theta$, $1 \le k \le \tilde{q}$ *and* $1 \le j \le n$, $\lambda_j(\tilde{p}^{(k)}; \theta) \ge \lambda_{\min}$ *and* $\sum_{j=1}^{n} \lambda_j(\tilde{p}^{(k)}; \theta) \le \lambda_{\max}$.

S3. *For all* $\theta \in \Theta$, $\mathcal{I}(\tilde{p}, \theta) \succeq c_f I$ *where* $\mathcal{I}(\tilde{p}, \theta) := [\mathcal{I}_{i,j}(\tilde{p}, \theta)] \in \mathbb{R}^{q \times q}$ *is a* q *by* q *matrix defined as*

$$\mathcal{I}_{i,j}(\tilde{p}, \theta) = \mathbb{E}_{\theta}^{\tilde{p}} \left[-\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log \mathbb{P}^{\tilde{p},\theta}(D_{1:\tilde{q}}) \right].$$

We call $\tilde{p}$ the *exploration prices*. Some comments are in order. S1 and S2 are crucial to guarantee that the estimation problem is well-defined, i.e., the seller is able to identify the true parameter vector by observing sufficient demand realizations under the exploration prices $\tilde{p}$. (If this is not the case, then the estimation problem is ill-defined and there is no hope for learning the true parameter vector.) The symmetric matrix $\mathcal{I}(\tilde{p}, \theta)$ defined in S3 is known as the *Fisher information* matrix in the literature, and it captures the amount of information that the seller obtains about the true parameter vector using the exploration prices $\tilde{p}$. S3 requires the Fisher matrix to be strongly positive definite; this is needed to guarantee that the seller's information about the underlying parameter vector strictly increases as he observes more demand realizations under $\tilde{p}$. All the results in this section require assumptions P1-P3 and S1-S3 to hold.

Remark 1. We want to point out that, given the demand function family, it is easy to find such exploration prices. For example, for linear and exponential demand function families, any $\tilde{q} = n + 1$ price vectors $\tilde{p}^{(1)}, \ldots, \tilde{p}^{(n+1)}$ constitute a set of exploration prices if (a) they are all in the interior of $\mathcal{P}$ and (b) the vectors $(1; \tilde{p}^{(1)})$, ..., $(1; \tilde{p}^{(n+1)}) \in \mathbb{R}^{n+1}$ are linearly independent. For logit demand function family, any $\tilde{q} = 2$ price vectors $\tilde{p}_1, \tilde{p}_2$ constitute a set of exploration prices if (a) they are both in the interior of $\mathcal{P}$ and (b) $\tilde{p}_i^{(1)} \neq \tilde{p}_i^{(2)}$ for all $i = 1, \ldots, n$. The choice of exploration prices is related to the literature of optimum experimental design. Although it is possible to "optimally" choose the exploration prices using techniques in optimal experiment design, it is beyond the scope of this paper. Interested readers are referred to Pzman (2013) for more details.

3.1. General Demand Function Family

We are now ready to discuss our heuristic for the general family of parametric demand. Our main result in this section is to show that PSC is *rate-optimal*, i.e., it attains the performance lower bound. It has been repeatedly shown in the literature (e.g., Besbes and Zeevi (2012), Broder and Rusmevichientong (2012), Wang et al. (2014)) that, in the most general setting, no admissible pricing control can have a better performance than $\Omega(\sqrt{k})$, i.e., $\rho^{\pi}(k) = \Omega(\sqrt{k})$ for all $\pi \in \Pi$. This obviously poses a fundamental limitation on the performance of any pricing control that we could hope for. An important question of both theoretical and practical interest is whether this lower bound is actually tight and whether there exists an easily implementable pricing control that guarantees a $\mathcal{O}(\sqrt{k})$ revenue loss. In the general parametric setting with only a *single* product and *without* capacity constraints (i.e., the uncapacitated setting), this question has been answered by Broder and Rusmevichientong (2012). If, on the other hand, the resources have limited capacity (i.e., the capacitated setting), Lei et al. (2014) recently propose a hybrid heuristic that guarantees a $\mathcal{O}(\sqrt{k})$ revenue loss. Thus, the question of the attainability of the lower bound in the single-product setting has been completely resolved. As for the general parametric setting with multiple products and capacity constraints, we are not aware of any result that guarantees a $\mathcal{O}(\sqrt{k})$ revenue loss. The heuristics analyzed in Wang et al. (2014) and Lei et al. (2014) are not easily generalizable to multiproduct setting. (This is because their heuristics exploit the structure of the optimal deterministic solution in the single-product setting. Unfortunately, no analogs of such structures exist in the multiproduct setting.) Moreover, the analysis of multiproduct setting with capacity constraints introduce new subtleties that do not previously exist in the uncapacitated setting. A family of self-adjusting controls, i.e., *Linear Rate Correction* (LRC), has been shown to perform very well in the capacitated multiproduct setting when the demand function is *known* to the seller (Jasin (2014)). Motivated by this result, we will adapt LRC and develop a family of self-adjusting controls called *Parametric Self-adjusting Control* (PSC) that can be employed in the unknown demand setting. We will show that PSC attains the best achievable revenue loss bound for the joint learning and pricing problem. We explain PSC below.

Parametric Self-adjusting Control. The idea behind PSC is to divide the selling season into two stages: the *exploration* stage, where we do price experimentations using the exploration prices, and the *exploitation* stage, where we apply LRC using the parameter estimate computed at the end of the exploration stage. The exploration stage lasts for L periods (L itself is a decision variable to be optimized) while the exploitation stage lasts for $T - L$ periods. Let $Q \in \mathbb{R}^{n\times n}$ be a real matrix satisfying $AQ = A$ and let $\hat{\theta}_L$ denote the ML estimate of θ^* computed at the end of the exploration stage. For all $t \geq L+1$, define $\hat{\Delta}_t := D_t - \lambda(p_t; \hat{\theta}_L)$. Let C_t denote the remaining capacity at the *end* of period t. The complete PSC procedure is given below.

Parametric Self-adjusting Control (PSC)

Tuning Parameter: L

Stage 1 (Exploration)
a. Set exploration prices $\{\tilde{p}^{(1)}, \tilde{p}^{(2)}, ..., \tilde{p}^{(\tilde{q})}\}$. (See below.)
b. For $t = 1$ to L, do:
- If $C_{t-1} \succ 0$, apply price $p_t = \tilde{p}^{(\lfloor (t-1)\tilde{q}/L \rfloor + 1)}$ in period t.
- Otherwise, for product $j = 1$ to n, do:
  - If product j requires any resource that has been depleted, set $p_{t,j} = p_j^{\infty}$.
  - Otherwise, set $p_{t,j} = p_{t-1,j}$.

Stage 2 (Exploitation)
a. Compute the ML estimate $\hat{\theta}_L$ given $p_{1:L}$ and $D_{1:L}$.
b. Solve the deterministic optimization $\mathrm{P}_{\lambda}(\hat{\theta}_L)$.
c. For $t = L+1$ to T, do:
- If $C_{t-1} \succ 0$, apply the following price in period t

$$p_t = p\left(\lambda^D(\hat{\theta}_L) - \sum_{s=L+1}^{t-1} \frac{Q\hat{\Delta}_s}{T-s}\,; \hat{\theta}_L\right).$$

- Otherwise, for product $j = 1$ to n, do:
  - If product j requires any resource that has been depleted, set $p_{t,j} = p_j^{\infty}$.
  - Otherwise, set $p_{t,j} = p_{t-1,j}$.

Please note that in the PSC the exploration prices that satisfy conditions S1-S3 are set as described in Remark 1 and, as we will show below, an optimal tuning parameter for L is to set $L = \lceil \sqrt{kT} \rceil$. In comparison to the original LRC, which uses $p_t = p(\lambda^D(\theta^*) - \sum_{s=1}^{t-1} \frac{Q\Delta_s}{T-s}\,; \theta^*)^4$, since the underlying parameter vector θ^* is not known and the sequence $\{\Delta_s\}$ is not observable, we use $\hat{\theta}_L$ and $\{\hat{\Delta}_s\}$ as their substitute in PSC. Intuitively, one would expect that if $\hat{\theta}_L$ is sufficiently close to θ^*, then PSC should retain the strong performance of LRC. This intuition, however, is *not* immediately obvious. It should be noted that while LRC only deals with the impact of *natural* randomness due to demand fluctuations, as captured in $\{\Delta_s\}$, PSC also introduces a sequence of *systematic* biases due to estimation error as captured in $\{\hat{\Delta}_s\}$ (by definition, $\mathbb{E}^{\pi}[\hat{\Delta}_s] \neq 0$). Thus, despite the strong performance of LRC, it is not a priori clear whether linear rate adjustments alone, without re-optimizations *and* re-estimations, is sufficient to reduce the impact of estimation error on revenue loss. Interestingly, the answer is yes. In fact, PSC is rate-optimal.

THEOREM 1. (RATE-OPTIMALITY OF PSC) *Suppose that we use $L = \lceil \sqrt{kT} \rceil$. Then, there exists a constant $M_1 > 0$ independent of $k \geq 1$ such that $\rho^{PSC}(k) \leq M_1\sqrt{k}$ for all $k \geq 1$.*

As a comparison, if we apply the same static price $p_t = p^D(\hat{\theta}_L)$ throughout the exploitation stage, subject to capacity constraints, then the optimal length of exploration stage is of the order $k^{2/3}$ and the resulting revenue loss is $\mathcal{O}(k^{2/3}\log^{0.5} k)$ (Besbes and Zeevi 2009). This underscores an important point that a simple and autonomous price update is sufficient to reduce the revenue loss from $\mathcal{O}(k^{2/3}\log^{0.5} k)$ to $\mathcal{O}(k^{1/2})$. Let $E(t) := ||\theta^* - \hat{\theta}_t||_2$ and define $\epsilon(t) := \mathbb{E}^\pi[E(t)^2]^{1/2}$. The proof of Theorem 1 depends crucially on the following lemmas.

Lemma 1. (Continuity of the Optimal Solutions) *There exist constants $\kappa > 0$ and $\bar{\delta} > 0$ independent of $k > 0$, such that for all $\theta \in \mathsf{Ball}(\theta^*, \bar{\delta})$,*

a. $p^D(\theta) \in \mathsf{Ball}(p^D(\theta^*), \phi/2)$, $\mathsf{Ball}(p^D(\theta), \phi/2) \subseteq \mathcal{P}$ and $||\lambda^D(\theta^*) - \lambda^D(\theta)||_2 \le \kappa||\theta^* - \theta||_2$,

b. $\mu^D(.) : \Theta \to \mathbb{R}^m_+$ is continuous at θ^*;

c. The capacity constraints of $P_\lambda(\theta)$ that correspond to the rows $\{i : \mu^D_i(\theta^*) > 0\}$ are binding.

Lemma 2. (Bounds for ML Estimator with I.I.D Observations) *There exist positive constants η_1, η_2, η_3 independent of $k > 0$, such that for all $\delta > 0$, we have $\mathbb{P}^\pi(E(L) > \delta) \le \eta_1 \exp(-\eta_2 L\delta^2)$ and $\epsilon(L) \le \eta_3/\sqrt{L}$.*

Lemma 3. (Exploitation revenue under PSC) *Let $\bar{\delta}$ be as defined in Lemma 1. Let $\hat{R}^{PSC}(k)$ denote the revenue under PSC during the exploitation stage. There exists a constant $M_0 > 0$ independent of $L > 0$ and $k \ge 3$ such that for all $k \ge 3$,*

$$\sum_{t=L+1}^{kT} r(\lambda^D(\theta^*); \theta^*) - \mathbb{E}^\pi\left[\hat{R}^{PSC}(k)\right] \le M_0\left[\epsilon(L)^2 k + \frac{\log k}{1 - \mathbb{P}^\pi(E(L) > \bar{\delta})} + L + \frac{1 + k\,\mathbb{P}^\pi\left(E(L) > \bar{\delta}\right)}{1 - \mathbb{P}^\pi(E(L) > \bar{\delta})}\right].$$

Some comments are in order. Lemma 1 tells us that the deterministic problem $\mathrm{P}(\hat{\theta}_L)$ is similar to the deterministic problem $\mathrm{P}(\theta^*)$ as long as the estimate $\hat{\theta}_L$ is sufficiently close to θ^*. In particular, the Lipschitz continuity of $\lambda^D(\theta)$ is useful to quantify the size of perturbation in the deterministic solution as a function of the estimation error. Lemma 2 is a typical statistical result that is needed to bound the size of the estimation error at the end of the exploration stage. Lemma 3 is the *key*. It characterizes the trade-off between exploration and exploitation by establishing the impact of the length of the exploration stage on the total revenue loss incurred during the exploitation stage; this, in turn, helps us to determine the optimal length of the exploration stage. We want to stress: The result of Lemma 3 is rather surprising. To see this, note that, if the true parameter vector is misestimated by a *small* error ϵ, then $\lambda^D(\hat{\theta}_L)$ is roughly ϵ away from $\lambda^D(\theta^*)$ as suggested by Lemma 1(a). If the seller simply uses the static price $p^D(\hat{\theta}_L)$ throughout the exploitation stage, then the one-period revenue loss is roughly $r(\lambda^D(\theta^*); \theta^*) - r(\lambda^D(\hat{\theta}_L); \theta^*) \approx \nabla_\lambda r(\lambda^D(\theta^*); \theta^*) \cdot (\lambda^D(\theta^*) -$

Table 1 Performance comparison of STA and PSC

	Revenue	STA		PSC	
k	upper bd.	RL(Std.)	% of RL	RL(Std.)	% of RL
100	24970	9876 (48)	39.5%	7711 (82)	30.9%
300	74911	20133 (169)	26.9%	14323 (205)	19.1%
1000	249702	45817 (443)	18.3%	29587 (437)	11.8%
3000	749107	97342 (1080)	13.0%	55633 (896)	7.4%
10000	2497023	223564 (2855)	9.0%	110542 (2012)	4.4%
30000	7491069	459024 (6274)	6.1%	205426 (4683)	2.7%
100000	24970230	1035790 (14572)	4.1%	371655 (9497)	1.5%
300000	74910689	2174142 (31567)	2.9%	702589 (21923)	0.9%

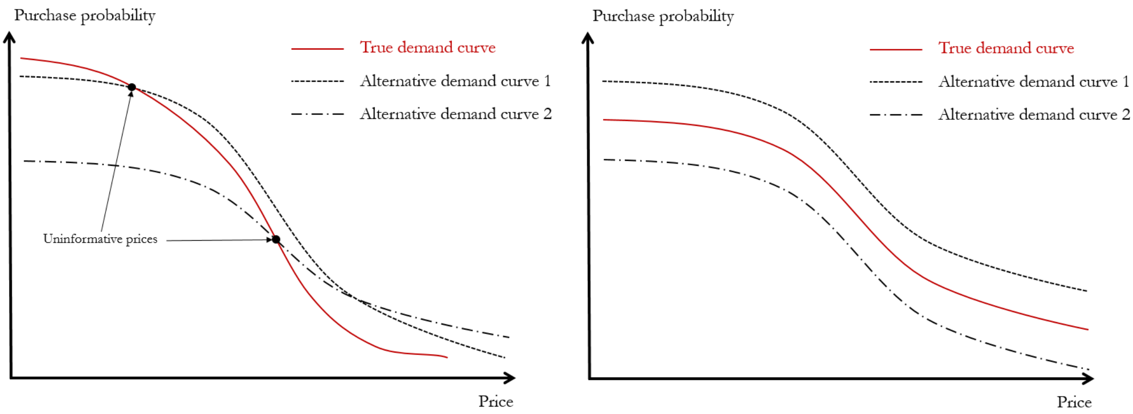
In this numerical example, we set $n=2, m=2$, $A=[1,1;0,2], C=[1;1]$. The demand model is a logit function, and $[\lambda_1(p_1,p_2);\lambda_2(p_1,p_2)] = (1+\exp(4-0.015p_1)+\exp(8-0.02p_2))^{-1}$ $[\exp(4-0.015p_1);\exp(8-0.02p_2)]$. For each heuristic, we vary the scale k from 100 to 300000 and run 1000 trials for each k.

$\lambda^D(\hat{\theta}_L)) \approx \Theta(\epsilon)$, which leads to a total revenue loss of $\mathcal{O}(\epsilon\, k)$. This is in contrast to the analysis in the *uncapacitated* setting where $\nabla_\lambda r(\lambda^D(\theta^*);\theta^*) = 0$ (because in this case $\lambda^D(\theta^*)$ is the global unconstrained optimizer of $r(\lambda;\theta^*)$), and thus a smaller revenue loss of order ϵ^2 is incurred in each period, which yields a total revenue loss of $\mathcal{O}(\epsilon^2 k)$ (see Broder and Rusmevichientong (2012)). This explains why the results in the uncapacitated setting are not directly applicable to the capacitated setting. In PSC, we use a feedback correction mechanism (i.e., the term $-\sum_{s=L+1}^{t-1}\frac{\hat{\Delta}_s}{T-s}$) that has the ability to mitigate the impact of systematic error ϵ on revenue loss. To further highlight the strength of self-adjusting price update, we report a numerical simulation in Table 1. Let *STA* denote the control that uses the deterministic price in the exploitation stage instead of adjusting prices using PSC's price update formula. (This control is the network RM version of the control in Besbes and Zeevi (2009).) Table 1 displays the revenue loss (RL) for *PSC* and *STA* and shows that *PSC* significantly outperforms *STA*. Finally, it should be noted that, although our analysis holds for all Q satisfying $AQ = A$, different choices of Q may lead to a different *non-asymptotic* performance. In particular, from the proof of Lemma 3, it can be seen that the constant M_0 is $\mathcal{O}(1+||Q||_2^2)$. Therefore, one approach to determine Q is to solve $\min\{||Q||_2 : s.t.\ AQ = A\}$. Note that this optimization is a convex program and A is known to the seller before the selling season; thus, the seller can solve the optimal Q off-line very efficiently.

3.2. Well-Separated Demand Function Family

The joint learning and pricing problem studied in §3.1 is very general: It allows both a general parametric demand form and an arbitrary number of unknown parameters. In this general case, the problem is naturally hard not only because active price experimentations are costly but also because, as it turns out, not all prices are equally informative. An example of the so-called *uninformative price* can be seen in Figure 1. Intuitively, if the seller experiments with an uninformative

Figure 1 Illustration of uninformative prices (left) and well-separated demand family (right)



Note. For a general demand function family (left), there may be uninformative prices at which the true demand curve and some alternative demand curves intersect. If the seller happens to use that price, he cannot statistically distinguish the true demand function from the alternative demand functions. This pathological phenomenon does not occur in well-separated demand function family (right).

price, then he will not be able to statistically distinguish the true demand curve from the wrong one regardless of the choice of the estimation procedure. Indeed, as pointed out by Broder and Rusmevichientong (2012), this is the reason why we cannot improve on the $\Omega(\sqrt{k})$ lower bound for revenue loss in general. To guarantee a stronger performance bound than $\Theta(\sqrt{k})$, we need to impose additional assumptions on the demand model. One condition that has been studied in the literature is the so-called *well-separatedness* of the family of demand functions proposed by Broder and Rusmevichientong (2012) (see Figure 1). They show that, for the case of the uncapacitated single-product RM, if the demand function family is well-separated, the $\Omega(\sqrt{k})$ lower bound on revenue loss can be reduced to $\Omega(\log k)$. This is a significant improvement in terms of the potentially achievable performance of an admissible pricing control. It is not, however, a priori clear whether a similar result also holds in the more general network RM setting with multiple products and capacity constraints. In what follows, we first provide the definition of well-separatedness condition in multidimensional parameter space, and then we discuss a heuristic called *Accelerated Parametric Self-adjusting Control* (APSC), which is specifically designed to address this setting.

Well-separated demand. To formalize the definition of well-separated demand, it is convenient to first consider the distribution of a sequence of demand realizations $D_{1:t} = d_{1:t}$ under a sequence of prices $p^{\pi}_{1:t} \in \mathcal{P}^t$ generated by an admissible control π, which is defined as

$$\mathbb{P}^{\pi,\theta}_t(d_{1:t}) \;=\; \mathbb{P}^{p^{\pi}_{1:t},\theta}_t(d_{1:t}) \;=\; \prod_{s=1}^{t}\left[\left(1-\sum_{j=1}^{n}\lambda_j(p^{\pi}_s;\theta)\right)^{(1-\sum_{j=1}^{n} d_{s,j})}\prod_{j=1}^{n}\lambda_j(p^{\pi}_s;\theta)^{d_{s,j}}\right].$$

Define $\mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max}) := \{p \in \mathcal{P} : \sum_{j=1}^n \lambda_j(p;\theta) \le \tilde{\lambda}_{\max}, \lambda_j(p;\theta) \ge \tilde{\lambda}_{\min}, j = 1, \ldots, n, \text{ for all } \theta \in \Theta\}$, for some $0 < \tilde{\lambda}_{\min} < \tilde{\lambda}_{\max} < 1$. We state the well-separated assumptions below. All the results in this subsection require these additional assumptions to hold.

Well-Separated Assumptions. *For any* $0 < \tilde{\lambda}_{\min} < \tilde{\lambda}_{\max} < 1$*, there exists* $c_f > 0$ *such that:*

W1. *For all* $p \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$*,* $\mathbb{P}^{p,\theta}(.) \neq \mathbb{P}^{p,\theta'}(.)$ *whenever* $\theta \neq \theta'$*;*

W2. *For all* $\theta \in \Theta$*,* $p \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$*,* $I(p,\theta) \succeq c_f I$ *for* $I(p,\theta) := [I_{i,j}(p,\theta)] \in \mathbb{R}^{q \times q}$ *defined as*

$$[I(p,\theta)]_{i,j} = \mathbb{E}_\theta^p\left[-\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log \mathbb{P}^{p,\theta}(D)\right] = \mathbb{E}_\theta^p\left[-\frac{\partial}{\partial\theta_i} \log \mathbb{P}^{p,\theta}(D) \frac{\partial}{\partial\theta_j} \log \mathbb{P}^{p,\theta}(D)\right].$$

W3. *For any* $p_{1:t} = (p_1, \ldots, p_t) \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})^t$*,* $\log \mathbb{P}_t^{p_{1:t},\theta}(D_{1:t})$ *is concave in* θ *on* Θ.

Assumptions W1 and W2 are the multiproduct multiparameter analogs of the well-separated condition given in Broder and Rusmevichientong (2012). A necessary condition for W1 to hold is that there is no "redundancy". This means that the number of products must be at least as many as the number of the unknown parameters. If the number of products is strictly smaller than the number of unknown parameters (i.e. $n < q$), then we are essentially trying to solve a system of n equations with q unknowns, which may result in the non-uniqueness of θ. Note that W2 is analogous to condition S3 and it ensures that seller's information about the parameter vector strictly increases as he observes more demand realizations under any $p \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$. The last condition W3 requires the log-likelihood function to behave nicely. This is easily satisfied by many commonly used demand functions such as linear, logit, and exponential demand functions. Note that this well-separatedness condition is not overly restrictive as it permits, for example general demand functions with unknown additive market size (i.e., for each product j, its demand is $\lambda_j(p) = a_j + g_j(p)$ where the market size a_j is unknown and $g_j : \mathcal{P} \to [0,1]$ is a known function) and general demand functions with unknown multiplicative market size (i.e., for each product j, its demand is $\lambda_j(p) = a_j g_j(p)$ where the market size a_j is unknown and $g_j : \mathcal{P} \to [0,1]$ is a known function). For more examples of well-separated demand in the single-product/single-parameter setting, see Broder and Rusmevichientong (2012).

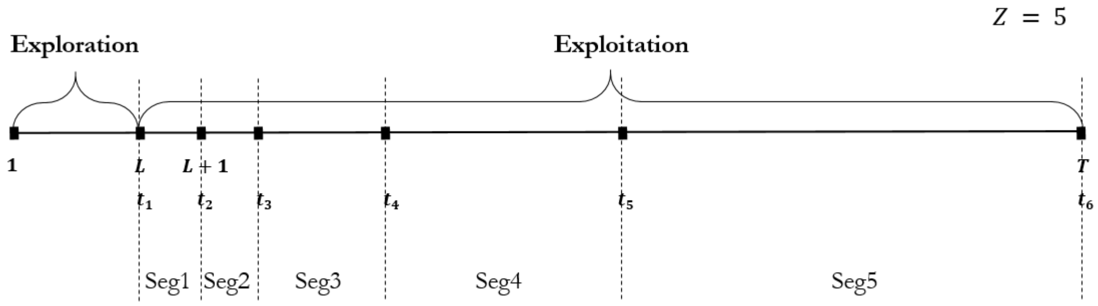
Passive learning with APSC. Estimating the unknown demand parameters from a family of well-separated candidate functions is considerably much easier than estimating the unknown parameters in the general setting. As discussed earlier, in the general parametric case, not all prices are equally informative. In contrast, under the well-separated condition, *all* prices are informative. This means that the demand data under *any* price will help improve the estimation, and the seller can continue to *passively* learn the demand parameter vector during the exploitation stage. The following result on ML estimation is the analog of Lemma 2 for non-i.i.d observations when the demand function family is well-separated.

Lemma 4. (Estimation Error of ML Estimator with Non-I.I.D Observations) *Fix some* $0<\tilde{\lambda}_{\min}<\tilde{\lambda}_{\max}<1$. *Suppose that an admissible control* π *satisfies* $p_s=\pi_s(D_{1:s-1})\in\mathcal{W}(\tilde{\lambda}_{\min},\tilde{\lambda}_{\max})$ *for all* $1\le s\le t$. *Then, under W1-W3, there exist constants* $\eta_4,\eta_5,\eta_6>0$, *such that* $\forall\delta>0,\mathbb{P}^{\pi}(E(t)>\delta)\le\eta_4 t^{q-1}\exp(-\eta_5 t\delta^2)$ *and* $\epsilon(t)\le\eta_6\sqrt{[(q-1)\log t+1]/t}$.

Remark 2. The result derived in Broder and Rusmevichientong (2012) (Theorem 4.7) can be viewed as a special case of ours. In particular, their result holds for the single product and single parameter setting whereas our result holds for a multidimensional setting with multiple products and multiple parameters. Although Hellinger distance and likelihood ratio are the common arguments used in deriving bounds in both results, we want to point out that the multidimensional parameter space is more complicated. To be precise, in the single dimension case, all candidate parameters lie on a line. Therefore, if ML estimator $\hat{\theta}_t$ is δ away from θ^*, then there are only two possibilities: Either $\hat{\theta}_t>\theta^*+\delta$ or $\hat{\theta}_t<\theta^*-\delta$. Thus deriving the tail bound reduces to bounding the probability that, given the observations, the likelihood of θ^* is smaller than either of the *two* points: $\theta^*-\delta$ and $\theta^*+\delta$. In contrast, in the multidimensional parameter case, if ML estimation error is larger than δ, one needs to bound the probability that the likelihood of θ^* is smaller than any of an *infinite* number of points that lie on the boundary of a multidimensional ball. This makes our extension nontrivial. Another observation is that as the dimension of the parameter space increases, the bounds deteriorate. This results in the different orders of regret bounds for the single parameter and the multiple parameters cases. However, since the bounds do not deteriorate too much, we are still able to attain a sharp performance bound for APSC when multiple parameters need to be estimated.

Accelerated Parametric Self-adjusting Control (APSC) divides the selling season into two stages similar to PSC: the initial exploration stage, which lasts L periods, and the exploitation stage, which lasts $T-L$ periods. However, unlike PSC, which stops learning the value of the underlying parameter vector once it exits the exploration stage, APSC continues to incorporate passive learning during its exploitation stage. To do this, APSC further divides the exploitation stage into small segments with increasing length (see Figure 2). Let t_z, $z=1,\ldots,Z+1$, be a sequence of *strictly* increasing integers satisfying $t_1=L,t_2=L+1$, $t_{Z+1}=T$, $t_z=\left\lceil\frac{t_{z+1}-L}{2}\right\rceil+L$ for all $z=2,\ldots,Z$, and let segment z contains all the periods in $(t_z,t_{z+1}]:=\{t_z+1,t_z+2,...,t_{z+1}\}$. (Note that when T and L are given, the sequence of integers is *uniquely* determined. It is not difficult to see that Z, the number of segments obtained under the procedure mentioned above, satisfies $Z\le\lceil\log_2(T-L+1)\rceil\le\lceil\log_2 T\rceil$.) The idea is to re-estimate the parameter vector at the beginning of each segment and use the new estimate to update the deterministic solution over time. The re-estimation

Figure 2 Illustration of APSC



Note. In this example, the first L periods are dedicated to exploration and the remaining periods are divided into five exploitation segments. The seller estimates the demand parameters and optimizes for the deterministic solution at the beginning of period t_1+1. The demand parameters are then re-estimated and the deterministic solution is updated accordingly at the beginning of periods $t_2+1, t_3+1, t_4+1, t_5+1$.

periods are spaced in a way that updates occur more frequently during the early part of the selling season, when our estimate is still highly inaccurate, and gradually phase out as the estimation accuracy improves. Once the parameter estimate is updated, ideally, the seller can update his deterministic solution by re-optimization. However, recall that frequent re-optimizations may still be computationally challenging for large-scale RM applications. To address this concern, we propose a re-optimization-free subroutine to update the deterministic solution at re-estimation points: (1) At the beginning of segment 1 (i.e., the beginning of period $L+1$), solve the deterministic optimization problem $\mathrm{P}(\hat{\theta}_1)$ to obtain the exact deterministic solution $\lambda^D(\hat{\theta}_1)$; (2) At the beginning of segment $z \geq 2$ (i.e., the beginning of period t_z+1), use Newton's method (see more details below) to obtain an approximate solution of $\mathrm{P}(\hat{\theta}_z)$. Since this procedure involves some subtleties, we discuss this subroutine below before laying out the full description of APSC.

To better explain the intuition behind the subroutine, we first briefly review Newton's method for the multi-variate equality constrained problem. Let $\mathcal{X}$ be a convex set in $\mathbb{R}^n$, f be a strongly concave function, and F and G be a matrix and a vector, respectively, with a proper dimension. We write down a nonlinear programming (NP) problem with equality constraints and its Karush-Kuhn-Tucker (KKT) conditions below:

$$(\mathrm{NP}) \quad \max_{x\in\mathcal{X}} \{f(x): \; Fx = G\}, \qquad (\mathrm{KKT}) \quad \{\nabla_x f(x^*) = F'\mu^*, Fx^* = G\},$$

where $(x^*;\mu^*)$ is the optimal pair of primal and dual solution. Since KKT conditions are both necessary and sufficient for the prescribed setting, to solve NP, we only need to solve the system of equations characterized by the KKT to which we will apply iterative Newton's method. To be precise, suppose that we have an approximate pair of primal and dual solution $(x_z;\mu_z)$. Then, our next pair of solution is given by $(x_{z+1};\mu_{z+1}) = (x_z;\mu_z) + (\Delta_x;\Delta_\mu)$, where the *Newton steps* Δ_x and Δ_μ are characterized by the following:

$$\begin{aligned} \nabla f(x_z+\Delta_x) &= F'(\mu_z+\Delta_\mu) \\ F(x_z+\Delta_x) &= G \end{aligned} \quad \approx \quad \begin{aligned} \nabla f(x_z)+\nabla^2 f(x_z)\Delta_x &= F'\mu_z + F'\Delta_\mu \\ Fx_z + F\Delta_x &= G \end{aligned}$$

$$\Leftrightarrow \begin{bmatrix} \Delta_x \\ \Delta_\mu \end{bmatrix} = \begin{bmatrix} -\nabla^2 f(x_z) & F' \\ F & 0 \end{bmatrix}^{-1} \begin{bmatrix} \nabla f(x_z) - F'\mu_z \\ G - Fx_z \end{bmatrix}.$$

The key result for Newton's method is that it has a locally quadratic convergence rate, i.e., there exists some positive constants γ and ξ such that if $||x_z - x^*||_2 \le \gamma$, then $||x_{z+1} - x^*||_2 \le \xi ||x_z - x^*||_2^2$ (see Boyd and Vandenberghe (2004) for details). Our idea is to tap into this locally quadratic convergence of Newton's method, coupled with the convergence result of ML estimator in Lemma 4, to develop a procedure for obtaining a sequence of solutions $\{\lambda_z^{NT}\}_{z=1}^Z$ that closely approximates $\{\lambda^D(\hat\theta_{t_z})\}_{z=1}^Z$. To implement this, we need to approximate $\mathrm{P}_\lambda(\hat\theta_{t_z})$ with an equality constrained problem $\mathrm{ECP}(\hat\theta_{t_z})$ (to be defined shortly) so that Newton's iteration can be properly applied. Let $C_i - (A\lambda^D(\hat\theta_{t_1}))_i$ denote the amount of slack for the i^{th} capacity constraint in $\mathrm{P}_\lambda(\hat\theta_{t_1})$ and define $\mathcal{B} := \{i : C_i/T - (A\lambda^D(\hat\theta_{t_1}))_i \le \eta\}$ to be the set of *potential* binding constraints in $\mathrm{P}_\lambda(\theta^*)$, where η is a threshold level to be chosen by the seller. (Since we do not know which constraints are actually binding in $\mathrm{P}_\lambda(\theta^*)$, we use $\mathcal{B}$ as our estimate. It can be shown that the constraints in $\mathcal{B}$ coincide with the binding constraints in $\mathrm{P}_\lambda(\theta^*)$ with a very high probability as $k \to \infty$ if η is properly chosen. We address how η should be chosen in Theorem 2 below.) Let B and C_B denote the submatrix of A and subvector of C with rows corresponding to the indices in $\mathcal{B}$ respectively. Similarly, let N and C_N denote the submatrix of A and subvector of C with rows corresponding to the indices *not* in $\mathcal{B}$ respectively. Define the *Equality Constrained Problem* (ECP) as follows:

$$\mathrm{ECP}(\theta) \qquad \max_{x\in\mathbb{R}^n} \left\{ r(x;\theta) : \; Bx = \frac{C_B}{T} \right\}$$

We denote by $x^D(\theta)$ the optimal solution of $\mathrm{ECP}(\theta)$. Note that if $\mathcal{B}$ coincides with the set of binding constraints of $\mathrm{P}_\lambda(\theta^*)$ at the optimal solution $\lambda^D(\theta^*)$, then not only $x^D(\theta^*)$ coincides with $\lambda^D(\theta^*)$, but also a stability result similar to Lemma 1(a) holds: there exist positive constants $\tilde\delta, \tilde\kappa$ such that for all $||\theta - \theta^*||_2 \le \tilde\delta$, $||x^D(\theta) - \lambda^D(\theta^*)||_2 = ||x^D(\theta) - x^D(\theta^*)||_2 \le \tilde\kappa ||\theta - \theta^*||_2$. This means that $\mathrm{ECP}(\theta)$ closely approximates $\mathrm{P}_\lambda(\theta^*)$ when θ is close to θ^*. We define the Newton iteration for $\mathrm{ECP}(\hat\theta_{t_z})$ in segment z as follows:

$$\begin{aligned} \mathsf{Newton}_z(x,\mu) := \begin{bmatrix} x+\Delta_x \\ \mu+\Delta_\mu \end{bmatrix} &= \begin{bmatrix} x \\ \mu \end{bmatrix} + \begin{bmatrix} -R^{-1} & B' \\ B & O \end{bmatrix}^{-1} \begin{bmatrix} G - B'\mu \\ C_B - Bx \end{bmatrix} \\ &= \begin{bmatrix} x \\ \mu \end{bmatrix} + \begin{bmatrix} -R + RB'S^{-1}BR & RB'S^{-1} \\ S^{-1}BR & S^{-1} \end{bmatrix} \begin{bmatrix} G - B'\mu \\ C_B - Bx \end{bmatrix} \end{aligned}$$

where $R = [\nabla^2_{\lambda\lambda} r(x;\hat{\theta}_{t_z})]^{-1}$, $G = \nabla_\lambda r(x;\hat{\theta}_{t_z})$, and $S = BRB'$. (This formula is derived using the formula for Newton step in multi-variate equality constrained problem and the block matrix inversion formula.) Let $\mathcal{S}_z := \Lambda_{\hat{\theta}_{t_z}} \cap \{\lambda \in \mathbb{R}^n : N\lambda \le C_N, B\lambda = C_B\}$ for $z = 1,\ldots,Z$. We can now state the *Deterministic Price Update Procedure* (DPUP) below which will be a part of the APSC described later.

Deterministic Price Update Procedure

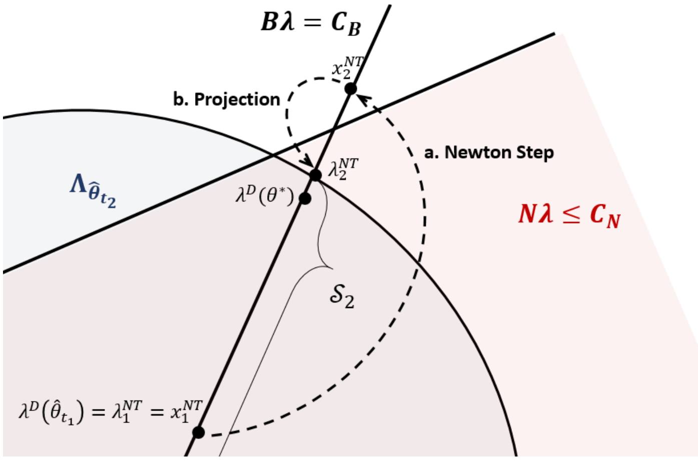
Tuning Parameter: η

For $z = 1$, do:
- a. Solve $\mathrm{P}_\lambda(\hat{\theta}_{t_1})$ and obtain $\lambda^D(\hat{\theta}_{t_1})$
- b. Identify $\mathcal{B} := \{i : C_i/T - (A\lambda^D(\hat{\theta}_{t_1}))_i \le \eta\}$
- c. Set $x_1^{NT} := \lambda^D(\hat{\theta}_{t_1})$, $\mu_1^{NT} = (BB')^{-1}B\,\nabla_\lambda r(x_1^{NT};\hat{\theta}_{t_1})$, and let $\lambda_1^{NT} := x_1^{NT}$.

For $z \ge 2$, do:
- a. Set $(x_z^{NT};\mu_z^{NT}) := \mathsf{Newton}_z(x_{z-1}^{NT}, \mu_{z-1}^{NT})$
- b. Let λ_z^{NT} be the projection of x_z^{NT} on $\mathcal{S}_z$, i.e., $\lambda_z^{NT} := \arg\min_{\lambda\in\mathcal{S}_z} ||x_z^{NT} - \lambda||_2$

We briefly explain the intuition behind DPUP. Recall that our goal is to obtain an approximate solution for each $\mathrm{P}_\lambda(\hat{\theta}_{t_z})$, $z = 1,\ldots,Z$, without re-optimization. Since $\mathrm{ECP}(\hat{\theta}_{t_z})$ and $\mathrm{P}_\lambda(\hat{\theta}_{t_z})$ are similar, the projection of $x^D(\hat{\theta}_{t_z})$ on $\mathcal{S}_z$ should be a very good approximation of $\lambda^D(\hat{\theta}_{t_z})$. Therefore, if we can find a good approximation of $x^D(\hat{\theta}_{t_z})$, say x_z, then by projecting x_z on $\mathcal{S}_z$, we can attain a good feasible approximation of $\lambda^D(\hat{\theta}_{t_z})$. This is where we need to apply Newton's method to approximately solve each $\mathrm{ECP}(\hat{\theta}_{t_z})$. In particular, segment 1 carries out two objectives: (1) We want to find the set of potential binding constraints $\mathcal{B}$ and (2) we need to compute an initial pair of approximate primal and dual solution $(x_1^{NT};\mu_1^{NT})$ to $\mathrm{ECP}(\hat{\theta}_{t_1})$. We use $\lambda^D(\hat{\theta}_{t_1})$ as our initial primal solution x_1^{NT}. The approximate dual solution μ_1^{NT} is computed using the formula proposed in Boyd and Vandenberghe (2004). (Naturally, since $\nabla_\lambda r(x^D(\hat{\theta}_{t_1});\hat{\theta}_{t_1}) = B'\mu^D(\hat{\theta}_{t_1})$ must hold at the optimal primal and dual solution of $\mathrm{ECP}(\hat{\theta}_{t_1})$, this suggests that we use $\mu_1^{NT} = (BB')^{-1}B\nabla_\lambda r(x_1^{NT};\hat{\theta}_{t_1})$.) For any later segment $z > 1$, we first use $(x_{z-1}^{NT};\mu_{z-1}^{NT})$ as an initial feasible point for $\mathrm{ECP}(\hat{\theta}_{t_z})$ and apply a single iteration of Newton update to obtain a *much better* (due to the locally quadratic convergence of Newton's method) approximate solution $(x_z^{NT};\mu_z^{NT})$ of $\mathrm{ECP}(\hat{\theta}_{t_z})$. Then, we project x_z^{NT} to $\mathcal{S}_z$ to obtain a feasible solution, λ_z^{NT}, to $\mathrm{P}_\lambda(\hat{\theta}_{t_z})$. By doing this, we manage to replace the full-scale re-optimization of $\mathrm{P}_\lambda(\hat{\theta}_{t_z})$ into one Newton update and one projection. It should be noted that, although it is theoretically possible to apply two (or more) iterations of Newton update, it is asymptotically unnecessary due to the locally quadratic convergence of Newton's method. Indeed, we show that $||x_z^{NT} - \lambda^D(\theta^*)||_2 = \Theta(||\hat{\theta}_{t_z} - \theta^*||_2)$. Thus, in light of Lemma 1(a),

Figure 3 Geometric illustration of DPUP for segment $z=2$



Note. In segment 2, step (a) is to apply Newton's method to the previous approximate solution x_1^{NT} to obtain a better solution to ECP($\hat{\theta}_{t_2}$), i.e., x_2^{NT}. This solution may not be feasible to $\mathrm{P}_\lambda(\hat{\theta}_{t_2})$, so in step (b), x_2^{NT} is projected on $\mathcal{S}_2$, which is a ray in this example, to obtain λ_2^{NT}.

x_z^{NT} approximates $\lambda^D(\theta^*)$ as well as $\lambda^D(\hat{\theta}_{t_z})$ in terms of the order of approximation error. (See Figure 3 for an illustration of DPUP.) Below, we provide the full description of APSC heuristic.

Accelerated Parametric Self-adjusting Control (APSC)

Tuning Parameters: L, η

Stage 1 (Exploration)
a. Set exploration prices $\{\tilde{p}^{(1)}, \tilde{p}^{(2)}, ..., \tilde{p}^{(\tilde{q})}\}$. (See below.)
b. For $t=1$ to L, do:
- If $C_{t-1} \succ 0$, apply price $p_t = \tilde{p}^{(\lfloor (t-1)\tilde{q}/L \rfloor + 1)}$ in period t,
- Otherwise, for product $j=1$ to n, do:
  - If product j requires any resource that has been depleted, set $p_{t,j} = p_j^\infty$.
  - Otherwise, set $p_{t,j} = p_{t-1,j}$.

Stage 2 (Exploitation)
For time segment $z=1$ to Z, do:
a. At the beginning of period t_z+1, compute ML estimate $\hat{\theta}_{t_z}$
b. Use DPUP(η) to obtain λ_z^{NT}.
c. For $t=t_z+1$ to t_{z+1}, do:
- If $C_{t-1} \succ 0$, apply the following price in period t

$$p_t := p\left(\lambda_z^{NT} - \sum_{s=t_1+1}^{t-1} \frac{Q\hat{\Delta}_s}{T-s}; \hat{\theta}_{t_z}\right),$$

- Otherwise, for product $j=1$ to n, do:
  - If product j requires any resource that has been depleted, set $p_{t,j} = p_j^\infty$.
  - Otherwise, set $p_{t,j} = p_{t-1,j}$.

Please note that in APSC the exploration prices that satisfy conditions S1-S3 are set as described in Remark 1. Moreover, under the choice of L, η described in the following theorem, APSC has a strong revenue performance as stated in the theorem below.

THEOREM 2. *Fix any $\epsilon > 0$. Suppose that we use $L = \lceil \log^{1+\epsilon}(kT) \rceil$ and $\eta = \log^{-\epsilon/4} k$. There exists a constant $M_2 > 0$ independent of $k \geq 3$ such that $\rho^{APSC}(k) \leq M_2 [\log^{1+\epsilon} k + (q-1)\log^2 k]$ for all $k \geq 3$.*

REMARK 3. Broder and Rusmevichientong (2012) has established that, under the well-separated case with one unknown parameter, the best achievable lower bound on the performance of any admissible pricing control in the *uncapacitated* single product case is $\Omega(\log k)$ and this bound is achievable by a heuristic called MLE-GREEDY. An open research question is whether this bound is also achievable in the more general case of capacitated network RM with well-separated demand. Our result gives a partial answer. We show that the revenue loss of APSC is worse than $\mathcal{O}(\log k)$ by a factor of $\log k$. However, in the case where there is only one parameter to estimate, the revenue loss of APSC is $\mathcal{O}(\log^{1+\epsilon} k)$. Since ϵ can be chosen to be arbitrarily small, APSC almost attains the best achievable performance bound for the special case with a single unknown parameter.

4. Nonparametric Demand Case

The results of §3 assume that the seller has a good prior knowledge of the functional form of the demand function. Although this is a justifiable assumption in many cases, in other cases such as new product launch where no historically relevant data is available, the seller is unlikely to know the structural form of demand. Blindly assuming a parametric demand model may be inappropriate and could potentially result in significant revenue loss if the parametric form is misspecified, e.g., a seller who uses linear model to fit the data generated by a logit model (see the numerical simulation in Besbes and Zeevi (2012)). This has motivated the study of the nonparametric approach in the literature. Recently, Wang et al. (2014) and Lei et al. (2014) propose novel nonparametric heuristics for the *single-leg* RM with $\mathcal{O}(\sqrt{k}\log^{4.5} k)$ and $\mathcal{O}(\sqrt{k})$ revenue loss, respectively. It is, however, not clear whether their heuristics can be extended to the *network* RM setting because the proposed nonparametric controls in both Wang et al. (2014) and Lei et al. (2014) heavily exploit the simple structure of the optimal deterministic solution for the single-leg RM problem, which cannot be generalized to the network setting. To the best of our knowledge, the only existing work in the literature that studies the nonparametric approach in the network setting is Besbes and Zeevi (2012). But, the performance of their heuristic quickly deteriorates when the number of products n is large due to the curse of dimensionality. This is a bad news for practitioners who have a large number of products to sell. Fortunately, not all is lost: It is known in the literature that if the underlying demand function has some additional properties, then the curse of dimensionality can be mitigated. For example, by exploiting the smoothness property, Besbes and Zeevi (2012) develop a heuristic based on Local Polynomial Approximation that attains a performance bound of $\mathcal{O}(k^{2/3+\epsilon}\log^{0.5} k)$ for any $\epsilon > 0$. Although this bound does not deteriorate when n is large, there is still a considerable gap with the $\Omega(\sqrt{k})$ lower bound on revenue loss. Is it actually possible to close this gap? Motivated by this question, in this section, we develop a nonparametric heuristic that uses spline estimation and demand linearization for the exploration stage and then uses self-adjusting control for the exploitation stage to further close the gap. It turns out that, if the underlying demand function is sufficiently smooth, our heuristic guarantees a performance bound of $\mathcal{O}(k^{1/2+\epsilon}\log k)$ for any $\epsilon > 0$, which almost attains the best achievable performance lower bound.

Nonparametric demand function and assumptions. Recall that we denote by $\lambda^*(.)$ the unknown demand function for the nonparametric case. Let $\bar{s}$ denote the largest integer such that $\left|\frac{\partial^{a_1,\ldots,a_n}\lambda_i^*(p)}{\partial p_1^{a_1}\ldots\partial p_n^{a_n}}\right|$ is uniformly bounded for all $0 \le a_1,\ldots,a_n \le \bar{s}$. We call $\bar{s}$ the *smoothness index*. We make the following smoothness condition.

Nonparametric Function Smoothness Assumptions.

N1. $\bar{s} \ge 2$.

N2. *There exists a constant $W > 0$ such that for all $i = 1,\ldots,n$ and $p \in \mathcal{P}$ and integers $0 \le a_1,\ldots,a_n \le \bar{s}$, $\left|\frac{\partial^{a_1,\ldots,a_n}\lambda_i^*(p)}{\partial p_1^{a_1}\ldots\partial p_n^{a_n}}\right| \le W$.*

The above assumptions are fairly mild and are satisfied by most commonly used demand functions, e.g., linear demand, polynomial demand with higher degree, logit demand, and exponential demand with a bounded domain of feasible prices. Note that the smoothness index reveals how difficult it is to estimate the corresponding demand function: The larger the value of $\bar{s}$, the smoother the demand function is, and it is easier to estimate its shape because the function value cannot have a drastic change locally.

Spline approximation of a deterministic function. To estimate a nonparametric function from noisy observations, we first study a simpler problem of approximating a *deterministic* function. To that end, we will use the results developed in *Spline Approximation.* (Although the Local Polynomial Approximation used in Besbes and Zeevi (2012) also utilizes the smoothness of the demand function to mitigate the curse of dimensionality in estimating multi-variate functions, we instead choose to use Spline Approximation method because this approach yields a differentiable demand function unlike the Local Polynomial Approximation. This differentiability not only enables a more efficient and stable computation for solving the deterministic optimization problem, but also facilitates the stability analysis of the deterministic optimization problem.) Spline functions have been widely used in engineering to approximate complicated functions, and their popularity is primarily due to their flexibility in effectively approximating complex curve shapes. This flexibility lies in the piecewise nature of spline functions – a spline function is constructed by attaching piecewise polynomial functions with a certain degree, and the coefficients of these polynomials are computed in a way such that a sufficiently high degree of smoothness is ensured in the places where the polynomials connect (the points where two piecewise polynomials are attached are called the *knots*). More formally, for all $l \in \{1, \dots n\}$, let $\underline{p}_l = x_{l,0} < x_{l,1} \cdots < x_{l,d} < x_{l,d+1} = \bar{p}_l$ be a partition that divides $[\underline{p}_i, \bar{p}_i]$ into $d+1$ subintervals of equal length. Let $\mathcal{G} := \otimes_{l=1}^n \mathcal{G}_l$ denote the knots grid where $\mathcal{G}_l = \{x_{l,i}\}_{i=0}^{d+1}$. We define the function space of *tensor-product polynomial splines of order* $(s; \dots; s) \in \mathbb{R}^n$ *with knots at points in* $\mathcal{G}$ as $\mathsf{S}(\mathcal{G}, s) = \otimes_{l=1}^n \mathsf{S}_l(\mathcal{G}_l, s)$ where $\mathsf{S}_l(\mathcal{G}_l, s) := \{f \in \mathsf{C}^{s-2}[\underline{p}_l, \bar{p}_l] :$ f is a single-variate polynomial of degree $s-1$ on each subinterval $[x_{l,i}, x_{l,i+1}), i = 0, \dots, d.\}$.

One of the key questions that the theory of Spline Approximation addresses is the following: given an arbitrary function f that satisfies N1-N2 and $\mathsf{S}(\mathcal{G}, s)$, find a spline function $g^* \in \mathsf{S}(\mathcal{G}, s)$ that approximates f well. Among the various approaches, one of the most popular approximations is using the *tensor-product B-Spline basis functions*. This approach is based on the key observation that $\mathsf{S}(\mathcal{G}, s)$ is a linear space of dimension $(d+s)^n$. This implies that there exists a set of $(d+s)^n$ basis functions (this set is not unique), and any function in $\mathsf{S}(\mathcal{G}, s)$ can be represented as a linear combination of the basis functions. We propose to use tensor-product B-Spline basis functions, denoted by $\{N_{i_1,\dots,i_n}(x_1, \dots, x_n)\}_{i_1=1,\dots,i_n=1}^{s+d,\dots,s+d}$, as the set of basis functions. These functions are defined formally in the Technical Details part (a) below, and are illustrated in Figure 4. Given the basis functions, for any spline function $g \in \mathsf{S}(\mathcal{G}, s)$, there exists a set of coefficients $\{c_{i_1,\dots,i_n}\}_{i_1=1,\dots,i_n=1}^{s+d,\dots,s+d}$ such that $g(x) = \sum_{i_1=1}^{s+d} \cdots \sum_{i_n=1}^{s+d} c_{i_1,\dots,i_n} N_{i_1,\dots,i_n}(x)$ for all $x \in \mathcal{P}$. Therefore, the problem of finding g^* is reduced to the problem of computing the coefficients for representing g^*, which we address below in the Technical Details part (b). Since the procedure of spline approximation essentially takes f as an input and outputs a function g^*, it can be viewed as a linear operator $\mathcal{L} : \mathsf{C}^0(\mathcal{P}) \to \mathsf{S}(\mathcal{G}, s)$. Lemma 5 highlights some useful properties of $\mathcal{L}$.

Technical Details for Spline Approximation: The B-Spline Approach

(a) Tensor-product B-Spline Basis Functions.

Step 1: For each $l = 1, \dots, n$, define an *extended partition* $\mathcal{G}_l^e := \{y_{l,i}\}_{i=1}^{2s+d}$, where

$$y_{l,1} = \cdots = y_{l,s} = x_{l,0},\ y_{l,s+1} = x_{l,1}, \dots, y_{l,s+d} = x_{l,d},\ y_{l,s+d+1} = \cdots = y_{l,2s+d} = x_{l,d+1}.$$

Step 2: For $1 \le i_1, \ldots, i_n \le s+d, l = 1, \ldots, n$, define the *tensor-product B-Spline basis function* as $N_{i_1,\ldots,i_n}(x_1,\ldots,x_n) = \prod_{l=1}^n N^s_{l,i_l}(x_l)$, where

$$N^s_{l,i}(x_l) = \begin{cases} (-1)^s (y_{l,i+s} - y_{l,i})[y_{l,i},\ldots,y_{l,i+s}](x_l - y)_+^{s-1}, & \text{if } x_{l,i} \le x_l < x_{l,i+1} \\ 0, & \text{otherwise} \end{cases}$$

for all $x_l \in [\underline{p}_l, \bar{p}_l]$ for all $l = 1, \ldots, n$ and for all $i = 1, \ldots, d+s$, where $(x_l - y)_+ = \max\{0, x_l - y\}$, and $[t_1, \ldots, t_{r+1}]f(y) := \sum_{i=1}^{r+1} f(t_i) \prod_{j=1, j\neq i}^{r+1} (t_i - t_j)^{-1}$ is the r^{th} *order divided difference* of a single variate real function f over the points $t_1, \ldots, t_{r+1}$.

(b) Calculating the Linear Coefficients.

Step 1: For $l = 1, \ldots, n, i = 1, \ldots, d+s$, let

$$\tau_{l,i,j} = y_{l,i} + (y_{l,i+s} - y_{l,i})\frac{j-1}{s-1} \quad \text{and} \quad \beta_{l,i,j} = \sum_{v=1}^{j} \frac{\xi^{(v)}_{l,i} \psi^{(v-1)}_{l,i,j}(0)}{(v-1)!}, \quad \text{for } j = 1, \ldots, s,$$

where

$$\xi^{(v)}_{l,i} = \frac{(-1)^{v-1}(v-1)!}{(s-1)!} \phi^{(s-v)}_{l,i,s}(0) \ \text{ and}$$
$$\phi_{l,i,s}(t) = \prod_{r=1}^{s-1} (t - y_{l,i+r}), \ \psi_{l,i,j}(t) = \prod_{r=1}^{j-1} (t - \tau_{l,i,r}), \ \psi_{l,i,1}(t) \equiv 1.$$

Step 2: For any $f = (f_1; \ldots; f_n) \in \mathsf{C}^0(\mathcal{P})$, let $\{\gamma_{l,i} : \mathsf{C}^0([\underline{p}_l, \bar{p}_l]) \to \mathbb{R}\}_{l=1,i=1}^{n,s+d}$ be a set of linear functionals defined as follows:

$$\gamma_{l,i} f_l = \sum_{j=1}^{s} \beta_{l,i,j} [\tau_{l,i,1}, \ldots, \tau_{l,i,j}] f_l.$$

Define another set of linear functionals $\{\gamma_{i_1,\ldots,i_n}\}_{i_1=1,\ldots,i_n=1}^{s+d,\ldots,s+d}$ such that

$$\gamma_{i_1,\ldots,i_n} f = \gamma_{1,i_1} \circ \gamma_{2,i_2} \cdots \circ \gamma_{n,i_n} f,$$

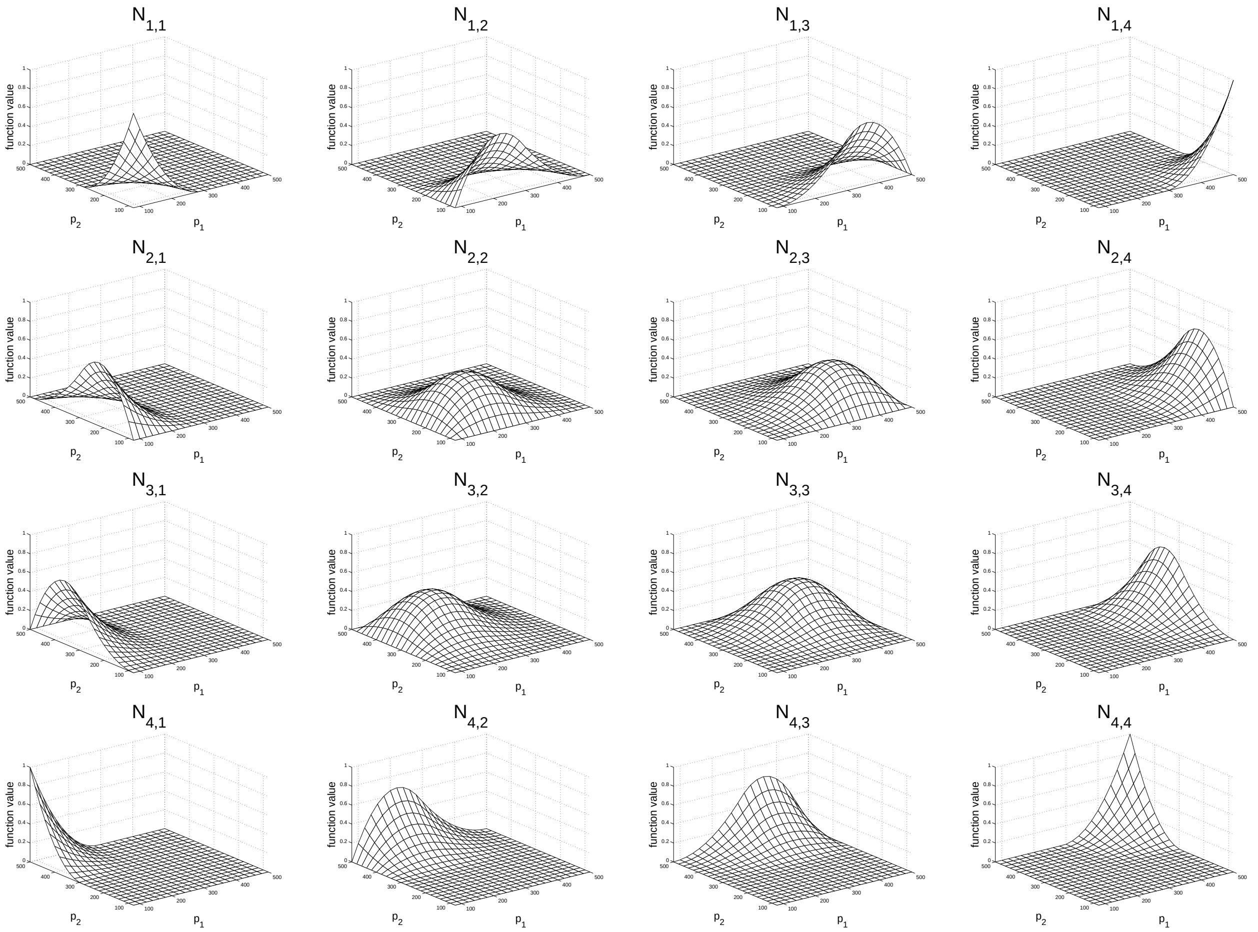
where γ_{l,i_l} is understood as being applied to f as a function of x_l. By the construction of γ_{l,i_l} and the definition of divided differences, basic algebra yields:

$$\gamma_{i_1,\ldots,i_n} f = \sum_{j_1=1}^{s} \sum_{r_1=1}^{j_1} \cdots \sum_{j_n=1}^{s} \sum_{r_n=1}^{j_n} \frac{\prod_{l=1}^n \beta_{l,i_l,j_l}}{\prod_{l=1}^n \prod_{s_l=1, s_l \neq r_l}^{j_l} (\tau_{l,i_l,r_l} - \tau_{l,i_l,s_l})} f(\tau_{1,i_1,r_1}, \ldots, \tau_{n,i_n,r_n}).$$

Step 3: Define a linear operator $\mathcal{L}_l : \mathsf{C}^0([\underline{p}_l, \bar{p}_l]) \to \mathsf{S}_l(\mathcal{G}_l, s)$ as $\mathcal{L}_l f(x_l) = \sum_{i=1}^{s+d} (\gamma_{l,i} f) N^s_{l,i}(x_l)$, for all $l = 1, \ldots, n$. Similarly, define a linear operator $\mathcal{L} : \mathsf{C}^0(\mathcal{P}) \to \mathsf{S}(\mathcal{G}, s)$ as

$$\mathcal{L} f(x_1, \ldots, x_n) = \sum_{i_1=1}^{s+d} \cdots \sum_{i_n=1}^{s+d} (\gamma_{i_1,\ldots,i_n} f) N_{i_1,\ldots,i_n}(x_1, \ldots, x_n).$$

Figure 4 Illustration of tensor-product B-Spline basis functions



Note. In this example, the domain is $[80,500]\times[80,500]$, and $s=3, d=1$. The knots grid $\mathcal{G}$ consists of $(d+2)^n = 3^2 = 9$ points, i.e., $(80,80)$, $(80,290)$, $(80,500)$, $(290,80)$, $(290,290)$, $(290,500)$, $(500,80)$, $(500,290)$, $(500,500)$, which slice the domain into 4 pieces (rectangles). A spline in $\mathcal{S}(\mathcal{G},3)$ is a biquadratic function on each piece, and is continuously differentiable on the places where different pieces connect. Per our construction, there are $(s+d)^n = 4^2 = 16$ basis functions. These hill-like basis functions are the building blocks for spline approximation.

Note that $\mathcal{L} = \mathcal{L}_1 \circ \mathcal{L}_2 \circ \cdots \circ \mathcal{L}_n$, where this composition of linear operators is understood as $\mathcal{L}_l$ being applied to a function of x_l.

Step 4: Set $g^* = \mathcal{L}f$.

LEMMA 5. *$\mathcal{L}$ is a bounded linear operator mapping $\mathsf{C}^0(\mathcal{P})$ to $\mathsf{S}(\mathcal{G},s)$. Also, $\mathcal{L}f = f$ for all $f \in \otimes_{l=1}^n \mathsf{P}^{s-1}([\underline{p}_l, \bar{p}_l])$.*

Spline approximation with noisy observations. We will now discuss the estimation of demand function $\lambda^*(.)$ by spline approximation with noisy observations. Let $\tilde{\mathcal{G}} := \{(\tau_{1,i_1,j_1};\ldots;\tau_{n,i_n,j_n}) : 1 \le i_1,\ldots,i_n \le s+d, 1 \le j_1,\ldots,j_n \le s\}$. Note that the constants $\{\gamma_{i_1,\ldots,i_n}\lambda_j^*\}_{i_1=1,\ldots,i_n=1}^{s+d,\ldots,s+d}$ depends on $\lambda_j^*(.)$

only via $\lambda^*_j(p), p \in \tilde{\mathcal{G}}$. So, if the seller could observe the demand rate of product j under prices in $\tilde{\mathcal{G}}$, he could construct an approximation of $\lambda^*_j(.)$ using a linear combination of tensor-product B-splines. In our problem, the seller cannot observe $\lambda^*_j(p)$ for $p \in \tilde{\mathcal{G}}$, but only its noisy observation $D_j(p) = \lambda^*_j(p) + \Delta_j$. To address this, we use empirical mean as a surrogate of $\lambda^*_j(p)$ and propose the following *Spline Estimation* algorithm to estimate the demand.

Spline Estimation

Input Parameters: $\tilde{L}_0, n, s$; **Tuning Parameter:** d

Algorithm:

Step 1: Estimate $\lambda^*(p)$ at points $p \in \tilde{\mathcal{G}}$. Set $L_0 = \tilde{L}_0 s^{-n}(s+d)^{-n}$. For each $p \in \tilde{\mathcal{G}}$

a. Apply price p L_0 times

b. Let $\tilde{\lambda}(p)$ be the sample mean of the L_0 observations.

Step 2: Construct spline approximation.

a. Calculate coefficients $c^j_{i_1,\ldots,i_n}, 1 \le i_1,\ldots,i_n \le s+d, j = 1,\ldots,n$ as:

$$c^j_{i_1,\ldots,i_n} = \sum_{j_1=1}^{s}\sum_{r_1=1}^{j_1}\cdots\sum_{j_n=1}^{s}\sum_{r_n=1}^{j_n} \frac{\tilde{\lambda}_j(\tau_{1,i_1,r_1},\ldots,\tau_{n,i_n,r_n})\ \prod_{l=1}^{n}\beta_{l,i_l,j_l}}{\prod_{l=1}^{n}\prod_{s_l=1,s_l\neq r_l}^{j_l}(\tau_{l,i_l,r_l}-\tau_{l,i_l,s_l})}.$$

b. Construct a tensor-product spline function $\tilde{\lambda}(p) = (\tilde{\lambda}_1(p);\ldots;\tilde{\lambda}_n(p))$, where

$$\tilde{\lambda}_j(p) = \sum_{i_1=1}^{s+d}\cdots\sum_{i_n=1}^{s+d} c^j_{i_1,\ldots,i_n} N_{i_1,\ldots,i_n}(p).$$

Note that $\tilde{L}_0 = L_0(s+d)^n s^n$ is the duration of the Spline Estimation algorithm. Let $a \wedge b$ denote $\min\{a,b\}$. The following important lemma states the errors of approximating $\lambda^*(p)$ and $\nabla\lambda^*(p)$ using $\tilde{\lambda}(p)$ and $\nabla\tilde{\lambda}(p)$ respectively. (Note that by choosing $s \ge 3$, $\nabla\tilde{\lambda}$ is well-defined.)

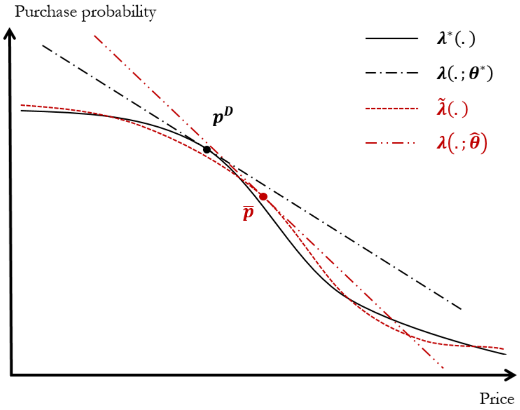
LEMMA 6. *Set $d = (\tilde{L}_0^{1/2}\log^{-1}k)^{1/(s+n)}$. If $\tilde{L}_0 \ge \log^3 k$ and $s \ge 3$, then there exist positive constants M_4 and M_5 independent of $k \ge 3$ such that for all $k \ge 3$,*

$$\mathbb{P}^\pi\left(||\lambda^*(.) - \tilde{\lambda}(.)||_\infty \ge M_4(\tilde{L}_0^{-1/2}\log k)^{\frac{s\wedge\bar{s}}{s+n}}\right) \le \frac{2}{k} \quad \textit{and}$$
$$\mathbb{P}^\pi\left(||(\nabla\lambda^*(.) - \nabla\tilde{\lambda}(.))'||_\infty \ge M_5(\tilde{L}_0^{-1/2}\log k)^{\frac{(s\wedge\bar{s})-1}{s+n}}\right) \le \frac{2}{k}.$$

Exploration algorithm. Using spline approximate $\tilde{\lambda}(p)$, we formulate an approximate deterministic problem as follows:

$$(\tilde{\mathrm{P}}) \qquad \tilde{r}^D := \max_{p\in\mathcal{P}}\left\{\tilde{r}(p):\ A\tilde{\lambda}(p) \preceq \frac{C}{T}\right\}$$

Figure 5 Illustration of locally linear approximation



Note. We propose to linearize the true nonparametric demand function $\lambda^*(.)$ at p^D and use this linear demand function $\lambda(.;\theta^*)$ as a surrogate for the true demand function. By doing this, we "transform" the nonparametric case into a parametric case with linear demand function family. Then, we linearize the estimated spline function $\tilde{\lambda}(.)$ at $\bar{p}$ to attain a linear demand function $\lambda(.;\hat{\theta})$, and view $\hat{\theta}$ as an estimate of the "true" θ^*.

where $\tilde{r}(p) = p \cdot \tilde{\lambda}(p)$. Let $\bar{p}$ denote an optimal solution of $\tilde{\mathrm{P}}$. Although $\bar{p}$ does not equal $p^D = \arg\max_{p\in\mathcal{P}}\{r(p) : s.t. A\lambda(p) \preceq C/T\}$ due to estimation error, $\bar{p}$ lies in close proximity of p^D when demand estimation error is small. The following lemma gives a Lipschitz-type "nonparametric" perturbation result for the deterministic pricing problem.

LEMMA 7. *There exists a positive constant M_6 independent of $||\lambda^*(.) - \tilde{\lambda}(.)||_\infty$ and $||(\nabla\lambda^*(.) - \nabla\tilde{\lambda}(.))'||_\infty$, such that $||p^D - \bar{p}||_\infty \le M_6 \max\{||\lambda^*(.) - \tilde{\lambda}(.)||_\infty, ||(\nabla\lambda^*(.) - \nabla\tilde{\lambda}(.))'||_\infty\}$.*

Since PSC is developed for the parametric demand case, to apply self-adjusting price update, we need to find an appropriate parametric demand family to approximate $\lambda^*(.)$. Note that we cannot use the spline function $\tilde{\lambda}(.)$ because its inverse function may not exist. A natural candidate is to use the linear function family $\lambda(p;\theta) = a + Bp$ where $a \in \mathbb{R}^n$, $B \in \mathbb{R}^{n\times n}$. Let $B'_1, \ldots, B'_n$ be the columns in B', and define $\theta = (a; B'_1; \ldots; B'_n) \in \mathbb{R}^{n(n+1)}$. Under the linear function family, the most proper candidate for the "true parameter vector" is $\theta^* = (\lambda^*(p^D) - \nabla\lambda^*(p^D) \cdot p^D; \nabla\lambda^*_1(p^D); \ldots; \nabla\lambda^*_n(p^D))$, which corresponds to the linearization of $\lambda^*(.)$ at p^D (i.e., $\lambda(.;\theta^*)$). Note that replacing $\lambda^*(.)$ with $\lambda(.,\theta^*)$ in the deterministic problem will not change the deterministic solution, i.e., $p^D = p^D(\theta^*)$, due to the KKT optimality condition. Therefore, as one may conjecture, if we can estimate θ^* well and use the corresponding estimated linear demand function to approximate the true demand function, we may be able to apply self-adjusting update and guarantee a strong revenue performance. However, we cannot simply use $(\tilde{\lambda}(p^D) - \nabla\tilde{\lambda}(p^D); \nabla\tilde{\lambda}_1(p^D); \ldots; \nabla\tilde{\lambda}_n(p^D))$ as an estimate of θ^*. This is because, even though $\tilde{\lambda}(.)$ can approximate $\lambda^*(.)$ well, we do not know p^D. That said, Lemma 7 tells us that

$\bar{p}$ lies in close proximity of p^D. This suggests that we use $\hat{\theta} = (\tilde{\lambda}(\bar{p}) - \nabla\tilde{\lambda}(\bar{p}) \cdot \bar{p}; \nabla\tilde{\lambda}_1(\bar{p}); \ldots; \nabla\tilde{\lambda}_n(\bar{p}))$ to approximate θ^*. (See Figure 5 for an illustration.) We state a lemma.

LEMMA 8. *There exists a constant M_7 independent of k and $\tilde{L}_0$ such that if $\tilde{L}_0 \geq \log^3 k$ and $s \geq 3$, then*

$$\mathbb{P}^\pi\left(||\theta^* - \hat{\theta}||_2 > M_7\epsilon(\tilde{L}_0)\right) \leq \frac{8}{k}$$

where $\epsilon(\tilde{L}_0) = (\log k/\sqrt{\tilde{L}_0})^{((s\wedge\bar{s})-1)/(s+n)}$.

Nonparametric self-adjusting control. We now introduce a heuristic that combines the self-adjusting price update with the aforementioned exploration algorithm. Since the duration of Spline Estimation is $\tilde{L}_0$ periods, the self-adjusting price updates will be applied starting from period $\tilde{L}_0 + 1$. Let $\tilde{\Delta}_t := \Delta_t + \lambda^*(p_t) - \lambda(p_t; \hat{\theta})$. Let $p^0(\theta)$ denote the optimal solution to the following optimization problem

$$(\mathrm{P}^0(\theta)) \qquad r^0 := \max_{p\in\mathcal{P}} \left\{ r(p;\theta) : \ A\lambda(p;\theta) \preceq \frac{C_{\tilde{L}_0}}{T - \tilde{L}_0} \right\}$$

where $r(p;\theta) = p \cdot \lambda(p;\theta)$, $C_{\tilde{L}_0}$ is the capacity level at the end of period $\tilde{L}_0$. Denote by $\lambda^0(\theta) = \lambda(p^0(\theta);\theta)$. The heuristic is outlined below.

Nonparametric Self-adjusting Control (NSC)

Input parameters: n, s, **Tuning Parameters:** d, L_0

Stage 1 (Exploration Phase 1 - Spline Estimation)

Apply *Spline Estimation* $\tilde{L}_0$ periods to get $\tilde{\lambda}(.)$.

Stage 2 (Exploration Phase 2 - Demand Linearization)

a. Solve $\tilde{\mathrm{P}}$ and obtain the optimizer $\bar{p}$.
b. Set $\hat{\theta} = (\tilde{\lambda}(\bar{p}) - \nabla\tilde{\lambda}(\bar{p}) \cdot \bar{p}; \nabla\tilde{\lambda}_1(\bar{p}); \ldots; \nabla\tilde{\lambda}_n(\bar{p}))$.
c. Let $\lambda(p;\hat{\theta}) = \hat{a} - \hat{B}p$, for all $p \in \mathcal{P}$, where $(\hat{a}; \hat{B}_1'; \ldots; \hat{B}_n') = \hat{\theta}$.

Stage 3 (Exploitation)

a. Solve $\mathrm{P}^0(\hat{\theta})$ for its static price $p^0(\hat{\theta})$
b. For $t = \tilde{L}_0 + 1$ to T, do:
- If $C_{t-1} \succ 0$, apply

$$p_t = p^0(\hat{\theta}) - \nabla p_\lambda(\lambda^0(\hat{\theta}); \hat{\theta}) \cdot \sum_{s=\tilde{L}_0+1}^{t-1} \frac{Q\tilde{\Delta}_s}{T - s}$$

- Otherwise, for product $j = 1$ to n, do:
  - If product j requires any resource that has been depleted, set $p_{t,j} = p_j^\infty$.
  - Otherwise, set $p_{t,j} = p_{t-1,j}$.

Table 2 Performance comparison of NSC and Misspecified PSC

	Revenue	NSC		Misspecified	
k	upper bd.	RL(Std.)	% of RL	RL(Std.)	% of RL
100	24970	10034 (29)	40.2%	17774 (52)	71.2%
300	74911	27441 (56)	36.6%	53489 (93)	71.4%
1000	249702	41106 (483)	16.5%	178192 (175)	71.4%
3000	749107	78433 (553)	10.5%	535224 (298)	71.4%
10000	2497023	167193 (794)	6.7%	1785524 (560)	71.5%
30000	7491069	349278 (1668)	4.7%	5359727 (989)	71.5%
100000	24970230	744175 (4938)	3.0%	17865978 (1725)	71.5%
300000	74910689	1532658 (7808)	2.0%	53593646 (2973)	71.5%

The setting of this numerical example is the same as in the one in Table 1. *NSC* is the NSC developed in this section with $s = 3$. *Misspecified* refers to the case where the seller uses PSC but wrongly assumes that the demand model comes from the linear function family whereas in fact it is a logit demand.

The following result states that when the tuning parameters are selected optimally, the performance of NSC is close to the best achievable performance.

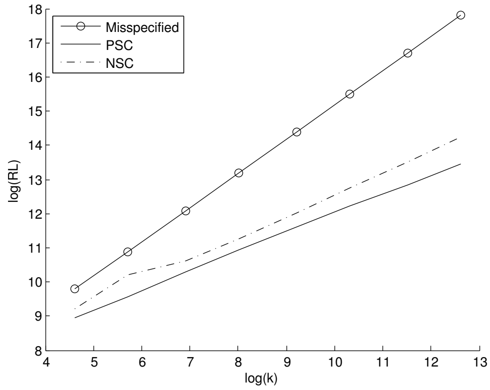
THEOREM 3. (NEAR RATE-OPTIMALITY OF NSC) *Let* $\tilde{L}_0 = k^{(s+n)/(2s+n-1)}(\log k)^{2(s-1)/(2s+n-1)}$. *In addition, we set* $d = (\tilde{L}_0^{-1/2} \log k)^{-1/(s+n)} = (\sqrt{k}\, \log^{-1} k)^{1/(2s+n-1)}$. *Then, there exists a constant* M_3 *independent of* $k > 3$ *such that for all* $s \geq 3$, *we have*

$$\rho^{NSC}(k) \leq M_3 k^{\frac{1}{2}+\epsilon(n,s,\bar{s})} \log k, \ \ \textit{where } \epsilon(n,s,\bar{s}) = \tfrac{1}{2}\left(\tfrac{2s-2(s\wedge\bar{s})+n+1}{2s+n-1}\right).$$

Note that unlike the heuristic proposed in Besbes and Zeevi (2012) which requires knowing $\bar{s}$ as input, our heuristic does not require the knowledge of the smoothness index $\bar{s}$. More interestingly, since most commonly used demand functions such as polynomial demand with arbitrary degree, logit demand, and exponential demand are infinitely differentiable (i.e., $\bar{s}$ can be arbitrarily large), for any fixed $\epsilon > 0$, we can select integers $s \geq (n+1)/(4\epsilon) - (n-1)/2$ such that the performance under *NSC* is $\mathcal{O}(k^{1/2+\epsilon} \log k)$. Since ϵ can be chosen to be arbitrarily small, the performance of NSC is very close to the best achievable performance lower bound $\Omega(\sqrt{k})$.

Per our discussions in §2, one drawback of the parametric approach is that the assumed demand function family may be misspecified; in particular, if the seller chooses the wrong functional form of the demand function and then blindly applies the parametric approach, the revenue loss can be huge. Given this drawback, our result provides an important insight: Since the asymptotic revenue loss gap between PSC and NSC is not too large when the demand function behaves nicely (i.e., the demand function is sufficiently smooth), if the seller is not very confident about the functional form of the demand function, he may be better off using the nonparametric approach. Indeed, our numerical illustration in Table 2 and Figure 6 show that model misspecification can potentially have a great impact on revenue.

Figure 6 Comparing the parametric approach and the nonparametric approach



Note. Using the data in Table 1-2, this log-log plot of revenue loss over scaling factor k compares the parametric and the nonparametric approaches. Note that the slope of the line represents the order of the revenue loss. The slopes for *Misspecified*, *PSC* and *NSC* are 1.0, 0.56 and 0.61 respectively in this graph.

5. Closing Remarks

We study the joint learning and pricing of the capacitated network RM problem. We develop heuristics for both the parametric and the nonparametric cases and evaluate their asymptotic performances. For the general parametric case, we develop the PSC heuristic, which first learns the demand function parameters by price experimentation and ML estimation, and then adjusts the price over time according to the realized demand. The heuristic is computationally easy to implement since it only requires one estimation and one optimization. Most strikingly, the heuristic achieves the best achievable asymptotic performance as its revenue loss rate is exactly $\mathcal{O}(\sqrt{k})$. This is the first known heuristic that attains the exact revenue loss lower bound for the capacitated network RM problems with general parametric demand. We also study the case where the family of the candidate demand functions satisfies the so-called "well-separatedness" condition. Under this condition the parameter estimation becomes much easier, and the seller can do exploitation while at the same time passively learn demand function. We develop the APSC heuristic, a modification of PSC, that reduces the revenue loss to $\mathcal{O}(\log^2 k)$. APSC is also a practical heuristic as it requires one optimization and $\Theta(\log_2 k)$ re-estimations.

Finally, we study the nonparametric case where the seller lacks the information of the functional form of demand. We develop a heuristic called NSC that uses Spline Estimation and demand linearization during the exploration stage to construct a linear demand function that closely approximates the nonparametric demand function around the optimal deterministic price. During the exploitation stage, we apply self-adjusting price updates. Although it is well-known that nonparametric learning in multidimensional problems suffers from the so-called "curse of dimensionality", we show that if the demand function is sufficiently smooth, then the performance under NSC is

$\mathcal{O}(k^{1/2+\epsilon}\log k)$ for any fixed $\epsilon > 0$. Since the family of most commonly used demand functions are infinitely differentiable, this result highlights an important point that not knowing the functional form of the demand function should not affect the revenue as much as one initially think.

In conclusion, two gaps in the literature have been significantly closed. By developing PSC with $\mathcal{O}(\sqrt{k})$ revenue loss bound, we close the gap between the revenue loss lower bound $\Omega(\sqrt{k})$ and the best revenue loss upper bound for existing heuristics under the general parametric case. By developing NSC with $\mathcal{O}(k^{1/2+\epsilon}\log k)$ revenue loss bound with arbitrarily small $\epsilon > 0$, we close the gap between the best performance bounds of the parametric case and the nonparametric case when the underlying demand function is sufficiently smooth. Our results suggest the wide applicability of self-adjusting controls in dynamic pricing problem. These simple self-adjusting controls can be used as a baseline for companies to develop more sophisticated dynamic pricing policies.

Endnotes

1. For example, in the airline industry, the benefit of using RM is roughly comparable to the airline's annual total profit, which is about 4%-5% of total revenue (Talluri and van Ryzin 2005).

2. A typical major US airline operates more than a thousand flights daily, each of which has more than ten different booking classes that are characterized by different combinations of service level and purchase restriction. Since passengers book tickets in advance, the airline needs to price not only the tickets for the same-day flights but also those with departure dates several months in the future. All these factors put together can easily translate into a daily pricing decision for *millions* of itineraries.

3. Although we implicitly assume that the demand function is stationary, our heuristics can be extended to accommodate some time-varying demand scenarios if the time-dependence of demand function has certain structural form. For example, in the fashion industry, irrespective of the condition of the market, the seller usually knows the fractions of the total sales that will be realized at multiple milestones over the selling season. This can be captured by incorporating in our demand model additively a time factor which is a known time-dependent fraction of an unknown total market size. Note that this model can be handled under the current stationary estimation framework by treating the unknown total market size as an additional parameter.

4. Jasin (2014) uses $Q = HA$ for some H satisfying $AH = I$.

References

Besbes, O., A. Zeevi. 2009. Dynamic pricing without knowing the demand function: Risk bound and near-optimal algorithms. *Oper. Res.* **57** 1407–1420.

Besbes, O., A. Zeevi. 2012. Blind network revenue management. *Oper. Res.* **60** 1537–1550.

Bitran, G., R. Caldentey. 2003. An overview of pricing models for revenue management. *Manufacturing Service Oper. Management* **5** 203–229.

Bonnans, J., A. Shapiro. 2000. *Perturbation Analysis of Optimization Problems*. Springer.

Borovkov, A. A. 1999. *Mathematical Statistics*. CRC Press.

Boyd, S., L. Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press.

Broder, J., P. Rusmevichientong. 2012. Dynamic pricing under a general parametric choice model. *Oper. Res.* **60** 965–980.

Chen, Q., S. Jasin, I. Duenyas. 2014. Realtime pricing with minimal and flexible price adjustment. *Under review* .

Elmaghraby, W., P. Keskinocak. 2003. Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Sci.* **49** 1287–1309.

Gallego, G., G. van Ryzin. 1994. Optimal dynamic pricing of inventory with stochastic demand over finite horizons. *Management Sci.* **40** 999–1020.

Gallego, G., G. van Ryzin. 1997. A multiproduct dynamic pricing problem and its applications to network yield management. *Oper. Res.* **45** 24–41.

Jasin, S. 2014. Reoptimization and self-adjusting price control for network revenue management. *Oper. Res.* **62** 1168–1178.

Lei, Y., S. Jasin, S. Amitabh. 2014. Near-optimal bisection search for nonparametric dynamic pricing with inventory constraint. *Under review* .

Maglaras, C., J. Meissner. 2006. Dynamic pricing strategies for multiproduct revenue management problems. *Manufacturing Service Oper. Management* **8** 136–148.

Pzman, A. 2013. *Design of experiments in nonlinear models : asymptotic normality, optimality criteria and small-sample properties*. Springer.

Schumaker, L. 2007. *Spline Functions: Basic Theory (Third Edition)*. Cambridge University Press.

Talluri, K., G. van Ryzin. 2005. *The theory and Practice of Revenue Management*. Springer.

Wang, Z., S. Deng, Y. Ye. 2014. Closing the gaps: A learning-while-doing algorithm for single-product revenue management problems. *Oper. Res.* **62** 318–331.

Proofs

A table of contents of the electronic companion is provided below. Note that some of the proofs in EC.1-EC.3 require some auxiliary results that can be found in EC.4.

Table of Contents

EC.1. Proof of Results in Section 3.1

In this section, we first prove Theorem 1 in §EC.1.1 using Lemma 1-3 and then we prove these lemmas in §§EC.1.2-EC.1.4 respectively. The proofs of other supporting lemmas which are used to prove Lemma 3 are deferred to §EC.1.5.

EC.1.1. Proof of Theorem 1

Throughout the proofs of this section, we fix $\pi = \text{PSC}$ and assume without loss of generality that $T=1$. Let $L = \lceil\sqrt{k}\rceil$. For $k \geq 3$, the total expected revenue loss under PSC is:

$$\rho^\pi(k) \leq L\bar{r} + M_0\left[\epsilon(L)^2 k + \frac{\log k}{1-\mathbb{P}^\pi(E(L)>\bar{\delta})} + L + \frac{1+k\,\mathbb{P}^\pi\left(E(L)>\bar{\delta}\right)}{1-\mathbb{P}^\pi(E(L)>\bar{\delta})}\right]$$

where the inequality follows because the revenue function in each period is bounded between 0 and $\bar{r}$ and also by Lemma 3. Since, by Lemma 2, $k\mathbb{P}^\pi(E(L)>\bar{\delta}) \leq k\eta_1 \exp(-\eta_2\bar{\delta}^2\lceil\sqrt{k}\rceil) \to 0$ as k tends to infinity, there exists a constant $K \geq 3$ such that for all $k > K$, we have $k\mathbb{P}^\pi(E(L)>\bar{\delta}) < \frac{1}{2}$ and $(1-\mathbb{P}^\pi(E(L)>\bar{\delta}))^{-1} < 2$. So, for all $k > K$, we can bound

$$\begin{aligned}\rho^\pi(k) &\leq \lceil\sqrt{k}\rceil\bar{r} + \frac{M_0\,\eta_3^2\,k}{\lceil\sqrt{k}\rceil} + 2M_0\log k + M_0\lceil\sqrt{k}\rceil + 3M_0 \\ &\leq 2\sqrt{k}\bar{r} + M_0\,\eta_3^2\sqrt{k} + 2M_0\sqrt{k} + 2M_0\sqrt{k} + 3M_0\sqrt{k} \\ &\leq (2\bar{r} + M_0\eta_3^2 + 7M_0)\sqrt{k},\end{aligned}$$

where η_3 is as in Lemma 2. As for $k < K$, we have $\rho^\pi(k) \leq K\bar{r}$. The result of Theorem 1 then follows by letting $M_1 = \max\{K\bar{r}, 2\bar{r} + M_0\eta_3^2 + 7M_0\}$. This completes the proof. $\square$

EC.1.2. Proof of Lemma 1

We will prove each part of the lemma in turn. Let $\bar{\delta} = \min\{\delta_1, \delta_2\}$ where δ_1 and δ_2 are strictly positive constants to be defined shortly.

Proof of part (a). This is an immediate corollary of Lemma 7 in §4. Note that, by assumption P2, we have $||\lambda(p;\theta^*) - \lambda(p;\theta)||_\infty \leq ||\lambda(p;\theta^*) - \lambda(p;\theta)||_2 \leq \omega||\theta^* - \theta||_2$ and $||(\nabla\lambda(p;\theta^*) - \nabla\lambda(p;\theta))'||_\infty = \max_{1\leq i\leq n}\sum_{j=1}^n |\frac{\partial\lambda_i}{\partial p_j}(p;\theta) - \frac{\partial\lambda_i}{\partial p_j}(p;\theta^*)| \leq n\omega||\theta^* - \theta||_2$ for all $\theta \in \Theta, p \in \mathcal{P}$. Hence, $||\lambda(.;\theta^*) - \lambda(.;\theta)||_\infty = \sup_{p\in\mathcal{P}}||\lambda(p;\theta^*) - \lambda(p;\theta)||_\infty \leq \omega||\theta^* - \theta||_2$ and $||(\nabla\lambda(.;\theta^*) - \nabla\lambda(.;\theta))'||_\infty = \sup_{p\in\mathcal{P}}||(\nabla\lambda(.;\theta^*) - \nabla\lambda(.;\theta))'||_\infty \leq n\omega||\theta^* - \theta||_2$. Therefore, by Lemma 7, $||p^D(\theta^*) - p^D(\theta)||_\infty \leq nM_6\omega||\theta^* - \theta||_2$. Let $\delta_1 = \phi(2n^{3/2}M_6\omega)^{-1}$. For all θ satisfying $||\theta - \theta^*||_2 \leq \bar{\delta} \leq \delta_1$, we have $||p^D(\theta^*) - p^D(\theta)||_2 \leq n^{1/2}||p^D(\theta^*) - p^D(\theta)||_\infty \leq n^{3/2}M_6\omega\delta_1 \leq \phi/2$. Hence, $p^D(\theta) \in \mathsf{Ball}(p^D(\theta^*), \phi/2)$. Since $\mathsf{Ball}(p^D(\theta^*), \phi) \subseteq \mathcal{P}$ by R4, we conclude that $\mathsf{Ball}(p^D(\theta), \phi/2) \subseteq \mathcal{P}$.

Since $\lambda(.;\theta^*)$ is continuously differentiable with respect to $p \in \mathcal{P}$ as implied by P3, and $\mathcal{P}$ is compact, there exists a constant $K > 0$ independent of $k > 0$ such that

$$\begin{aligned}||\lambda^D(\theta^*) - \lambda^D(\theta)||_2 &= ||\lambda(p^D(\theta^*);\theta^*) - \lambda(p^D(\theta);\theta)||_2 \\ &\leq ||\lambda(p^D(\theta^*);\theta^*) - \lambda(p^D(\theta);\theta^*)||_2 + ||\lambda(p^D(\theta);\theta^*) - \lambda(p^D(\theta);\theta)||_2 \\ &\leq K||p^D(\theta^*) - p^D(\theta)||_2 + \omega||\theta^* - \theta||_2 \\ &\leq (\omega + n^{3/2}KM_6\omega)||\theta^* - \theta||_2,\end{aligned}$$

where the second inequality also follows by P2. The result follows by letting $\kappa = \omega + n^{3/2}KM_6\omega$. Part (a) is proved.

Proof of part (b). Since $\mathrm{P}_\lambda(\theta)$ is a convex program for all $\theta \in \Theta$, by the Karush-Kuhn-Tucker optimality condition, $\nabla_\lambda r(\lambda^D(\theta);\theta) = A'\mu^D(\theta)$. By our assumption, A has full row rank. Thus, there exists some m by n matrix $\bar{A}$ such that $\mu^D(\theta) = \bar{A}\nabla_\lambda r(\lambda^D(\theta);\theta)$. Since the right hand side is continuous at θ^*, we conclude that $\mu^D(.)$ is continuous at θ^* as well. Part (b) is proved.

Proof of part (c). Let $\underline{\mu} = \min_{1\le i\le n}\{\mu_i^D(\theta^*) : \mu_i^D(\theta^*) > 0\}$. Since $\mu^D(.)$ is continuous at θ^* by part (b), there exists $\delta_2 > 0$ such that $||\mu^D(\theta) - \mu^D(\theta^*)||_2 < \underline{\mu}$ for all $\theta \in \mathsf{Ball}(\theta^*, \delta_2)$. This means for $\theta \in \mathsf{Ball}(\theta^*, \bar{\delta})$, we also have $\mu_i^D(\theta) > 0$ whenever $\mu_i^D(\theta^*) > 0$, which implies that the corresponding constraints in $\mathrm{P}(\theta)$ are binding due to Karush-Kuhn-Tucker condition. Part (c) is proved. □

EC.1.3. Proof of Lemma 2

The proof of Lemma 2 is a multiproduct extension of Lemma 3.7 in Broder and Rusmevichientong (2012), and is based on a well-known result in Maximum Likelihood Theory. We state this result in Theorem EC.1 (see §EC.4.1).

To apply Theorem EC.1 to our setting, we simply need to verify conditions (i)-(iv). First, note that Θ is a compact subset of $\mathbb{R}^q$ and $\mathcal{D}^{\tilde{q}}$ is a discrete-valued sample space. Conditions (i) and (iv), they are immediately satisfied because of S1 and S3. As for conditions (ii) and (iii), recall that

$$\left|\left|\nabla_\theta \log \mathbb{P}^{\tilde{p},\theta}(D_{1:\tilde{q}})\right|\right|_2 = \left|\left|\sum_{s=1}^{\tilde{q}}\left[\left(1-\sum_{j=1}^{n} D_{s,j}\right)\nabla_\theta \log\left(1-\sum_{j=1}^{n}\lambda_j(\tilde{p}^{(s)};\theta)\right) + \sum_{j=1}^{n} D_{s,j}\nabla_\theta \log \lambda_j(\tilde{p}^{(s)};\theta)\right]\right|\right|_2$$
$$\le \sum_{s=1}^{\tilde{q}}\left(\left|\left|\nabla_\theta \log\left(1-\sum_{j=1}^{n}\lambda_j(\tilde{p}^{(s)};\theta)\right)\right|\right|_2 + \sum_{j=1}^{n}\left|\left|\nabla_\theta \log \lambda_j(\tilde{p}^{(s)};\theta)\right|\right|_2\right).$$

By Assumption P1 and S2, for all $1 \le s \le \tilde{q}$ and $1 \le j \le n$, $\lambda_j(\tilde{p}^{(s)};.) \in \mathsf{C}^1(\Theta)$ and is bounded away from zero, and $\sum_{j=1}^n \lambda_j(\tilde{p}^{(s)};.) \in \mathsf{C}^1(\Theta)$ is also bounded away from one. These imply that $||\nabla_\theta \log\left(1-\sum_{j=1}^n \lambda_j(\tilde{p}^{(s)};.)\right)||_2$ and $||\nabla_\theta \log \lambda_j(\tilde{p}^{(s)};.)||_2$, $j = 1,\ldots,n$, are both continuous functions of θ for $s = 1,\ldots,\tilde{q}$ and are, due to compactness of Θ, bounded. So, (ii) follows. As for (iii), note that $\mathbb{P}^{\tilde{p},\theta}(D_{1:\tilde{q}})$ is continuous in θ and it is also bounded away from zero. (In fact, $\mathbb{P}^{\tilde{p},\theta}(D_{1:\tilde{q}}) \ge [\lambda_{\min}^n(1-\lambda_{\max})]^{\tilde{q}}$ by S2.) So, $\theta \to \sqrt{\mathbb{P}^{\tilde{p},\theta}(D_{1:\tilde{q}})}$ is differentiable on Θ for all $D_{1:\tilde{q}} \in \mathcal{D}^{\tilde{q}}$. We have thus verified all the conditions of Theorem EC.1.

We will now use Theorem EC.1 to prove Lemma 2. A direct application of Theorem EC.1 leads to $\mathbb{P}^\pi(E(L) > \delta) \le \eta_1 \exp(-\eta_2 L\delta^2)$. Also, since $\epsilon(L)^2 = \mathbb{E}^\pi\left[E(L)^2\right] = \int_0^\infty \mathbb{P}^\pi(E(L)^2 \ge x)dx = \int_0^\infty \mathbb{P}^\pi(E(L) \ge \sqrt{x})dx \le \int_0^\infty \eta_1 e^{-\eta_2 Lx}dx = \eta_1/(\eta_2 L)$, the result follows by taking $\eta_3 = \sqrt{\eta_1/\eta_2}$. □

EC.1.4. Proof of Lemma 3

Fix $\pi =$ PSC. Without loss of generality, we assume that $T=1$. Let $\mathcal{A}$ denote the event that $E(L)\le\bar{\delta}$. We first define a stopping time and show some useful properties of this stopping time on the event of $\mathcal{A}$. Let $\lambda_L>0$ be such that $A\lambda_L\mathbf{e}\prec C$. Define $\psi=\frac{\min\{\phi,2\lambda_L\}}{\max\{2,4\omega||Q||_2\}}$ and define the cumulative demand at the end of period t as $S_t:=\sum_{s=1}^t D_s$. Let τ be the minimum of k and the first time $t\ge L+1$ the following condition is violated:

$$(C1)\qquad \psi > \left|\left|\sum_{s=L+1}^{t}\frac{\hat{\Delta}_s}{k-s}\right|\right|_2+\left|\left|\frac{S_L-L\lambda_L\mathbf{e}}{k-t}\right|\right|_2.$$

Let C_t denote the available capacity level at the end of period t. We denote by $\hat{\lambda}_t:=\lambda^D(\hat{\theta}_L)-\sum_{s=L+1}^{t-1}\frac{Q\hat{\Delta}_s}{k-s}$ the demand rate that the seller *believes* he induces in period t, and by $\lambda_t:=\lambda(p(\hat{\lambda}_t;\hat{\theta}_L);\theta^*)$ the actual induced demand rate upon applying price $p(\hat{\lambda}_t;\hat{\theta}_L)$ in period t. Note that, by definition, we can also write $\hat{\lambda}_t=\lambda(p(\hat{\lambda}_t;\hat{\theta}_L);\hat{\theta}_L)$. We state two useful lemmas.

LEMMA EC.1. *For sample paths in $\mathcal{A}$, we have $C_t\succ 0$ and $\hat{\lambda}_t\in\Lambda_{\hat{\theta}_L}$ for all $L+1\le t<\tau$.*

LEMMA EC.2. *There exists $K_0>0$ independent of $k\ge 3$ such that for all $k\ge 3$*

$$\mathbb{E}^\pi[k-\tau|\mathcal{A}]\;\le\;K_0\left[\frac{\log k}{1-\mathbb{P}^\pi(E(L)>\bar{\delta})}+\epsilon(L)^2k+L\right].$$

Lemma EC.1 essentially says that, on $\mathcal{A}$, the remaining capacity C_t is always positive and the price $p(\hat{\lambda}_t;\hat{\theta}_L)$ is always feasible before τ, and Lemma EC.2 establishes a bound for the expected remaining time after τ. Define $r^D(\theta^*):=r(\lambda^D(\theta^*);\theta^*)$ and let R_t^π denote the revenue earned during period t under policy π. Let $\bar{\Delta}_t:=R_t^\pi-r(\lambda_t;\theta^*)$. We have:

$$\begin{aligned}
&\sum_{t=L+1}^{k} r^D(\theta^*)-\mathbb{E}^\pi\left[\hat{R}^\pi(k)\right]\\
&=\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left(r^D(\theta^*)-R_t^\pi\right)+\sum_{t=\tau}^{k}\left(r^D(\theta^*)-R_t^\pi\right)\right]\\
&=\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left(r^D(\theta^*)-r(\lambda_t;\theta^*)\right)+\sum_{t=\tau}^{k}\left(r^D(\theta^*)-R_t^\pi\right)\right]-\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\bar{\Delta}_t\right]\\
&\le\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left(r^D(\theta^*)-r(\lambda_t;\theta^*)\right)+\sum_{t=\tau}^{k}\left(r^D(\theta^*)-R_t^\pi\right)\middle|\mathcal{A}\right]\mathbb{P}^\pi(\mathcal{A})+\bar{r}k\mathbb{P}^\pi(\mathcal{A}^c)-\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\bar{\Delta}_t\right]\\
&\le\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left(r^D(\theta^*)-r(\lambda_t;\theta^*)\right)+\sum_{t=\tau}^{k}\left(r^D(\theta^*)-R_t^\pi\right)\middle|\mathcal{A}\right]+\bar{r}k\mathbb{P}^\pi(\mathcal{A}^c)+\bar{r}\\
&=\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left(r^D(\theta^*)-r(\lambda_t;\theta^*)\right)+\sum_{t=\tau}^{k}\left(r^D(\theta^*)-R_t^\pi\right)\middle|\mathcal{A}\right]+\bar{r}+\bar{r}k\mathbb{P}^\pi(E(L)>\bar{\delta}). \qquad \text{(EC.1)}
\end{aligned}$$

The first inequality follows because $\bar{r}$ is the upper bound on revenue rate for each period, which is also the maximum possible revenue loss for a single period on average. As for the second inequality, note that $\{\bar{\Delta}_t\}_{t=L+1}^{k-1}$ is a martingale difference sequence with respect to $\{\mathcal{H}_t\}_{t=L+1}^{k-1}$. Thus, by the Optional Stopping Time Theorem, we have $-\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\bar{\Delta}_t\right] = -\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau}\bar{\Delta}_t\right] + \mathbb{E}^\pi\left[\bar{\Delta}_\tau\right] \le \bar{r}$, so the second inequality holds. We now analyze the first two terms in (EC.1). By Taylor expansion and R3,

$$\begin{aligned}
&\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left(r^D(\theta^*) - r(\lambda_t;\theta^*)\right) + \sum_{t=\tau}^{k}\left(r^D(\theta^*) - R_t^\pi\right)\middle|\mathcal{A}\right] + \bar{r} \\
&\quad \le \mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\nabla r(\lambda^D(\theta^*);\theta^*)\cdot(\lambda^D(\theta^*) - \lambda_t)\middle|\mathcal{A}\right] + \frac{\bar{v}}{2}\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}||\lambda^D(\theta^*) - \lambda_t||_2^2\middle|\mathcal{A}\right] \\
&\quad\quad + \bar{r}\left(\mathbb{E}^\pi[k-\tau|\mathcal{A}] + 2\right)
\end{aligned} \tag{EC.2}$$

By Lemma EC.1, $\hat{\lambda}_t = \lambda^D(\hat{\theta}_L) - Q\sum_{s=L+1}^{t-1}\frac{\hat{\Delta}_s}{k-s} \in \Lambda_{\hat{\theta}_L}$ before τ. Also, recall that, by definition, $\hat{\Delta}_t = D_t - \hat{\lambda}_t = \Delta_t + \lambda_t - \hat{\lambda}_t$. So, we can write the first term in (EC.2) as follows:

$$\begin{aligned}
&\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\nabla r(\lambda^D(\theta^*);\theta^*)\cdot(\lambda^D(\theta^*) - \lambda_t)\middle|\mathcal{A}\right] \\
&\quad = \mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\mu^D(\theta^*)'A\left(\lambda^D(\theta^*) - \hat{\lambda}_t + \hat{\lambda}_t - \lambda_t\right)\middle|\mathcal{A}\right] \\
&\quad = \mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\mu^D(\theta^*)'\left(A\lambda^D(\theta^*) - A\lambda^D(\hat{\theta}_L) + \sum_{s=L+1}^{t-1}\frac{A\hat{\Delta}_s}{k-s} + A\Delta_t - A\hat{\Delta}_t\right)\middle|\mathcal{A}\right] \\
&\quad = \mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\mu^D(\theta^*)'\left(A\lambda^D(\theta^*) - A\lambda^D(\hat{\theta}_L)\right)\middle|\mathcal{A}\right] \\
&\quad\quad + \mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\mu^D(\theta^*)'\left(\sum_{s=L+1}^{t-1}\frac{A\hat{\Delta}_s}{k-s} + A\Delta_t - A\hat{\Delta}_t\right)\middle|\mathcal{A}\right].
\end{aligned} \tag{EC.3}$$

By Lemma 1(c), for all sample paths on $\mathcal{A}$, the set of constraints of $\mathrm{P}(\theta^*)$ that have nonzero optimal dual variables also have nonzero optimal dual variables in $\mathrm{P}(\hat{\theta}_L)$ and are thus binding at the optimal solution $\lambda^D(\hat{\theta}_L)$. This implies that the first expectation in (EC.3) is zero because, for all i, either we have $\mu_i^D(\theta^*) = 0$ or $(A\lambda^D(\theta^*))_i - (A\lambda^D(\hat{\theta}_L))_i = 0$. As for the second term of (EC.3), we can further write:

$$\begin{aligned}
&\mathbb{E}^\pi\left[\mu^D(\theta^*)'\sum_{t=L+1}^{\tau-1}\left(\sum_{s=L+1}^{t-1}\frac{A\hat{\Delta}_s}{k-s} + A\Delta_t - A\hat{\Delta}_t\right)\middle|\mathcal{A}\right] \\
&\quad = \mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\mu^D(\theta^*)'A\Delta_t\middle|\mathcal{A}\right] + \mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left(\frac{\tau-t-1}{k-t} - 1\right)\mu^D(\theta^*)'A\hat{\Delta}_t\middle|\mathcal{A}\right].
\end{aligned}$$

Since $\{\Delta_t\}_{t=L+1}^{k-1}$ is a martingale difference sequence with respect to $\{\mathcal{H}_t\}_{t=L+1}^{k-1}$, we can bound:

$$\begin{aligned}
&\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\mu^D(\theta^*)'A\Delta_t\middle|\mathcal{A}\right]\\
&\quad= \frac{\mu^D(\theta^*)'A}{\mathbb{P}^\pi(\mathcal{A})}\left\{\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\Delta_t\right]-\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\Delta_t\middle|\mathcal{A}^c\right]\mathbb{P}^\pi(\mathcal{A}^c)\right\}\\
&\quad\le \mu^D(\theta^*)'A\mathbf{e}\frac{1+k\mathbb{P}^\pi(E(L)>\bar{\delta})}{1-\mathbb{P}^\pi(E(L)>\bar{\delta})},
\end{aligned}$$

where the inequality follows because $\mathbb{E}^\pi[\sum_{t=L+1}^{\tau-1}\Delta_t]=\mathbb{E}^\pi[\sum_{t=L+1}^{\tau}\Delta_t]-\mathbb{E}^\pi[\Delta_\tau]\prec\mathbf{e}$ (by Optional Stopping Time Theorem) and the fact that $|\Delta_t|\prec\mathbf{e}$. As for the second term, note that, by (C1) in the definition of τ,

$$\begin{aligned}
&\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left(\frac{\tau-t-1}{k-t}-1\right)\mu^D(\theta^*)'A\hat{\Delta}_t\middle|\mathcal{A}\right]\\
&\quad\le \mathbb{E}^\pi\left[(k-\tau+1)\left|\mu^D(\theta^*)'\sum_{t=L+1}^{\tau-1}\frac{A\hat{\Delta}_t}{k-t}\right|\middle|\mathcal{A}\right]\\
&\quad\le \mathbb{E}^\pi\left[(k-\tau+1)\left|\left|\mu^D(\theta^*)\right|\right|_2||A||_2\left|\left|\sum_{t=L+1}^{\tau-1}\frac{\hat{\Delta}_t}{k-t}\right|\right|_2\middle|\mathcal{A}\right]\\
&\quad\le \psi\left|\left|\mu^D(\theta^*)\right|\right|_2||A||_2\left(\mathbb{E}^\pi\left[k-\tau|\mathcal{A}\right]+1\right).
\end{aligned}$$

Putting this together with Lemma EC.2, for the first term in (EC.2), we have:

$$\begin{aligned}
&\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\nabla_\lambda r(\lambda^D(\theta^*);\theta^*)\cdot(\lambda^D(\theta^*)-\lambda_t)\middle|\mathcal{A}\right]\\
&\quad\le K_1\left[\frac{\log k}{1-\mathbb{P}^\pi(E(L)>\bar{\delta})}+\epsilon(L)^2k+L+\frac{1+k\mathbb{P}^\pi(E(L)>\bar{\delta})}{1-\mathbb{P}^\pi(E(L)>\bar{\delta})}\right],
\end{aligned}$$

where the constant $K_1=\mu^D(\theta^*)'A\mathbf{e}+(1+K_0)\psi||\mu^D(\theta^*)||_2||A||_2$ is independent of $k\ge 3$.

We now bound the second term in (EC.2). Note that

$$\begin{aligned}
&\frac{\bar{v}}{2}\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}||\lambda^D(\theta^*)-\lambda_t||_2^2\middle|\mathcal{A}\right]\\
&\quad\le \bar{v}\,\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left|\left|\hat{\lambda}_t-\lambda_t\right|\right|_2^2\middle|\mathcal{A}\right]+\bar{v}\,\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left|\left|\lambda^D(\theta^*)-\hat{\lambda}_t\right|\right|_2^2\middle|\mathcal{A}\right]. \qquad \text{(EC.4)}
\end{aligned}$$

Since $\lambda_t=\lambda(p(\hat{\lambda}_t;\hat{\theta}_L);\theta^*)$ and $\hat{\lambda}_t=\lambda(p(\hat{\lambda}_t;\hat{\theta}_L);\hat{\theta}_L)$, by P2, we have

$$\bar{v}\mathbb{E}^\pi[\sum_{t=L+1}^{\tau-1}||\hat{\lambda}_t-\lambda_t||_2^2|\mathcal{A}]\le\bar{v}\omega^2k\,\mathbb{E}^\pi[||\theta^*-\hat{\theta}_L||_2^2\,|\,\mathcal{A}]\le\bar{v}\omega^2k\,\mathbb{E}^\pi[||\theta^*-\hat{\theta}_L||_2^2]\le\bar{v}\omega^2\epsilon(L)^2k.$$

(By definition of $\mathcal{A}$, $\mathbb{E}^\pi[||\theta^* - \hat{\theta}_L||_2^2 \,|\, \mathcal{A}] \le \mathbb{E}^\pi[||\theta^* - \hat{\theta}_L||_2^2]$.) As for the second term in (EC.4),

$$\begin{aligned}
&\bar{v}\,\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left|\left|\lambda^D(\theta^*) - \hat{\lambda}_t\right|\right|_2^2 \,\middle|\, \mathcal{A}\right] \\
&\quad= \bar{v}\,\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left|\left|\lambda^D(\theta^*) - \lambda^D(\hat{\theta}_L) + Q\sum_{s=L+1}^{t-1}\frac{\hat{\Delta}_s}{k-s}\right|\right|_2^2 \,\middle|\, \mathcal{A}\right] \\
&\quad\le 2\bar{v}k\mathbb{E}^\pi\left[\left|\left|\lambda^D(\theta^*) - \lambda^D(\hat{\theta}_L)\right|\right|_2^2 \,\middle|\, \mathcal{A}\right] + 2\bar{v}\,||Q||_2^2\,\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left|\left|\sum_{s=L+1}^{t-1}\frac{\hat{\Delta}_s}{k-s}\right|\right|_2^2 \,\middle|\, \mathcal{A}\right] \\
&\quad\le 2\bar{v}\kappa^2\epsilon(L)^2 k + 2\bar{v}\,||Q||_2^2\,\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left|\left|\sum_{s=L+1}^{t-1}\frac{\hat{\Delta}_s}{k-s}\right|\right|_2^2 \,\middle|\, \mathcal{A}\right] \qquad \text{(EC.5)}
\end{aligned}$$

where the second inequality follows by Lemma 1(a). Using $\hat{\Delta}_t = D_t - \hat{\lambda}_t = \Delta_t + \lambda_t - \hat{\lambda}_t$, we can bound the second term in (EC.5) as follows:

$$\begin{aligned}
&\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left|\left|\sum_{s=L+1}^{t-1}\frac{\hat{\Delta}_s}{k-s}\right|\right|_2^2 \,\middle|\, \mathcal{A}\right] \\
&\quad\le 2\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left|\left|\sum_{s=L+1}^{t-1}\frac{\Delta_s}{k-s}\right|\right|_2^2 \,\middle|\, \mathcal{A}\right] + 2\mathbb{E}^\pi\left[\sum_{t=L+1}^{\tau-1}\left|\left|\sum_{s=L+1}^{t-1}\frac{\lambda_s - \hat{\lambda}_s}{k-s}\right|\right|_2^2 \,\middle|\, \mathcal{A}\right] \\
&\quad\le \frac{2}{\mathbb{P}^\pi(\mathcal{A})}\mathbb{E}^\pi\left[\sum_{t=L+1}^{k-1}\left|\left|\sum_{s=L+1}^{t-1}\frac{\Delta_s}{k-s}\right|\right|_2^2\right] + 2\mathbb{E}^\pi\left[\sum_{t=L+1}^{k-1}\left(\sum_{s=L+1}^{t-1}\frac{\omega E(L)}{k-s}\right)^2 \,\middle|\, \mathcal{A}\right] \\
&\quad\le \frac{2}{\mathbb{P}^\pi(\mathcal{A})}\mathbb{E}^\pi\left[\sum_{t=L+1}^{k-1}\left|\left|\sum_{s=L+1}^{t-1}\frac{\Delta_s}{k-s}\right|\right|_2^2\right] + 2\sum_{t=L+1}^{k-1}\left[\sum_{s=L+1}^{t-1}\frac{\sqrt{\mathbb{E}^\pi\left[\omega^2 E(L)^2 | \mathcal{A}\right]}}{k-s}\right]^2 \\
&\quad\le \frac{2}{\mathbb{P}^\pi(\mathcal{A})}\mathbb{E}^\pi\left[\sum_{t=L+1}^{k-1}\sum_{s=L+1}^{t-1}\frac{||\Delta_s||_2^2}{(k-s)^2}\right] + 2\sum_{t=L+1}^{k-1}\left(\sum_{s=L+1}^{t-1}\frac{\omega\epsilon(L)}{k-s}\right)^2 \\
&\quad\le \frac{16}{1 - \mathbb{P}^\pi(E(L) > \bar{\delta})}\log k + 6\omega^2\epsilon(L)^2 k \qquad \text{(EC.6)}
\end{aligned}$$

where the second inequality holds by the law of total expectation and P2, the third inequality follows by first expanding the square of the sum and then applying Cauchy-Swartz inequality to the cross-terms, the fourth inequality follows by the orthogonality of martingale differences $\{\Delta_s\}$ and $\mathbb{E}^\pi[E(L)^2|\mathcal{A}] \le \epsilon(L)^2$, and the fifth inequality holds by integral approximation. In particular, the first term after the fourth inequality can be bounded using $||\Delta_s||_2 = ||D_s - \lambda_s||_2 \le ||D_s||_2 + ||\lambda_s||_2 \le 2$ and $\sum_{t=L+1}^{k-1}\sum_{s=L+1}^{t-1}\frac{1}{(k-s)^2} \le \sum_{t=L+1}^{k-1}\frac{1}{k-t} \le 1 + \log k \le 2\log k$ (recall that $k \ge 3$) whereas the second term can be bounded using the following integral comparison:

$$\sum_{t=L+1}^{k-1}\left(\sum_{s=L+1}^{t-1}\frac{1}{k-s}\right)^2 \le \sum_{t=1}^{k-1}\left(\sum_{s=1}^{t-1}\frac{1}{k-s}\right)^2 \le \sum_{t=1}^{k-1}\left(\int_1^t \frac{1}{k-s}ds\right)^2 \le \sum_{t=1}^{k-1}\log^2\left(\frac{k}{k-t}\right)$$

$$\leq \log^2 k + \int_1^{k-1} \log^2\left(\frac{k}{k-t}\right) dt \leq \log^2 k + 2k \leq 3k, \tag{EC.7}$$

where the last inequality follows because $\log^2 k < k$ for $k \geq 1$.

Thus, for the second term in (EC.2), we have

$$\frac{\bar{v}}{2}\mathbb{E}^\pi\left[\sum_{t=L+1}^{k-1} ||\lambda^D(\theta^*) - \lambda_t||_2^2 \middle| \mathcal{A}\right] \;\leq\; K_2\left[\frac{\log k}{1-\mathbb{P}^\pi(E(L) > \bar{\delta})} + \epsilon(L)^2 k\right],$$

where $K_2 = \bar{v}\omega^2 + 2\bar{v}\kappa^2 + 32\bar{v}||Q||_2^2 + 12\omega^2\bar{v}||Q||_2^2$. Combining all results together, we conclude that

$$\sum_{t=L+1}^{k} r^D(\theta^*) - \mathbb{E}^\pi\left[\hat{R}^\pi(k)\right] \leq M_0\left[\frac{\log k}{1-\mathbb{P}^\pi(E(L) > \bar{\delta})} + \epsilon(L)^2 k + L + \frac{1 + k\,\mathbb{P}^\pi\left(E(L) > \bar{\delta}\right)}{1-\mathbb{P}^\pi(E(L) > \bar{\delta})}\right],$$

where $M_0 = K_1 + K_2 + \bar{r}K_0 + 3\bar{r}$. (Note that the last term in (EC.1) can be bounded by $\bar{r}(1 + k\mathbb{P}^\pi(E(L) > \bar{\delta}))/(1 - \mathbb{P}^\pi(E(L) > \bar{\delta}))$.) This completes the proof of Lemma 3. □

EC.1.5. Proof of Supporting Lemmas

Proof of Lemma EC.1. As in the proof of Lemma 3, we assume without loss of generality that $T = 1$. First, note that $\hat{\lambda}_t \in \Lambda_{\hat{\theta}_L}$ is equivalent to $p_t \in \mathcal{P}$. Consider sample paths on $\mathcal{A}$. If $\tau \leq L+1$, then there is nothing to prove. Suppose that $\tau > L+1$, we will use an induction argument to establish the result. Since $E(L) \leq \bar{\delta}$ on $\mathcal{A}$, by Lemma 1(a), $\mathsf{Ball}(p^D(\hat{\theta}_L), \frac{\phi}{2}) \subseteq \mathcal{P}$. For $t = L+1$, $||p_{L+1} - p^D(\hat{\theta}_L)||_2 = 0 < \frac{\phi}{2}$, so $p_{L+1} \in \mathcal{P}$ and hence $\hat{\lambda}_{L+1} \in \Lambda_{\hat{\theta}_L}$. In addition, we also have:

$$\begin{aligned}
C_{L+1} = C_L - AD_{L+1} &= kC - AS_L - A\left(\hat{\lambda}_{L+1} + \hat{\Delta}_{L+1}\right) \\
&= kC - LC + LC - AS_L - A\left(\lambda^D(\hat{\theta}_L) + \hat{\Delta}_{L+1}\right) \\
&\succeq (k-L-1)C + LC - AS_L - A\hat{\Delta}_{L+1} \\
&\succeq (k-L-1)A\lambda_L\mathbf{e} + LA\lambda_L\mathbf{e} - AS_L - A\hat{\Delta}_{L+1} \\
&= (k-L-1)A\left(\lambda_L\mathbf{e} - \frac{S_L - L\lambda_L\mathbf{e}}{k-L-1} - \frac{\hat{\Delta}_{L+1}}{k-L-1}\right) \\
&\succeq (k-L-1)A\left(\lambda_L\mathbf{e} - \left|\left|\frac{S_L - L\lambda_L\mathbf{e}}{k-L-1}\right|\right|_2\mathbf{e} - \left|\left|\frac{\hat{\Delta}_{L+1}}{k-L-1}\right|\right|_2\mathbf{e}\right) \\
&\succ (k-L-1)\left(\lambda_L - \psi\right)A\mathbf{e} \\
&\succeq 0
\end{aligned}$$

(recall that $S_t = \sum_{s=1}^t D_s$) where the first inequality follows because $A\lambda^D(\hat{\theta}_L) \preceq C$, the second inequality follows because $A\lambda_L\mathbf{e} \preceq C$ by definition of λ_L, the fourth (strict) inequality follows by

(C1) and $A\mathbf{e} \succ 0$, and the last inequality follows by the definition of ψ. This is our base case. Now, suppose that $C_s \succ 0, \hat{\lambda}_s \in \Lambda_{\hat{\theta}_L}$ for all $s = L+1, L+2, \ldots, t-1$, and $t-1 < \tau$. If $t \geq \tau$, we have finished the induction. If, on the other hand, $t < \tau$,

$$\left|\left|p_t - p^D(\hat{\theta}_L)\right|\right|_2 \leq \omega ||Q||_2 \left|\left|\sum_{s=L+1}^{t-1} \frac{\hat{\Delta}_s}{k-s}\right|\right|_2 < \omega ||Q||_2 \psi \leq \frac{\phi}{4}$$

where the first inequality follows by $p_t = p(\hat{\lambda}_t; \hat{\theta}_L), p^D(\hat{\theta}_L) = p(\lambda^D(\hat{\theta}_L); \hat{\theta}_L)$ and P1, the second inequality follows by (C1) and the last inequality follows by the definition of ψ. So, by Lemma 1(a), we still have $p_t \in \mathcal{P}$ and hence $\hat{\lambda}_t \in \Lambda_{\hat{\theta}_L}$. As for the remaining capacity level C_t, by similar argument as before, we have

$$\begin{aligned}
C_t = C_L - \sum_{s=L+1}^{t} AD_s &= kC - AS_L - \sum_{s=L+1}^{t} A\left(\hat{\lambda}_s + \hat{\Delta}_s\right) \\
&= kC - tC + tC - AS_L - \sum_{s=L+1}^{t} A\left(\lambda^D(\hat{\theta}_L) - Q\sum_{v=L+1}^{s-1} \frac{\hat{\Delta}_v}{k-v} + \hat{\Delta}_s\right) \\
&\succeq (k-t)C + LC - AS_L - \sum_{s=L+1}^{t}\left(A\hat{\Delta}_s - \sum_{v=L+1}^{s-1} \frac{A\hat{\Delta}_v}{k-v}\right) \\
&\succeq (k-t)A\lambda_L\mathbf{e} + LA\lambda_L\mathbf{e} - AS_L - \sum_{s=L+1}^{t}\left(A\hat{\Delta}_s - \sum_{v=L+1}^{s-1} \frac{A\hat{\Delta}_v}{k-v}\right) \\
&= (k-t)A\left(\lambda_L\mathbf{e} - \frac{S_L - L\lambda_L\mathbf{e}}{k-t} - \sum_{s=L+1}^{t} \frac{\hat{\Delta}_s}{k-s}\right) \\
&\succeq (k-t)A\left(\lambda_L\mathbf{e} - \left|\left|\frac{S_L - L\lambda_L\mathbf{e}}{k-t}\right|\right|_2 \mathbf{e} - \left|\left|\sum_{s=L+1}^{t} \frac{\hat{\Delta}_s}{k-s}\right|\right|_2 \mathbf{e}\right) \\
&\succ (k-t)A(\lambda_L - \psi)\mathbf{e} \\
&\succeq 0.
\end{aligned}$$

This completes the induction. □

Proof of Lemma EC.2. As in the proof of Lemma 3, we assume without loss of generality that $T = 1$. Because τ is non-negative, we can write $\mathbb{E}^\pi[k-\tau|\mathcal{A}] = k - \sum_{t=0}^{k-1} \mathbb{P}^\pi(\tau > t|\mathcal{A}) = \sum_{t=1}^{k-1} \mathbb{P}^\pi(\tau \leq t|\mathcal{A})$. We now bound $\mathbb{P}^\pi(\tau \leq t|\mathcal{A})$. By the union bound, we have

$$\begin{aligned}
\mathbb{P}^\pi(\tau \leq t|\mathcal{A}) &= \mathbb{P}^\pi\left(\max_{L+1\leq s\leq t}\left\{\left|\left|\frac{S_L - L\lambda_L\mathbf{e}}{k-s}\right|\right|_2 + \left|\left|\sum_{v=L+1}^{s} \frac{\hat{\Delta}_v}{k-v}\right|\right|_2\right\} \geq \psi \,\middle|\, \mathcal{A}\right) \\
&\leq \mathbb{P}^\pi\left(\max_{L+1\leq s\leq t}\left|\left|\sum_{v=L+1}^{s} \frac{\hat{\Delta}_v}{k-v}\right|\right|_2 \geq \frac{\psi}{2} \,\middle|\, \mathcal{A}\right) + \mathbb{P}^\pi\left(\max_{L+1\leq s\leq t}\left|\left|\frac{S_L - L\lambda_L\mathbf{e}}{k-s}\right|\right|_2 \geq \frac{\psi}{2} \,\middle|\, \mathcal{A}\right).
\end{aligned} \tag{EC.8}$$

We first bound the first term in (EC.8) below.

$$
\begin{aligned}
&\mathbb{P}^\pi\left(\max_{L+1\le s\le t}\left|\left|\sum_{v=L+1}^{s}\frac{\hat{\Delta}_v}{k-v}\right|\right|_2 \ge \frac{\psi}{2}\middle|\mathcal{A}\right)\\
&\le \mathbb{P}^\pi\left(\max_{L+1\le s\le t}\left|\left|\sum_{v=L+1}^{s}\frac{\Delta_v}{k-v}\right|\right|_2^2 \ge \frac{\psi^2}{16}\middle|\mathcal{A}\right)+\mathbb{P}^\pi\left(\max_{L+1\le s\le t}\sum_{v=L+1}^{s}\frac{||\lambda_v-\hat{\lambda}_v||_2}{k-v}\ge\frac{\psi}{4}\middle|\mathcal{A}\right)\\
&\le \frac{1}{\mathbb{P}^\pi(\mathcal{A})}\,\mathbb{P}^\pi\left(\max_{L+1\le s\le t}\left|\left|\sum_{v=L+1}^{s}\frac{\Delta_v}{k-v}\right|\right|_2^2 \ge \frac{\psi^2}{16}\right)+\mathbb{P}^\pi\left(\left(\sum_{s=L+1}^{t}\frac{||\lambda_s-\hat{\lambda}_s||_2}{k-s}\right)^2\ge\frac{\psi^2}{16}\middle|\mathcal{A}\right)\\
&\le \frac{16}{\psi^2\mathbb{P}^\pi(\mathcal{A})}\,\mathbb{E}^\pi\left[\left|\left|\sum_{s=L+1}^{t}\frac{\Delta_s}{k-s}\right|\right|_2^2\right]+\frac{16}{\psi^2}\,\mathbb{E}^\pi\left[\left(\sum_{s=L+1}^{t}\frac{||\lambda_s-\hat{\lambda}_s||_2}{k-s}\right)^2\middle|\mathcal{A}\right]\\
&\le \frac{16}{\psi^2\mathbb{P}^\pi(\mathcal{A})}\,\mathbb{E}^\pi\left[\sum_{s=L+1}^{t}\frac{||\Delta_s||_2^2}{(k-s)^2}\right]+\frac{16}{\psi^2}\left(\sum_{s=L+1}^{t}\frac{\sqrt{\mathbb{E}^\pi[||\lambda_s-\hat{\lambda}_s||_2^2|\mathcal{A}]}}{k-s}\right)^2\\
&\le \frac{16}{\psi^2\mathbb{P}^\pi(\mathcal{A})}\left[\frac{4}{(k-t)^2}+\frac{4}{k-t}\right]+\frac{16}{\psi^2}\left[\frac{2\omega^2\epsilon(L)^2}{(k-t)^2}+2\omega^2\epsilon(L)^2\left(\log\left(\frac{k}{k-t}\right)\right)^2\right],
\end{aligned}
$$

where the first inequality follows by the definition of $\hat{\Delta}_v$, the triangle inequality of the norms and union bound, the second inequality follows by the law of total probability for the first term and the monotonicity of max-operator for the second term, the third inequality follows by the Doob's sub-martingale inequality for the first term and Markov's inequality for the second term, the fourth inequality follows by the orthogonality of martingale differences for the first term and Cauchy-Schwartz inequality for the second term, and the last inequality follows by $\mathbb{E}^\pi[E(L)^2|\mathcal{A}]\le\epsilon(L)^2$ and the same integral approximation bound as in (EC.6).

As for the second term in (EC.8), we can apply Markov's inequality and get:

$$
\begin{aligned}
\mathbb{P}^\pi\left(\max_{1\le s\le t}\left|\left|\frac{S_L-L\lambda_L\mathbf{e}}{k-s}\right|\right|_2\ge\frac{\psi}{2}\middle|\mathcal{A}\right)&\le\mathbb{P}^\pi\left(\frac{||S_L-L\lambda_L\mathbf{e}||_2^2}{(k-t)^2}\ge\frac{\psi^2}{4}\middle|\mathcal{A}\right)\\
&\le\max\left\{1,\frac{4}{\psi^2}\mathbb{E}^\pi\left[\frac{||S_L-L\lambda_L\mathbf{e}||_2^2}{(k-t)^2}\middle|\mathcal{A}\right]\right\}\\
&\le\max\left\{1,\frac{4n(1+\lambda_L)^2L^2}{\psi^2(k-t)^2}\right\},
\end{aligned}
$$

where the last inequality follows because $||S_L-L\lambda_L\mathbf{e}||_2\le||L\mathbf{e}+L\lambda_L\mathbf{e}||_2=\sqrt{n}(1+\lambda_L)L$. Putting all the bounds together, we have for all $k\ge3$:

$$
\begin{aligned}
\mathbb{E}^\pi[k-\tau|\mathcal{A}]\le&\sum_{t=1}^{k-1}\left\{\frac{16}{\psi^2\mathbb{P}^\pi(\mathcal{A})}\left[\frac{4}{(k-t)^2}+\frac{4}{k-t}\right]+\frac{16}{\psi^2}\left[\frac{2\omega^2\epsilon(L)^2}{(k-t)^2}+2\omega^2\epsilon(L)^2\log^2\left(\frac{k}{k-t}\right)\right]\right\}\\
&+\sum_{t=1}^{k-1}\max\left\{1,\frac{4n(1+\lambda_L)^2L^2}{\psi^2(k-t)^2}\right\}
\end{aligned}
$$

$$\leq \frac{128}{\psi^2\mathbb{P}^\pi(\mathcal{A})}\sum_{t=1}^{k-1}\frac{1}{k-t}+\frac{32\omega^2\epsilon(L)^2}{\psi^2}\sum_{t=1}^{k-1}\frac{1}{(k-t)^2}+\frac{32\omega^2\epsilon(L)^2}{\psi^2}\sum_{t=1}^{k-1}\left(\log\left(\frac{k}{k-t}\right)\right)^2$$
$$+\sum_{t=1}^{k-L-1}\frac{4n(1+\lambda_L)^2L^2}{\psi^2(k-t)^2}+L$$
$$\leq \frac{128}{\psi^2\mathbb{P}^\pi(\mathcal{A})}(1+\log k)+\frac{64\omega^2}{\psi^2}\epsilon(L)^2+\frac{96\omega^2}{\psi^2}\epsilon(L)^2k+\left(\frac{4n(1+\lambda_L)^2}{\psi^2}+1\right)L$$
$$\leq \frac{256}{\psi^2}\frac{\log k}{1-\mathbb{P}^\pi(E(L)>\bar{\delta})}+\frac{160\omega^2}{\psi^2}\epsilon(L)^2k+\left(\frac{4n(1+\lambda_L)^2}{\psi^2}+1\right)L$$

where the third inequality follows by integral approximation. The result follows by letting $K_0=\frac{256}{\psi^2}+\frac{160\omega^2}{\psi^2}+\frac{4n(1+\lambda_L)^2}{\psi^2}+1$. □

EC.2. Proof of Results in Section 3.2

In this section, we first prove Lemma 4 in §EC.2.1, followed by the proof of Theorem 2 in §EC.2.2. All the supporting lemmas which are used to prove Lemma 4 and Theorem 2 are proved in §EC.2.3.

EC.2.1. Proof of Lemma 4

We first illustrate the idea using Figure EC.1. Note that $E(t)>\epsilon$ is equivalent to the event that ML estimator $\hat{\theta}_t$ is in the outside of the ball $V:=\mathsf{Ball}(\theta^*,\epsilon)$. In addition, under the concavity assumption of the log-likelihood, $\hat{\theta}_t\neq\mathsf{Ball}(\theta^*,\epsilon)$ implies that at least one point on the surface of a hypercube S, which is centered at θ^* and is a subset of V, has a larger log-likelihood than the log-likelihood at θ^*. The probability of this event is a valid upper bound of $\mathbb{P}^\pi(E(t)>\epsilon)$. However, the challenge is that there are a continuum of such potential points. The idea of the proof is to consider a grid of points on the surface of that hypercube S, and the granularity of the grid is set to be fine enough so that any point on the surface of that hypercube can be closely approximated by one point on the grid. We will show that the existence of a point on the surface of S with a higher log-likelihood than the true parameter vector θ^* is extremely unlikely. We now rigorously prove this lemma.

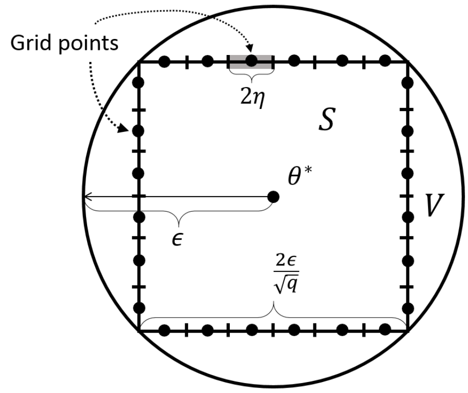
Step 1

Fix some $0<\tilde{\lambda}_{\min}<\tilde{\lambda}_{\max}<1$. First, we will show that for all $D\in\mathcal{D}$, for all $p\in\mathcal{W}(\tilde{\lambda}_{\min},\tilde{\lambda}_{\max})$ and for all $\theta\in\Theta$, $\nabla_\theta\log\mathbb{P}_1^{p,\theta}(D)$ is jointly continuous in θ and p. Recall that $\nabla_\theta\log\mathbb{P}_1^{p,\theta}(D)=((\partial/\partial\theta_1)\log\mathbb{P}_1^{p,\theta}(D);\ldots;(\partial/\partial\theta_n)\log\mathbb{P}_1^{p,\theta}(D))$ where for all $1\leq k\leq n$,

$$\frac{\partial\log\mathbb{P}_1^{p,\theta}(D)}{\partial\theta_k}=-\frac{(1-\sum_{j=1}^n D_j)\log\left(1-\sum_{j=1}^n\lambda_j(p;\theta)\right)}{1-\sum_{j=1}^n\lambda_j(p;\theta)}\left(\sum_{j=1}^n\frac{\partial\lambda_j(p;\theta)}{\partial\theta_k}\right)+\sum_{j=1}^n\frac{D_j\log(\lambda_j(p;\theta))}{\lambda_j(p;\theta)}\frac{\partial\lambda_j(p;\theta)}{\partial\theta_k}.$$

Since $\lambda_j(p;.)\in\mathsf{C}^1(\Theta)$ by P1, $\lambda(.;\theta)\in\mathsf{C}^2(\mathcal{P})$ by P3 and the denominators are strictly greater than zero, $\nabla_\theta\log\mathbb{P}_1^{p,\theta}(D)$ is jointly continuous in θ and p.

Figure EC.1 Geometric illustration of Lemma 4



Note. This illustrates the case when there are two parameters to estimate ($q = 2$). V denotes the disk (ball) centered at θ^* with radius ϵ. Note that the event of $||\theta^* - \hat{\theta}_t||_2 > \epsilon$ corresponds to the event when $\hat{\theta}_t$ lies in the exterior of V. In this example, the surface of the rectangle(hypercube) S consists of four edges.

Step 2

Since Θ and $\mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ are compact, $\mathcal{D}$ is finite and $\nabla_\theta \log \mathbb{P}_1^{p,\theta}(D)$ is jointly continuous in θ and p for all $D \in \mathcal{D}$, there exists a constant $c_g > 0$ independent of θ, p, D such that for all $\theta \in \Theta$, $p \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$, and $v \in \mathbb{R}^q$ satisfying $||v||_2 = 1$, $\nabla_\theta \log \mathbb{P}_1^{p,\theta}(D) \cdot v < c_g$. Therefore, for any $v, ||v||_2 = 1$, if $p_s^\pi \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ for $1 \le s \le t$, then we have:

$$\nabla_\theta \log \mathbb{P}_t^{\pi,\theta}(D_{1:t}) \cdot v = \sum_{s=1}^{t} \nabla_\theta \log \mathbb{P}_1^{p_s^\pi,\theta}(D_s) \cdot v < c_g t. \tag{EC.9}$$

Now, fix $\epsilon > 0$ and consider a hypercube $S \in \mathbb{R}^q$ centered at the origin with edge $2\epsilon/\sqrt{q}$. Let ∂S denote the surface of S, its area is given by $c_q(\epsilon/\sqrt{q})^{q-1}$ for a constant c_q that depends only on q. Cover ∂S with a set of identical hypercubes in $\mathbb{R}^{q-1}$ with edge 2η (see Figure EC.1 for an illustration) and denote by N the number of cubes needed to cover ∂S. Then, $N = (\epsilon/(\sqrt{q}\eta))^{q-1}$. Let $v_j \in \partial S, j = 1, \dots, N$ denote the center of those $2\eta-$cubes. These points constitute a set of grid points on the surface. Then for any $x \in \partial S$, $\min_{j=1,\dots,N} ||x - v_j||_2 \le \sqrt{q}\eta$. By W3, we have that for any $\theta' \in S + \theta^*$ and any $j = 1, \dots, N$,

$$\log \mathbb{P}_t^{\pi,\theta'}(D_{1:t}) - \log \mathbb{P}_t^{\pi,\theta^*+v_j}(D_{1:t}) \le \nabla_\theta \log \mathbb{P}_t^{\pi,\theta^*+v_j}(D_{1:t}) \cdot (\theta' - \theta^* - v_j)$$

Let $j^*(\theta) = \arg\min_{j=1,\dots,N} ||\theta - \theta^* - v_j||_2$. We then have

$$\log \mathbb{P}_t^{\pi,\theta'}(D_{1:t}) - \log \mathbb{P}_t^{\pi,\theta^*+v_{j^*(\theta')}}(D_{1:t}) \le c_g t ||\theta' - \theta^* - v_{j^*(\theta')}||_2 \le c_g \sqrt{q}\eta t. \tag{EC.10}$$

where the first inequality follows by (EC.9). The following is the key argument for this proof:

$$\left\{ ||\hat{\theta}_t - \theta^*||_2 > \epsilon \right\} \subseteq \left\{ ||\hat{\theta}_t - \theta^*||_\infty > \frac{\epsilon}{\sqrt{q}} \right\}$$

$$\subseteq \left\{\log \mathbb{P}_t^{\pi,\theta^*+v}(D_{1:t}) \geq \log \mathbb{P}_t^{\pi,\theta^*}(D_{1:t}), \text{for some } v \text{ with } ||v||_\infty = \tfrac{\epsilon}{\sqrt{q}}\right\}$$
$$\subseteq \left\{\log \mathbb{P}_t^{\pi,\theta^*+v_{j^*(\theta^*+v)}}(D_{1:t}) + c_g\sqrt{q}\eta t \geq \log \mathbb{P}_t^{\pi,\theta^*}(D_{1:t}), \text{for some } v \text{ with } ||v||_\infty = \tfrac{\epsilon}{\sqrt{q}}\right\}$$
$$\subseteq \cup_{j=1}^N \left\{\log \mathbb{P}_t^{\pi,\theta^*+v_j}(D_{1:t}) + c_g\sqrt{q}\eta t \geq \log \mathbb{P}_t^{\pi,\theta^*}(D_{1:t})\right\}$$
$$= \cup_{j=1}^N \left\{Z_t^\pi(v_j, D_{1:t}) \geq \exp(-c_g\sqrt{q}\eta t)\right\},$$

where $Z_t^\pi(u, D_{1:t}) := \mathbb{P}_t^{\pi,\theta^*+u}(D_{1:t})/\mathbb{P}_t^{\pi,\theta^*}(D_{1:t})$ is the likelihood ratio for any $u \in \Theta - \theta^*$. The first inclusion follows by norm inequality, the second inclusion follows by the concavity of the log-likelihood function and the definition of ML estimator, the third inclusion follows by (EC.10), the fourth inequality follows because by definition $j^*(\theta^*+v) \in \{1, \ldots, N\}$ for all v. We state a lemma below.

Lemma EC.3. *Fix some $0 < \tilde{\lambda}_{\min} < \tilde{\lambda}_{\max} < 1$. Suppose that an admissible control π satisfies $p_s = \pi_s(D_{1:s-1}) \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ for all $1 \leq s \leq t$. Then there exists a constant $c_h > 0$ such that for all π and for all $u \in \Theta - \theta^*$, $\mathbb{E}^\pi[\sqrt{Z_t^\pi(u, D_{1:t})}] \leq \exp(-c_h||u||_2^2 t/2)$.*

By Lemma EC.3, the following holds

$$\begin{aligned}
\mathbb{P}^\pi\left(||\hat{\theta}_t - \theta^*||_2 > \epsilon\right) &\leq \sum_{j=1}^N \mathbb{P}^\pi\left(Z_t^\pi(v_j, D_{1:t}) \geq \exp(-c_g\sqrt{q}\eta t)\right) \\
&\leq \sum_{j=1}^N \exp\left(\frac{c_g\sqrt{q}\eta t}{2}\right)\mathbb{E}^\pi\left[\sqrt{Z_t^\pi(v_j, D_{1:t})}\right] \\
&\leq \sum_{j=1}^N \exp\left(\frac{c_g\sqrt{q}\eta t}{2} - \frac{c_h||v_j||_2^2 t}{2}\right) \\
&\leq \left(\frac{\epsilon}{\sqrt{q}\eta}\right)^{q-1} \exp\left(-\frac{c_h\epsilon^2 t}{2q} + \frac{c_g\sqrt{q}\eta t}{2}\right),
\end{aligned}$$

where the second inequality follows by the Markov's inequality, the third inequality follows by Lemma EC.3, and the last inequality follows because $N = (\epsilon/(\sqrt{q}\eta))^{q-1}$ and $\min_{j=1,\ldots,N}||v_j||_2 \geq \min_{j=1,\ldots,N}||v_j||_\infty \geq \epsilon/\sqrt{q}$. Now, let $\eta = \epsilon/t$, then we have

$$\mathbb{P}^\pi\left(||\hat{\theta}_t - \theta^*||_2 > \epsilon\right) \leq \min\left\{1, q^{-\frac{q-1}{2}} t^{q-1} \exp\left(-\frac{c_h\epsilon^2 t}{2q} + \frac{c_g\sqrt{q}\epsilon}{2}\right)\right\}.$$

Note that when $\epsilon \leq 1, \exp((-c_h\epsilon^2 q^{-1}t + c_g\sqrt{q}\epsilon)/2) \leq \exp(c_g\sqrt{q}/2)\exp(-c_h\epsilon^2 q^{-1}t/4)$. Note also that when $\epsilon > 1$, there exists $M > 0$ independent of ϵ such that $\exp((-c_h\epsilon^2 q^{-1}t + c_g\sqrt{q}\epsilon)/2) \leq \exp(-c_h\epsilon^2 q^{-1}t/4), \forall t > M$. With these two observations, we consider two cases below.

Case 1: $t > M$. In this case, we have $\mathbb{P}^\pi\left(||\hat{\theta}_t - \theta^*||_2 > \epsilon\right) \leq \tilde{\eta}_4 t^{q-1}\exp(-\eta_5 t\epsilon^2)$, where $\tilde{\eta}_4 = q^{-(q-1)/2}\max\{1, \exp(c_g\sqrt{q}/2)$, and $\eta_5 = c_h q^{-1}/4$.

Case 2: $t \le M$. Let $\bar{\theta}$ be the largest distance between any two points in Θ. ($\bar{\theta} < \infty$ because Θ is bounded.) Then, we claim that for this case, $\mathbb{P}^\pi\left(||\hat{\theta}_t - \theta^*||_2 > \epsilon\right) \le \bar{\eta}_4 t^{q-1}\exp(-\eta_5 t\epsilon^2)$ where η_5 is defined as in Case 1 and $\bar{\eta}_4 = \exp(\eta_5 M\bar{\theta}^2)$. The claim is true because: if $\epsilon > \bar{\theta}$, $\mathbb{P}^\pi\left(||\hat{\theta}_t - \theta^*||_2 > \epsilon\right) = 0$, so the bound holds; if $\epsilon \le \bar{\theta}$, $\mathbb{P}^\pi\left(||\hat{\theta}_t - \theta^*||_2 > \epsilon\right) \le 1 = \bar{\eta}_4\exp(-\eta_5 M\bar{\theta}^2) \le \bar{\eta}_4 t^{q-1}\exp(-\eta_5 t\epsilon^2)$.

Combining the two cases above, we conclude that $\mathbb{P}^\pi\left(||\hat{\theta}_t - \theta^*||_2 > \epsilon\right) \le \min\{1, \eta_4 t^{q-1}\exp(-\eta_5 t\epsilon^2)\}$ where $\eta_4 = \max\{\tilde{\eta}_4, \bar{\eta}_4\}$. Hence,

$$
\begin{aligned}
\mathbb{E}^\pi\left[||\hat{\theta}_t - \theta^*||_2^2\right] &= \int_0^\infty \mathbb{P}^\pi\left(||\hat{\theta}_t - \theta^*||_2^2 \ge x\right)dx \\
&= \int_0^\infty \min\left\{1, \eta_4 t^{q-1}\exp\left(-\eta_5 tx\right)\right\}dx \\
&\le \int_0^{\frac{2(q-1)\log t}{\eta_5 t}} dx + \int_{\frac{2(q-1)\log t}{\eta_5 t}}^\infty \left[\eta_4 t^{q-1}\exp\left(-\frac{\eta_5 tx}{2}\right)\right]\exp\left(-\frac{\eta_5 tx}{2}\right)dx \\
&\le \frac{2(q-1)\log t}{\eta_5 t} + \eta_4\int_{\frac{2(q-1)\log t}{\eta_5 t}}^\infty \exp\left(-\frac{\eta_5 tx}{2}\right)dx \\
&\le \frac{2(q-1)\log t}{\eta_5 t} + \frac{2\eta_4}{\eta_5 t} \\
&\le \frac{2\max\{1,\eta_4\}}{\eta_5}\frac{(q-1)\log t + 1}{t}
\end{aligned}
$$

where the fourth inequality holds because for all $x \ge \frac{2(q-1)\log t}{\eta_5 t}$, $\eta_4 t^{q-1}\exp\left(-\frac{\eta_5 tx}{2}\right) \le 1$. We complete the proof by letting $\eta_6 = \sqrt{2\max\{1,\eta_4\}/\eta_5}$. □

EC.2.2. Proof of Theorem 2

We first state an analog of Lemma 1(a) for ECP(θ) below.

Lemma EC.4. *Suppose that $\mathcal{B}$ coincides with the set of binding constraints of $P(\theta^*)$ at the optimal solution. There exist $\tilde{\delta} > 0$ and $\tilde{\kappa} > 0$ independent of $k > 0$ such that for all $\theta \in \mathsf{Ball}(\theta^*, \tilde{\delta})$, $||x^D(\theta^*) - x^D(\theta)||_2 \le \tilde{\kappa}||\theta^* - \theta||_2$.*

The proof of Lemma EC.4 is similar to the proof of Lemma 1 and so is omitted. We now proceed to prove Theorem 2 in several steps.

Step 1

Fix $\pi =$ APSC and let $k \ge 3$ throughout the proof. Throughout this section, we will assume that $T = 1$. (This is without loss of generality.) Set $L = \lceil(\log k)^{1+\epsilon}\rceil$ and $\eta = (\log k)^{-\epsilon/4}$. We first show that the set of binding constraints of P(θ^*) at the optimal solution can be *correctly* identified with a very high probability. Let $\mathcal{E}_i := \{C_i = (A\lambda^D(\theta^*))_i,\, i \notin \mathcal{B}\} \cup \{C_i > (A\lambda^D(\theta^*))_i,\, i \in \mathcal{B}\}$ denote the event that the i^{th} capacity constraint is wrongly classified. (The event $\mathcal{E}_i$ is a union of two events:

either the i^{th} constraint is actually binding but not included in $\mathcal{B}$ or it is not binding but is included in $\mathcal{B}$.) By definition of η,

$$\begin{aligned}\mathbb{P}^\pi\left(C_i=(A\lambda^D(\theta^*))_i,\, i\notin\mathcal{B}\right) &= \mathbb{P}^\pi\left(C_i=(A\lambda^D(\theta^*))_i,\, C_i-(A\lambda^D(\hat\theta_{t_1}))_i>\eta\right)\\ &= \mathbb{P}^\pi\left((A\lambda^D(\theta^*)-A\lambda^D(\hat\theta_{t_1}))_i>\eta\right)\\ &\le \mathbb{P}^\pi\left(\kappa||A||_2E(t_1)>\eta\right)\\ &\le \eta_1\exp\left(-\eta_2 t_1\frac{\eta^2}{\kappa^2||A||_2^2}\right)\;\le\;\eta_1\exp\left(-\frac{\eta_2}{\kappa^2||A||_2^2}(\log k)^{1+\frac{\epsilon}{2}}\right),\end{aligned}$$

where the first inequality follows by Lemma 1(a), the second inequality follows by Lemma 2, and the last inequality holds by definition of t_1 and η. Define $\underline{s}:=\min\{C_i-(A\lambda^D(\theta^*))_i : C_i-(A\lambda^D(\theta^*))_i>0, i=1,\dots,m\}$. Since $\underline{s}$ does not scale with k, there exists a constant $\Omega_0>0$ such that $\eta<\underline{s}/2$ for all $k\ge\Omega_0$. So, for $k\ge\Omega_0$, by Lemmas 1(a) and 2, we can bound:

$$\begin{aligned}\mathbb{P}^\pi\left(C_i>(A\lambda^D(\theta^*))_i,\, i\in\mathcal{B}\right) &= \mathbb{P}^\pi\left(C_i\ge(A\lambda^D(\theta^*))_i+\underline{s},\, C_i-(A\lambda^D(\hat\theta_{t_1}))_i\le\eta\right)\\ &\le \mathbb{P}^\pi\left((A\lambda^D(\hat\theta_{t_1})-A\lambda^D(\theta^*))_i\ge\underline{s}-\eta\right)\\ &\le \mathbb{P}^\pi\left(\kappa||A||_2E(t_1)\ge\underline{s}-\eta\right)\\ &\le \eta_1\exp\left(-\eta_2 t_1\frac{(\underline{s}-\eta)^2}{\kappa^2||A||_2^2}\right)\;\le\;\eta_1\exp\left(-\frac{\eta_2\,\underline{s}^2}{4\kappa^2||A||_2^2}\log^{1+\epsilon}k\right).\end{aligned}$$

Putting the above two bounds together, for $k\ge\Omega_0$, the probability of wrongly identifying the binding constraints can be bounded as follows:

$$\begin{aligned}\mathbb{P}^\pi\left(\cup_{i=1}^m\mathcal{E}_i\right) &\le \sum_{i=1}^m\left[\mathbb{P}^\pi\left(C_i=(A\lambda^D(\theta^*))_i,\, i\notin\mathcal{B}\right)+\mathbb{P}^\pi\left(C_i>(A\lambda^D(\theta^*))_i,\, i\in\mathcal{B}\right)\right]\\ &\le m\,\eta_1\left[\exp\left(-\frac{\eta_2}{\kappa^2||A||_2^2}(\log k)^{1+\frac{\epsilon}{2}}\right)+\exp\left(-\frac{\eta_2\,\underline{s}^2}{4\kappa^2||A||_2^2}(\log k)^{1+\epsilon}\right)\right].\end{aligned}\quad\text{(EC.11)}$$

Step 2

Let τ be the minimum of k and the first time $t\ge t_1+1$ such that the following condition (C1) is violated: $\psi>||\sum_{s=t_1+1}^t\frac{\hat\Delta_s}{k-s}||_2+||\frac{S_L-L\lambda_L\mathbf{e}}{k-t}||_2$, where ψ is as defined in the proof of Theorem 1 and $\hat\Delta_s=D_s-\lambda(p_s;\hat\theta_{t_z})$ for $s\in(t_z,t_{z+1}]$ and $1\le z\le Z$. Define $\mathcal{A}:=\{\cap_{i=1}^m\mathcal{E}_i^c\}\cap\left\{E(t_z)\le\min\{\hat\delta,(\log t_z)^{-\epsilon/4}\},\text{for all } t_z<\tau\right\}$, where $\hat\delta=\min\{\bar\delta,\tilde\delta,\phi/(2\omega\kappa)\}$ and $\bar\delta$ and $\tilde\delta$ are as defined in Lemma 1 and Lemma EC.4 respectively. (Event $\mathcal{A}$ can be interpreted as the event where all binding constraints are correctly identified and the size of all subsequent estimation errors are sufficiently small.)

Note that for $t_z<\tau$, $\lambda^D(\theta^*)\in\Lambda_{\hat\theta_{t_z}}$ on $\mathcal{A}$. This is because $||p(\lambda^D(\theta^*);\hat\theta_{t_z})-p(\lambda^D(\hat\theta_{t_z});\hat\theta_{t_z})||_2\le\omega||\lambda^D(\theta^*)-\lambda^D(\hat\theta_{t_z})||_2\le\omega\kappa||\theta^*-\hat\theta_{t_z}||_2\le\phi/2$, where the first inequality follows by P1, the second inequality follows by Lemma 1(a) and the fact that $\hat{\delta} \le \bar{\delta}$, and the last inequality follows since $\hat{\delta} \le \phi/(2\omega\kappa)$. We then have $\lambda^D(\theta^*) \in \Lambda_{\hat{\theta}_{t_z}}$ since $p(\lambda^D(\theta^*); \hat{\theta}_{t_z}) \in \mathsf{Ball}(p^D(\hat{\theta}_{t_z}), \phi/2) \subseteq \mathcal{P}$, where the last inclusion follows by Lemma 1(a). The two important lemmas below establish the approximation error of DPUP and some important properties of the stopping time τ.

LEMMA EC.5. *There exist positive constants γ and ξ independent of $\theta \in \Theta$ such that if $||x^D(\theta) - x_{z-1}^{NT}||_2 \le \gamma$, then $||x^D(\theta) - x_z^{NT}||_2 \le \xi ||x^D(\theta) - x_{z-1}^{NT}||_2^2$.*

LEMMA EC.6. *There exist positive constants $0 < \tilde{\lambda}_{\min} < \tilde{\lambda}_{\max} < 1$, Ω_1, and constants Γ_1 and Γ_2 independent of $k \ge \Omega_1$, such that $\tilde{\lambda}_{\min} \le \lambda_{\min}, \tilde{\lambda}_{\max} \ge \lambda_{\max}$, and for all $k \ge \Omega_1$ and all sample paths on $\mathcal{A}$:*

(a) $||x^D(\hat{\theta}_{t_z}) - x_z^{NT}||_2^2 \le \Gamma_1 (\log t_z)^{-\epsilon/2}$ *for* $t_z < \tau$.

(b) $C_t \succ 0$, $p_t \in \mathsf{Ball}(p^D(\theta^*), 7\phi/8) \subseteq \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ *and* $\hat{\lambda}_t \in \Lambda_{\hat{\theta}_{t_z}}$ *for all* $t \in (t_z, t_{z+1}] \cap [t_1, \tau)$.

(c) $\mathbb{E}^\pi[||x^D(\hat{\theta}_{t_z}) - x_z^{NT}||_2^2 \mathbf{1}_{\{t_z < \tau\}} \mid \mathcal{A}] \le \Gamma_2 / t_z$

Lemma EC.5 essentially establishes a *uniform* locally quadratic convergence of the Newton's method for solving ECP($\hat{\theta}_{t_z}$) for all z, which is used for proving Lemma EC.6(a) and (c). Lemma EC.6(a) and (c) establish the approximation errors between x_Z^{NT} and the deterministic optimal solution $x^D(\hat{\theta}_{t_z})$. Note that Lemma EC.6(b) states that $p_t \in \mathsf{Ball}(p^D(\theta^*), 7\phi/8) \subseteq \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ for all $t_1 \le t < \tau$. In addition, for $t \le t_1$, $p_t \in \{\tilde{p}^{(1)}, \dots, \tilde{p}^{(\tilde{q})}\} \subseteq \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ due to $\tilde{\lambda}_{\min} \le \lambda_{\min}$, $\tilde{\lambda}_{\max} \ge \lambda_{\max}$ and S2. Therefore, the condition for Lemma 4 is satisfied. There exists a constant $\Omega_2 \ge \max\{\Omega_0, \Omega_1\}$ such that, for all $k \ge \Omega_2$,

$$
\begin{aligned}
k\mathbb{P}^\pi(\mathcal{A}^c) &\le k \sum_{z=1}^{Z} \left[\mathbb{P}^\pi(E(t_z) > \hat{\delta}) + \mathbb{P}^\pi(E(t_z) > (\log t_z)^{-\frac{\epsilon}{4}} \right] + k\, \mathbb{P}^\pi \left(\cup_{i=1}^m \mathcal{E}_i \right) \\
&\le k \sum_{z=1}^{Z} \eta_4 t_z^{q-1} \left[\exp\left(-\eta_5 t_z \hat{\delta}^2 \right) + \exp\left(-\frac{\eta_5 t_z}{(\log t_z)^{\frac{\epsilon}{2}}} \right) \right] + k\, \mathbb{P}^\pi \left(\cup_{i=1}^m \mathcal{E}_i \right) \\
&\le 2k(\log_2 k) \left[\exp\left(-\frac{\eta_5 (\log k)^{1+\epsilon} \hat{\delta}^2}{2} \right) + \exp\left(-\frac{\eta_5 (\log k)^{1+\epsilon}}{2 (\log k)^{\frac{\epsilon}{2}}} \right) \right] + k\, \mathbb{P}^\pi \left(\cup_{i=1}^m \mathcal{E}_i \right) \\
&\le 2k(\log_2 k) \left[\exp\left(-\frac{\eta_5 (\log k)^{1+\epsilon} \hat{\delta}^2}{2} \right) + \exp\left(-\frac{\eta_5 (\log k)^{1+\frac{\epsilon}{2}}}{2} \right) \right] \\
&\quad + m \eta_1 k \left[\exp\left(-\frac{\eta_2}{\kappa^2 ||A||_2^2} (\log k)^{1+\frac{\epsilon}{2}} \right) + \exp\left(-\frac{\eta_2 \underline{s}^2}{4\kappa^2 ||A||_2^2} (\log k)^{1+\epsilon} \right) \right] \le \frac{1}{2},
\end{aligned}
$$

where the second inequality follows by Lemma 4, the third inequality follows by a combination of $\eta_4 t_z^{q-1} \exp(-\eta_5 t_z \hat{\delta}^2/2) \to 0$ and $\eta_4 t_z^{q-1} \exp(-\eta_5 t_z (\log t_z)^{-\epsilon/2}/2) \to 0$ as $k \to \infty$, $t_z \ge t_1 \ge (\log k)^{1+\epsilon}$ for $z \ge 1$, and $Z \le \lceil \log_2 k \rceil \le 2 \log_2 k$, the fourth inequality follows by (EC.11), and the last inequality follows because the formula after the fourth inequality goes to zero as $k \to \infty$. Note that the above inequality also implies $\mathbb{P}^\pi(\mathcal{A}) > \frac{1}{2}$ when $k \geq \Omega_2$. Define $\Psi_\epsilon := \sum_{t=t_1+1}^{k-1} \left(\sum_{s=t_1+1}^{t-1} \frac{\bar{\epsilon}(s)}{k-s} \right)^2$ and $\Phi_\epsilon := \sum_{t=t_1+1}^{k-1} \bar{\epsilon}(s)^2$, where $\bar{\epsilon}(s) := \eta_6 \sqrt{[(q-1)\log t_z + 1]/t_z}$ for all $s \in (t_z, t_{z+1}]$. By Lemma 4, $\mathbb{E}^\pi[||\hat{\theta}_t - \theta^*||_2^2 \mathbf{1}_{\{t<\tau\}} | \mathcal{A}] \leq \bar{\epsilon}(t)^2$. The following result is useful to derive our bounds later.

Lemma EC.7. *Under APSC, there exists a constant $K_3 > 0$ independent of $k \geq 1$ such that $\Psi_\epsilon < K_3(1 + (q-1)\log k)$ and $\Phi_\epsilon < K_3[1 + \log k + (q-1)(\log k)^2]$.*

Step 3

Let $K = \max\{\Omega_0, \Omega_1, \Omega_2, 3\}$. If $k < K$, the total expected revenue loss can be bounded by $K\bar{r}$. So, we will focus on the case $k \geq K$. By the same arguments as in (EC.1) and (EC.2), $\rho^\pi(k) \leq L\bar{r} + \sum_{t=t_1+1}^{k} r^D(\theta^*) - \mathbb{E}[\hat{R}^\pi(k)]$ and

$$\sum_{t=t_1+1}^{k} r^D(\theta^*) - \mathbb{E}\left[\hat{R}^\pi(k)\right] \leq \mathbb{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1} \mu^D(\theta^*)' A(\lambda^D(\theta^*) - \lambda_t) \middle| \mathcal{A}\right] + \frac{\bar{v}}{2}\mathbb{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1} ||\lambda^D(\theta^*) - \lambda_t||_2^2\right] + \bar{r}\,\mathbb{E}^\pi[k-\tau|\mathcal{A}] + 2\bar{r} + \bar{r}\,k\mathbb{P}^\pi(\mathcal{A}^c) \tag{EC.12}$$

Note that on $\mathcal{A}$ we have $\mu^D(\theta^*)' A\lambda^D(\theta^*) = \mu^D(\theta^*)' A\lambda_z^{NT}$ (because $B\lambda^D(\theta^*) = C_B = B\lambda_z^{NT}$ and $\mu^D(\theta^*)_i = 0$ for all $i \notin \mathcal{B}$ by KKT conditions). Therefore, similar to the proof of Lemma 3, we can bound the first term in (EC.12) with $K_4\mathbb{E}^\pi[k-\tau+1|\mathcal{A}]$ where $K_4 := 3\mu^D(\theta^*)' A\mathbf{e} + \psi||\mu^D(\theta)||_2||A||_2$ is independent of $k \geq K$.

As for the second term in (EC.12), recall that $\hat{\lambda}_t = \lambda_{z(t)}^{NT} - Q\sum_{s=t_1+1}^{t-1} \frac{\hat{\Delta}_s}{k-s}$ denotes the demand rate that the seller believes he is inducing during period t where $z(t)$ is the unique integer z such that $t \in (t_z, t_{z+1}]$. Note that (EC.4) still holds. We can bound two term in (EC.4) respectively using: $\bar{v}\mathbb{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1} ||\hat{\lambda}_t - \lambda_t||_2^2 \middle| \mathcal{A}\right] = \bar{v}\sum_{t=t_1+1}^{k-1} \mathbb{E}^\pi\left[\omega^2||\hat{\theta}_t - \theta^*||_2^2\,\mathbf{1}_{\{t<\tau\}} \middle| \mathcal{A}\right] \leq \bar{v}\omega^2\sum_{t=t_1+1}^{k-1} \bar{\epsilon}(t)^2 \leq \bar{v}\omega^2\Phi_\epsilon$ (by P2), and

$$\begin{aligned}
&\bar{v}\mathbb{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1} ||\lambda^D(\theta^*) - \hat{\lambda}_t||_2^2 \middle| \mathcal{A}\right] \\
&\leq 2\bar{v}\mathbb{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1} ||\lambda^D(\theta^*) - \lambda_{z(t)}^{NT}||_2^2 \middle| \mathcal{A}\right] + 2\bar{v}\mathbb{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1} \left|\left| Q\sum_{s=t_1+1}^{t-1} \frac{\hat{\Delta}_s}{k-s} \right|\right|_2^2 \middle| \mathcal{A}\right] \\
&\leq 2\bar{v}\mathbb{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1} ||\lambda^D(\theta^*) - \lambda_{z(t)}^{NT}||_2^2 \middle| \mathcal{A}\right] \\
&\quad + 2\bar{v}||Q||_2^2\left(\mathbb{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1} \left|\left| \sum_{s=t_1+1}^{t-1} \frac{\Delta_s}{k-s} \right|\right|_2^2 \middle| \mathcal{A}\right] + \mathbb{E}^\pi\left[\sum_{t=t_1+1}^{k-1} \left(\sum_{s=t_1+1}^{t-1} \frac{\omega E(s)\mathbf{1}_{\{s<\tau\}}}{k-s}\right)^2 \middle| \mathcal{A}\right]\right)
\end{aligned}$$

$$\leq 2\bar{v}\mathbb{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1}||\lambda^D(\theta^*)-\lambda^{NT}_{z(t)}||_2^2\middle|\mathcal{A}\right]+2\bar{v}||Q||_2^2\left(\frac{16}{\mathbb{P}^\pi(\mathcal{A})}\log k+\sum_{t=t_1+1}^{k-1}\left[\sum_{s=t_1+1}^{t-1}\frac{\sqrt{\mathbb{E}^\pi\left[\omega^2E(s)^2\mathbf{1}_{\{s<\tau\}}\middle|\mathcal{A}\right]}}{k-s}\right]^2\right)$$

$$\leq 2\bar{v}\mathbb{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1}||\lambda^D(\theta^*)-\lambda^{NT}_{z(t)}||_2^2\middle|\mathcal{A}\right]+2\bar{v}||Q||_2^2(32\log k+\omega^2\Psi_\epsilon)$$

$$\leq K_5(\Psi_\epsilon+\log k)+2\bar{v}\mathbb{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1}||\lambda^D(\theta^*)-\lambda^{NT}_{z(t)}||_2^2\middle|\mathcal{A}\right]$$

for some constant $K_5>0$ independent of $k\geq K$ (the second and the third inequalities follow by the same argument as in (EC.5) and (EC.6) and recall that $K\geq 3$), and the fourth inequality follows since $\mathbb{E}^\pi\left[E(s)^2\mathbf{1}_{\{s<\tau\}}\middle|\mathcal{A}\right]\leq\bar{\epsilon}(s)^2$. We now analyze the last term of the above. Note that, on $\mathcal{A}$, we have for all $t<\tau$

$$\begin{aligned}\left|\left|\lambda^D(\theta^*)-\lambda^{NT}_{z(t)}\right|\right|_2 &\leq \left|\left|\lambda^D(\theta^*)-x^{NT}_{z(t)}\right|\right|_2+\left|\left|\lambda^{NT}_{z(t)}-x^{NT}_{z(t)}\right|\right|_2\\ &\leq 2\left|\left|\lambda^D(\theta^*)-x^{NT}_{z(t)}\right|\right|_2\\ &= 2\left|\left|x^D(\theta^*)-x^{NT}_{z(t)}\right|\right|_2\\ &\leq 2\left|\left|x^D(\theta^*)-x^D(\hat{\theta}_{t_{z(t)}})\right|\right|_2+2\left|\left|x^D(\hat{\theta}_{t_{z(t)}})-x^{NT}_z\right|\right|_2,\end{aligned}\tag{EC.13}$$

where the second inequality follows because $\lambda^D(\theta^*)$ lies in $\mathcal{S}_{z(t)}$ where $x^{NT}_{z(t)}$ is projected into (note that on $\mathcal{A}$, (1) $\hat{\theta}_{t_{z(t)}}\in\mathsf{Ball}(\theta^*,\phi/(2\omega\kappa))$ for $t<\tau$ which implies that, as shown previously, $\lambda^D(\theta^*)\in\Lambda_{\hat{\theta}_{t_{z(t)}}}$, and (2) the binding constraints of $\mathrm{P}(\theta^*)$ at $\lambda^D(\theta^*)$ are correctly identified which means that $B\lambda^D(\theta^*)=C_B$ and $N\lambda^D(\theta^*)\leq C_N$) and the equality follows because $\lambda^D(\theta^*)=x^D(\theta^*)$ on $\mathcal{A}$ due to the strongly concavity of the objective and the fact that $\lambda^D(\theta^*)$ is an interior solution. By Lemma EC.4

$$\mathbb{E}^\pi\left[\sum_{s=t_1+1}^{\tau-1}||x^D(\theta^*)-x^D(\hat{\theta}_{t_{z(s)}})||_2^2\middle|\mathcal{A}\right]=\sum_{s=t_1+1}^{k-1}\mathbb{E}^\pi\left[\tilde{\kappa}^2||\theta^*-\hat{\theta}_{t_{z(s)}}||_2^2\mathbf{1}_{\{s<\tau\}}\middle|\mathcal{A}\right]\ \leq\ \tilde{\kappa}^2\Phi_\epsilon$$

Furthermore, by Lemma EC.6(a) and the fact that $t_{z+1}-t_z\leq 2t_z$ for all z, we have

$$\begin{aligned}\mathbb{E}^\pi\left[\sum_{s=t_1+1}^{\tau-1}||x^D(\hat{\theta}_{t_{z(s)}})-x^{NT}_{z(s)}||_2^2\middle|\mathcal{A}\right] &= \sum_{s=t_1+1}^{k-1}\mathbb{E}^\pi\left[||x^D(\hat{\theta}_{t_{z(s)}})-x^{NT}_{z(s)}||_2^2\mathbf{1}_{\{s<\tau\}}\middle|\mathcal{A}\right]\\ &\leq \sum_{z=1}^{Z}(t_{z+1}-t_z)\frac{\Gamma_2}{t_z}\ \leq\ 2Z\Gamma_2\ \leq\ 4\Gamma_2\log_2 k.\end{aligned}$$

Combining the inequalities above, the second term of (EC.12) can be bounded as follows:

$$\begin{aligned}\frac{\bar{v}}{2}\mathbb{E}^\pi\left[\sum_{t=t_1+1}^{\tau-1}||\lambda^D(\theta^*)-\lambda_t||_2^2\right] &\leq \bar{v}\omega^2\Phi_\epsilon+K_5(\Psi_\epsilon+\log k)+4\bar{v}\tilde{\kappa}^2\Phi_\epsilon+16\bar{v}\Gamma_2\log_2\\ &\leq K_6(1+\log k+(q-1)\log^2 k)\end{aligned}$$

for $K_6 = (\bar{v}\omega^2 + 4\bar{v}\tilde{\kappa}^2 + K_5)K_3 + K_5 + 16\bar{v}\Gamma_2$. To bound the third term in (EC.12), the following lemma is useful.

LEMMA EC.8. *There exists a constant $K_7 > 0$ independent of $k \geq K$ such that for all $k \geq K$, $\mathbb{E}^\pi[k - \tau|\mathcal{A}] \leq K_7(\log k + L)$.*

Combining all the above and recalling that $L = \lceil (\log k)^{1+\epsilon} \rceil$, for all $k \geq K$, we have:

$$
\begin{aligned}
\rho^\pi(k) &\leq 2\bar{r}(\log k)^{1+\epsilon} + (K_4 + \bar{r})(\mathbb{E}^\pi[k-\tau|\mathcal{A}] + 1) + K_6(1 + \log k + (q-1)\log^2 k) + \frac{5}{2}\bar{r} \\
&\leq \left(2\bar{r} + K_4 + \bar{r} + K_6 + \frac{5}{2}\bar{r}\right)\left[1 + (\log k)^{1+\epsilon} + (q-1)\log^2 k\right] \\
&\leq K_8[(\log k)^{1+\epsilon} + (q-1)\log^2 k],
\end{aligned}
$$

for some constant K_8 independent of $k \geq K$. The result of Theorem 2 follows by using $M_2 = \max\{\bar{r}K, K_8\}$. □

EC.2.3. Proof of Supporting Lemmas

Proof of Lemma EC.3. Recall that $\mathcal{D} = \{D \in \{0,1\}^n : \sum_{j=1}^n D_j \leq 1\}$. We define the conditional Hellinger distance as follows:

$$
H_t^\pi(\theta_1, \theta_2, D_t|D_{1:t-1}) := \sum_{D_t \in \mathcal{D}} \left(\sqrt{\mathbb{P}_t^{\pi,\theta_1}(D_t|D_{1:t-1})} - \sqrt{\mathbb{P}_t^{\pi,\theta_2}(D_t|D_{1:t-1})}\right)^2.
$$

We state a lemma and postpone its proof to the end of this subsection.

LEMMA EC.9. *Fix some $0 < \tilde{\lambda}_{\min} < \tilde{\lambda}_{\max} < 1$. Suppose that an admissible control π satisfies $p_s = \pi_s(D_{1:s-1}) \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ for all $1 \leq s \leq t$. Then there exists a positive constant c_h such that $H_t^\pi(\theta_1, \theta_2, D_t|D_{1:t-1}) \geq c_h||\theta_1 - \theta_2||_2^2$ for all $\theta_1, \theta_2 \in \Theta$.*

For $u \in \Theta - \theta^*$, define $Z_t^\pi(u, D_t|D_{1:t-1}) := \mathbb{P}_t^{\pi,\theta^*+u}(D_t|D_{1:t-1})/\mathbb{P}_t^{\pi,\theta^*}(D_t|D_{1:t-1})$. Using Lemma EC.9, we can derive a bound for its moment below:

$$
\begin{aligned}
\mathbb{E}^\pi\left[\sqrt{Z_t^\pi(u, D_t|D_{1:t-1})}\right] &= \sum_{D_t \in \mathcal{D}} \sqrt{\frac{\mathbb{P}_t^{\pi,\theta^*+u}(D_t|D_{1:t-1})}{\mathbb{P}_t^{\pi,\theta^*}(D_t|D_{1:t-1})}}\,\mathbb{P}_t^{\pi,\theta^*}(D_t|D_{1:t-1}) \\
&= \sum_{D_t \in \mathcal{D}} \sqrt{\mathbb{P}_t^{\pi,\theta^*+u}(D_t|D_{1:t-1})\mathbb{P}_t^{\pi,\theta^*}(D_t|D_{1:t-1})} \\
&= 1 - \frac{H_t^\pi(\theta^*, \theta^*+u, D_t|D_{1:t-1})}{2} \\
&\leq \exp\left(-\frac{H_t^\pi(\theta^*, \theta^*+u, D_t|D_{1:t-1})}{2}\right) \leq \exp\left(-\frac{c_h||u||_2^2}{2}\right).
\end{aligned}
$$

The result of Lemma EC.3 can now be proved by repeated conditioning: by definition,

$$\begin{aligned}
\mathbb{E}^{\pi}\left[\sqrt{Z_t^{\pi}(u,D_{1:t})}\right] &= \mathbb{E}^{\pi}\left[\mathbb{E}^{\pi}\left[\sqrt{Z_t^{\pi}(u,D_{1:t})}\,\middle|\, D_{1:t-1}\right]\right] \\
&= \mathbb{E}^{\pi}\left[\sqrt{Z_{t-1}^{\pi}(u,D_{1:t-1})}\;\mathbb{E}^{\pi}\left[\sqrt{Z_k^{\pi}(u,D_t|D_{1:t-1})}\right]\right] \\
&\le \mathbb{E}^{\pi}\left[\sqrt{Z_{t-1}^{\pi}(u,D_{1:t-1})}\right]\exp\left(-\frac{c_h||u||_2^2}{2}\right) \\
&\le \exp\left(-\frac{c_h||u||_2^2 t}{2}\right).
\end{aligned}$$

This completes the proof. □

Proof of Lemma EC.5. Fix $\theta\in\Theta$. Note that ECP(θ) is a convex optimization with linear equality constraints. Let m_B denote the number of columns of B, and define F to be an n by $n-m_B$ matrix whose columns are linearly independent and $BF=0$. (In case there are multiple matrices that satisfy this condition, pick any one of them.) Then $\{x: Bx=C_B/T\}=\{x: x=Fz+\hat{x}, z\in\mathbb{R}^{n-m_B}\}$ where $\hat{x}$ satisfies $B\hat{x}=C_B/T$. Hence, ECP(θ) is equivalent to an unconstrained optimization problem $\max_{z\in\mathbb{R}^{n-m_B}} g(z;\theta):=r(Fz+\hat{x};\theta)$ in the sense that there is a one-to-one mapping between the optimizer of ECP(θ) $x^D(\theta)$ and the optimizer of the unconstrained problem $z^D(\theta)$: (1) $x^D(\theta)=Fz^D(\theta)+\hat{x}$, and (2) $z^D(\theta)=(F'F)^{-1}F'(x^D(\theta)-\hat{x})$. In addition, by Section 10.2.3 in Boyd and Vandenberghe (2004), if a feasible point of ECP(θ) $x^{(k)}$ and a feasible point of the unconstrained problem $z^{(k)}$ satisfy $x^{(k)}=Fz^{(k)}+\hat{x}$, then the Newton steps for ECP(θ) (to obtain a new feasible point $x^{(k+1)}$) and the unconstrained problem (to obtain a new feasible point $z^{(k+1)}$) coincide in the sense that $x^{(k+1)}=Fz^{(k+1)}+\hat{x}$. This relationship enables us to analyze the behavior of $x^{(k)}$ by studying $z^{(k)}$ whose convergence behavior is characterized by Theorem EC.2 (see §EC.4.2).

Before applying Theorem EC.2, we first show that the conditions in Theorem EC.2 hold. Note that since Λ_θ is compact, the linear transformation of it, $\mathcal{Z}_\theta:=\{z: z=(F'F)^{-1}F'(x-\hat{x}), x\in\Lambda_\theta\}$ is also compact. Also note that since $p(.;\theta)\in\mathsf{C}^2(\Lambda_\theta)$ by P3, $r(.;\theta)\in\mathsf{C}^2(\Lambda_\theta)$ and $g(.;\theta)\in\mathsf{C}^2(\mathcal{Z}_\theta)$. Hence condition (i) holds: there exists some constant L such that $||\nabla^2_{zz}g(z;\theta)-\nabla^2_{zz}g(y;\theta)||_2\le L||z-y||_2$. Denote by $\sigma_{\min}(.),\sigma_{\max}(.)$ the smallest and the largest eigenvalues of a squared matrix. Since $\nabla^2_{zz}g(z;\theta)=F'\nabla^2_{\lambda\lambda}r(Fz+\hat{x};\theta)F$ and $-MI\preceq\nabla^2_{\lambda\lambda}r(Fz+\hat{x};\theta)\preceq -mI$ by P3, we conclude that (ii) holds: $-\bar{M}I\preceq\nabla^2_{zz}g(z;\theta)\preceq-\bar{m}I$ where $\bar{M}=M\sigma_{\max}(F'F)$ and $\bar{m}=m\sigma_{\min}(F'F)$. Then, by Theorem EC.2, we have that there exists a constant $\eta=\min\{1,3(1-2\alpha)\}\bar{m}^2/L$ for some $\alpha\in(0,0.5)$ independent of θ such that if $||\nabla_z g(z^{(k)};\theta)||_2<\eta$, then $||\nabla_z g(z^{(k+1)};\theta)||_2<\frac{L}{2\bar{m}}||\nabla_z g(z^{(k)};\theta)||_2^2$. Note that by strong convexity of $g(.;\theta)$, $\bar{M}^{-1}||\nabla_z g(z;\theta)||_2\le||z-z^D(\theta)||_2\le 2\bar{m}^{-1}||\nabla_z g(z;\theta)||_2$. Also note that for $x = Fz + \hat{x}$, $||x - x^D(\theta)||_2 \leq ||F||_2 ||z - z^D(\theta)||_2$ and $||z - z^D(\theta)||_2 \leq ||(F'F)^{-1}F'||_2 ||x - x^D(\theta)||_2$. Therefore,

$$\begin{aligned}||x^{(k+1)} - x^D(\theta)||_2 &\leq ||F||_2 ||z^{(k+1)} - z^D(\theta)||_2 \leq 2\bar{m}^{-1}||F||_2 ||\nabla_z g(z^{(k+1)};\theta)||_2 \\ &\leq L\bar{m}^{-2}||F||_2 ||\nabla_z g(z^{(k)};\theta)||_2^2 \leq L\bar{m}^{-2}\bar{M}||F||_2 ||z^{(k)} - z^D(\theta)||_2^2 \\ &\leq L\bar{m}^{-2}\bar{M}||F||_2 ||(F'F)^{-1}F'||_2^2 ||x^{(k)} - x^D(\theta)||_2^2\end{aligned}$$

Let $\gamma = \eta$ and $\xi = L\bar{m}^{-2}\bar{M}||F||_2||(F'F)^{-1}F'||_2^2$. Note that they are both independent of θ. The result follows by letting $x^{(k+1)} = x_z^{NT}$ and $x^{(k)} = x_{z-1}^{NT}$. $\square$

Proof of Lemma EC.6. Let $\Omega_1 = \max_{i=1,..,4}\{V_i\}$, where V_i's are positive constants to be defined later. We prove the results one by one.

(a) Let $\bar{\kappa} = \max\{\kappa, \tilde{\kappa}\}$ where κ and $\tilde{\kappa}$ are defined in Lemma 1 and Lemma EC.1 (see §EC.2.2) respectively. Let $\Gamma_1 = \max\{1, 4\bar{\kappa}^2\}$. We proceed by induction. If $t_1 \geq \tau$, there is nothing to prove, so we consider the case when $t_1 < \tau$. Note that by DPUP algorithm, $x_1^{NT} = \lambda^D(\hat{\theta}_{t_1})$ and $x^D(\theta^*) = \lambda^D(\theta^*)$ on $\mathcal{A}$. Thus, when $t_1 < \tau$ we have

$$\begin{aligned}||x^D(\hat{\theta}_{t_1}) - x_1^{NT}||_2^2 &= ||x^D(\hat{\theta}_{t_1}) - \lambda^D(\hat{\theta}_{t_1})||_2^2 \\ &\leq \Big(||x^D(\hat{\theta}_{t_1}) - x^D(\theta^*)||_2 + ||\lambda^D(\theta^*) - \lambda^D(\hat{\theta}_{t_1})||_2\Big)^2 \\ &\leq 4\bar{\kappa}^2 E(t_1)^2 \ \leq \ \Gamma_1 (\log t_1)^{-\frac{\epsilon}{2}}\end{aligned}$$

where the last inequality follows by the definition of $\mathcal{A}$. This is our base case. We now do the inductive step. Suppose that $t_{z-1} < \tau$ and $||x^D(\hat{\theta}_{t_{z-1}}) - x_{z-1}^{NT}||_2^2 \leq \Gamma_1(\log t_{z-1})^{-\epsilon/2}$. If $t_z \geq \tau$ there is nothing to prove. If $t_z < \tau$, then we need to show that $||x^D(\hat{\theta}_{t_z}) - x_z^{NT}||_2^2 \leq \Gamma_1(\log t_z)^{-\epsilon/2}$ also holds. Let $V_1 > 0$ be the smallest integer satisfying $\lceil(\log V_1)^{1+\epsilon}\rceil > e^2$. Then, for $k \geq \Omega_1 \geq V_1$, we have

$$\begin{aligned}\left|\left|x^D(\hat{\theta}_{t_z}) - x_{z-1}^{NT}\right|\right|_2^2 &\leq 3\left|\left|x^D(\hat{\theta}_{t_z}) - x^D(\theta^*)\right|\right|_2^2 + 3\left|\left|x^D(\theta^*) - x^D(\hat{\theta}_{t_{z-1}})\right|\right|_2^2 + 3\left|\left|x^D(\hat{\theta}_{t_{z-1}}) - x_{z-1}^{NT}\right|\right|_2^2 \\ &\leq \frac{3\bar{\kappa}^2}{(\log t_z)^{\frac{\epsilon}{2}}} + \frac{3\bar{\kappa}^2}{(\log t_{z-1})^{\frac{\epsilon}{2}}} + \frac{3\Gamma_1}{(\log t_{z-1})^{\frac{\epsilon}{2}}} \\ &\leq \frac{3\bar{\kappa}^2}{(\log t_z)^{\frac{\epsilon}{2}}} + \frac{3\bar{\kappa}^2}{(\log \sqrt{t_z})^{\frac{\epsilon}{2}}} + \frac{3\Gamma_1}{(\log \sqrt{t_z})^{\frac{\epsilon}{2}}} \\ &\leq 3\left[\bar{\kappa}^2 + 2^{\frac{\epsilon}{2}}(\bar{\kappa}^2 + \Gamma_1)\right]\frac{1}{(\log t_z)^{\frac{\epsilon}{2}}},\end{aligned}$$

where the second inequality follows by definition of $\mathcal{A}$ and induction hypothesis, the third inequality follows because $t_{z-1} \geq \frac{t_z}{2} \geq \sqrt{t_z} \geq \sqrt{t_1} = \sqrt{\lceil(\log k)^{1+\epsilon}\rceil} > e$ when $k \geq \Omega_1 \geq V_1$. Let $V_2 \geq V_1$ be such that for all $k \geq V_2$ and $z = 1, \ldots, Z$, the following holds: (1) $(\log t_z)^{\epsilon/2} \geq 3\gamma^{-2}\left[\bar{\kappa}^2 + 2^{\epsilon/2}(\bar{\kappa}^2 + \Gamma_1)\right]$ and (2) $9\xi^2\left[\bar{\kappa}^2 + 2^{\epsilon/2}(\bar{\kappa}^2 + \Gamma_1)\right]^2(\log t_z)^{-\epsilon/2} \leq 1 \leq \Gamma_1$. (Recall that γ and ξ are the constants for the locally quadratic convergence of Newton's method defined in Lemma EC.5.) Inequality (1) ensures that $||x^D(\hat{\theta}_{t_z}) - x^{NT}_{z-1}||_2 \le \gamma$ for all $k \ge \Omega_1 \ge V_2$ and inequality (2) ensures, by the locally quadratic convergence of the Newton's method, that $||x^D(\hat{\theta}_{t_z}) - x^{NT}_z||_2^2 \le \xi^2 ||x^D(\hat{\theta}_{t_z}) - x^{NT}_{z-1}||_2^4 \le \Gamma_1 (\log t_z)^{-\epsilon/2}$. This completes the induction.

(b) First, we claim that there exist $0 < \tilde{\lambda}_{\min} < \tilde{\lambda}_{\max} < 1$ such that (1) $\tilde{\lambda}_{\min} \le \lambda_{\min}$ and $\tilde{\lambda}_{\max} \ge \lambda_{\max}$, and (2) if $p_t \in \mathsf{Ball}(p^D(\theta^*), 7\phi/8)$ for $t \in [t_1+1, \tau)$, then $p_t \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ for all $1 \le t < \tau$, which will be used to prove Lemma EC.6(c). If this is true, then Lemma 4 can be used to bound $E(t_z)$ as long as $t_z < \tau$. We now find such $\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max}$ below.

We first consider $p \in \mathsf{Ball}(p^D(\theta^*), 7\phi/8)$. Define $V_p := \mathsf{Ball}(p^D(\theta^*), 7\phi/8)$ (note that by our notation, V_p is a *closed* ball) and $V_\lambda(\theta) := \{x \in \Lambda_\theta : x \in \lambda(p;\theta), p \in V_p\}$. Also, define $O_p := \{p \in \mathcal{P} : ||p - p^D(\theta^*)||_2 < \phi\}$ (note that this is an *open* ball) and $O_\lambda(\theta) := \{x \in \Lambda_\theta : x \in \lambda(p;\theta), p \in O_p\}$. Note that $V_p \subseteq O_p \subseteq \mathcal{P}$ by R4. This implies that $V_\lambda(\theta) \subseteq O_\lambda(\theta) \subseteq \Lambda_\theta$. In addition, since $p(.;\theta)$ is continuous in λ by P3 and O_p is an open set, $O_\lambda(\theta)$ is an open set. Therefore, $O_\lambda(\theta)$ lies in the interior of Λ_θ, and hence, $V_\lambda(\theta) \subseteq O_\lambda(\theta)$ also lies in the interior of Λ_θ. This implies that for any $\theta \in \Theta$, $\lambda_{\min}(\theta) := \inf_{p \in V_p} \min_{1 \le j \le n} \lambda_j(p;\theta) > 0$ and $\lambda_{\max}(\theta) := \sup_{p \in V_p} \sum_{j=1}^n \lambda_j(p;\theta) < 1$. Since Θ is compact and $\lambda_{\min}(\theta)$ and $\lambda_{\max}(\theta)$ are continuous functions, there exists some $\theta', \theta'' \in \Theta$ such that $\sup_{\theta \in \Theta} \lambda_{\max}(\theta) = \lambda_{\max}(\theta') < 1$ and $\inf_{\theta \in \Theta} \lambda_{\min}(\theta) = \lambda_{\min}(\theta'') > 0$. Hence, for all $p \in \mathsf{Ball}(p^D(\theta^*), 7\phi/8) = V_p$ and for all θ, $1 - \sum_{j=1}^n \lambda_j(p;\theta) \ge 1 - \sup_{\theta \in \Theta} \sup_{p \in V_p} \sum_{j=1}^n \lambda_j(p;\theta) = 1 - \sup_{\theta \in \Theta} \lambda_{\max}(\theta) = 1 - \lambda_{\max}(\theta') > 0$ and $\lambda_j(p;\theta) \ge \inf_{\theta \in \Theta} \inf_{p \in V_p} \min_{1 \le j \le n} \lambda_j(p;\theta) \ge \inf_{\theta \in \Theta} \lambda_{\min}(\theta) \ge \lambda_{\min}(\theta'') > 0$ for all $1 \le j \le n$. Set $\tilde{\lambda}_{\max} = \max\{\lambda_{\max}, \lambda_{\max}(\theta')\}, \tilde{\lambda}_{\min} = \min\{\lambda_{\min}, \lambda_{\min}(\theta'')\}$ where $\lambda_{\max}$ and $\lambda_{\min}$ are as defined in S2. Note that by S2, for $p \in \{\tilde{p}^{(1)}, \ldots, \tilde{p}^{(\tilde{q})}\}$, $1 - \sum_{j=1}^n \lambda_j(p;\theta) \ge 1 - \lambda_{\max} \ge 1 - \tilde{\lambda}_{\max}$ and $\lambda_j(p;\theta) \ge \lambda_{\min} \ge \tilde{\lambda}_{\min}$ for all $1 \le j \le n$ and for all $\theta \in \Theta$. This completes the proof of the claim: if $t \le t_1$, then $p_t \in \{\tilde{p}^{(1)}, \ldots, \tilde{p}^{(n)}\} \subseteq \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$; if $t_1 < t < \tau$, then $p_t \in \mathsf{Ball}(p^D(\theta^*), 7\phi/8) \subseteq \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$.

Note that $\hat{\lambda}_t \in \Lambda_{\hat{\theta}_t}$ is equivalent to $p_t \in \mathcal{P}$ which is immediately satisfied if $p_t \in \mathsf{Ball}(p^D(\theta^*), 7\phi/8) \subseteq \mathsf{Ball}(p^D(\theta^*), \phi) \subseteq \mathcal{P}$ (the last inequality follows by R4). This means that we only need to show $C_t \succ 0$ and $p_t \in \mathsf{Ball}(p^D(\theta^*), 7\phi/8)$ for $t_1 \le t < \tau$. Let $V_3 \ge V_2$ be such that for all $k \ge V_3$ and $z = 1, \ldots, Z$, $\left(2\sqrt{\Gamma_1} + 3\kappa\right)(\log t_z)^{-\epsilon/4} < \phi/(8\omega)$. We now prove the result by induction. If $\tau \le t_1 + 1$, then there is nothing to prove. Suppose that $\tau > t_1 + 1$. Since $E(t_1) \le \bar{\delta}$ on $\mathcal{A}$, by Lemma 1(a), $p^D(\hat{\theta}_1) \in \mathsf{Ball}(p^D(\theta^*), \phi/2)$. For $t = t_1 + 1$, we then have $||p_{t_1+1} - p^D(\theta^*)||_2 = ||p^D(\hat{\theta}_1) - p^D(\theta^*)||_2 \le \phi/2$, so $p_{t_1+1} \in \mathcal{P}$. In addition, similar to Lemma EC.1, we also have $C_{t_1+1} = kC - LC + LC - AS_L - A(\lambda_1^{NT} + \hat{\Delta}_{t_1+1}) \succeq (k - L - 1)C + LC - AS_L - A\hat{\Delta}_{t_1+1} \succ 0$ where the first inequality follows by the fact that $A\lambda_1^{NT} \preceq C$, and the second inequality follows by the same argument as in Lemma EC.1. This is the base case. Now suppose $C_s \succ 0, p_s \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ for all $s \le t - 1$ for some $t - 1 < \tau$ with $t - 1 \in [t_z, t_{z+1})$. If $t \ge \tau$, there is nothing to prove. So we only need to show that $C_t \succ 0, p_t \in \mathcal{W}(\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max})$ when $t < \tau$. Note that when $t < \tau$, we have $t_z \le t < \tau$. Hence, by definition of $\mathcal{A}$, we have

$$\begin{aligned}
||p_t - p^D(\theta^*)||_2 &\le ||p_t - p(\lambda_z^{NT}; \hat{\theta}_{t_z})||_2 + ||p(\lambda_z^{NT}; \hat{\theta}_{t_z}) - p^D(\hat{\theta}_{t_z})||_2 + ||p^D(\hat{\theta}_{t_z}) - p^D(\theta^*)||_2 \\
&\le w||Q||_2 \left|\left| \sum_{s=t_1+1}^{t-1} \frac{\hat{\Delta}_s}{k-s} \right|\right|_2 + ||p(\lambda_z^{NT}; \hat{\theta}_{t_z}) - p(\lambda^D(\hat{\theta}_{t_z}); \hat{\theta}_{t_z})||_2 + \frac{\phi}{2} \\
&\le \frac{\phi}{4} + \omega||\lambda_z^{NT} - \lambda^D(\hat{\theta}_{t_z})||_2 + \frac{\phi}{2} \le \frac{\phi}{4} + \frac{\phi}{8} + \frac{\phi}{2} = \frac{7\phi}{8}
\end{aligned}$$

where the second inequality follows by Lemma 1(a) and the fact that $E(t_z) < \bar{\delta}$ on $\mathcal{A}$, the last inequality results from the following inequality

$$\begin{aligned}
\left|\left|\lambda_z^{NT} - \lambda^D(\hat{\theta}_{t_z})\right|\right|_2 &\le \left|\left|\lambda_z^{NT} - \lambda^D(\theta^*)\right|\right|_2 + \left|\left|\lambda^D(\theta^*) - \lambda^D(\hat{\theta}_{t_z})\right|\right|_2 \\
&\le 2\left|\left|x_z^{NT} - x^D(\hat{\theta}_{t_z})\right|\right|_2 + 2\left|\left|x^D(\hat{\theta}_{t_z}) - x^D(\theta^*)\right|\right|_2 + \left|\left|\lambda^D(\theta^*) - \lambda^D(\hat{\theta}_{t_z})\right|\right|_2 \\
&\le 2\sqrt{\Gamma_1}(\log t_z)^{-\frac{\epsilon}{4}} + 3\bar{\kappa}E(t_z) \\
&\le \left(2\sqrt{\Gamma_1} + 3\bar{\kappa}\right)(\log t_z)^{-\frac{\epsilon}{4}} \; < \; \frac{\phi}{8\omega},
\end{aligned}$$

where the second inequality follows by (EC.13) and the fourth inequality follows by the definition of $\mathcal{A}$. Hence, $p_t \in \mathsf{Ball}(p^D(\theta^*), 7\phi/8)$. For C_t, by a similar argument to Lemma EC.1, we have $C_t = kC - tC + tC - AS_L - \sum_{s=t_1+1}^{t} A(\lambda_{z(s)}^{NT} - Q\sum_{v=t_1+1}^{s-1} \frac{\hat{\Delta}_v}{k-v} + \hat{\Delta}_s) \succeq (k-t)C + LC - AS_L - \sum_{s=t_1+1}^{t}(A\hat{\Delta}_s - \sum_{v=t_1+1}^{s-1} \frac{A\hat{\Delta}_v}{k-v}) \succ 0$. This completes the induction.

(c) Let $V_4 \ge V_3$ be such that $27\xi^2\left(5\bar{\kappa}^4\left[8\eta_4 + 4(q-1)^2(\log t_z)^2\right]/(\eta_5^2 t_z) + 2\Gamma_1\Gamma_2/(\log t_{z-1})^{\frac{\epsilon}{2}}\right) < 1$ for all $k \ge V_4$ and $z = 1, \ldots, Z$, where $\Gamma_2 = \max\{1, 4\bar{\kappa}^2\eta_3^2\}$, η_4 and η_5 are as in Lemma 4. Again, we show by induction. For $z = 1$, we have:

$$\begin{aligned}
&\mathbb{E}^\pi[||x^D(\hat{\theta}_{t_1}) - x_1^{NT}||_2^2 \mathbf{1}_{\{t_1<\tau\}}|\mathcal{A}] = \mathbb{E}^\pi[||x^D(\hat{\theta}_{t_1}) - \lambda^D(\hat{\theta}_{t_1})||_2^2 \mathbf{1}_{\{t_1<\tau\}}|\mathcal{A}] \\
&\le 2\,\mathbb{E}^\pi[||x^D(\hat{\theta}_{t_1}) - x^D(\theta^*)||_2^2 \mathbf{1}_{\{t_1<\tau\}}|\mathcal{A}] + 2\,\mathbb{E}^\pi[||\lambda^D(\hat{\theta}_{t_1}) - \lambda^D(\theta^*)||_2^2 \mathbf{1}_{\{t_1<\tau\}}|\mathcal{A}] \\
&\le 4\bar{\kappa}^2 \frac{\eta_3^2}{t_1} \; \le \; \frac{\Gamma_2}{t_1},
\end{aligned}$$

where the second to the last inequality follows by Lemma 2. This is our base case. We now do the inductive step. Suppose that $\mathbb{E}^\pi[||x^D(\hat{\theta}_{t_s}) - x_s^{NT}||_2^2 \mathbf{1}_{\{t_s<\tau\}}|\mathcal{A}] \le \Gamma_2 t_s^{-1}$ holds for $s = z-1$, we need to show that same thing holds for $s = z$. Then, for $k \ge \Omega_1 \ge V_4$, we have:

$$\begin{aligned}
&\mathbb{E}^\pi\left[\left|\left|x^D(\hat{\theta}_{t_z}) - x_z^{NT}\right|\right|_2^2 \mathbf{1}_{\{t_z<\tau\}}\middle|\mathcal{A}\right] \le \xi^2\mathbb{E}^\pi\left[\left|\left|x^D(\hat{\theta}_{t_z}) - x_{z-1}^{NT}\right|\right|_2^4 \mathbf{1}_{\{t_z<\tau\}}\middle|\mathcal{A}\right] \\
&\le 27\xi^2\left\{\mathbb{E}^\pi\left[\left|\left|x^D(\hat{\theta}_{t_z}) - x^D(\theta^*)\right|\right|_2^4 \mathbf{1}_{\{t_z<\tau\}}\middle|\mathcal{A}\right] + \mathbb{E}^\pi\left[\left|\left|x^D(\theta^*) - x^D(\hat{\theta}_{t_{z-1}})\right|\right|_2^4 \mathbf{1}_{\{t_z<\tau\}}\middle|\mathcal{A}\right]\right. \\
&\qquad \left. + \mathbb{E}^\pi\left[\left|\left|x^D(\hat{\theta}_{t_{z-1}}) - x_{z-1}^{NT}\right|\right|_2^4 \mathbf{1}_{\{t_z<\tau\}}\middle|\mathcal{A}\right]\right\}
\end{aligned}$$

$$
\begin{aligned}
&\le 27\xi^2\left\{\bar{\kappa}^4\mathbb{E}^\pi[E(t_z)^4\mathbf{1}_{\{t_z<\tau\}}|\mathcal{A}]+\bar{\kappa}^4\mathbb{E}^\pi_{\theta^*}[E(t_{z-1})^4\mathbf{1}_{\{t_z<\tau\}}|\mathcal{A}]+\frac{\Gamma_1}{(\log t_{z-1})^{\frac{\epsilon}{2}}}\frac{\Gamma_2}{t_{z-1}}\right\}\\
&\le 27\xi^2\left\{\frac{8\eta_4+4(q-1)^2(\log t_z)^2}{\eta_5^2t_z^2}\bar{\kappa}^4+\frac{8\eta_4+4(q-1)^2(\log t_{z-1})^2}{\eta_5^2t_{z-1}^2}\bar{\kappa}^4+\frac{\Gamma_1}{(\log t_{z-1})^{\frac{\epsilon}{2}}}\frac{2\Gamma_2}{t_z}\right\}\\
&\le 27\xi^2\left\{\frac{5\bar{\kappa}^4\left[8\eta_4+4(q-1)^2(\log t_z)^2\right]}{\eta_5^2t_z}+\frac{2\Gamma_1\Gamma_2}{(\log t_{z-1})^{\frac{\epsilon}{2}}}\right\}\frac{1}{t_z}\\
&\le \frac{1}{t_z}\le\frac{\Gamma_2}{t_z},
\end{aligned}
$$

where the first inequality follows by Lemma EC.6(a), the third inequality follows by Lemma EC.4, Lemma EC.6(a) and the induction hypothesis, and the fourth inequality holds because Lemma EC.6(b) shows that $p_s\in\mathcal{W}(\tilde{\lambda}_{\min},\tilde{\lambda}_{\max})$ for $s<\tau$ which means that the condition for Lemma 4 is satisfied, so

$$
\begin{aligned}
\mathbb{E}^\pi\left[E(t)^4\mathbf{1}_{\{t<\tau\}}\middle|\mathcal{A}\right]&\le\int_0^\infty\mathbb{P}^\pi\left(||\hat{\theta}_t-\theta^*||_2^4\ge x\right)dx\\
&\le\int_0^\infty\min\left\{1,\eta_4t^{q-1}\exp\left(-\eta_5t\sqrt{x}\right)\right\}dx\\
&\le\int_0^{\left(\frac{2(q-1)\log t}{\eta_5t}\right)^2}dx+\int_{\left(\frac{2(q-1)\log t}{\eta_5t}\right)^2}^\infty\left[\eta_4t^{q-1}\exp\left(-\frac{\eta_5t\sqrt{x}}{2}\right)\right]\exp\left(-\frac{\eta_5t\sqrt{x}}{2}\right)dx\\
&\le\frac{4(q-1)^2(\log t)^2}{\eta_5^2t^2}+\eta_4\int_{\left(\frac{2(q-1)\log t}{\eta_5t}\right)^2}^\infty\exp\left(-\frac{\eta_5t\sqrt{x}}{2}\right)dx\\
&\le\frac{4(q-1)^2(\log t)^2}{\eta_5^2t^2}+\eta_4\int_0^\infty\exp\left(-\frac{\eta_5t\sqrt{x}}{2}\right)dx\\
&\le\frac{8\eta_4+4(q-1)^2(\log t)^2}{\eta_5^2t^2}.
\end{aligned}
$$

This completes the induction. □

Proof of Lemma EC.7. We first derive a bound for Φ_ϵ. By definition $t_z=\lceil(t_{z+1}-L)/2\rceil+L$ for $z>1$, so $t_z-L\ge(t_{z+1}-L)/2$. This implies that $t_{z+1}-t_z\le t_z$ for all $z>1$. For $z=1$, we also have $t_2-t_1=1\le L=t_1$. Recall that $Z\le\lceil\log_2k\rceil\le 2\log_2k$. Thus, we can bound Φ_ϵ as follows:

$$
\begin{aligned}
\Phi_\epsilon=\sum_{s=t_1+1}^{k-1}\bar{\epsilon}(s)^2=\sum_{z=1}^{Z}(t_{z+1}-t_z)\bar{\epsilon}(t_z)^2&\le\sum_{z=1}^{Z}(t_{z+1}-t_z)\,\eta_6^2\,\frac{(q-1)\log t_z+1}{t_z}\\
&\le\eta_6^2Z[(q-1)\log k+1]\\
&\le K_\Phi[1+\log k+(q-1)\log^2k]
\end{aligned}
$$

for some positive constant K_Φ independent of $k\ge1$.

We now derive a bound for Ψ_ϵ. To do that, we first show that there exists a constant $K>3$ such that for all $k\ge K$, we have (1)$(\log k)^{1+\epsilon}/k<1/19$, (2)$Z\ge3$ and (3)$t_{Z-2}\le k/3$. Note that as

$k \to \infty$, we have $(\log k)^{1+\epsilon}/k \to 0$, $Z \to \infty$ and $t_{z+1} - L \to \infty$ for $z = Z-2, Z-1, Z$. This implies that $t_z - L = \lceil (t_{z+1} - L)/2 \rceil \leq 2(t_{z+1} - L)/3$ for $z = Z-2, Z-1, Z$ when k is large. Therefore, there exists a constant $K > 3$ such that for all $k \geq K$, we have $(\log k)^{1+\epsilon}/k < 1/19$, $Z \geq 3$ and $t_{Z-2} \leq \frac{8}{27}(t_{Z+1} - L) + L = \frac{8}{27}k + \frac{19}{27}(\log k)^{1+\epsilon} < \frac{k}{3}$.

Since $\bar{\epsilon}(t_z) = \eta_6 \sqrt{[(q-1)\log t_z + 1]/t_z} \leq \eta_6 \sqrt{q}$, we conclude that for $k < K$, $\Psi_\epsilon \leq k(k\eta_6\sqrt{q})^2 \leq K^3 \eta_6^2 q$. We now focus on the case when $k \geq K$. Note that,

$$\Psi_\epsilon = \sum_{t=t_1+1}^{k-1} \left(\sum_{s=t_1+1}^{t-1} \frac{\bar{\epsilon}(s)}{k-s} \right)^2 \leq 2 \sum_{t=t_1+1}^{k-1} \left(\sum_{s=t_1+1}^{t_{Z-2}} \frac{\bar{\epsilon}(s)}{k-s} \right)^2 + 2 \sum_{t=t_{Z-2}+1}^{k-1} \left(\sum_{s=t_{Z-2}+1}^{t-1} \frac{\bar{\epsilon}(s)}{k-s} \right)^2. \tag{EC.14}$$

Since $t_{Z-2} > k/4$ (recall that $t_{z+1} \leq 2t_z$ and $t_{Z+1} = k$), we have $\bar{\epsilon}(s) < \eta_6 \sqrt{4[(q-1)\log k + 1]/k}$ for all $s > t_{Z-2}$. So, for all $k \geq K$, the second term in (EC.14) can be bounded by

$$\frac{8\eta_6^2[1 + (q-1)\log k]}{k} \sum_{t=t_{Z-2}+1}^{k-1} \left(\sum_{s=t_{Z-2}+1}^{t-1} \frac{1}{k-s} \right)^2 \leq \frac{8\eta_6^2[1 + (q-1)\log k]}{k} 3k \leq K_{\Psi,2}[1 + (q-1)\log k]$$

for some positive constant $K_{\Psi,2} = 24\eta_6^2$ independent of $k \geq K$, where the first inequality follows by a similar argument as in (EC.7) and $k \geq K > 3$. As for the first term in (EC.14), for all $k \geq K$, we have

$$\begin{aligned}
2 \sum_{t=t_1+1}^{k-1} \left(\sum_{s=t_1+1}^{t_{Z-2}} \frac{\bar{\epsilon}(s)}{k-s} \right)^2 &\leq 2k \left(\sum_{s=t_1+1}^{t_{Z-2}} \frac{\bar{\epsilon}(s)}{k-s} \right)^2 \\
&\leq 2k \left(\sum_{z=1}^{Z-3} \frac{t_{z+1} - t_z}{k - t_{z+1}} \eta_6 \sqrt{\frac{1 + (q-1)\log t_z}{t_z}} \right)^2 \\
&\leq 4k\eta_6^2 \left(\sum_{z=1}^{Z-3} \frac{t_{z+1} - t_z}{k - t_{z+1}} \sqrt{\frac{1 + (q-1)\log k}{t_{z+1}}} \right)^2 \\
&\leq 4k\eta_6^2 [1 + (q-1)\log k] \left(\int_1^{t_{Z-2}} \frac{1}{k-x} \sqrt{\frac{1}{x}} dx \right)^2 \\
&\leq 4k\eta_6^2 [1 + (q-1)\log k] \left(\frac{2\log(\frac{\sqrt{2}}{\sqrt{2}-1})}{\sqrt{k}} \right)^2 \leq K_{\Psi,1}[1 + (q-1)\log k]
\end{aligned}$$

where $K_{\Psi,1} = 16\eta_6^2 \log^2(\frac{\sqrt{2}}{\sqrt{2}-1})$. The second inequality follows by Lemma 4. The third inequality follows because $t_{z+1} \leq 2t_z$. Note that the function $f(x) = \frac{1}{(k-x)\sqrt{x}}$ is decreasing when $x < \frac{k}{3}$. Since $t_{Z-2} < \frac{k}{3}$, the fourth inequality holds by integral approximation. The fifth inequality follows by

$$\int_1^{t_{Z-2}} \frac{1}{k-x} \sqrt{\frac{1}{x}} dx = \frac{1}{\sqrt{k}} \int_1^{t_{Z-2}} \left(\frac{1}{\sqrt{k} - \sqrt{x}} + \frac{1}{\sqrt{k} + \sqrt{x}} \right) d\sqrt{x} \leq \frac{2}{\sqrt{k}} \log \left(\frac{\sqrt{k} - 1}{\sqrt{k} - \sqrt{t_{Z-2}}} \right) \leq \frac{2\log(\frac{\sqrt{2}}{\sqrt{2}-1})}{\sqrt{k}}.$$

Thus, we conclude that there exists some positive constant K_Ψ independent of $k \ge 1$ such that $\Psi_\epsilon \le \max\{(K_{\Psi,1}+K_{\Psi,2})[1+(q-1)\log k], K^3\eta_6^2 q\} \le K_\Psi[1+(q-1)\log k]$. We complete the proof by letting $K_3 = \max\{K_\Phi, K_\Psi\}$. $\square$

Proof of Lemma EC.8. The proof of Lemma EC.8 is very similar to that of Lemma EC.2, with some nontrivial twists. Per the proof of Lemma EC.2, we only need to bound $\mathbb{P}^\pi(\tau \le t|\mathcal{A})$. Note that we have

$$\begin{aligned}
\mathbb{P}^\pi(\tau \le t|\mathcal{A}) &\le \mathbb{P}^\pi\left(\max_{L+1\le s\le t}\left\{\left|\left|\frac{S_L - L\lambda_L\mathbf{e}}{k-s}\right|\right|_2 + \left|\left|\sum_{v=L+1}^{s}\frac{\Delta_v}{k-v}\right|\right|_2 + \sum_{v=L+1}^{s}\frac{||\lambda_v-\hat{\lambda}_v||_2\mathbf{1}_{\{v\le\tau\}}}{k-v}\right\} \ge \psi \middle| \mathcal{A}\right)\\
&\le \mathbb{P}^\pi\left(\max_{L+1\le s\le t}\left|\left|\frac{S_L - L\lambda_L\mathbf{e}}{k-s}\right|\right|_2 \ge \frac{\psi}{2}\middle|\mathcal{A}\right) + \mathbb{P}^\pi\left(\max_{L+1\le s\le t}\left|\left|\sum_{v=L+1}^{s}\frac{\Delta_v}{k-v}\right|\right|_2 \ge \frac{\psi}{4}\middle|\mathcal{A}\right)\\
&\quad + \mathbb{P}^\pi\left(\max_{L+1\le s\le t}\sum_{v=L+1}^{s}\frac{||\lambda_v-\hat{\lambda}_v||_2\mathbf{1}_{\{v\le\tau\}}}{k-v} \ge \frac{\psi}{4}\middle|\mathcal{A}\right)\\
&\le \max\left\{1, \frac{4n(1+\lambda_L)^2L^2}{\psi^2(k-t)^2}\right\} + \frac{16}{\psi^2\mathbb{P}^\pi(\mathcal{A})}\left[\frac{4}{(k-t)^2}+\frac{4}{k-t}\right]\\
&\quad + \mathbb{P}^\pi\left(\max_{L+1\le s\le t}\sum_{v=L+1}^{s}\frac{||\lambda_v-\hat{\lambda}_v||_2\mathbf{1}_{\{v\le\tau\}}}{k-v} \ge \frac{\psi}{4}\middle|\mathcal{A}\right) \qquad \text{(EC.15)}
\end{aligned}$$

where the last inequality follows by the same argument in Lemma EC.2. We now bound the last term in (EC.15):

$$\begin{aligned}
&\mathbb{P}^\pi\left(\max_{L+1\le s\le t}\sum_{v=L+1}^{s}\frac{||\lambda_v-\hat{\lambda}_v||_2\mathbf{1}_{\{v\le\tau\}}}{k-v} \ge \frac{\psi}{4}\middle|\mathcal{A}\right) \le \frac{16}{\psi^2}\left(\sum_{s=L+1}^{t}\frac{\sqrt{\mathbb{E}^\pi[||\lambda_s-\hat{\lambda}_s||_2^2\mathbf{1}_{\{\tau\le s\}}|\mathcal{A}]}}{k-s}\right)^2\\
&\le \frac{16}{\psi^2}\left(\sum_{s=L+1}^{t}\frac{\sqrt{\mathbb{E}^\pi[||\lambda_s-\hat{\lambda}_s||_2^2\mathbf{1}_{\{\tau< s\}}|\mathcal{A}]}}{k-s} + \frac{\sqrt{\mathbb{E}^\pi[||\lambda_s-\hat{\lambda}_s||_2^2\mathbf{1}_{\{\tau= s\}}|\mathcal{A}]}}{k-s}\right)^2\\
&\le \frac{32}{\psi^2}\left(\sum_{s=L+1}^{t}\frac{\sqrt{\mathbb{E}^\pi[||\lambda_s-\hat{\lambda}_s||_2^2\mathbf{1}_{\{\tau< s\}}|\mathcal{A}]}}{k-s}\right)^2 + \frac{32}{\psi^2}\left(\sum_{s=L+1}^{t}\frac{\sqrt{\mathbb{E}^\pi[||\lambda_s-\hat{\lambda}_s||_2^2\mathbf{1}_{\{\tau= s\}}|\mathcal{A}]}}{k-s}\right)^2\\
&\le \frac{32\omega^2}{\psi^2}\left(\sum_{s=L+1}^{t}\frac{\bar{\epsilon}(s)}{k-s}\right)^2 + \frac{32}{\psi^2}\left(\sum_{s=L+1}^{t}\frac{\sqrt{2}\sqrt{\mathbb{E}^\pi[\mathbf{1}_{\{\tau=s\}}|\mathcal{A}]}}{k-s}\right)^2\\
&\le \frac{32\omega^2}{\psi^2}\left(\sum_{s=L+1}^{t}\frac{\bar{\epsilon}(s)}{k-s}\right)^2 + \frac{128}{\psi^2}\left(\frac{1}{k-t}\right)
\end{aligned}$$

where the first inequality follows the same argument as in the proof of Lemma EC.2, the fourth inequality follows by Lemma 4 and the fact that for any two points $x_1, x_2 \in \Delta^{n-1}$ we have $||x_1 - x_2||_2^2 \le 2$, and the last inequality follows because by Cauchy-Schwartz inequality,

$$\left(\sum_{s=L+1}^{t}\frac{\sqrt{\mathbb{E}^\pi[\mathbf{1}_{\{\tau=s\}}|\mathcal{A}]}}{k-s}\right)^2 \le \left(\sum_{s=L+1}^{t}\frac{1}{(k-s)^2}\right)\left(\sum_{s=L+1}^{t}\mathbb{E}^\pi[\mathbf{1}_{\{\tau=s\}}|\mathcal{A}]\right) \le \frac{1}{(k-t)^2}+\frac{1}{k-t} \le \frac{2}{k-t}$$

Finally, we have for all $k \geq K \geq \Omega_2 \geq 3$,

$$\begin{aligned}
\mathbb{E}^\pi[k-\tau|\mathcal{A}] &= \sum_{t=1}^{k-1} \mathbb{P}^\pi(\tau \leq t|\mathcal{A}) \leq \frac{256}{\psi^2} \frac{\log k}{1-\mathbb{P}^\pi(E(L) > \bar{\delta})} + \left(\frac{4n(1+\lambda_L)^2}{\psi^2} + 1\right) L \\
&\quad + \frac{32\omega^2}{\psi^2} \sum_{t=1}^{k-1} \left(\sum_{s=L+1}^{t} \frac{\bar{\epsilon}(s)}{k-s}\right)^2 + \frac{128}{\psi^2} \sum_{t=1}^{k-1} \left(\frac{1}{k-t}\right) \\
&\leq \frac{256}{\psi^2} \frac{\log k}{1-\mathbb{P}^\pi(E(L) > \bar{\delta})} + \left(\frac{4n(1+\lambda_L)^2}{\psi^2} + 1\right) L \\
&\quad + \frac{64\omega^2}{\psi^2} \sum_{t=1}^{k-1} \left(\sum_{s=L+1}^{t-1} \frac{\bar{\epsilon}(s)}{k-s}\right)^2 + \frac{64\omega^2}{\psi^2} \sum_{t=1}^{k-1} \frac{\bar{\epsilon}(t)^2}{(k-t)^2} + \frac{128}{\psi^2} \sum_{t=1}^{k-1} \left(\frac{1}{k-t}\right) \\
&\leq \frac{512}{\psi^2} \log k + \left(\frac{4n(1+\lambda_L)^2}{\psi^2} + 1\right) L + \frac{64K_3\omega^2 q}{\psi^2} \log k + \frac{128\omega^2\eta_6^2 q}{\psi^2} + \frac{128}{\psi^2} \leq K_7(\log k + L)
\end{aligned}$$

where $K_7 = 640/\psi^2 + 64K_3\omega^2 q/\psi^2 + 128\omega^2\eta_6^2 q/\psi^2 + (4n(1+\lambda_L)^2/\psi^2 + 1)$, the first inequality follows by a similar argument as in Lemma EC.2, and the third inequality follows by Lemma EC.7 and the fact that $\bar{\epsilon}(t) \leq \eta_6\sqrt{q}$. □

Proof of Lemma EC.9. Note that, for any $\theta_1, \theta_2 \in \Theta, \theta_1 \neq \theta_2$, by Fatou's lemma, we have

$$\begin{aligned}
\liminf_{\theta' \to \theta_1, \theta'' \to \theta_2} \frac{H_t^\pi(\theta', \theta'', D_t|D_{1:t-1})}{||\theta' - \theta''||_2^2} &= \liminf_{\theta' \to \theta_1, \theta'' \to \theta_2} \sum_{D_t \in \mathcal{D}} \frac{\left(\sqrt{\mathbb{P}_t^{\pi,\theta'}(D_t|D_{1:t-1})} - \sqrt{\mathbb{P}_t^{\pi,\theta''}(D_t|D_{1:t-1})}\right)^2}{||\theta' - \theta''||_2^2} \\
&\geq \sum_{D_t \in \mathcal{D}} \liminf_{\theta' \to \theta_1, \theta'' \to \theta_2} \frac{\left(\sqrt{\mathbb{P}_t^{\pi,\theta'}(D_t|D_{1:t-1})} - \sqrt{\mathbb{P}_t^{\pi,\theta''}(D_t|D_{1:t-1})}\right)^2}{||\theta' - \theta''||_2^2} \\
&= \frac{H_t^\pi(\theta_1, \theta_2, D_t|D_{1:t-1})}{||\theta_1 - \theta_2||_2^2} \ > \ 0, \qquad \text{(EC.16)}
\end{aligned}$$

where the last inequality follows by W1. Let $\underline{\sigma}(.)$ denote the smallest eigenvalues of a real symmetric matrix. If we now set $\theta_1 = \theta_2 = \theta$, since $\sqrt{\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})}$ is continuously differentiable in θ, there exists $\tilde{\theta}$ on the line segment connecting θ' and θ'' such that

$$\begin{aligned}
&\liminf_{\theta' \to \theta, \theta'' \to \theta} \frac{H_t^\pi(\theta', \theta'', D_t|D_{1:t-1})}{||\theta' - \theta''||_2^2} \\
&\quad \geq \sum_{D_t \in \mathcal{D}} \liminf_{\theta' \to \theta, \theta'' \to \theta} \left[\left(\frac{\partial}{\partial\theta}\sqrt{\mathbb{P}_t^{\pi,\tilde{\theta}}(D_t|D_{1:t-1})}\right)' \frac{\theta' - \theta''}{||\theta' - \theta''||_2}\right]^2 \\
&\quad = \sum_{D_t \in \mathcal{D}} \liminf_{\theta' \to \theta, \theta'' \to \theta} \frac{(\theta' - \theta'')'}{||\theta' - \theta''||_2} \left(\frac{\partial}{\partial\theta}\sqrt{\mathbb{P}_t^{\pi,\tilde{\theta}}(D_t|D_{1:t-1})}\right) \left(\frac{\partial}{\partial\theta}\sqrt{\mathbb{P}_t^{\pi,\tilde{\theta}}(D_t|D_{1:t-1})}\right)' \frac{\theta' - \theta''}{||\theta' - \theta''||_2} \\
&\quad \geq \sum_{D_t \in \mathcal{D}} \liminf_{\theta' \to \theta, \theta'' \to \theta} \underline{\sigma}\left(\left(\frac{\partial}{\partial\theta}\sqrt{\mathbb{P}_t^{\pi,\tilde{\theta}}(D_t|D_{1:t-1})}\right) \left(\frac{\partial}{\partial\theta}\sqrt{\mathbb{P}_t^{\pi,\tilde{\theta}}(D_t|D_{1:t-1})}\right)'\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{D_t \in \mathcal{D}} \underline{\sigma}\left(\left(\frac{\partial}{\partial\theta}\sqrt{\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})}\right)\left(\frac{\partial}{\partial\theta}\sqrt{\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})}\right)'\right)\\
&= \sum_{D_t \in \mathcal{D}} \frac{\underline{\sigma}\left((\frac{\partial}{\partial\theta}\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1}))(\frac{\partial}{\partial\theta}\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1}))'\right)}{4\,\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})}\\
&= \frac{1}{4}\sum_{D_t \in \mathcal{D}} \underline{\sigma}\left(\left(\frac{\partial}{\partial\theta}\log \mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})\right)\left(\frac{\partial}{\partial\theta}\log \mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})\right)'\right)\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})\\
&\geq \frac{c_f}{4} \; > \; 0 \qquad\qquad \text{(EC.17)}
\end{aligned}
$$

where the first inequality follows by Fatou's Lemma as in (EC.16) and the Mean Value Theorem, and the third equality follows because

$$
\frac{\partial}{\partial\theta}\sqrt{\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})} = \frac{\frac{\partial}{\partial\theta}\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})}{2\sqrt{\mathbb{P}_t^{\pi,\theta}(D_t|D_{1:t-1})}}
$$

(by chain rule) and the last two inequalities follow by the definition of Fisher information and W2. To prove Lemma EC.9, it suffices to show that, for any $\theta_1, \theta_2 \in \Theta, H_t^\pi(\theta_1, \theta_2, D_t|D_{1:t-1})/||\theta_1 - \theta_2||_2^2 \geq c_h$ for some $c_h > 0$ independent of θ_1, θ_2. (If $\theta_1 = \theta_2$, the ratio is to be understood as its limit.) Suppose not, since the ratio is always non-negative, there exists two sequences $\theta_1^n \to \theta_1, \theta_2^n \to \theta_2$ such that $\liminf_{n\to\infty} H_t^\pi(\theta_1^n, \theta_2^n, D_t|D_{1:t-1})/||\theta_1^n - \theta_2^n||_2^2 = 0$. But, this contradicts with (EC.16) when $\theta_1 \neq \theta_2$ and with (EC.17) when $\theta_1 = \theta_2$. This completes the proof. □

EC.3. Proof of results in Section 4

EC.3.1. Proof of Lemma 5

We first show that $\mathcal{L}$ is a bounded linear operator. For all $f \in \mathsf{C}^0(\mathcal{P})$, there exists $\underline{p}_l \leq x_l^* \leq \bar{p}_l$ for all $l = 1, \ldots, n$ such that

$$
\begin{aligned}
||\mathcal{L}f(.)||_\infty &= \sup_{x\in\mathcal{P}} |\mathcal{L}f(x)| = \sup_{x_1\in[\underline{p}_1,\bar{p}_1]} \cdots \sup_{x_n\in[\underline{p}_n,\bar{p}_n]} |\mathcal{L}_1\circ\cdots\circ\mathcal{L}_n f(x_1,\ldots,x_n)|\\
&\leq \sup_{x_1\in[\underline{p}_1,\bar{p}_1]} \cdots \sup_{x_{n-1}\in[\underline{p}_{n-1},\bar{p}_{n-1}]} |\mathcal{L}_1\circ\cdots\circ\mathcal{L}_{n-1} f(x_1,\ldots,x_n^*)|(2s)^s\\
&\leq \ldots \leq (2s)^{ns} f(x_1^*,\ldots,x_n^*) \leq (2s)^{ns}||f(.)||_\infty,
\end{aligned}
$$

where the inequalities follow by Theorem EC.3. We now prove that $\mathcal{L}f = f$ for all $f \in \otimes_{l=1}^n \mathsf{P}^{s-1}[\underline{p}_l, \bar{p}_l]$. Note that $\mathcal{L} = \mathcal{L}_1 \circ \mathcal{L}_2 \circ \cdots \circ \mathcal{L}_n$. Applying Theorem EC.3 iteratively n times, we obtain that $\mathcal{L}f = \mathcal{L}_1 \circ\cdots\circ \mathcal{L}_n f = \mathcal{L}_1\circ\cdots\circ\mathcal{L}_{n-1} f = \cdots = f$. □

EC.3.2. Proof of Lemma 6

We will proceed in several steps.

Step 1

Let $\bar{\Delta}_l := (\bar{p}_l - \underline{p}_l)/(d+1)$. Fix some $s \le \tilde{i}_1, \dots, \tilde{i}_n \le s+d$. Define a hypercube $H_{\tilde{i}_1,\dots,\tilde{i}_n} := \{p = (p_1; \dots; p_n) : y_{l,\tilde{i}_l} \le p_l \le y_{l,\tilde{i}_l+1}, 1 \le l \le n\}$. Note that for any $p \in H_{\tilde{i}_1,\dots,\tilde{i}_n}$,

$$||\lambda^*(p) - \tilde{\lambda}(p)||_\infty \le ||\lambda^*(p) - \mathcal{L}\lambda^*(p)||_\infty + ||\mathcal{L}\lambda^*(p) - \tilde{\lambda}(p)||_\infty.$$

By Corollary EC.1 (see §EC.4.3), there exists $f_1 = (f_{1,1}; \dots; f_{1,n}) \in \otimes_{l=1}^n \mathsf{P}^{(s\wedge\bar{s})-1}[\underline{p}_l, \bar{p}_l]$ such that

$$\begin{aligned}||\lambda^*(.) - f_1(.)||_\infty &= \sup_{p \in H_{\tilde{i}_1,\dots,\tilde{i}_n}} ||\lambda^*(p) - f_1(p)||_\infty = \sup_{p \in H_{\tilde{i}_1,\dots,\tilde{i}_n}} \max_{1\le j\le n} |\lambda_j^*(p) - f_{1,j}(p)| \\ &\le \sup_{p \in H_{\tilde{i}_1,\dots,\tilde{i}_n}} \max_{1\le j\le n} \left\{ C_{n,(s\wedge\bar{s})} \sum_{i=1}^n \bar{\Delta}_i^{(s\wedge\bar{s})} |(\partial^{(s\wedge\bar{s})}/\partial p_i^{(s\wedge\bar{s})})\lambda_j^*(p)| \right\} \le C_{n,(s\wedge\bar{s})} W \sum_{i=1}^n \bar{\Delta}_i^{(s\wedge\bar{s})},\end{aligned}$$

where $C_{n,(s\wedge\bar{s})}$ is a positive constant that only depends on $n, s\wedge\bar{s}$ and the last inequality follows by assumption N2. Then,

$$\begin{aligned}||\lambda^*(p) - \mathcal{L}\lambda^*(p)||_\infty &\le ||\lambda^*(p) - f_1(p)||_\infty + ||f_1(p) - \mathcal{L}f_1(p)||_\infty + ||\mathcal{L}f_1(p) - \mathcal{L}\lambda^*(p)||_\infty \\ &= ||\lambda^*(p) - f_1(p)||_\infty + ||\mathcal{L}(\lambda^*(p) - f_1(p))||_\infty \le [1 + (2s)^{ns}]\, ||\lambda^*(.) - f_1(.)||_\infty \\ &\le C_{n,(s\wedge\bar{s})} W [1 + (2s)^{ns}] \left(\sum_{i=1}^n \bar{\Delta}_i^{(s\wedge\bar{s})} \right) \\ &\le n C_{n,(s\wedge\bar{s})} W [1 + (2s)^{ns}] \left(\max_{1\le l\le n} \{\bar{p}_l - \underline{p}_l\} \right)^{(s\wedge\bar{s})} \frac{1}{d^{(s\wedge\bar{s})}},\end{aligned}$$

where the first equality and the second inequality follows by Lemma 5 (note that $s\wedge\bar{s} \le s$). Also, we have that for any $1 \le i_1, \dots, i_n \le s+d$, $1 \le j \le n$,

$$\begin{aligned}\left|\gamma_{i_1,\dots,i_n}\lambda_j^* - c^j_{i_1,\dots,i_n}\right| &\le \sum_{j_1=1}^s \sum_{r_1=1}^{j_1} \cdots \sum_{j_n=1}^s \sum_{r_n=1}^{j_n} \frac{\left(\prod_{l=1}^n \beta_{l,i_l,j_l}\right) |\lambda_j^*(\tau_{1,i_1,r_1}, \dots, \tau_{n,i_n,r_n}) - \tilde{\lambda}_j(\tau_{1,i_1,r_1}, \dots, \tau_{n,i_n,r_n})|}{\prod_{l=1}^n \prod_{s_l=1, s_l\ne r_l}^{j_l} (\tau_{l,i_l,r_l} - \tau_{l,i_l,r_l})} \\ &\le \sum_{j_1=1}^s \sum_{r_1=1}^{j_1} \cdots \sum_{j_n=1}^s \sum_{r_n=1}^{j_n} \frac{\prod_{l=1}^n (\bar{\Delta}_l s)^{j_l-1}}{\prod_{l=1}^n (\bar{\Delta}_l / s)^{j_l-1}} |\lambda_j^*(\tau_{1,i_1,r_1}, \dots, \tau_{n,i_n,r_n}) - \tilde{\lambda}_j(\tau_{1,i_1,r_1}, \dots, \tau_{n,i_n,r_n})| \\ &\le \sum_{j_1=1}^s \sum_{r_1=1}^{j_1} \cdots \sum_{j_n=1}^s \sum_{r_n=1}^{j_n} s^{2(\sum_{l=1}^n j_l - n)} |\lambda_j^*(\tau_{1,i_1,r_1}, \dots, \tau_{n,i_n,r_n}) - \tilde{\lambda}_j(\tau_{1,i_1,r_1}, \dots, \tau_{n,i_n,r_n})| \\ &\le \left(\frac{s+s^2}{2} \right)^n s^{2(ns-n)}\, \xi^j_{i_1,\dots,i_n},\end{aligned}$$

where the second inequality follows by Theorem EC.5 (see §EC.4.3), and $\xi^j_{i_1,\ldots,i_n} := \max_{1\le r_1,\ldots,r_n\le s}\{|\lambda^*_j(\tau_{1,i_1,r_1},\ldots,\tau_{n,i_n,r_n}) - \tilde{\lambda}_j(\tau_{1,i_1,r_1},\ldots,\tau_{n,i_n,r_n})|\}$. Hence:

$$\begin{aligned}|\mathcal{L}\lambda^*_j(p) - \tilde{\lambda}_j(p)| &\le \sum_{i_1=1}^{s+d}\cdots\sum_{i_n=1}^{s+d}\left|\gamma_{i_1,\ldots,i_n}\lambda^*_j - c^j_{i_1,\ldots,i_n}\right|\cdot|N_{i_1,\ldots,i_n}(p)|\\ &\le \left(\frac{s+s^2}{2}\right)^n s^{2(ns-n)}\max_{1\le i_1,\ldots,i_n\le s+d}\left\{\xi^j_{i_1,\ldots,i_n}\right\},\end{aligned}$$

where the last inequality holds by Corollary EC.2 (see §EC.4.3). This implies: For all $p\in H_{\tilde{i}_1,\ldots,\tilde{i}_n}$, there exists a constant $C^1_{n,s,\bar{s},\mathcal{P}}>0$ depending only on $n,s,\bar{s},\mathcal{P}$ such that $||\lambda^*(p)-\tilde{\lambda}(p)||_\infty \le C^1_{n,s,\bar{s},\mathcal{P}}\left(Wd^{-(s\wedge\bar{s})}+\max_{1\le i_1,\ldots,i_n\le s+d,1\le j\le n}\left\{\xi^j_{i_1,\ldots,i_n}\right\}\right)$. Note that term after the inequality does not depend on $\tilde{i}_1,\ldots,\tilde{i}_n$; so, we have:

$$||\lambda^*(.)-\tilde{\lambda}(.)||_\infty \le \sup_{p\in\mathcal{P}}||\lambda^*(p)-\tilde{\lambda}(p)||_\infty \le C^1_{n,s,\bar{s},\mathcal{P}}\left(\frac{W}{d^{(s\wedge\bar{s})}}+\max_{\substack{1\le i_1,\ldots,i_n\le s+d\\ 1\le j\le n}}\left\{\xi^j_{i_1,\ldots,i_n}\right\}\right) \tag{EC.18}$$

Step 2

Following similar arguments as in Step 1, we now derive a bound for $||\nabla\lambda^*_j(.)-\nabla\tilde{\lambda}_j(.)||_1$. Consider the hypercube $H_{\tilde{i}_1,\ldots,\tilde{i}_n}$ defined in Step 1. Note that $||\nabla\lambda^*_j(p)-\nabla\tilde{\lambda}_j(p)||_1 \le ||\nabla\lambda^*_j(p)-\mathcal{L}\nabla\lambda^*_j(p)||_1 + ||\mathcal{L}\nabla\lambda^*_j(p)-\nabla\tilde{\lambda}_j(p)||_1$. By Corollary EC.1 (see §EC.4.3), there exists $f_2=(f_{2,1};\ldots;f_{2,n})\in \otimes^n_{l=1}\mathsf{P}^{(s\wedge\bar{s})-1}[\underline{p}_l,\bar{p}_l]$ such that for all $p\in H_{\tilde{i}_1,\ldots,\tilde{i}_n}$, we have

$$||\nabla\lambda^*_j(p)-\nabla f_{2,j}(p)||_1 \le \max_{1\le j\le n} C_{n,(s\wedge\bar{s})-1}\sum_{i=1}^n\bar{\Delta}_i^{(s\wedge\bar{s})-1}|(\partial^{(s\wedge\bar{s})}/\partial p_i^{(s\wedge\bar{s})})\lambda^*_j(p)| \le C_{n,(s\wedge\bar{s})-1}W\sum_{i=1}^n\bar{\Delta}_i^{(s\wedge\bar{s})-1},$$

where $C_{n,(s\wedge\bar{s})-1}$ is a positive constant that only depends on n and $(s\wedge\bar{s})-1$. Then, we have

$$||\nabla\lambda^*(p)-\mathcal{L}\nabla\lambda^*(p)||_1 \le nC_{n,(s\wedge\bar{s})-1}W\left[1+(2s)^{ns}\right]\left(\max_{1\le l\le n}\{\bar{p}_l-\underline{p}_l\}\right)^{(s\wedge\bar{s})-1} d^{-((s\wedge\bar{s})-1)}.$$

Now, by Corollary EC.3 (see §EC.4.3),

$$||\nabla N_{i_1,\ldots,i_n}(p)||_1 \;\le\; n||\nabla N_{i_1,\ldots,i_n}(p)||_\infty \;\le\; n(s-1)\max_{1\le l\le n}\{\bar{\Delta}_l^{-1}\} \;\le\; 2n(s-1)\max_{1\le l\le n}\{(\bar{p}_l-\underline{p}_l)^{-1}\}d.$$

This implies:

$$\begin{aligned}||\mathcal{L}\nabla\lambda^*_j(p)-\nabla\tilde{\lambda}_j(p)||_1 &\le \sum_{i_1=1}^{s+d}\cdots\sum_{i_n=1}^{s+d}\left|\gamma_{i_1,\ldots,i_n}\lambda^*_j - c^j_{i_1,\ldots,i_n}\right|\cdot||\nabla N_{i_1,\ldots,i_n}(p)||_1\\ &\le 2n\left(\frac{s+s^2}{2}\right)^n s^{2(ns-n)}(s+d)^n(s-1)\max_{1\le l\le n}\{(\bar{p}_l-\underline{p}_l)^{-1}\}d\max_{1\le i_1,\ldots,i_n\le s+d}\left\{\xi^j_{i_1,\ldots,i_n}\right\}.\end{aligned}$$

We conclude that there exists a constant $C^2_{n,s,\bar{s},\mathcal{P}} > 0$ depending only on $n, s, \bar{s}, \mathcal{P}$ such that

$$||(\nabla\lambda^*(.) - \nabla\tilde{\lambda}(.))'||_\infty \leq \max_{1\leq j\leq n}\sup_{p\in\mathcal{P}} ||\nabla\lambda_j^*(p) - \nabla\tilde{\lambda}_j(p)||_1$$
$$\leq C^2_{n,s,\bar{s},\mathcal{P}}\left(\frac{W}{d^{(s\wedge\bar{s})-1}} + \max_{\substack{1\leq i_1,\ldots,i_n\leq s+d\\ 1\leq j\leq n}}\left\{\xi^j_{i_1,\ldots,i_n}\right\}d\right). \quad \text{(EC.19)}$$

Step 3

We now analyze $\max_{(i_1,\ldots,i_n)\in\mathcal{G},1\leq j\leq n}\left\{\xi^j_{i_1,\ldots,i_n}\right\}$. Let $\bar{\mathcal{G}} = \{p = (p_1;\ldots;p_n)\in\mathcal{P} : p_l = \tau_{l,i_l,r_l}, 1\leq i_l\leq s+d, 1\leq r_l\leq s, l = 1,\ldots,n\}$. Then, $\max_{1\leq i_1,\ldots,i_n\leq s+d}\left\{\xi^j_{i_1,\ldots,i_n}\right\} = \max_{p\in\bar{\mathcal{G}}}|\lambda_j^*(p) - \tilde{\lambda}_j(p)|$. Note that, for all $x\geq 0$, we can bound

$$\mathbb{P}\left(\max_{p\in\bar{\mathcal{G}}}|\lambda_j^*(p) - \tilde{\lambda}_j(p)|\geq x\right)\leq \mathbb{P}\left(\max_{p\in\bar{\mathcal{G}}}\{\tilde{\lambda}_j(p) - \lambda_j^*(p)\}\geq x\right) + \mathbb{P}\left(\max_{p\in\bar{\mathcal{G}}}\{\lambda_j^*(p) - \tilde{\lambda}_j(p)\}\geq x\right)$$

We now bound the terms on the right hand side of the inequality separately. For any $x\geq 0, t\geq 0$,

$$\mathbb{P}\left(\max_{p\in\bar{\mathcal{G}}}\{\tilde{\lambda}_j(p) - \lambda_j^*(p)\}\geq x\right)\leq \exp(-tx)\mathbb{E}\left[\exp\left(t\max_{p\in\bar{\mathcal{G}}}\{\tilde{\lambda}_j(p) - \lambda_j^*(p)\}\right)\right]$$
$$\leq \exp(-tx)\mathbb{E}\left[\exp\left(t\sum_{p\in\bar{\mathcal{G}}}(\tilde{\lambda}_j(p) - \lambda_j^*(p))\right)\right]$$
$$\leq \exp(-tx)\left[\max_{p\in\bar{\mathcal{G}}}\left\{\mathbb{E}\left[\exp\left(t(\tilde{\lambda}_j(p) - \lambda_j^*(p))\right)\right]\right\}\right]^{s^n(s+d)^n}.$$

Note that there exists a $p^*\in\bar{\mathcal{G}}$ such that the maximum is attained. So, for all $0\leq t\leq L_0$:

$$\max_{p\in\bar{\mathcal{G}}}\left\{\mathbb{E}\left[\exp\left(t(\tilde{\lambda}_j(p) - \lambda_j^*(p))\right)\right]\right\} = \mathbb{E}\left[\exp\left(t(\tilde{\lambda}_j(p^*) - \lambda_j^*(p^*))\right)\right]$$
$$= \exp(-t\lambda_j^*(p^*))\mathbb{E}\left[\exp\left(\frac{t}{L_0}\sum_{i=1}^{L_0}\mathsf{Bin}(\lambda_j^*(p^*))\right)\right]$$
$$= \exp(-t\lambda_j^*(p^*))\left\{\mathbb{E}\left[\exp\left(\frac{t}{L_0}\mathsf{Bin}(\lambda_j^*(p^*))\right)\right]\right\}^{L_0}$$
$$= \exp(-t\lambda_j^*(p^*))\left\{1 - \lambda_j^*(p^*) + \lambda_j^*(p^*)\exp\left(\frac{t}{L_0}\right)\right\}^{L_0}$$
$$\leq \exp(-t\lambda_j^*(p^*))\left\{\exp\left(\lambda_j^*(p^*)\left[\exp\left(\frac{t}{L_0}\right) - 1\right]\right)\right\}^{L_0}$$
$$= \exp(-t\lambda_j^*(p^*))\exp\left(\lambda_j^*(p^*)L_0\sum_{j=1}^{\infty}\frac{1}{j!}\left(\frac{t}{L_0}\right)^j\right)$$
$$= \exp\left(\lambda_j^*(p^*)L_0\sum_{j=2}^{\infty}\frac{1}{j!}\left(\frac{t}{L_0}\right)^j\right)$$
$$\leq \exp\left(\lambda_j^*(p^*)t^2/L_0\right)\ \leq\ \exp(t^2/L_0),$$

where the last inequality follows by the fact that $\sum_{j=2}^{\infty}(j!)^{-1}(t/L_0)^j \le (t/L_0)^2\sum_{j=2}^{\infty}[j(j-1)]^{-1} \le (t/L_0)^2$. Hence, we have that for all $0 \le t \le L_0$, $\mathbb{P}(\max_{p\in\bar{\mathcal{G}}}\{\tilde{\lambda}_j(p)-\lambda_j^*(p)\} \ge x) \le \exp(s^n(s+d)^n t^2/L_0 - tx)$. Following a similar argument, we can show that for all $0 \le t \le L_0$, there exists some $q^* \in \bar{\mathcal{G}}$ such that

$$\begin{aligned}\mathbb{P}\left(\max_{p\in\bar{\mathcal{G}}}\{\lambda_j^*(p)-\tilde{\lambda}_j(p)\} \ge x\right) &\le \exp(-tx)\left[\max_{p\in\bar{\mathcal{G}}}\left\{\mathbb{E}\left[\exp\left(t(\lambda_j^*(p)-\tilde{\lambda}_j(p))\right)\right]\right\}\right]^{s^n(s+d)^n}\\ &\le \exp(-tx)\left[\exp(t\lambda_j^*(q^*))\exp\left(\lambda_j^*(q^*)L_0\sum_{j=1}^{\infty}\frac{(-1)^j}{j!}\left(\frac{t}{L_0}\right)^j\right)\right]^{s^n(s+d)^n}\\ &\le \exp(s^n(s+d)^n\lambda_j^*(q^*)t^2/L_0 - tx) \le \exp(s^n(s+d)^n t^2/L_0 - tx).\end{aligned}$$

Pick $t=\sqrt{L_0}s^{-n/2}(s+d)^{-n/2}\log k$ (note that $t/L_0=(L_0 s^n(s+d)^n\log^{-2}k)^{-1/2}=(\tilde{L}_0\log^{-2}k)^{-1/2}<1$) and $x=2L_0^{-1/2}(s+d)^{n/2}s^{n/2}\log k$, we then have for $k\ge 3$:

$$\mathbb{P}\left(\max_{p\in\bar{\mathcal{G}}}\{\tilde{\lambda}_j(p)-\lambda_j^*(p)\} \ge \frac{2\log k}{\sqrt{L_0}}s^{\frac{n}{2}}(s+d)^{\frac{n}{2}}\right) \le 2\exp\left(-\log^2 k\right) \le 2\exp(-\log k) = \frac{2}{k}.$$

Note that $\tilde{L}_0 = L_0(s+d)^n s^n$. Hence, $L_0=\tilde{L}_0 s^{-n}(s+d)^{-n}$. Combine with the results derive in Step 1 and 2, we then have that there exists constant $C^3_{n,s,\bar{s},\mathcal{P}}, C^4_{n,s,\bar{s},\mathcal{P}}$ depending on $n,s,\bar{s},\mathcal{P}$ only, such that for all $k\ge 3$:

$$\begin{aligned}&\mathbb{P}\left(\left|\left|\lambda^*(.)-\tilde{\lambda}(.)\right|\right|_\infty \ge C^3_{n,s,\bar{s},\mathcal{P}}\left(\frac{W}{d^{(s\wedge\bar{s})}}+\frac{2\log k}{\sqrt{\tilde{L}_0}}d^n\right)\right)\\ \le\ &\mathbb{P}\left(\left|\left|\lambda^*(.)-\tilde{\lambda}(.)\right|\right|_\infty \ge C^1_{n,s,\bar{s},\mathcal{P}}\left(\frac{W}{d^{(s\wedge\bar{s})}}+\frac{2\log k}{\sqrt{L_0}}s^{\frac{n}{2}}(s+d)^{\frac{n}{2}}\right)\right)\le\frac{2}{k},\ \text{and}\\ &\mathbb{P}\left(\left|\left|\left(\nabla\lambda^*(.)-\nabla\tilde{\lambda}(.)\right)'\right|\right|_\infty \ge C^4_{n,s,\bar{s},\mathcal{P}}\left(\frac{W}{d^{(s\wedge\bar{s})-1}}+\frac{2\log k}{\sqrt{\tilde{L}_0}}d^{n+1}\right)\right)\\ \le\ &\mathbb{P}\left(\left|\left|\left(\nabla\lambda^*(.)-\nabla\tilde{\lambda}(.)\right)'\right|\right|_\infty \ge C^2_{n,s,\bar{s},\mathcal{P}}\left(\frac{W}{d^{(s\wedge\bar{s})-1}}+\frac{2\log k}{\sqrt{L_0}}s^{\frac{n}{2}}(s+d)^{\frac{n}{2}}d\right)\right)\le\frac{2}{k}.\end{aligned}$$

Let $d=(\tilde{L_0}^{-1/2}\log k)^{-1/(s+n)}$. We conclude that there exist constants M_4, M_5 independent of $k\ge 3$ such that for all $k\ge 3$,

$$\mathbb{P}\left(\left|\left|\lambda^*(.)-\tilde{\lambda}(.)\right|\right|_\infty \ge M_4\left(\frac{\log k}{\sqrt{\tilde{L}_0}}\right)^{\frac{s\wedge\bar{s}}{s+n}}\right)\le\frac{2}{k}\ \text{and}\ \mathbb{P}\left(\left|\left|\left(\nabla\lambda^*(.)-\nabla\tilde{\lambda}(.)\right)'\right|\right|_\infty \ge M_5\left(\frac{\log k}{\sqrt{\tilde{L}_0}}\right)^{\frac{(s\wedge\bar{s})-1}{s+n}}\right)\le\frac{2}{k}.$$

□

EC.3.3. Proof of Lemma 7

By assumption, $\Lambda_{\lambda^*} := \{x : x = \lambda^*(p), p \in \mathcal{P}\}$ is convex. Recall that by R1, $\lambda^*(.)$ is invertible and its inverse function $p^* : \Lambda_{\lambda^*} \to \mathcal{P}$ satisfies that $\lambda^*(p^*(x)) = x, \forall x \in \Lambda_{\lambda^*}$. Let $\delta(p) := \tilde{\lambda}(p) - \lambda^*(p)$. Note that $||\delta(p)||_\infty \le ||\lambda^*(p) - \tilde{\lambda}(p)||_\infty \le ||\lambda^*(.) - \tilde{\lambda}(.)||_\infty$ for all $p \in \mathcal{P}$. Since the optimal solution of P_λ is stationary, i.e., $\lambda_t^D = \lambda^D$ for all t, we can formulate an equivalent "one-period" version P'_λ, and also construct an auxiliary optimization problem P^{ac}_λ:

$$(\text{P}'_\lambda) \quad r^D := \max_{\lambda \in \Lambda_{\lambda^*}} r^*(\lambda) \quad \text{s.t. } A\lambda \preceq \frac{C}{T} \quad \text{and}$$
$$(\text{P}^{ac}_\lambda) \quad \bar{r}^D := \max_{\lambda \in \Lambda_{\lambda^*}} r^*(\lambda) \quad \text{s.t. } A\lambda + A\delta(p^*(\lambda)) \preceq \frac{C}{T}.$$

Note that $\lambda^D = \lambda^*(p^D)$ is the unique optimizer of P'_λ. Let λ^{ac} denote an optimizer of P^{ac}_λ (note that λ^{ac} may not be unique). By the strong concavity of $r^*(.)$ as a function of λ, λ^{ac} is the unique optimizer of the following optimization problem whose optimal value remains to be $\bar{r}^D$:

$$(\bar{\text{P}}^{ac}_\lambda) \quad \bar{r}^D := \max_{\lambda \in \Lambda_{\lambda^*}} r^*(\lambda) \quad \text{s.t. } A\lambda \preceq \frac{C}{T} - A\delta(p^*(\lambda^{ac})).$$

Note that if we view the term $A\delta(p^*(\lambda^{ac}))$ on the right hand side of the inequality as a perturbation of the term C/T in P'_λ, optimization $(\bar{\text{P}}^{ac}_\lambda)$ is equivalent to

$$(\text{P}'_\lambda(\epsilon)) \quad r^D(\epsilon) := \max_{\lambda \in \Lambda_{\lambda^*}} r^*(\lambda) \quad \text{s.t. } A\lambda \preceq \frac{C}{T} - \epsilon,$$

where $\epsilon = -A\delta(p^*(\lambda^{ac}))$. In light of Corollary EC.4 (see §EC.4.3), there exists a constant $K_{13} > 0$ independent of $||\epsilon||_\infty$ such that $||\lambda^D - \lambda^{ac}||_\infty \le K_{13}||\epsilon||_\infty$. Now, let $\tilde{\text{P}}_\lambda$ denote the optimization problem: $\max_{\lambda \in \Lambda_{\lambda^*}} \{\tilde{r}(\lambda) : \text{s.t. } A\lambda + A\delta(p^*(\lambda)) \preceq C/T\}$. Let $\bar{\lambda} := \lambda^*(\bar{p})$. (This optimization problem is emanated from $\tilde{\text{P}}$. The only difference is that $\tilde{\text{P}}_\lambda$ optimizes over λ instead of p as in $\tilde{\text{P}}$.) Since $\bar{p}$ is an optimizer of $\tilde{\text{P}}$, $\bar{\lambda}$ is an optimizer of $\tilde{\text{P}}_\lambda$. Note that the constraints of $\tilde{\text{P}}_\lambda$ and P^{ac}_λ are identical. Thus, $\bar{\lambda}$ is feasible to P^{ac}_λ and λ^{ac} is feasible to $\tilde{\text{P}}_\lambda$. By the optimality condition of P^{ac}_λ and the fact that $r^*(.)$ is strongly concave with respect to λ and the eigenvalues of the Hessian matrix of $r^*(.)$ are bounded from above by $-\underline{v}$, we have $r^*(\bar{\lambda}) \le r^*(\lambda^{ac}) + \nabla r^*(\lambda^{ac}) \cdot (\bar{\lambda} - \lambda^{ac}) - \frac{\underline{v}}{2}||\lambda^{ac} - \bar{\lambda}||_2^2 \le r^*(\lambda^{ac}) - \frac{\underline{v}}{2}||\lambda^{ac} - \bar{\lambda}||_\infty^2$ (here, we use the fact that $\nabla r^*(\lambda^{ac}) \cdot (\bar{\lambda} - \lambda^{ac}) \le 0$, by the optimality of λ^{ac}). By the optimality condition of $\tilde{\text{P}}_\lambda$, we have $\tilde{r}(\lambda^{ac}) \le \tilde{r}(\bar{\lambda})$; so,

$$\begin{aligned}
\frac{\underline{v}}{2}||\lambda^{ac} - \bar{\lambda}||_\infty^2 &\le r^*(\lambda^{ac}) - r^*(\bar{\lambda}) \\
&\le [r^*(\lambda^{ac}) - \tilde{r}(\lambda^{ac})] - \left[r^*(\bar{\lambda}) - \tilde{r}(\bar{\lambda})\right] \\
&\le ||(\nabla_\lambda r^*(\xi) - \nabla_\lambda \tilde{r}(\xi))'||_\infty ||\lambda^{ac} - \bar{\lambda}||_\infty,
\end{aligned}$$

for some ξ, where the last inequality follows by Mean Value Theorem and norm inequality. This indicates that $||\lambda^{ac} - \bar{\lambda}||_\infty \leq 2\underline{v}^{-1}||(\nabla_\lambda r^*(\xi) - \nabla_\lambda \tilde{r}(\xi))'||_\infty$. Since $p(.)$ is continuously differentiable by R1 and Λ_{λ^*} is compact, $||\nabla p^*(\lambda)||_\infty$ is uniformly bounded by some constant $K > 0$. Note that $r^*(\lambda) = \lambda \cdot p^*(\lambda)$ and $\tilde{r}(\lambda) = (\lambda + \delta(p^*(\lambda))) \cdot p^*(\lambda)$, so $r^*(\lambda) - \tilde{r}(\lambda) = p^*(\lambda) \cdot \delta(p^*(\lambda))$ and

$$\begin{aligned}
||(\nabla_\lambda r^*(\xi) - \nabla_\lambda \tilde{r}(\xi))'||_\infty &= ||\nabla_\lambda r^*(\xi) - \nabla_\lambda \tilde{r}(\xi)||_1 \\
&= ||\nabla_\lambda p^*(\xi)\,\delta(p^*(\xi)) + \nabla_\lambda p^*(\xi)\,\nabla_p \delta(p^*(\xi))\,p^*(\xi)||_1 \\
&\leq ||\nabla_\lambda p^*(\xi)||_1 ||\delta(p^*(\xi))||_1 + ||\nabla_\lambda p^*(\xi)||_1 ||\nabla_p \delta(p^*(\xi))||_1 ||p^*(\xi)||_1 \\
&\leq nK||\lambda^*(.) - \tilde{\lambda}(.)||_\infty + nK\left(\sum_{l=1}^n \bar{p}_l\right)||\nabla\lambda^*(.) - \nabla\tilde{\lambda}(.)||_\infty.
\end{aligned}$$

Finally, since $p^*(.)$ is in $\mathsf{C}^1(\Lambda_{\theta^*})$ and Λ_{θ^*} is compact, there exists some constant K' such that $||p^D - \bar{p}||_\infty \leq K'||\lambda^D - \bar{\lambda}||_\infty$. This implies $||p^D - \bar{p}||_\infty \leq K'||\lambda^D - \bar{\lambda}||_\infty \leq K'(||\lambda^D - \lambda^{ac}||_\infty + ||\lambda^{ac} - \bar{\lambda}||_\infty) \leq K'[K_{13}||\epsilon||_\infty + 2v^{-1}nK||\lambda^*(.) - \tilde{\lambda}(.)||_\infty + 2v^{-1}nK(\sum_{l=1}^n \bar{p}_l)||(\nabla\lambda^*(.) - \nabla\tilde{\lambda}(.))'||_\infty] \leq M_6 \max\{||\lambda^*(.) - \tilde{\lambda}(.)||_\infty, ||\nabla\lambda^*(.) - \nabla\tilde{\lambda}(.)||_\infty\}$ for $M_6 = K'[K_{13} + 2v^{-1}nK + 2v^{-1}nK(\sum_{l=1}^n \bar{p}_l)]$ that is independent of $||\lambda^*(.) - \tilde{\lambda}(.)||_\infty$ and $||(\nabla\lambda^*(.) - \nabla\tilde{\lambda}(.))'||_\infty$. □

EC.3.4. Proof of Lemma 8

Recall that $\epsilon(\tilde{L}_0) := (\log k/\sqrt{\tilde{L}_0})^{((s\wedge\bar{s})-1)/(s+n)}$. Define $\bar{\theta} = (\lambda^*(\bar{p}) - \nabla\lambda^*(\bar{p}) \cdot \bar{p}; \nabla\lambda_1^*(\bar{p}); \ldots; \nabla\lambda_n^*(\bar{p}))$. We first bound $\mathbb{P}(||\theta^* - \bar{\theta}||_2 > (C^1_{n,\mathcal{P}} + 1)\epsilon(\tilde{L}_0))$ for some $C^1_{n,\mathcal{P}}$ defined later. Let $\mathcal{E} = \{||p^D - \bar{p}||_2 > \sqrt{n}M_6 \max\{M_4, M_5\}\epsilon(\tilde{L}_0)\}$. Since $\theta^* = (\lambda^*(p^D) - \nabla\lambda^*(p^D) \cdot p^D; \nabla\lambda_1^*(p^D); \ldots; \nabla\lambda_n^*(p^D))$, by continuity of $\lambda^*(.)$ and $\nabla\lambda^*(.)$, there exists $C^1_{n,\mathcal{P}} \geq 0$ depending only on n and $\mathcal{P}$ such that, conditioning on $\mathcal{E}^c$, we have:

$$\begin{aligned}
&||\theta^* - \bar{\theta}||_2^2 = \left|\left|\lambda^*(p^D) - \nabla\lambda^*(p^D) \cdot p^D - \lambda^*(\bar{p}) + \nabla\lambda^*(\bar{p}) \cdot \bar{p}\right|\right|_F^2 + \left|\left|\nabla\lambda^*(p^D) - \nabla\lambda^*(\bar{p})\right|\right|_F^2 \\
&= \sum_{i=1}^n \left(\lambda_i^*(p^D) - \sum_{j=1}^n p_j^D \frac{\partial\lambda_i^*}{\partial p_j}(p^D) - \lambda_i^*(\bar{p}) + \sum_{j=1}^n \bar{p}_j \frac{\partial\lambda_i^*}{\partial p_j}(\bar{p})\right)^2 + \sum_{i=1}^n\sum_{j=1}^n \left(\frac{\partial\lambda_i^*}{\partial p_j}(p^D) - \frac{\partial\lambda_i^*}{\partial p_j}(\bar{p})\right)^2 \\
&\leq (C^1_{n,\mathcal{P}}\,\epsilon(\tilde{L}_0))^2.
\end{aligned} \tag{EC.20}$$

where the first equality follows since for $\theta = (a; B_1'; \ldots; B_n')$, $||\theta||_2^2 = ||a||_F^2 + ||B||_F^2$. By the law of total probability, we have:

$$\begin{aligned}
&\mathbb{P}\left(\left|\left|\theta^* - \bar{\theta}\right|\right|_2 > (C^1_{n,\mathcal{P}} + 1)\epsilon(\tilde{L}_0)\right) \\
&\quad\leq \mathbb{P}\left(\left|\left|\theta^* - \bar{\theta}\right|\right|_2 > (C^1_{n,\mathcal{P}} + 1)\epsilon(\tilde{L}_0)\,\middle|\,\mathcal{E}\right)\mathbb{P}(\mathcal{E}) \\
&\qquad+ \mathbb{P}\left(\left|\left|\theta^* - \bar{\theta}\right|\right|_2 > (C^1_{n,\mathcal{P}} + 1)\epsilon(\tilde{L}_0)\,\middle|\,\mathcal{E}^c\right)\mathbb{P}(\mathcal{E}^c) \\
&\quad\leq \mathbb{P}(\mathcal{E}) + \mathbb{P}\left(\left|\left|\theta^* - \bar{\theta}\right|\right|_2 > (C^1_{n,\mathcal{P}} + 1)\epsilon(\tilde{L}_0)\,\middle|\,\mathcal{E}^c\right) \leq \frac{4}{k}
\end{aligned} \tag{EC.21}$$

where the last inequality follows by (EC.20) and the inequality below:

$$\mathbb{P}(\mathcal{E}) \leq \mathbb{P}\left(\sqrt{n}\left|\left|p^D - \bar{p}\right|\right|_\infty > \sqrt{n}M_6 \max\{M_4, M_5\}\epsilon(\tilde{L}_0)\right)$$
$$\leq \mathbb{P}\left(\sqrt{n}M_6 \max\{||\lambda^*(.) - \tilde{\lambda}(.)||_\infty, ||(\nabla\lambda^*(.) - \nabla\tilde{\lambda}(.))'||_\infty\} > \sqrt{n}M_6 \max\{M_4, M_5\}\left(\frac{\log k}{\sqrt{\tilde{L}_0}}\right)^{\frac{(s\wedge\bar{s})-1}{s+n}}\right)$$
$$\leq \mathbb{P}\left(||\lambda^*(.) - \tilde{\lambda}(.)||_\infty > M_4\left(\frac{\log k}{\sqrt{\tilde{L}_0}}\right)^{\frac{s\wedge\bar{s}}{s+n}}\right) + \mathbb{P}\left(||(\nabla\lambda^*(.) - \nabla\tilde{\lambda}(.))'||_\infty > M_5\left(\frac{\log k}{\sqrt{\tilde{L}_0}}\right)^{\frac{(s\wedge\bar{s})-1}{s+n}}\right) \leq \frac{4}{k},$$

where the second inequality follows from Lemma 7, the third inequality follows by the union bound and the fact that $\sqrt{\tilde{L}_0} \geq \log^{3/2} k > \log k$, and the last inequality follows by Lemma 6. We now bound $\mathbb{P}(||\bar{\theta} - \hat{\theta}|| > C^2_{n,\mathcal{P}} \max\{M_4, M_5\}\epsilon(\tilde{L}_0))$, for some $C^2_{n,\mathcal{P}}$ defined below. Note that there exists a constant $C^2_{n,\mathcal{P}} > 0$ depending only on n and $\mathcal{P}$ such that:

$$||\bar{\theta} - \hat{\theta}||_2^2 = \left|\left|\lambda^*(\bar{p}) - \nabla\lambda^*(\bar{p})\cdot\bar{p} - \tilde{\lambda}(\bar{p}) + \nabla\tilde{\lambda}(\bar{p})\cdot\bar{p}\right|\right|_F^2 + \left|\left|\nabla\lambda^*(\bar{p}) - \nabla\tilde{\lambda}(\bar{p})\right|\right|_F^2$$
$$\leq (C^2_{n,\mathcal{P}} \max\{||\lambda^*(.) - \tilde{\lambda}(.)||_\infty, ||\nabla\lambda^*(.) - \nabla\tilde{\lambda}(.)||_\infty\})^2.$$

So, we can bound:

$$\mathbb{P}\left(||\bar{\theta} - \hat{\theta}||_2 > C^2_{n,\mathcal{P}} \max\{M_4, M_5\}\epsilon(\tilde{L}_0)\right)$$
$$\leq \mathbb{P}\left(C^2_{n,\mathcal{P}} \max\{||\lambda^*(.) - \tilde{\lambda}(.)||_\infty, ||(\nabla\lambda^*(.) - \nabla\tilde{\lambda}(.))'||_\infty\} > C^2_{n,\mathcal{P}} \max\{M_4, M_5\}\epsilon(\tilde{L}_0)\right)$$
$$\leq \mathbb{P}\left(||\lambda^*(.) - \tilde{\lambda}(.)||_\infty > M_4\epsilon(\tilde{L}_0)\right) + \mathbb{P}\left(||(\nabla\lambda^*(.) - \nabla\tilde{\lambda}(.))'||_\infty > M_5\epsilon(\tilde{L}_0)\right) \leq \frac{4}{k}, \quad \text{(EC.22)}$$

where the last inequality follows from Lemma 6 and the fact that $\sqrt{\tilde{L}_0} \geq \log^{3/2} k > \log k$. Finally, by combining (EC.21) and (EC.22) and letting $M_7 = C^1_{n,\mathcal{P}} + C^2_{n,\mathcal{P}} \max\{M_4, M_5\} + 1$, we have

$$\mathbb{P}\left(||\theta^* - \hat{\theta}||_2 > M_7\epsilon(\tilde{L}_0)\right) \leq \mathbb{P}\left(||\theta^* - \tilde{\theta}||_2 > (C^1_{n,\mathcal{P}} + 1)\epsilon(\tilde{L}_0)\right) + \mathbb{P}\left(||\tilde{\theta} - \hat{\theta}||_2 > C^2_{n,\mathcal{P}} \max\{M_4, M_5\}\epsilon(\tilde{L}_0)\right) \leq \frac{8}{k}.$$

□

EC.3.5. Proof of Theorem 3

Throughout this section, we fix $\pi=$ NLRC and assume that $T = 1$ (this is without lost of generality). Let τ be the minimum of k and the first time $t \geq \tilde{L}_0 + 1$ such that the following condition $(C1)$ is violated: $\psi > ||\sum_{s=\tilde{L}_0+1}^{t-1} \frac{\tilde{\Delta}_s}{k-s}||_2$, where $\psi := \sqrt{\epsilon(\tilde{L}_0)} = (k^{-1/4}\log^{1/2} k)^{((s\wedge\bar{s})-1)/(2s+n-1)}$ and $\tilde{\Delta}_s = \Delta_s + \lambda^*(p_s) - \lambda(p_s;\hat{\theta})$. Define $\mathcal{A} := \{||\theta^* - \hat{\theta}||_2 \leq M_7\epsilon(\tilde{L}_0)\}$, where M_7 and $\epsilon(\tilde{L}_0)$ are as defined in Lemma 8. Note that, by Lemma 8, we have $k\mathbb{P}(\mathcal{A}^c) \leq 8$; so, for all $k \geq \Omega_3 = 17$, $\mathbb{P}(\mathcal{A}^c) < 1/2$.

Moreover, since $\epsilon(\tilde{L}_0)=o(1)$ (recall that $\tilde{L}_0 \geq \log^3 k$), there exists a constant $\Omega_4>0$ independent of k such that for all $k\geq \Omega_4$ and all sample paths on $\mathcal{A}$, we have $||\theta^*-\hat{\theta}||_2 \leq M_7\epsilon(\tilde{L}_0) < \bar{\delta}$ where $\bar{\delta}$ is as defined in Lemma 1. We will suppress the dependency of $\epsilon(\tilde{L}_0)$ on $\tilde{L}_0$ for notational brevity whenever there is no confusion. Now, define $\lambda_t := \lambda^*(p_t)$ and $\hat{\lambda}_t := \lambda(p_t;\hat{\theta})$. As long as $p_t \in \mathcal{P}$, we have $\hat{\lambda}_t = \lambda(p_t;\hat{\theta}) = \lambda^0(\hat{\theta}) - \sum_{s=1}^{t-1}\frac{Q\tilde{\Delta}_s}{k-s}$. Similar to the proof of Theorem 1, we state two lemmas.

Lemma EC.10. *There exists some constant $\Omega_5>0$ independent of k such that for all $k\geq\Omega_5$ and for all sample paths on $\mathcal{A}$, $p_t\in\mathcal{P}$ and $C_t \succ 0$ for all $t<\tau$.*

Lemma EC.11. *There exists some constant K_9 independent of $k \geq \max\{\Omega_3,\Omega_4,\Omega_5\}$ such that for all $k\geq \max\{\Omega_3,\Omega_4,\Omega_5\}$, $\mathbb{E}[k-\tau|\mathcal{A}] \leq K_9(\epsilon^2 k + \epsilon^{-1}\log k + \epsilon^{-2})$*

Let $\hat{R}^{\pi}_{\lambda^*}(k)$ denote the revenue during exploitation stage. Since the one period revenue loss is bounded by $\bar{r}$, we have $\rho^{\pi}(k) \leq \tilde{L}_0\bar{r} + \sum_{t=\tilde{L}_0+1}^{k} r^*(\lambda^D) - \mathbb{E}^{\pi}_{\lambda^*}[\hat{R}^{\pi}_{\lambda^*}(k)] = \tilde{L}_0\bar{r} + \sum_{t=\tilde{L}_0+1}^{k} r^*(\lambda^D(\theta^*)) - \mathbb{E}^{\pi}_{\lambda^*}[\hat{R}^{\pi}_{\lambda^*}(k)]$. (Note that $\lambda^D = \arg\max_{\lambda\in\Lambda_{\lambda^*}}\{r^*(\lambda) : A\lambda \preceq C\}$ and $\lambda^D(\theta^*) = \arg\max_{\lambda\in\Lambda_{\lambda^*}}\{r(\lambda;\theta^*) : A\lambda \preceq C\}$. Recall that by construction of θ^*, we have $\lambda^D = \lambda^D(\theta^*)$.). The following result is useful for bounding the revenue loss later.

Lemma EC.12. *There exist some constant Ω_6 independent of k and some constant $K_{10}>0$ independent of $k\geq\Omega_6$ and $\hat{\theta}$ such that for all $k\geq\Omega_6$, $||\lambda^D(\hat{\theta}) - \lambda^0(\hat{\theta})||_2 \leq K_{10}\epsilon(\tilde{L}_0)^2$*

Define $K := \max\{\Omega_3,\Omega_4,\Omega_5,\Omega_6\}$. For $k\leq K, \rho^{\pi}(k)\leq \bar{r}K$. We now consider the case when $k>K$. By similar arguments as in (EC.1) and (EC.2), we have that

$$\sum_{t=\tilde{L}_0+1}^{k} r^*(\lambda^D(\theta^*)) - \mathbb{E}^{\pi}_{\lambda^*}\left[\hat{R}^{\pi}_{\lambda^*}(k)\right] \leq \mathbb{E}^{\pi}_{\lambda^*}\left[\sum_{t=\tilde{L}_0+1}^{\tau-1} \nabla r^*(\lambda^D(\theta^*))\cdot(\lambda^D(\theta^*)-\lambda_t)\middle|\mathcal{A}\right]$$
$$+\frac{\bar{v}}{2}\mathbb{E}^{\pi}_{\lambda^*}\left[\sum_{t=\tilde{L}_0+1}^{\tau-1}||\lambda^D(\theta^*)-\lambda_t||_2^2\middle|\mathcal{A}\right] + \bar{r}\,\mathbb{E}^{\pi}_{\lambda^*}[k-\tau|\mathcal{A}] + 2\bar{r} + \bar{r}k\mathbb{P}(\mathcal{A}^c) \quad \text{(EC.23)}$$

Note that $\nabla r^*(\lambda^D)\cdot(\lambda^D(\theta^*)-\lambda_t) = \nabla r(\lambda^D(\theta^*);\theta^*)\cdot(\lambda^D(\theta^*)-\lambda_t) = \mu^D(\theta^*)'A(\lambda^D(\theta^*)-\lambda^D(\hat{\theta}) + \lambda^D(\hat{\theta}) - \lambda^0(\hat{\theta}) + \lambda^0(\hat{\theta}) - \hat{\lambda}_t + \hat{\lambda}_t - \lambda_t)$. Therefore, for the first term of (EC.23), we have for $k<K$:

$$\mathbb{E}^{\pi}_{\lambda^*}\left[\sum_{t=\tilde{L}_0+1}^{\tau-1}\nabla r^*(\lambda^D(\theta^*))\cdot(\lambda^D(\theta^*)-\lambda_t)\middle|\mathcal{A}\right] = \mathbb{E}^{\pi}_{\lambda^*}\left[\sum_{t=\tilde{L}_0+1}^{\tau-1}\mu^D(\theta^*)'(A\lambda^D(\theta^*)-A\lambda^D(\hat{\theta}))\middle|\mathcal{A}\right]$$
$$+\mathbb{E}^{\pi}_{\lambda^*}\left[\sum_{t=\tilde{L}_0+1}^{\tau-1}\mu^D(\theta^*)'(A\lambda^D(\hat{\theta})-A\lambda^0(\hat{\theta}))\middle|\mathcal{A}\right] + \mathbb{E}^{\pi}_{\lambda^*}\left[\sum_{t=\tilde{L}_0+1}^{\tau-1}\mu^D(\theta^*)'A(\lambda^0(\hat{\theta})-\hat{\lambda}_t+\hat{\lambda}_t-\lambda_t)\middle|\mathcal{A}\right]$$
$$\leq K_{10}\mu^D(\theta^*)'A\mathbf{e}\epsilon^2 k + \mathbb{E}^{\pi}_{\lambda^*}\left[\sum_{t=\tilde{L}_0+1}^{\tau-1}\mu^D(\theta^*)'\left(\sum_{s=\tilde{L}_0+1}^{t-1}\frac{A\tilde{\Delta}_s}{k-s} + A\Delta_t - A\tilde{\Delta}_t\right)\middle|\mathcal{A}\right]$$

$$\leq K_{10}\mu^D(\theta^*)'A\mathbf{e}\,\epsilon^2 k + \mu^D(\theta^*)'A\mathbf{e}\frac{1+k\mathbb{P}^\pi_{\lambda^*}(\mathcal{A}^c)}{\mathbb{P}^\pi_{\lambda^*}(\mathcal{A})} + \psi||\mu^D(\theta^*)||_2||A||_2(\mathbb{E}^\pi_{\lambda^*}[k-\tau|\mathcal{A}]+1)$$
$$\leq K_{10}\mu^D(\theta^*)'A\mathbf{e}\,\epsilon^2 k + 18\mu^D(\theta^*)'A\mathbf{e} + ||\mu^D(\theta^*)||_2||A||_2 K_9(\epsilon^2 k + \epsilon^{-1}\log k + \epsilon^{-2}) + ||\mu^D(\theta^*)||_2||A||_2$$
$$\leq K_{11}(1+\epsilon^2 k + \epsilon^{-1}\log k + \epsilon^{-2}) \tag{EC.24}$$

where $K_{11} = (18+K_{10})\mu^D(\theta^*)'A\mathbf{e} + ||\mu^D(\theta^*)||_2||A||_2(1+K_9)$, the first inequality follows by the fact that $\mu^D(\theta^*)'A(\lambda^D(\theta^*)-\lambda^D(\hat\theta))$ on $\mathcal{A}$ (see the paragraph after (EC.3) for explanation) and Lemma EC.12, the second inequality follows by a similar argument in the proof of Lemma 3, the third inequality follows by Lemma EC.11 and the fact that $\psi < 1$.

We now bound the second term of (EC.23). A key observation is that there exists some constant κ_0 such that $||\lambda^*(p_t)-\lambda(p_t;\hat\theta)||_2 = ||\lambda^*(p^D)+\nabla\lambda^*(p^D)\cdot(p_t-p^D)+(p_t-p^D)'\nabla^2\lambda^*(\xi)(p_t-p^D) - \lambda(p^D;\theta^*)-\nabla\lambda(p^D;\theta^*)\cdot(p_t-p^D)||_2 = ||(p_t-p^D)'\nabla^2\lambda^*(\xi)(p_t-p^D)||_2 \leq \kappa_0||p_t-p^D||_2^2$, where the second equality follows by the construction of θ^*. So, conditioning on $\mathcal{A}$, for all $t<\tau$,

$$\begin{aligned}
\left|\left|\lambda_t-\hat\lambda_t\right|\right|_2 &= \left|\left|\lambda^*(p_t)-\lambda(p_t;\hat\theta)\right|\right|_2 \leq ||\lambda^*(p_t)-\lambda(p_t;\theta^*)||_2 + \left|\left|\lambda(p_t;\theta^*)-\lambda(p_t;\hat\theta)\right|\right|_2\\
&\leq \kappa_0||p_t-p^D||_2^2+\omega M_7\epsilon = \kappa_0||p_t-p^0(\hat\theta)+p^0(\hat\theta)-p^D(\hat\theta)+p^D(\hat\theta)-p^D(\theta^*)||_2^2+\omega M_7\epsilon\\
&\leq 3\kappa_0||p_t-p^0(\hat\theta)||_2^2+3\kappa_0||p^0(\hat\theta)-p^D(\hat\theta)||_2^2+3\kappa_0||p^D(\hat\theta)-p^D(\theta^*)||_2^2+\omega M_7\epsilon\\
&\leq 3\kappa_0||\nabla p_\lambda(\lambda^0(\hat\theta);\hat\theta)||_2^2||Q||_2^2\left|\left|\sum_{s=\tilde L_0+1}^{t-1}\frac{\tilde\Delta_s}{k-s}\right|\right|_2^2+3\kappa_0\omega^2K_{10}^2\epsilon^4+3\kappa_0\left(n^{3/2}M_6\omega||\theta^*-\hat\theta||_2\right)^2+\omega M_7\epsilon\\
&\leq 3\kappa_0\omega^2||Q||_2^2\psi^2+3\kappa_0\omega^2K_{10}^2\epsilon^4+3\kappa_0n^3M_6^2M_7^2\omega^2\epsilon^2+\omega M_7\epsilon\leq\omega_0\epsilon
\end{aligned} \tag{EC.25}$$

where $\omega_0 = 3\kappa_0\omega^2||Q||^2+3\kappa_0\omega^2K_{10}^2+3\kappa_0n^3M_6^2M_7^2\omega^2+\omega M_7$, the fourth inequality follows by Lemma EC.12 and Lemma 7, and the fifth inequality follows by the definition of τ and $\mathcal{A}$. We now bound the second term in (EC.23) below.

$$\begin{aligned}
&\frac{\bar v}{2}\mathbb{E}^\pi_{\lambda^*}\left[\sum_{t=\tilde L_0+1}^{\tau-1}\left|\left|\lambda^D(\theta^*)-\lambda_t)\right|\right|_2^2|\mathcal{A}\right]\\
&\leq \bar v\mathbb{E}^\pi_{\lambda^*}\left[\sum_{t=\tilde L_0+1}^{\tau-1}\left|\left|\lambda^D(\theta^*)-\lambda^0(\hat\theta)\right|\right|_2^2|\mathcal{A}\right]+\bar v\mathbb{E}^\pi_{\lambda^*}\left[\sum_{t=\tilde L_0+1}^{\tau-1}\left|\left|\lambda^0(\hat\theta)-\lambda_t\right|\right|_2^2|\mathcal{A}\right]\\
&\leq 2\bar vk\left(\mathbb{E}^\pi_{\lambda^*}\left[\left|\left|\lambda^D(\theta^*)-\lambda^D(\hat\theta)\right|\right|_2^2|\mathcal{A}\right]+\mathbb{E}^\pi_{\lambda^*}\left[\left|\left|\lambda^D(\hat\theta)-\lambda^0(\hat\theta)\right|\right|_2^2|\mathcal{A}\right]\right)+\bar v\mathbb{E}^\pi_{\lambda^*}\left[\sum_{t=\tilde L_0+1}^{\tau-1}\left|\left|\lambda^0(\hat\theta)-\lambda_t\right|\right|_2^2|\mathcal{A}\right]\\
&\leq 2\bar vk\left(\kappa^2M_7^2\epsilon^2+K_{10}^2\epsilon^4\right)+\bar v\mathbb{E}^\pi_{\lambda^*}\left[\sum_{t=\tilde L_0+1}^{\tau-1}\left|\left|\lambda^0(\hat\theta)-\lambda_t\right|\right|_2^2|\mathcal{A}\right]\\
&\leq 2\bar vK_{12}\log k+(2\bar vK_{12}+2\bar vK_{10}^2+2\bar v\kappa^2M_7^2)\epsilon^2k
\end{aligned}$$

where the last inequality follows because:

$$
\begin{aligned}
&\frac{1}{2}\mathbb{E}^{\pi}_{\lambda^*}\left[\sum_{t=\tilde{L}_0+1}^{\tau-1}\left|\left|\lambda^0(\hat{\theta})-\lambda_t\right|\right|_2^2|\mathcal{A}\right]\le \mathbb{E}^{\pi}_{\lambda^*}\left[\sum_{t=\tilde{L}_0+1}^{\tau-1}\left|\left|\lambda^0(\hat{\theta})-\hat{\lambda}_t\right|\right|_2^2|\mathcal{A}\right]+\mathbb{E}^{\pi}_{\lambda^*}\left[\sum_{t=\tilde{L}_0+1}^{\tau-1}\left|\left|\hat{\lambda}_t-\lambda_t\right|\right|_2^2|\mathcal{A}\right]\\
&\le \mathbb{E}^{\pi}_{\lambda^*}\left[\sum_{t=\tilde{L}_0+1}^{\tau-1}\left|\left|\sum_{s=\tilde{L}_0+1}^{t-1}\frac{\tilde{\Delta}_s}{k-s}\right|\right|^2|\mathcal{A}\right]+\mathbb{E}^{\pi}_{\lambda^*}\left[\sum_{t=\tilde{L}_0+1}^{\tau-1}\left|\left|\hat{\lambda}_t-\lambda_t)\right|\right|_2^2|\mathcal{A}\right]\\
&\le 2\,\mathbb{E}^{\pi}_{\lambda^*}\left[\sum_{t=\tilde{L}_0+1}^{\tau-1}\left|\left|\sum_{s=\tilde{L}_0+1}^{t-1}\frac{\Delta_s}{k-s}\right|\right|^2|\mathcal{A}\right]+2\,\mathbb{E}^{\pi}_{\lambda^*}\left[\sum_{t=\tilde{L}_0+1}^{\tau-1}\left(\sum_{s=\tilde{L}_0+1}^{t-1}\frac{||\lambda_s-\hat{\lambda}_s||_2}{k-s}\right)^2|\mathcal{A}\right]\\
&\qquad+\mathbb{E}^{\pi}_{\lambda^*}\left[\sum_{t=\tilde{L}_0+1}^{\tau-1}\left|\left|\hat{\lambda}_t-\lambda_t)\right|\right|_2^2|\mathcal{A}\right]\\
&\le \frac{2}{\mathbb{P}^{\pi}_{\lambda^*}(\mathcal{A})}\mathbb{E}^{\pi}_{\lambda^*}\left[\sum_{t=1}^{\tau-1}\left|\left|\sum_{s=1}^{t-1}\frac{\Delta_s}{k-s}\right|\right|^2\right]+2\,\mathbb{E}^{\pi}_{\lambda^*}\left[\sum_{t=1}^{\tau-1}\left(\sum_{s=1}^{t-1}\frac{\omega_0\epsilon}{k-s}\right)^2|\mathcal{A}\right]+\omega_0^2k\epsilon^2\\
&\le 32\log k+6\omega_0^2\epsilon^2k+\omega_0^2\epsilon^2k\le K_{12}(\log k+\epsilon^2k)=K_{12}(\log k+\epsilon^2k)
\end{aligned}\tag{EC.26}
$$

where $K_{12}=32+7\omega_0^2$. Combine Lemma EC.11, (EC.23), (EC.24) and (EC.26), we have that there exists some constant $M_8>0$ independent of $k>K$ such that for all $k>K$,

$$
\sum_{t=\tilde{L}_0+1}^{k} r^*(\lambda^D(\theta^*))-\mathbb{E}^{\pi}_{\lambda^*}\left[\hat{R}^{\pi}_{\lambda^*}(k)\right]\le M_8\left(\epsilon(\tilde{L}_0)^2k+\epsilon(\tilde{L}_0)^{-1}\log k+\epsilon(\tilde{L}_0)^{-2}\right)
$$

Putting things together, we then conclude that there exists some constant M_9 independent of $k>K$ such that for all $k>K$:

$$
\begin{aligned}
\rho^{\pi}(k)&\le M_8\left[k\left(\frac{\log^2 k}{\tilde{L}_0}\right)^{\frac{(s\wedge\bar{s})-1}{s+n}}+\log k\left(\frac{\sqrt{\tilde{L}_0}}{\log k}\right)^{\frac{(s\wedge\bar{s})-1}{s+n}}+\left(\frac{k}{\log^2 k}\right)^{\frac{(s\wedge\bar{s})-1}{2s+n-1}}\right]+\bar{r}\tilde{L}_0\\
&\le M_8\left(k^{\frac{2s-(s\wedge\bar{s})+n}{2s+n-1}}(\log k)^{\frac{2((s\wedge\bar{s})-1)}{2s+n-1}}+k^{\frac{(s\wedge\bar{s})-1}{2(2s+n-1)}}(\log k)^{\frac{2s-(s\wedge\bar{s})+n}{2s+n-1}}+k^{\frac{(s\wedge\bar{s})-1}{2s+n-1}}\right)+\bar{r}k^{\frac{s+n}{2s+n-1}}(\log k)^{\frac{2(s-1)}{2s+n-1}}\\
&\le M_9k^{\frac{2s-(s\wedge\bar{s})+n}{2s+n-1}}\log k
\end{aligned}
$$

where the last inequality holds by letting $\tilde{L}_0=k^{(s+n)/(2s+n-1)}(\log k)^{2(s-1)/(2s+n-1)}$. The result then follows by letting $M_3=M_9+\bar{r}K$. □

EC.3.6. Proof of supporting lemmas

Proof of Lemma EC.10. Let $\Omega_5=\max\{\Omega_4,C_1,C_2\}$ where C_1 and C_2 are constants to be defined later. Assume without loss of generality that $T=1$. Recall that $p^D=\arg\max_{p\in\mathcal{P}}\{r^*(p):A\lambda^*(p)\preceq C\}=\arg\max_{p\in\mathcal{P}}\{r(p;\theta^*):A\lambda(p;\theta^*)\preceq C\}$ and $p^D(\hat{\theta})=\arg\max_{p\in\mathcal{P}}\{r(p;\hat{\theta}):A\lambda(p;\hat{\theta})\preceq C\}$. Note that for $k\ge\Omega_5\ge\Omega_4$, we have $||\theta^*-\hat{\theta}||_2<\bar{\delta}$. Hence, by Lemma 1(a), we conclude that

$||p^D - p^D(\hat\theta)||_2 < \phi/2$. Recall that $p^0(\hat\theta) = \arg\max_{p\in\mathcal{P}}\{r(p;\hat\theta) : A\lambda(p;\hat\theta) \preceq C_{\tilde L_0}/(k-\tilde L_0)\}$. Since $C - C_{\tilde L_0}/(k-\tilde L_0) = (\tilde L_0 C - AS_{\tilde L_0})/(k-\tilde L_0) \to 0$ as $k\to\infty$, by a similar argument as in Lemma EC.12, there exists $C_1 > 0$ independent of $\hat\theta \in \mathsf{Ball}(\theta^*, \bar\delta)$ such that for all $k \geq C_1$, $||p^D(\hat\theta) - p^0(\hat\theta)||_2 < \phi/8$. Since $\omega||Q||_2\psi = \omega||Q||_2\sqrt{\epsilon(\tilde L_0)} \to 0$ as $k\to\infty$, there exists a constant $C_2 > C_1$ such that for all $k \geq C_2$, $\omega||Q||_2\psi \leq \phi/4$. The rest of the proof goes by induction. Fix some $k \geq \Omega_5$. If $\tau \leq \tilde L_0 + 1$, there is nothing to prove. Suppose $\tau > \tilde L_0 + 1$. $p_{\tilde L_0+1} = p^0(\hat\theta) \in \mathsf{Ball}(p^D, 5\phi/8) \subseteq \mathcal{P}$. Following the same argument in the proof of Lemma EC.1, we have $C_{\tilde L_0+1} \succ 0$. This is our induction base. Suppose that $C_s \succ 0, p_s \in \mathcal{P}$ for all $s = \tilde L_0+1, \tilde L_0+2, \ldots, t-1$ and $t-1 < \tau$. If $t \geq \tau$, we have finished the induction. Otherwise, $||p_t - p^0(\hat\theta))||_2 = ||\nabla p(\lambda^0(\hat\theta);\hat\theta)\cdot\sum_{s=\tilde L_0+1}^{t-1}\frac{Q\tilde\Delta_s}{k-s}||_2 \leq \omega||Q||_2||\sum_{s=\tilde L_0+1}^{t-1}\frac{\tilde\Delta_s}{k-s}||_2 \leq \omega||Q||_2\psi \leq \phi/4$ where the last inequality follows as $k \geq \Omega_5 \geq C_2$. So $p_t \in \mathsf{Ball}(p^D, 7\phi/8) \subseteq \mathcal{P}$. $C_t \succ 0$ can be show in the same way as in the proof of Lemma EC.1. This completes the induction. □

Proof of Lemma EC.11. Assume without loss of generality that $T = 1$ and fix $k \geq \max\{\Omega_3, \Omega_4, \Omega_5\}$. Similar to (EC.15) in the proof of Lemma EC.8, we have

$$\begin{aligned}\mathbb{E}^\pi_{\lambda^*}[k-\tau|\mathcal{A}] = \sum_{t=1}^{k-1}\mathbb{P}^\pi_{\lambda^*}(\tau\leq t|\mathcal{A}) &\leq \sum_{t=1}^{k-1}\mathbb{P}^\pi_{\lambda^*}\left(\max_{\tilde L_0+1\leq s\leq t}\left|\left|\sum_{v=\tilde L_0+1}^{s}\frac{\Delta_v}{k-v}\right|\right|_2 \geq \frac{\psi}{2}\middle|\mathcal{A}\right)\\ &+\sum_{t=1}^{k-1}\mathbb{P}^\pi_{\lambda^*}\left(\max_{\tilde L_0+1\leq s\leq t}\left\{\sum_{v=\tilde L_0+1}^{s}\frac{||\lambda_v-\hat\lambda_v||_2\mathbf{1}_{\{v\leq\tau\}}}{k-v}\right\}\geq\frac{\psi}{2}\middle|\mathcal{A}\right).\end{aligned}\tag{EC.27}$$

By the argument in the proof of Lemma EC.8, the first term in (EC.27) can be bounded by $\psi^{-2}C_1\log k = C_1\epsilon(\tilde L_0)^{-1}\log k$ for some constant C_1 independent of $k \geq \max\{\Omega_3,\Omega_4,\Omega_5\}$ (note that $k \geq \Omega_3 = 17 > 3$). We now bound the second term in (EC.27):

$$\begin{aligned}&\mathbb{P}^\pi_{\lambda^*}\left(\max_{\tilde L_0+1\leq s\leq t}\left\{\sum_{v=\tilde L_0+1}^{s}\frac{||\lambda_v-\hat\lambda_v||_2\mathbf{1}_{\{v\leq\tau\}}}{k-v}\right\}\geq\frac{\psi}{2}\middle|\mathcal{A}\right) = \mathbb{P}^\pi_{\lambda^*}\left(\sum_{s=\tilde L_0+1}^{t}\frac{||\lambda_s-\hat\lambda_s||_2\mathbf{1}_{\{s\leq\tau\}}}{k-s}\geq\frac{\psi}{2}\middle|\mathcal{A}\right)\\ &\leq \frac{16}{\psi^4}\mathbb{E}^\pi_{\lambda^*}\left[\left(\sum_{s=\tilde L_0+1}^{t}\frac{||\lambda_s-\hat\lambda_s||_2\mathbf{1}_{\{s<\tau\}}}{k-s}+\sum_{s=\tilde L_0+1}^{t}\frac{||\lambda_s-\hat\lambda_s||_2\mathbf{1}_{\{s=\tau\}}}{k-s}\right)^4\middle|\mathcal{A}\right]\\ &\leq \frac{128}{\psi^4}\mathbb{E}^\pi_{\lambda^*}\left[\left(\sum_{s=\tilde L_0+1}^{t}\frac{||\lambda_s-\hat\lambda_s||_2\mathbf{1}_{\{s<\tau\}}}{k-s}\right)^4\middle|\mathcal{A}\right]+\frac{128}{\psi^4}\mathbb{E}^\pi_{\lambda^*}\left[\left(\sum_{s=\tilde L_0+1}^{t}\frac{||\lambda_s-\hat\lambda_s||_2\mathbf{1}_{\{s=\tau\}}}{k-s}\right)^4\middle|\mathcal{A}\right]\\ &\leq \frac{128\omega_0^4\epsilon(\tilde L_0)^4}{\psi^4}\log^4\left(\frac{k}{k-t}\right)+\frac{128}{\psi^4}\left(\frac{\sqrt2}{k-t}\right)^4 \leq \frac{128\omega_0^4\epsilon(\tilde L_0)^4}{\psi^4}\log^4\left(\frac{k}{k-t}\right)+\frac{512}{\psi^4}\left(\frac{1}{k-t}\right)^4\\ &= 128\omega_0^4\epsilon(\tilde L_0)^2\log^4\left(\frac{k}{k-t}\right)+512\epsilon(\tilde L_0)^{-2}\left(\frac{1}{k-t}\right)^4\end{aligned}\tag{EC.28}$$

where the third inequality follows by (EC.25) and the fact that $||\lambda_t - \hat{\lambda}_t||_2^2 \le 2$, and the last equality follows since $\psi = \sqrt{\epsilon(\tilde{L}_0)}$. Note that $\sum_{t=1}^{k-1} \log^s(\frac{k}{k-t}) \le s!k$ for some constant M_s only depending on s, there exists some constant C_2 independent of $k \ge \max\{\Omega_3, \Omega_4, \Omega_5\}$ such that:

$$\sum_{t=1}^{k-1} \mathbb{P}_{\lambda^*}^{\pi}\left(\max_{\tilde{L}_0+1\le s\le t}\left\{\sum_{v=\tilde{L}_0+1}^{s} \frac{||\lambda_v - \hat{\lambda}_v||_2 \mathbf{1}_{\{v\le\tau\}}}{k-v}\right\} \ge \frac{\psi}{2}\middle| \mathcal{A}\right)$$
$$\le \sum_{t=1}^{k-1} 128\omega_0^4 \epsilon(\tilde{L}_0)^2 \log^4\left(\frac{k}{k-t}\right) + 512\epsilon(\tilde{L}_0)^{-2}\sum_{t=1}^{k-1}\left(\frac{1}{k-t}\right)^4 \le C_2\left(\epsilon(\tilde{L}_0)^2 k + \epsilon(\tilde{L}_0)^{-2}\right)$$

Therefore, we have that for all $k \ge \max\{\Omega_3, \Omega_4, \Omega_5\}$, $\mathbb{E}[k-\tau|\mathcal{A}] \le C_1\epsilon(\tilde{L}_0)^{-1}\log k + C_2\epsilon(\tilde{L}_0)^2 k + C_2\epsilon(\tilde{L}_0)^{-2} \le K_9(\epsilon(\tilde{L}_0)^2 k + \epsilon(\tilde{L}_0)^{-1}\log k + \epsilon(\tilde{L}_0)^{-2})$ for some constant K_9 independent of $k \ge \max\{\Omega_3, \Omega_4, \Omega_5\}$. □

Proof of Lemma EC.12. Without loss of generality, assume $T = 1$. Define $\hat{\zeta} = C - kC_{\tilde{L}_0}/(k - \tilde{L}_0)$. Note that $\tilde{L}_0/k \to 0$ as $k \to \infty$. Hence, there exists some constant $\Omega_6 \ge \Omega_4$ such that for all $k \ge \Omega_6$,

$$||\hat{\zeta}||_\infty = \left|\left|C - \frac{C_{\tilde{L}_0}}{k-\tilde{L}_0}\right|\right|_\infty = \left|\left|\frac{(kC - \tilde{L}_0 C) - (kC - AS_{\tilde{L}_0})}{k-\tilde{L}_0}\right|\right|_\infty = \left|\left|\frac{AS_{\tilde{L}_0} - \tilde{L}_0 C}{k-\tilde{L}_0}\right|\right|_\infty$$
$$\le 2\left(||A\mathbf{e}||_\infty + ||C||_\infty\right)\frac{\tilde{L}_0}{k} = 2\left(||A\mathbf{e}||_\infty + ||C||_\infty\right)\left(\frac{\log^2 k}{k}\right)^{\frac{s-1}{2s+n-1}}$$
$$\le 2\left(||A\mathbf{e}||_\infty + ||C||_\infty\right)\epsilon(\tilde{L}_0)^2$$

where the first inequality follows since $k - \tilde{L}_0 > k/2$ for large k. Note that $\lambda^0(\hat{\theta}) = \arg\max_{\lambda\in\Lambda_{\hat{\theta}}}\{r(\lambda;\hat{\theta}) : A\lambda \preceq C_{\tilde{L}_0}/(k-\tilde{L}_0)\}$ and $\lambda^D(\hat{\theta}) = \arg\max_{\lambda\in\Lambda_{\hat{\theta}}}\{r(\lambda;\hat{\theta}) : A\lambda \preceq C\}$. Hence, by Corollary EC.4, there exists a constant $M_{\hat{\theta}}$ independent of $\hat{\zeta}$ but dependent on $\hat{\theta}$ such that $||\lambda^0(\hat{\theta}) - \lambda^D(\hat{\theta})||_\infty \le M_{\hat{\theta}}||\hat{\zeta}||_\infty = 2(||A\mathbf{e}||_\infty + ||C||_\infty)M_{\hat{\theta}}\epsilon(\tilde{L}_0)^2$. Note that, as we will show below, both $\lambda^0(\theta)$ and $\lambda^D(\theta)$ are continuous in θ. This indicates that M_θ can be chosen to be continuous in θ for all $\theta \in \Theta$. The result is then proven by letting $K_{10} = 2(||A\mathbf{e}||_\infty + ||C||_\infty)\sup_{\theta\in\Theta} M_\theta < \infty$. We now prove the continuity of $\lambda^0(\theta)$ and $\lambda^D(\theta)$ below.

Recall that by Lemma 7 and the argument in Lemma 1, $p^D(\theta)$ is continuous in θ. Since $p^0(\theta) = \arg\max_{p\in\mathcal{P}}\{r(p;\theta) : A\lambda(p;\theta) \preceq C + \hat{\zeta}\}$, by a similar argument we have that $p^0(\theta)$ is also continuous in θ for all $\hat{\zeta}$. Note that $\lambda^0(\theta) = \lambda(p^0(\theta);\theta)$ and $\lambda^D(\theta) = \lambda(p^D(\theta);\theta)$ and $\lambda(p;\theta)$ is continuous in both p and θ. Therefore, both $\lambda^0(\theta)$ and $\lambda^D(\theta)$ are continuous in θ. □

EC.4. Auxiliary Results

EC.4.1. Results for Maximum Likelihood Theory

THEOREM EC.1. (TAIL INEQUALITY FOR MLE BASED ON IID SAMPLES, THEOREM 36.3 IN BOROVKOV (1999)) *Let $\Theta \in \mathbb{R}^q$ be compact and convex, and let $\{\mathbb{P}^\theta : \theta \subseteq \Theta\}$ be a family of dis-*

tributions on a discrete sample space $\mathcal{Y}$. Suppose Y is a random variable taking value in $\mathcal{Y}$ with distribution $\mathbb{P}^{\theta}$, and the following conditions hold:

(i) $\mathbb{P}^{\theta} \neq \mathbb{P}^{\theta'}$ whenever $\theta \neq \theta'$;

(ii) For some $r > q$, $\sup_{\theta\in\Theta} \mathbb{E}_{\theta}[||\nabla_{\theta} \log \mathbb{P}^{\theta}(Y)||_2^r] = \gamma < \infty$;

(iii) The function $\theta \to \sqrt{\mathbb{P}^{\theta}(Y)}$ is differentiable on Θ for any $Y \in \mathcal{Y}$;

(iv) The Fisher information matrix, whose $(i,j)^{th}$ entry is given by $\mathbb{E}_{\theta}\left[-\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log \mathbb{P}^{\theta}(Y)\right]$, is positive definite.

If $Y_1, Y_2, \ldots$ is a sequence of i.i.d. random variables taking value in $\mathcal{Y}$ with distribution $\mathbb{P}^{\theta}$, and $\hat{\theta}(t) = \arg\max_{\theta\in\Theta} \prod_{l=1}^{t} \mathbb{P}^{\theta}(Y_l)$ is the maximum likelihood estimate based on t i.i.d. samples, then, there exist constants $\eta_1 > 0$ and $\eta_2 > 0$ depending only on r, q, $\mathbb{P}^{\theta}$ and Θ such that for all $t \geq 1$ and all $\delta \geq 0, \mathbb{P}^{\theta}(||\hat{\theta}(t) - \theta||_2 > \delta) \leq \eta_1 \exp(-t\eta_2\delta^2)$.

EC.4.2. Results for Newton's Method

THEOREM EC.2. (QUADRATIC CONVERGENCE OF NEWTON'S METHOD FOR CONVEX UNCONSTRAINED OPTIMIZATION PROBLEMS, SECTION 9.5.3 IN BOYD AND VANDENBERGHE (2004)) *Suppose $g(z)$ is a concave function whose unconstrained optimizer is x^*. Let $\{x^{(k)}\}_{k=1}^{\infty}$ be a sequence of points obtained by Newton's method. Assume there exist positive constants m, M, L such that*

(i) $||\nabla^2 g(z) - \nabla^2 g(y)||_2 \leq L||z-y||_2$, and

(ii) $-MI \preceq \nabla^2 g(z) \preceq -mI$.

Then, there exists constant $\eta = \min\{1, 3(1-2\alpha)\}m^2/L$ where $\alpha \in (0, 0.5)$ such that if $||\nabla g(x^{(k)})||_2 < \eta$, then $||\nabla g(x^{(k+1)})||_2 \leq \frac{L}{2m^2}||\nabla g(x^{(k)})||_2^2$.

EC.4.3. Results for Spline Approximation

THEOREM EC.3. (THEOREM 6.18 AND THEOREM 6.22 IN SCHUMAKER (2007)) *Let $\mathsf{B}([\underline{p}_l, \bar{p}_l])$ be the set of bounded functions on $[\underline{p}_l, \bar{p}_l]$. Then for $l = 1, \ldots, n, \mathcal{L}_l$ is a linear operator mapping $\mathsf{B}([\underline{p}_l, \bar{p}_l])$ into $\mathsf{S}_l(\mathcal{G}_l, s)$. Moreover, $\mathcal{L}_l f = f$ for all $f \in \mathsf{P}^{s-1}([\underline{p}_l, \bar{p}_l])$. In addition, for every $g \in \mathsf{C}^0([\underline{p}_l, \bar{p}_l])$, $||(\mathcal{L}_l g)(.)||_{\infty} \leq (2s)^s ||g(.)||_{\infty}$.*

THEOREM EC.4. (THEOREM 13.20 IN SCHUMAKER (2007)) *Let Λ be a complete set of multiple-indices and let $0 < \epsilon < 1$. Then there exists a constant C depending only on $n, \epsilon, \psi, \Lambda$ such that for all $f \in \mathsf{L}_p^{\partial\Lambda}(\Omega)$,*

$$||D^{\beta}(f - T_{\psi}^{\Lambda} f)||_q \leq C\delta^{1/q-1/p} \sum_{\alpha\in\partial(\Lambda-\beta)} \delta^{\alpha} ||D^{\alpha} D^{\beta} f||_p$$

for any β *and for all* $1 \le q \le p \le \infty$ *satisfying* $\epsilon \le \max\{\lfloor |\alpha|/d \rfloor, 1/q - 1/p + |\alpha|/d, \min\{1-1/p, 1/q\}\}$ *for all* $\alpha \in \partial(\Lambda - \beta)$.

(Note: For definition of complete multiple-indices Λ *and its* boundary $\partial\Lambda$, *see Definition 13.15 on pg. 510 of Schumaker (2007). For the* Sobolev Space $\mathsf{L}_p^{\Lambda}(\Omega)$, *see Definition 13.3 on pg.504 of Schumaker (2007). See (13.34) for the definition of the* tensor Taylor expansion $T_{\psi}^{\Lambda} f$, *and (13.48) for the definition of* δ. *Finally, see (13.9) - (13.12) for definitions of* $|\alpha|, \delta^{\alpha}$, *and* D^{α}. *)*

Corollary EC.1. *Let* $f : \mathbb{R}^n \to [0,1]^n$ *be a function that satisfies N1-N2. Then for any* $s \ge 3$, *there exists* $g \in \otimes_{l=1}^{n} \mathsf{P}^{(s\wedge\bar{s})-1}([\underline{p}_l, \bar{p}_l])$ *such that:*

$$||(\nabla_{x_1^{\beta_1},\dots,x_n^{\beta_n}})(f-g)(.)||_\infty \le C \sum_{i=1}^{n} \bar{\Delta}_i^{(s\wedge\bar{s})-\beta_i} \left|\left| \frac{\partial^{(s\wedge\bar{s})}}{\partial p_i^{(s\wedge\bar{s})}} f(.) \right|\right|_\infty$$

for $0 = \beta_1 = \cdots = \beta_n$ *and* $1 = \beta_1 = \cdots = \beta_n$.

Proof. This result follows by Theorem EC.4. Let $\Lambda = \{\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{Z}_+^n : 0 \le \alpha_i \le (s \wedge \bar{s}) - 1, 1 \le i \le n\}$ for $s \ge 3$ and $\Omega = H_{\tilde{i}_1,\dots,\tilde{i}_n}$. Also, let $p = \infty, q = \infty$. Note that for all j, $s \wedge \bar{s} - \beta_j \ge 1$ since $\bar{s} \ge 2$ by N1 and $s \ge 3$ and $\beta_j \le 1$ for all j. This ensures that there exists some $\epsilon \in (0,1)$ such that $\epsilon \le \max\{\lfloor |\alpha|/n \rfloor, 1/q - 1/p + |\alpha|/n, \min\{-1/p, 1/q\}\} = \max_{1\le j \le n}\{((s\wedge\bar{s}) - \beta_j)/n\}$ for all $\alpha \in \partial\Lambda = \{(s \wedge \bar{s} - \beta_j)\mathbf{e}_j : 1 \le j \le n\}$. Since Ω is a compact set and N2 ensures that for f, all its derivatives of order $(\bar{s}, \dots, \bar{s})$ or lower are uniformly bounded, we conclude that $f \in \mathsf{L}_\infty^{(s\wedge\bar{s},\dots,s\wedge\bar{s})}(H_{\tilde{i}_1,\dots,\tilde{i}_n})$. The result follows by letting $g = T_{\psi}^{\Lambda} f \in \otimes_{l=1}^{n} \mathsf{P}^{(s\wedge\bar{s})-1}([\underline{p}_l, \bar{p}_l])$. $\square$

Theorem EC.5. (Lemma 6.19 in Schumaker (2007)) *For* $\{\beta_{l,i,j}\}_{l=1,i=1,j=1}^{n,s+d,s}$, *we have* $|\beta_{l,i,j}| \le (y_{l,i+s} - y_{l,i})^{j-1} \le (s\bar{\Delta}_l)^{j-1}$.

Theorem EC.6. (Theorem 12.4 in Schumaker (2007)) *Let* $Y_{i_1,\dots,i_n} = \otimes_{l=1}^{n}(y_{l,i_l}, y_{l,i_l+s})$ *for all* $1 \le i_1, \dots, i_n \le s+d$. *Then,* $N_{i_1,\dots,i_d}(p) > 0$ *for* $p = (p_1, \dots, p_n) \in Y_{i_1,\dots,i_n}$, $N_{i_1,\dots,i_d}(p) = 0$ *for* $p = (p_1, \dots, p_n) \notin Y_{i_1,\dots,i_n}$, *and* $\sum_{v_1=i_1+s-1}^{i_1} \cdots \sum_{v_n=i_n+s-1}^{i_n} N_{v_1,\dots,v_n}(p) \equiv 1$ *for* $p \in Y_{i_1,\dots,i_n}$.

Corollary EC.2. *We have that* $\sum_{i_1=1}^{s+d} \cdots \sum_{i_n=1}^{s+d} |N_{i_1,\dots,i_n}(p)| = 1$ *for all* $p \in \mathcal{P}$.

Proof. Let $Y_{i_1,\dots,i_n} = \otimes_{l=1}^{n}(y_{l,i_l}, y_{l,i_l+s})$ for all $1 \le i_1, \dots, i_n \le s+d$. By Theorem EC.6, $\sum_{v_1=i_1+1-s}^{i_1} \cdots \sum_{v_n=i_n+1-s}^{i_n} N_{v_1,\dots,v_n}(p) \equiv 1$ for $p \in Y_{i_1,\dots,i_n}$. Since in addition, we also have by the same Theorem that $N_{i_1,\dots,i_n}(p) > 0$ for $p = (p_1, \dots, p_n) \in Y_{i_1,\dots,i_n}$, and $N_{i_1,\dots,i_n}(p) = 0$ for $p = (p_1, \dots, p_n) \notin Y_{i_1,\dots,i_n}$. We thus conclude that $\sum_{i_1=1}^{s+d} \cdots \sum_{i_n=1}^{s+d} |N_{i_1,\dots,i_n}(p)| = \sum_{i_1=1}^{s+d} \cdots \sum_{i_n=1}^{s+d} N_{i_1,\dots,i_n}(p) = \sum_{v_1=i_1+1-s}^{i_1} \cdots \sum_{v_n=i_n+1-s}^{i_n} N_{v_1,\dots,v_n}(p) \equiv 1$ for $p \in Y_{i_1,\dots,i_n}$ for all $1 \le i_1, \dots, i_n \le s+d$. The result then follows. $\square$

Theorem EC.7. (Theorem 4.22 in Schumaker (2007)) *Fix* $l = 1,\dots,n$*, and let* $s \geq 2$*. Suppose* k *and* p_l *are such that* $y_{l,k} \leq p_l < y_{l,k+1}$*, and define* $\Delta_{i_l,k,j} = \min\{(y_{l,v+j}, y_{l,v}) : y_{l,i_l} \leq y_{l,v} \leq y_{l,k} < y_{l,k+1} \leq y_{l,v+j} \leq y_{l,i_l+s}\}$*, for* $j = 1,\dots,s$*. Suppose* $\sigma > 0$ *and that* $\Delta_{i_l,k,s-\sigma+1} > 0$*. Then* $|D^\sigma N^s_{l,i_l}(p_l)| \leq \Gamma_{s,\sigma}/(\prod_{q=1}^{\sigma} \Delta_{i_l,l,s-q})$ *where* $\Gamma_{s,\sigma} = \frac{(s-1)!}{(s-\sigma-1)!}\begin{pmatrix}\sigma\\ \lfloor \sigma/2 \rfloor\end{pmatrix}$. [5]

Corollary EC.3. $||\nabla N_{i_1,\dots,i_n}(p)||_\infty \leq (s-1)\max_{l=1}^n\{\bar{\Delta}_l^{-1}\}$ *for all* $p \in \mathcal{P}$*.*

Proof. The result is a direct corollary of Theorem EC.7. Let $\sigma = 1$, and then we have that for $p_l \in [y_{l,k}, y_{l,k+1})$, $|\nabla_{p_l} N^s_{l,i_l}(p_l)| \leq (s-1)/\Delta_{i_l,k,s-1} \leq (s-1)/\bar{\Delta}_l$, where the last inequality follows since $\Delta_{i_l,k,s-1} \geq \bar{\Delta}_l$. Since $(s-1)/\bar{\Delta}_l$ does not depend on k, we conclude that $|\nabla_{p_l} N^s_{l,i_l}(p_l)| \leq (s-1)/\bar{\Delta}_l$ for all $p_l \in [\underline{p}_l, \bar{p}_l]$. Hence, $||\nabla N_{i_1,\dots,i_n}(p)||_\infty = \max_{l=1}^n |\nabla_{i_l} N^s_{l,i_l}(p)| \leq (s-1)\max_{l=1}^n\{\bar{\Delta}_l^{-1}\}$. □

EC.4.4. Results for Stability Analysis of Optimization Problems

Consider a family of parameterized nonlinear programs as follows:

$$
\begin{aligned}
(P_u) \qquad & \min\nolimits_{x\in\mathbb{R}^n} \; f(x,u) \\
& \quad s.t. \quad g_i(x,u) = 0, i = 1,\dots,k, \\
& \qquad\qquad g_i(x,u) \leq 0, i = k+1,\dots,p,
\end{aligned}
$$

with $u \in \mathcal{U} \subseteq \mathbb{R}^q$ being the parameter vector. When $u = u_0$, the above problem P_{u_0} is called the unperturbed problem. The Lagrangian function associated with P_u is $L(x,\mu,u) := f(x,u) + \sum_{i=1}^p \mu_i g_i(x,u)$, and they denote by $M(x,u)$ the set of Lagrange multipliers at a point x for (P_u). They denote by $I(x,u)$ the set of inequality constraints active at x. Let $d \in \mathbb{R}^q$ and define $\bar{u}(t) := u_0 + td$.

Definition EC.1. (Definition 3.2 in Bonnans and Shapiro (2000)) For $\epsilon \geq 0, u \in \mathcal{U}$, we say that $\bar{x}(u)$ is an ϵ-optimal solution of (P_u) if $\bar{x}(u)$ is feasible and $f(\bar{x},u) \leq \inf_{x\in\mathbb{R}^n} f(x,u) + \epsilon$.

Definition EC.2. (Gollan's Condition, (5.111) in Bonnans and Shapiro (2000)) We say that Gollan's condition holds in direction $d \in \mathbb{R}^q$ if the following holds:

(a) $\nabla_x g_i(x_0,u_0), i = 1,\dots,k$, are linearly independent,

(b) $\exists h \in \mathbb{R}^n$ such that $\nabla g_i(x_0,u_0)\cdot(h,d) = 0, i = 1,\dots,k$, and $\nabla g_i(x_0,u_0)\cdot(h,d) < 0, i \in I(x_0,u_0)$.

The following is a stronger condition of the *Strong Second Order Sufficient Optimality Condition* in (5.120) in Bonnans and Shapiro (2000). In other words, if the following holds, then Strong Second Order Sufficient Optimality Condition holds automatically.

Definition EC.3. We say that a *stronger* strong second order sufficient optimality condition holds in a direction d if $\sup_{\mu\in\mathbb{R}^q} h'\nabla^2_{xx}L(x_0,\mu,u_0)h > 0, \forall h \in \mathbb{R}^n\backslash\{0\}$.

We now state their main sensitivity result for parameterized nonlinear programs.

Theorem EC.8. (Theorem 5.53(a) in Bonnans and Shapiro (2000)) *Suppose that:*

(i) the unperturbed problem (P_{u_0}) has unique optimal solution x_0,

(ii) Gollan's condition holds in the direction d,

(iii) the set $M(x_0,u_0)$ of Lagrange multipliers is nonempty,

(iv) the strong second order sufficient conditions are satisfied,

(v) for all $t>0$ small enough the feasible set of ($P_{u(t)}$) is nonempty and uniformly bounded.

Then for any $o(t^2)$-optimal solution $\bar{x}(t)$ of ($P_{u(t)}$), where $t\geq 0$, $\bar{x}(t)$ is Lipschitz stable at x_0, i.e., $||\bar{x}(t)-x_0|| = \mathcal{O}(t)$.

Corollary EC.4. *Consider $P_\lambda(\epsilon)$: $J^D := \max_{\lambda\in\Lambda}\{r^*(\lambda) : s.t., A\lambda \preceq C/T-\epsilon\}$. Denote by $x^*(\epsilon)$ the optimal solution to $P_\lambda(\epsilon)$. Then, $||x^*(0)-x^*(\epsilon)||_\infty = K_{13}||\epsilon||_\infty$ for some positive K_{13} independent of $||\epsilon||_\infty$.*

Proof. We now verify the conditions (i)-(v) for $\mathrm{P}_\lambda(\epsilon)$. For the unperturbed problem $\mathrm{P}_\lambda(0)$, by strict concavity assumption, we conclude that it has a unique optimal solution $x^*(0)$ and thus (i) holds. For (ii), note that we don't have equality constraints, so we only need to verify the second part of (b) in Definition EC.2, which immediately follows because the derivative of those constraints are a subset of the rows of A which are linearly independent. Note also that the constraints do not depend on ϵ. So what we have showed is that Gollan's condition holds for all direction d. (ii) holds. By duality theory of convex optimization, there exists Lagrange multipliers $\mu^*(0)$, so (iii) holds. Note that $\nabla_{\lambda\lambda}L(x^*(0),\mu^*(0),\epsilon) = \nabla_{\lambda\lambda}r^*(x^*(0))$ is negative definite by the strict concavity assumption of the revenue function. Note that our problem is a maximization problem whereas Theorem EC.8 is for minimization problems, so (iv) holds. Because the feasible set of $\mathrm{P}_\lambda(\epsilon)$ is nonempty and uniformly bounded, and the feasible set doesn't depend on ϵ, so (v) holds. The optimal solution of $\mathrm{P}_\lambda(\epsilon)$ is definitely $o(t^2)$-optimal to $\mathrm{P}_\lambda(\epsilon)$. Hence, Lipschitz continuity holds for the optimal solution. □