A. Additional Theoretical Results and Examples

A.1. Different Cases for $\arg\max_{\gamma\in[0,1]} \mathrm{OPT}(\mathcal{P},\gamma)$

Take the case of $T=1$, $\mathcal{Z}=\{[0,1]\}$, $\mathcal{W}=\{w_1,w_2\}$ with equal probability of occurring, $b=1$, and $\alpha=0.5$. Call $\Pi(\cdot\in A)$ to the function that takes the value of 0 if condition A holds and $-\infty$ otherwise. We show examples in which $\arg\max_{\gamma\in[0,1]} \mathrm{OPT}(\mathcal{P},\gamma)$ match the different cases mentioned in the paper. In most of the examples below the upper bound cost constraint hold trivially, reason why we do not "enforce" it using $\Pi(\cdot\leq 1)$, with the only exception on the $\gamma=\frac{1}{2}$ example.

Infinite solutions. $f(z;\theta^*,w_1)=z$, $c(z;\theta^*,w_1)=z$, $f(z;\theta^*,w_2)=z$, $c(z;\theta^*,w_2)=z$. In this case $\mathbb{E}[f(z;\theta^*,w)]=z$ and $\mathbb{E}[c(z;\theta^*,w)]=z$. Then, for any $\gamma\in[0,1]$ we have

$$\mathrm{OPT}(\mathcal{P},\gamma)=\frac{1}{2}\left(\max_{z\in[0,1]}\left\{z+\Pi(\tfrac{1}{2}\leq z)\right\}+\max_{z\in[0,1]}\left\{z+\Pi(\tfrac{1}{2}\leq z)\right\}\right)$$

The equality comes directly from the definition of $\mathrm{OPT}(\mathcal{P},\gamma)$. Is direct to see that $z=1$ maximizes both optimization problems and that $\mathrm{OPT}(\mathcal{P})=\mathrm{OPT}(\mathcal{P},\gamma)$ for all $\gamma\in[0,1]$.

No solution. $f(z;\theta^*,w_1)=z$, $c(z;\theta^*,w_1)=0$, $f(z;\theta^*,w_2)=0$, $c(z;\theta^*,w_2)=0$. Since the cost terms are always zero, the cost lower bound 0.5 is never achieved and no feasible solution exist.

$\gamma=\frac{1}{2}$ **as unique solution**. $f(z;\theta^*,w_1)=z$, $c(z;\theta^*,w_1)=0$, $f(z;\theta^*,w_2)=-z$, $c(z;\theta^*,w_2)=2z$. In this case $\mathbb{E}[f(z;\theta^*,w)]=0$ and $\mathbb{E}[c(z;\theta^*,w)]=z$. Then, for any $\gamma\in[0,1]$ we have

$$\begin{aligned}\mathrm{OPT}(\mathcal{P},\gamma)=&\frac{1}{2}\left(\max_{z\in[0,1]}\left\{(1-\gamma)z+\Pi(\tfrac{1}{2}\leq\gamma z)\right\}+\max_{z\in[0,1]}\left\{-(1-\gamma)z+\Pi(\tfrac{1}{2}\leq(2-\gamma)z)+\Pi((2-\gamma)z\leq 1)\right\}\right)\\ =&\frac{1}{2}\left((1-\gamma)+\Pi(\tfrac{1}{2}\leq\gamma)+\max_{z\in[0,1]}\left\{-(1-\gamma)z+\Pi(\tfrac{1}{2}\leq(2-\gamma)z)+\Pi((2-\gamma)z\leq 1)\right\}\right)\end{aligned}$$

The second equality uses that the first optimization problem has $z=1$ as its unique optimal solution whenever $\gamma\neq 1$ and that $0=\mathrm{OPT}(\mathcal{P},1)<\mathrm{OPT}(\mathcal{P},0.5)=\frac{1}{6}$. Is direct from the result above that $\mathrm{OPT}(\mathcal{P},\gamma)=-\infty$ for any $\gamma<0.5$. Then, we have:

$$\begin{aligned}\mathrm{OPT}(\mathcal{P})=&\frac{1}{2}\left(\max_{z\in[0,1],\gamma\in[0.5,1)}(1-\gamma)-(1-\gamma)z+\Pi(\tfrac{1}{2}\leq(2-\gamma)z)+\Pi((2-\gamma)z\leq 1)\right)\\ =&\frac{1}{2}\left(\max_{\gamma\in[0.5,1)}(1-\gamma)-\tfrac{1-\gamma}{2(2-\gamma)}\right)\end{aligned}$$

The first equality uses the definition of $\mathrm{OPT}(\mathcal{P})$ and that we have restricted γ to be in $[0.5,1)$. The second equality uses that for any $\gamma\in[0.5,1)$ the unique optimal is $z(\gamma)=\frac{1}{2(2-\gamma)}$ as it maximizes the term $-(1-\gamma)z$ by taking the smallest feasible z value that satisfies the cost lower bound. Finally, the function $\xi(\gamma):=(1-\gamma)-\frac{1-\gamma}{2(2-\gamma)}$ is differentiable on $\gamma\in[0.5,1]$ and has strictly negative derivative on $\gamma\in[0.5,1]$, which implies $\xi(0.5)>\xi(\gamma)$ for any $\gamma\in[0.5,1]$, proving that $\gamma=0.5$ is the unique optimal solution.

$\gamma=0$ **as unique solution**. $f(z;\theta^*,w_1)=z^2$, $c(z;\theta^*,w_1)=z$, $f(z;\theta^*,w_2)=-z$, $c(z;\theta^*,w_2)=1-z$. In this case $\mathbb{E}[f(z;\theta^*,w)]=0.5(z^2-z)$ and $\mathbb{E}[c(z;\theta^*,w)]=0.5$. Then, for any $\gamma\in[0,1]$ we have

$$\begin{aligned}\mathrm{OPT}(\mathcal{P},\gamma)=&\frac{1}{2}\Bigg(\max_{z\in[0,1]}\left\{z^2(1-\tfrac{\gamma}{2})-z\tfrac{\gamma}{2}+\Pi(\tfrac{1}{2}\leq(1-\gamma)z+\tfrac{\gamma}{2})\right\}\\ &+\max_{z\in[0,1]}\left\{\tfrac{\gamma}{2}z^2-z(1-\tfrac{\gamma}{2})+\Pi(\tfrac{1}{2}\leq(1-\gamma)(1-z)+\tfrac{\gamma}{2})\right\}\Bigg)\end{aligned}$$

To understand why $\gamma=0$ is the unique solution let us analyze both maximization problems separately. The expression $\frac{\gamma}{2}z^2-z(1-\frac{\gamma}{2})$ in the second maximization problem is non-positive in $(z,\gamma)\in[0,1]^2$ as we can write it as $(\frac{\gamma}{2}z^2-\frac{1}{2}z)-z(\frac{1}{2}-\frac{\gamma}{2})$ where each term is non-positive. Then, an optimal solution for it is $(z,\gamma)=(0,0)$ which also satisfies the lower cost constraints. Similarly, the expression $z^2(1-\frac{\gamma}{2})-z\frac{\gamma}{2}$ in $(z,\gamma)\in[0,1]^2$ of the first maximization problem has a maximum in $(z,\gamma)=(1,0)$, optimal pair which also satisfies the lower cost constraints.

$\gamma = 1$ **as unique solution**. $f(z;\theta^*,w_1) = z$, $c(z;\theta^*,w_1) = 0$, $f(z;\theta^*,w_2) = z$, $c(z;\theta^*,w_2) = z$. In this case $\mathbb{E}[f(z;\theta^*,w)] = z$ and $\mathbb{E}[c(z;\theta^*,w)] = 0.5z$. Then, for any $\gamma \in [0,1]$ we have

$$\mathrm{OPT}(\mathcal{P},\gamma) = \frac{1}{2}\left(\max_{z\in[0,1]}\left\{z + \Pi(\tfrac{1}{2} \le \tfrac{\gamma}{2}z)\right\} + \max_{z\in[0,1]}\left\{z + \Pi(\tfrac{1}{2} \le (1-\tfrac{\gamma}{2})z)\right\}\right)$$

The result is direct as $(z,\gamma) = (1,1)$ is the only pair in $[0,1]^2$ which makes the first optimization problem feasible.

A.2. Bound on Δ_{Learn}

Before stating this subsection result, we define an stricter version of Assumption 3.1

Assumption A.1 ((Stricter) Bounded Dual Iterates)**.** *There is an absolute constant $C'_h > 0$ such that $\|\lambda^t\|_1 \le C'_h$ for all $t \in [T]$ almost surely.*

Proposition A.1. *Run Algorithm 2 with a constant "step-size" rule $\eta_t \leftarrow \eta$ for all $t \ge 1$ where $\eta > 0$. Suppose that Assumption A.1 holds and that $c(\cdot;\cdot,\cdot)$ is Lipschitz on its θ argument, in particular, that it exists $L_c > 0$, such that $\|c(z;\theta,w) - c(z;\theta',w)\|_\infty \le L_c\|\theta - \theta'\|_\theta$ for any $(z,w,\theta,\theta') \in \mathcal{Z}\times\mathcal{W}\times\Theta\times\Theta$. Then, for any distribution $\mathcal{P}$ over $w \in \mathcal{W}$, it holds that*

$$\Delta_{Learn} \le L_c\,(1 + C'_h)\,\mathbb{E}\left[\sum_{t=1}^{\tau_A}\|\theta^* - \theta^t\|_\theta\right].$$

Proof. The proof is obtained directly by bounding each term of Δ_{Learn} separately. First,

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{\tau_A} c(z^t;\theta^*,w^t) - c(z^t;\theta^t,w^t)\right] \le& \mathbb{E}\left[\sum_{t=1}^{\tau_A}\|c(z^t;\theta^*,w^t) - c(z^t;\theta^t,w^t))\|_\infty\right]\\ \le& L_c\mathbb{E}\left[\sum_{t=1}^{\tau_A}\|\theta^* - \theta^t\|_\theta\right],\end{aligned}$$

where we have used above that $c(\cdot;\cdot,\cdot)$ its Lipschitz on its θ argument. Now, for any pair x, y of real vectors of same dimension it holds $|x^T y| \le \|x\|_\infty\|y\|_1$. Using the latter fact and again that $c(\cdot;\cdot,\cdot)$ is Lipschitz on its θ argument, we have

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{\tau_A}(c(z^t;\theta^*,w^t) - c(z^t;\theta^t,w^t))^T\lambda^t\right] \le& \mathbb{E}\left[\sum_{t=1}^{\tau_A}|(c(z^t;\theta^*,w^t) - c(z^t;\theta^t,w^t))^T\lambda^t|\right]\\ \le& \mathbb{E}\left[\sum_{t=1}^{\tau_A}\|c(z^t;\theta^*,w^t) - c(z^t;\theta^t,w^t))\|_\infty\|\lambda^t\|_1\right]\\ \le& L_c\mathbb{E}\left[\sum_{t=1}^{\tau_A}\|\lambda^t\|_1\|\theta^* - \theta^t\|_\theta\right]\\ \le& L_c C'_h\mathbb{E}\left[\sum_{t=1}^{\tau_A}\|\theta^* - \theta^t\|_\theta\right].\end{aligned}$$

□

A.3. Proof That $\mathrm{OPT}(\mathcal{P}) = \mathrm{OPT}(\mathcal{P},0)$ in the Linear Contextual Bandits Experiment and Solving it Efficiently.

This appendix subsection shows the following three results. 1. That for any $\rho \ge 0.5$ we have $\mathrm{OPT}(\mathcal{P},\gamma) > -\infty$ for all $\gamma \in [0,1]$. 2. That $\mathrm{OPT}(\mathcal{P},\gamma) \le \mathrm{OPT}(\mathcal{P},0)$ for all $\gamma \in (0,1]$. 3. How to efficiently solve $\mathrm{OPT}(\mathcal{P},0)$. Take $\mathcal{Z} = \{z \in \mathbb{R}^K_+ : \sum_{i=1}^K z_i \le 1\}$ and $\gamma \in [0,1]$ arbitrary. As notation, here we use superscripts to denote time (but also use $\cdot^T$ to denote dot operation between vectors when need), use subscripts to denote row indexes, and use W, W', W^t, W'^t to represent matrices of size $d \times n$. Also, to shorten notation, we write $\mathbf{W}$ to define a sequence $\{W^1,\ldots,W^T\}$ of W^t matrices (analogous for $\mathbf{W}'$). The traditional multiplication between a matrix A of size $d \times n$ and a vector x of size n is

written as $Ax = ((A_1)^T x, \ldots, (A_d)^T x)$. The term inside the outer expectation of $\mathrm{OPT}(\mathcal{P}, \gamma)$ corresponds to (for $\gamma = 1$ the outer expectation can be removed)

$$O(\mathbf{W}, \gamma) := \max_{z^t \in \mathcal{Z}: t \in [T]} \ (1-\gamma) \sum_{t=1}^{T} (W^t \theta^*)^T z^t + \gamma \mathbb{E}_{W' \sim \mathcal{P}}[(W'\theta^*)^T z^t]$$
$$\text{s.t. } 0.5 * T \leq \rho \sum_{t=1}^{T} \sum_{i=1}^{d} z_i^t \leq T.$$

Notice that a solution $\mathbf{z} = \{z^1, \ldots, z^T\}$ is either feasible or infeasible independently of the context vector arrivals $\mathbf{W} = \{W^1, \ldots, W^T\}$ and γ. For any $\rho \geq 0.5$ and $\gamma \in [0, 1]$, it holds $\mathrm{OPT}(\mathcal{P}, \gamma) > -\infty$ as we can choose $\mathbf{z}$ satisfying $\sum_{i=1}^{d} z_i^t = 0.5/\rho$ for all $t \in [T]$ (our problem setup uses $\rho = 4$). A direct application of Jensen inequality shows $\mathrm{OPT}(\mathcal{P}, 1) \leq \mathrm{OPT}(\mathcal{P}, 0)$, so let us take $\gamma \in (0, 1)$ arbitrary. For any sequence $\mathbf{W}$, let $\mathbf{z}_\gamma(\mathbf{W})$ be an optimal solution of $O(\mathbf{W}, \gamma)$, we have

$$\begin{aligned}
\mathrm{OPT}(\mathcal{P}, \gamma) &= \mathbb{E}_{\mathbf{W} \sim \mathcal{P}^T} \left[(1-\gamma) \sum_{t=1}^{T} (W^t \theta^*)^T z_\gamma^t(\mathbf{W}) + \gamma \mathbb{E}_{W' \sim \mathcal{P}} \left[(W'\theta^*)^T z_\gamma^t(\mathbf{W}) \right] \right] \\
&= \mathbb{E}_{\mathbf{W} \sim \mathcal{P}^T} \left[(1-\gamma) \sum_{t=1}^{T} (W^t \theta^*)^T z_\gamma^t(\mathbf{W}) \right] + \mathbb{E}_{\mathbf{W} \sim \mathcal{P}^T} \left[\gamma \sum_{t=1}^{T} \mathbb{E}_{W' \sim \mathcal{P}} [(W'\theta^*)^T z_\gamma^t(\mathbf{W})] \right] \\
&= \mathbb{E}_{\mathbf{W} \sim \mathcal{P}^T} \left[(1-\gamma) \sum_{t=1}^{T} (W^t \theta^*)^T z_\gamma^t(\mathbf{W}) \right] + \mathbb{E}_{\mathbf{W} \sim \mathcal{P}^T} \left[\mathbb{E}_{\mathbf{W}' \sim \mathcal{P}^T} \left[\gamma \sum_{t=1}^{T} (W'^t \theta^*)^T z_\gamma^t(\mathbf{W}) \right] \right] \\
&= \mathbb{E}_{\mathbf{W} \sim \mathcal{P}^T} \left[(1-\gamma) \sum_{t=1}^{T} (W^t \theta^*)^T z_\gamma^t(\mathbf{W}) \right] + \mathbb{E}_{\mathbf{W} \sim \mathcal{P}^T, \mathbf{W}' \sim \mathcal{P}^T} \left[\gamma \sum_{t=1}^{T} ((W^t)^T \theta^*)^T z_\gamma^t(\mathbf{W}') \right] \\
&= \mathbb{E}_{\mathbf{W} \sim \mathcal{P}^T} \left[\sum_{t=1}^{T} (W^t \theta^*)^T \left((1-\gamma) z_\gamma^t(\mathbf{W}) + \gamma \mathbb{E}_{\mathbf{W}' \sim \mathcal{P}^T} [z_\gamma^t(\mathbf{W}')] \right) \right] \\
&\leq \mathbb{E}_{\mathbf{W} \sim \mathcal{P}^T} \left[\sum_{t=1}^{T} (W^t \theta^*)^T z_0^t(\mathbf{W}) \right] = \mathrm{OPT}(\mathcal{P}, 0).
\end{aligned}$$

The second equality uses the linearity of the expectation operator, the third uses that each W'^t is sampled i.i.d. from $\mathcal{P}$, the fourth that $\mathbf{W}$ and $\mathbf{W}'$ are i.i.d. and can be exchanged, the fifth uses the linearity of the expectation operator again, and the final inequality uses the definition of $\mathbf{z}_0(\mathbf{W})$. In particular, the last inequality uses that $(1-\gamma)\mathbf{z}_\gamma(\mathbf{W}) + \gamma \mathbb{E}_{\mathbf{W}' \sim \mathcal{P}^T}[\mathbf{z}_\gamma(\mathbf{W}')]$ is a feasible solution of $O(\mathbf{W}, 0)$. Finally, notice that for any given $\mathbf{W}$ solving $O(\mathbf{W}, 0)$ is equivalent to solving the following knapsack problem

$$O(\mathbf{W}, 0) = \max_{y^t \in [0,1]: t \in T} \ \sum_{t=1}^{T} \left(\max_{i \in [d]} (W_i^t)^T \theta^* \right) y^t$$
$$\text{s.t. } 0.5 * T \leq \rho \sum_{t=1}^{T} y^t \leq T.$$

Let $\{m_1, \ldots, m_T\}$ represent the sequence $\{\max_{i \in [d]} (W_i^t)^T \theta^*\}_{t=1}^T$ ordered from biggest to smallest value. Then, is not hard to see that

$$O(\mathbf{W}, 0) = \max_{i_{\max} \in \left[\left\lceil \frac{T}{2\rho} \right\rceil, \left\lfloor \frac{T}{\rho} \right\rfloor \right]} \ \sum_{i=1}^{i_{max}} m_i,$$

where $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ are the traditional ceiling and floor integer functions respectively.

B. Proofs

B.1. Proof of Proposition 2.1

Proof. Let $\mathcal{P}^T$ be the distribution from which the $(w^1, \dots, w^T)$ vectors are sampled, with each w^t being sampled $i.i.d.$ from $\mathcal{P}$. For any $\gamma \in [0, 1]$

$$
\begin{aligned}
&\mathrm{OPT}(\mathcal{P}, \gamma) \\
&= \mathbb{E}_{\mathcal{P}^T}\left[\begin{array}{c}\max\limits_{z^t \in \mathcal{Z}: t\in[T]} \ \sum_{t=1}^T (1-\gamma) f(z^t;\theta^*, w^t) + \gamma \mathbb{E}_{\mathcal{P}}[f(z^t;\theta^*, w)] \\ \text{s.t. } T\alpha_k b_k \le \sum_{t=1}^T (1-\gamma) c_k(z^t;\theta^*, w^t) + \gamma \mathbb{E}_{\mathcal{P}}[c_k(z^t;\theta^*, w)] \le T b_k \text{ for all } k \in [K]\end{array}\right] \\
&\le \mathbb{E}_{\mathcal{P}^T}\left[\max_{z^t \in \mathcal{Z}: t\in[T]} \left\{ \sum_{t=1}^T (1-\gamma)\left(f(z^t;\theta^*, w^t) - \lambda^T c(z^t;\theta^*, w^t)\right) + \gamma \mathbb{E}_{\mathcal{P}}[f(z^t;\theta^*, w) - \lambda^T c(z^t;\theta^*, w)] \right\} + Tp(\lambda)\right] \\
&= \mathbb{E}_{\mathcal{P}^T}\left[\sum_{t=1}^T \max_{z^t \in \mathcal{Z}: t\in T} (1-\gamma)\left(f(z^t;\theta^*, w^t) - \lambda^T c(z^t;\theta^*, w^t)\right) + \gamma \mathbb{E}_{\mathcal{P}}[f(z^t;\theta^*, w) - \lambda^T c(z^t;\theta^*, w)]\right] + Tp(\lambda) \\
&\le (1-\gamma)\mathbb{E}_{\mathcal{P}^T}\left[\sum_{t=1}^T \max_{z^t \in \mathcal{Z}: t\in T} f(z^t;\theta^*, w^t) - \lambda^T c(z^t;\theta^*, w^t)\right] \\
&+ \gamma \mathbb{E}_{\mathcal{P}^T}\left[\sum_{t=1}^T \max_{z^t \in \mathcal{Z}: t\in T} \mathbb{E}_{\mathcal{P}}[f(z^t;\theta^*, w) - \lambda^T c(z^t;\theta^*, w)]\right] + Tp(\lambda) \\
&\le (1-\gamma) T \mathbb{E}_{\mathcal{P}}\left[\max_{z\in\mathcal{Z}} f(z;\theta^*, w) - \lambda^T c(z;\theta^*, w)\right] + \gamma T \max_{z\in\mathcal{Z}} \mathbb{E}_{\mathcal{P}}\left[f(z;\theta^*, w) - \lambda^T c(z;\theta^*, w)\right] + Tp(\lambda) \\
&\le (1-\gamma) T \mathbb{E}_{\mathcal{P}}\left[\varphi(\lambda;\theta^*, w)\right] + \gamma T \mathbb{E}_{\mathcal{P}}\left[\max_{z\in\mathcal{Z}} f(z;\theta^*, w) - \lambda^T c(z;\theta^*, w)\right] + Tp(\lambda) \\
&= T\mathbb{E}_{\mathcal{P}}\left[\varphi(\lambda;\theta^*, w)\right] + Tp(\lambda) \\
&= TD(\lambda;\theta^*)
\end{aligned}
$$

The first equality is the definition of $\mathrm{OPT}(\mathcal{P}, \gamma)$, the first inequality uses Lagrangian duality for both the lower and upper bounds constraints, the second equality uses that $p(\lambda)$ can be moved outside the expectation and that the sum can be changed with the maximization operator as there is no constraint linking the z^t variables. The second inequality uses that for any $a(\cdot)$ and $b(\cdot)$ real valued functions we have $max_{z\in\mathcal{Z}}\{a(z) + b(z)\} \le max_{z\in\mathcal{Z}} a(z) + max_{z\in\mathcal{Z}} b(z)$, the third inequality uses that all w^t are $i.i.d$ sampled, that all maximization problems are the same in the first term, and that the outer expectation can be removed from the second term. The fourth inequality uses the definition of $\varphi(\cdot;\cdot,\cdot)$ and that $\max_{z\in\mathcal{Z}} \mathbb{E}_{\mathcal{P}}[\cdot] \le \mathbb{E}_{\mathcal{P}}[\max_{z\in\mathcal{Z}} \cdot]$. Finally, we use the definition of $\varphi(\cdot;\cdot,\cdot)$ again and the fact that $\gamma + (1-\gamma) = 1$. □

B.2. Proof of Proposition 2.2

Proof. First note that the $p(\cdot)$ function used inside $D(\cdot;\cdot)$ is convex since $b \ge 0$ and $\alpha \in [-1, 1)^K$. We need to prove that $D(\lambda;\theta) + \mathbb{E}_{\mathcal{P}}[\tilde{g}(\lambda;\theta, w)]^T(\lambda' - \lambda) \le D(\lambda';\theta)$ for any $\lambda \in \Lambda$ and $\lambda' \in \Lambda$. Let p' be any member of $\partial p(\lambda)$, we have

$$
\begin{aligned}
D(\lambda;\theta) + \mathbb{E}_{\mathcal{P}}[\tilde{g}(\lambda;\theta, w)]^T(\lambda' - \lambda) =& \mathbb{E}_{\mathcal{P}}[\varphi(\lambda;\theta, w) + p(\lambda) + \tilde{g}(\lambda;\theta, w)^T(\lambda' - \lambda)] \\
=& \mathbb{E}_{\mathcal{P}}[f(z(\lambda;\theta, w);\theta, w) - (\lambda')^T c(z(\lambda;\theta, w);\theta, w) + p(\lambda) + p'^T(\lambda' - \lambda)] \\
\le& \mathbb{E}_{\mathcal{P}}[f(z(\lambda;\theta, w);\theta, w) - (\lambda')^T c(z(\lambda;\theta, w);\theta, w) + p(\lambda')] \\
\le& D(\lambda';\theta).
\end{aligned}
$$

The first equality uses the definition of $D(\lambda;\theta)$, the second equality uses the definition of $\tilde{g}(\lambda;\theta, w)$, the first inequality uses the subgradient inequlity for $p(\cdot)$, and the second inequality uses the definition of $D(\lambda';\theta)$. □

B.3. Intermediate Results

The following propositions were not mentioned in the paper. Proposition B.1 shows an inequality that holds for Step 7. of Algorithm 2 under the conditions given for Λ and $h(\cdot)$ on the paper. Propositions B.2 and B.3 are intermediate steps to prove

Theorem 3.1. Proposition B.2 bounds $T - \tau_A$ in expectation. Proposition B.3 shows an upper bound for the regret that Algorithm 2 up to period τ_A. Proposition B.4 is the key result needed to prove Proposition 3.1.

Proposition B.1. *Let $\Lambda \subseteq \mathbb{R}^K$ be a set which can be defined separately for each dimension $k \in [K]$, either being $\Lambda_k = \mathbb{R}$ or $\Lambda_k = \mathbb{R}_+$. Let $h(\cdot) : \Lambda \to \mathbb{R}$ be a function that satisfies $h(\lambda) = \sum_{k=1}^K h_k(\lambda_k)$, with $h_k(\cdot)$ being a strongly convex univariate differentiable function for all $k \in [K]$. Given arbitrary $\lambda' \in \Lambda$, $\tilde{g} \in \mathbb{R}^K$, and $\eta > 0$ define $\lambda^+ = \arg\min_{\lambda \in \Lambda} \lambda^T \tilde{g}^t + \frac{1}{\eta} V_h(\lambda, \lambda')$. Then, for all $k \in [K]$ it holds*

1. *If $\Lambda_k = \mathbb{R}$, then $\dot{h}_k(\lambda_k^+) = \dot{h}_k(\lambda_k') - \eta \tilde{g}_k$.*
2. *If $\Lambda_k = \mathbb{R}_+$, then $\dot{h}_k(\lambda_k^+) = \dot{h}_k(\lambda_k') - \eta \tilde{g}_k$ if $\lambda_k^+ > 0$ or $\dot{h}_k(\lambda_k^+) \geq \dot{h}_k(\lambda_k') - \eta \tilde{g}_k$ if $\lambda_k^+ = 0$.*

Therefore, proving that $\nabla h(\lambda^+) \geq \nabla h(\lambda') - \eta \tilde{g}$.

Proof. Notice that $\min_{\lambda \in \Lambda} \lambda^T \tilde{g}^t + \frac{1}{\eta} V_h(\lambda, \lambda') = \sum_{k \in [K]} \min_{\lambda_k \in \Lambda_k} \phi_k(\lambda_k; \lambda_k', \tilde{g}_k)$ with $\phi_k(\lambda_k; \lambda_k', \tilde{g}_k) := \tilde{g}_k \lambda_k + \frac{1}{\eta}(h_k(\lambda_k) - h_k(\lambda_k') - \dot{h}_k(\lambda_k')(\lambda_k - \lambda_k'))$ for all $k \in [K]$. Then, independently per coordinate we minimize a strongly convex function under a non-empty closed convex set, which shows that λ_k^+ exists for each $k \in [K]$. Also, λ_k^+ can be found using first order necessary optimality conditions for each $k \in [K]$. Taking $k \in [K]$ arbitrary, we split the proof in two cases.

$\Lambda_k = \mathbb{R}$. By first order optimality conditions we immediately obtain $\dot{h}_k(\lambda_k^+) = \dot{h}(\lambda_k') - \eta \tilde{g}_k$.

$\Lambda_k = \mathbb{R}_+$. Define $\Pi_+(\cdot) : \mathbb{R} \to \{0\} \cup \{\infty\}$ as the convex function that takes the value of 0 if its input is non-negative and ∞ otherwise. Then, the minimization problem for dimension k can be re-written as $\min_{\lambda_k \in \Lambda_k} \phi_k(\lambda_k; \lambda_k', \tilde{g}_k) + \Pi_+(\lambda_k)$. First order necessary optimality conditions imply $0 \in \partial(\phi_k(\lambda_k^+; \lambda_k', \tilde{g}_k) + \Pi_+(\lambda_k^+))$. Then, there exists $y \in \partial(\Pi_+(\lambda_k^+))$, such that $\dot{h}_k(\lambda_k^+) = \dot{h}(\lambda_k') - \eta \tilde{g}_k - \eta y$. The result is obtained directly using that $\partial(\Pi_+(\lambda_k))$ is equal to $\{0\}$ when $\lambda_k > 0$ and equal to $\mathbb{R}_-$ when $\lambda_k = 0$. □

Proposition B.2. *Run Algorithm 2 with a constant "step-size" rule $\eta_t \leftarrow \eta$ for all $t \geq 1$ where $\eta > 0$. Suppose that Assumption A.1 holds and take τ_A as in Definition 3.1. Then,*

$$\mathbb{E}[T - \tau_A] \leq \frac{\bar{C}}{\underline{b}} + \frac{C_h + \|\nabla h(\lambda^1)\|_\infty}{\eta \underline{b}} + \frac{\|\mathbb{E}[\sum_{t=1}^{\tau_A} c(z^t; \theta^*, w^t) - c(z^t; \theta^t, w^t)]\|_\infty}{\underline{b}}.$$

Proof. Let $k' \in [K]$ be the index of the first violated upper cost bound, *i.e.* the index which activates the stop time τ_A. Here we assume that some upper cost bound constraint is violated, *i.e.* that $\tau_A < T$, if not the result is trivial. Step 6. of Algorithm 2 defines $\tilde{g}_{k'}^t = -c_{k'}(z^t; \theta^t, w^t) + b_{k'}(\mathbb{1}(\lambda_{k'} \geq 0) + \alpha_{k'} \mathbb{1}(\lambda_{k'} < 0))$, which can be upper bounded by $\tilde{g}_{k'}^t \leq -c_{k'}(z^t; \theta^t, w^t) + b_{k'}$. Using the definition of τ_A and $\tilde{g}_{k'}^t$ we have

$$\begin{aligned}
\sum_{t=1}^{\tau_A} \tilde{g}_{k'}^t \leq & b_{k'} \tau_A - \sum_{t=1}^{\tau_A} c_{k'}(z^t; \theta^*, w^t) + \left(\sum_{t=1}^{\tau_A} (c_{k'}(z^t; \theta^*, w^t) - c_{k'}(z^t; \theta^t, w^t)) \right) \\
\leq & b_{k'} \tau_A - b_{k'} T + \bar{C} + \left(\sum_{t=1}^{\tau_A} (c_{k'}(z^t; \theta^*, w^t) - c_{k'}(z^t; \theta^t, w^t)) \right) \\
\Rightarrow T - \tau_A \leq & \frac{1}{b_{k'}} \left(\bar{C} - \sum_{t=1}^{\tau_A} \tilde{g}_{k'}^t \right) + \frac{1}{b_{k'}} \left(\sum_{t=1}^{\tau_A} (c_{k'}(z^t; \theta^*, w^t) - c_{k'}(z^t; \theta^t, w^t)) \right).
\end{aligned}$$

Using that our update rule satisfies $\dot{h}_{k'}(\lambda_{k'}^{t+1}) \geq \dot{h}_{k'}(\lambda_{k'}^t) - \eta \tilde{g}_{k'}^t$ for all $t \leq \tau_A$ and the definitions of $\underline{b}$ and C_h, we get

$$\begin{aligned}
-\sum_{t=1}^{\tau_A} \tilde{g}_{k'}^t & \leq \frac{1}{\eta} \left(\dot{h}_{k'}(\lambda_{k'}^{\tau_A + 1}) - \dot{h}_{k'}(\lambda_{k'}^1) \right) \\
\Rightarrow T - \tau_A & \leq \frac{\bar{C}}{b_{k'}} + \frac{\dot{h}_{k'}(\lambda_{k'}^{\tau_A+1}) - \dot{h}_{k'}(\lambda_{k'}^1)}{\eta b_{k'}} + \left(\frac{\sum_{t=1}^{\tau_A} (c_{k'}(z^t; \theta^*, w^t) - c_{k'}(z^t; \theta^t, w^t))}{b_{k'}} \right) \\
\Rightarrow \mathbb{E}[T - \tau_A] & \leq \frac{\bar{C}}{\underline{b}} + \frac{C_h + \|\nabla h(\lambda^1)\|_\infty}{\eta \underline{b}} + \left(\frac{\|\mathbb{E}[\sum_{t=1}^{\tau_A} c(z^t; \theta^*, w^t) - c(z^t; \theta^t, w^t)]\|_\infty}{\underline{b}} \right)
\end{aligned}$$

□

Proposition B.3. *Run Algorithm 2 with a constant "step-size" rule $\eta_t \leftarrow \eta$ for all $t \geq 1$ where $\eta > 0$. Denote $\bar{\lambda}^{\tau_A} = \frac{\sum_{t=1}^{\tau_A} \lambda^t}{\tau_A}$ (τ_A as in Definition 3.1). It holds*

$$\mathbb{E}\left[\tau_A D(\bar{\lambda}^{\tau_A}; \theta^*) - \sum_{t=1}^{\tau_A} f(z^t; \theta^t, w^t)\right] \leq \frac{2(\bar{C}^2 + \bar{b}^2)}{\sigma_1}\eta\mathbb{E}[\tau_A] + \frac{1}{\eta}V_h(\lambda, \lambda^1) + \mathbb{E}\left[\sum_{t=1}^{\tau_A}(c(z^t; \theta^*, w^t) - c(z^t; \theta^t, w^t))^T\lambda^t\right].$$

Proof. For clarity, we sometimes use $\mathbb{E}_w[\cdot]$, $\mathbb{E}_{w^t}[\cdot]$, or $\mathbb{E}_{\mathcal{H}^{t-1}}[\cdot]$ to indicate the random variable over which the expectation is taken. Using $\mathbb{E}[\cdot]$ indicates that the expectation is taken over the "whole" randomness of Algorithm 2. Call $\tilde{g}^t$ the vector obtained in Step 6. and define $\mathbb{E}[\tilde{g}^t] = g^t$. The proof is composed of three steps. 1. Bounding $\tilde{g}^t$. 2. Upper bounding $\mathbb{E}\left[\sum_{s=1}^{\tau_A}(g^s)^T(\lambda^s - \lambda)\right]$. 3. Lower bounding $\mathbb{E}\left[\sum_{s=1}^{\tau_A}(g^s)^T(\lambda^s - \lambda)\rangle\right]$. The upper and lower bounds match the left and right hand side of the terms in Proposition B.3.

Step 1. Upper bound for $\mathbb{E}[\|\tilde{g}^t\|_\infty^2]$.

$$\mathbb{E}[\|\tilde{g}^t\|_\infty^2] \leq \mathbb{E}[(\|c(z^t; \theta^t, w^t)\|_\infty + \|b\|_\infty)^2] \leq 2\mathbb{E}[\|c(z^t; \theta^t, w^t)\|_\infty^2 + \|b\|_\infty^2] \leq 2(\bar{C}^2 + \bar{b}^2)$$

Step 2. Upper bound for $\mathbb{E}\left[\sum_{s=1}^{\tau_A}(g^s)^T(\lambda^s - \lambda)\right]$. Notice

$$\begin{aligned}
&\mathbb{E}_{w^t}[(\tilde{g}^t)^T(\lambda^t - \lambda)|\lambda^t, \theta^t] \\
\leq&\mathbb{E}_{w^t}\left[(\tilde{g}^t)^T(\lambda^t - \lambda^{t+1}) + \frac{1}{\eta}V_h(\lambda, \lambda^t) - \frac{1}{\eta}V_h(\lambda, \lambda^{t+1}) - \frac{1}{\eta}V_h(\lambda^{t+1}, \lambda^t)\big|\lambda^t, \theta^t\right] \\
\leq&\mathbb{E}_{w^t}\left[(\tilde{g}^t)^T(\lambda^t - \lambda^{t+1}) + \frac{1}{\eta}V_h(\lambda, \lambda^t) - \frac{1}{\eta}V_h(\lambda, \lambda^{t+1}) - \frac{\sigma_1}{2\eta}\|\lambda^{t+1} - \lambda^t\|_1^2\big|\lambda^t, \theta^t\right] \\
\leq&\mathbb{E}_{w^t}\left[\frac{\eta}{\sigma_1}\|\tilde{g}^t\|_\infty^2 + \frac{1}{\eta}V_h(\lambda, \lambda^t) - \frac{1}{\eta}V_h(\lambda, \lambda^{t+1})\big|\lambda^t, \theta^t\right] \\
\leq&\frac{2\eta}{\sigma_1}(\bar{C}^2 + \bar{b}^2) + \frac{1}{\eta}V_h(\lambda, \lambda^t) - \mathbb{E}_{w^t}\left[\frac{1}{\eta}V_h(\lambda, \lambda^{t+1})\big|\lambda^t, \theta^t\right],
\end{aligned} \tag{4}$$

where the first inequality is due to the three point property (Lemma 4.1 of (Beck & Teboulle, 2003)), the second uses $V_h(\lambda^{t+1}, \lambda^t) \geq \frac{\sigma_1}{2}\|\lambda^{t+1} - \lambda^t\|_1^2$ given that $h(\cdot)$ is σ_1-strongly convex with respect to the $\|\cdot\|_1$ norm, the third uses that for any two vectors a^1 and a^2 of same dimension it holds $(a^1)^T a^2 + 0.5\|a^1\|_\infty^2 \geq -0.5\|a^2\|_1^2$, and the final inequality is just understanding which terms are constant under the conditional expectation. Taking $E_{\mathcal{H}^{t-1}}[\cdot]$ over both sides of equation (4) and using the law of total expectation we get

$$\mathbb{E}[\eta(g^t)^T(\lambda^t - \lambda)] \leq \frac{2(\bar{C}^2 + \bar{b}^2)}{\sigma_1}\eta^2 + \mathbb{E}\left[V_h(\lambda, \lambda^t)\right] - \mathbb{E}\left[V_h(\lambda, \lambda^{t+1})\right], \tag{5}$$

since the pair (λ^t, θ^t) is completely determined by $\mathcal{H}^{t-1} \cup \{w^t\}$ and that w^t is independent of $\mathcal{H}^{t-1}$. Then, regardless of the value of τ_A, using the telescopic property and that $V_h(\cdot, \cdot)$ is non-negative we obtain

$$\mathbb{E}\left[\sum_{s=1}^{\tau_A}(g^s)^T(\lambda^s - \lambda)\right] \leq \frac{2(\bar{C}^2 + \bar{b}^2)}{\sigma_1}\eta\mathbb{E}[\tau_A] + \frac{V_h(\lambda, \lambda^1)}{\eta}.$$

Step 3. Lower bounds for $\mathbb{E}\left[\sum_{s=1}^{\tau_A}(g^s)^T(\lambda^s - \lambda)\right]$. By definition of g^t, using the subgradient inequality we get

$$(g^t)^T(\lambda^t - \lambda) \geq D(\lambda^t; \theta^t) - D(\lambda; \theta^t) \geq D(\lambda^t; \theta^t) - \left(\mathbb{E}_w[\varphi(\lambda; \theta^t, w)] + \sum_{k\in[K]} b_k([\lambda_k]_+ - \alpha_k[-\lambda_k]_+)\right).$$

For any $w \in \mathcal{W}$ we have $f(z(\lambda^t;\theta^t,w);\theta^t,w) - \lambda^T c(z(\lambda^t;\theta^t,w);\theta^t,w) \leq \varphi(\lambda;\theta^t,w)$ as by definition $z(\lambda^t;\theta^t,w)$ is an optimal solution of $\varphi(\lambda^t;\theta^t,w)$ not of $\varphi(\lambda;\theta^t,w)$. Defining $\bar{\lambda}^{\tau_A} := \frac{1}{\tau_A}\sum_{t=1}^{\tau_A}\lambda^t$, taking $\lambda = (0,0,\ldots,0)$, and summing from one to τ_A we get

$$
\begin{aligned}
&\sum_{t=1}^{\tau_A}(g^t)^T(\lambda^t - 0)\\
&\geq \sum_{t=1}^{\tau_A} D(\lambda^t;\theta^t) - \mathbb{E}_w[f(z(\lambda^t;\theta^t,w);\theta^t,w)]\\
&\geq \sum_{t=1}^{\tau_A}\left(D(\lambda^t;\theta^*) - \mathbb{E}_w[f(z(\lambda^t;\theta^t,w);\theta^*,w)]\right) + \sum_{t=1}^{\tau_A}\left(D(\lambda^t;\theta^t) - D(\lambda^t;\theta^*)\right)\\
&\quad + \sum_{t=1}^{\tau_A}\left(\mathbb{E}_w[f(z(\lambda^t;\theta^t,w);\theta^*,w) - f(z(\lambda^t;\theta^t,w);\theta^t,w)]\right)\\
&\geq \left(\tau_A D(\bar{\lambda}^{\tau_A};\theta^*) - \sum_{t=1}^{\tau_A}\mathbb{E}_w[f(z(\lambda^t;\theta^*,w);\theta^*,w)]\right) + \sum_{t=1}^{\tau_A}\left(D(\lambda^t;\theta^t) - D(\lambda^t;\theta^*)\right)\\
&\quad + \sum_{t=1}^{\tau_A}\left(\mathbb{E}_w[f(z(\lambda^t;\theta^t,w);\theta^*,w) - f(z(\lambda^t;\theta^t,w);\theta^t,w)]\right). \qquad (6)
\end{aligned}
$$

Taking expectation over (6) and using the results from Step 2. we get

$$
\begin{aligned}
&\mathbb{E}\left[\tau_A D(\bar{\lambda}^{\tau_A};\theta^*) - \sum_{t=1}^{\tau_A}\mathbb{E}_w\left[f(z(\lambda^t;\theta^t,w);\theta^*,w)\right]\right] \leq \frac{2(\bar{C}^2+\bar{b}^2)}{\sigma_1}\eta\mathbb{E}[\tau_A] + \frac{1}{\eta}V_h(0,\lambda^1)\\
&+\mathbb{E}\left[\sum_{t=1}^{\tau_A}\mathbb{E}_w[c(z(\lambda^t;\theta^t,w);\theta^t,w)]^T\lambda^t\right] - \mathbb{E}\left[\sum_{t=1}^{\tau_A}\mathbb{E}_w[c(z(\lambda^t;\theta^t,w);\theta^*,w)]^T\lambda^t\right], \qquad (7)
\end{aligned}
$$

where we have used the definition of $D(\cdot,\cdot)$ to reduce the second line of (7) to use only the cost functions. Equation (7) almost matches the conclusion of Theorem 3.1 except that (7) uses a $\mathbb{E}[\sum_{t=1}^{\tau_A}\mathbb{E}_w[\cdot]]$ term, while the theorem uses $\mathbb{E}[\sum_{t=1}^{\tau_A}\cdot]$. The previous issue is solved using the Optional Stopping Theorem. We prove now that $\mathbb{E}\left[\sum_{t=1}^{\tau_A} f(z(\lambda^t;\theta^t,w^t);\theta^*,w^t)\right]$ equals $\mathbb{E}\left[\sum_{t=1}^{\tau_A}\mathbb{E}_w\left[f(z(\lambda^t;\theta^t,w);\theta^*,w)\right]\right]$ (the analysis for the cost terms appearing in the second line of (7) is analogous). First notice

$$\mathbb{E}_w\left[f(z(\lambda;\theta,w);\theta^*,w)|\lambda=\lambda^t,\theta=\theta^t\right] = \mathbb{E}_w\left[f(z(\lambda^t;\theta^t,w);\theta^*,w)|\mathcal{H}^{t-1}\right].$$

Define the martingale $M^t = \sum_{s=1}^{t} f(z(\lambda^s;\theta^s,w^s);\theta^*,w^s) - \mathbb{E}_w[f(z(\lambda^s;\theta^s,w);\theta^*,w)|\mathcal{H}^{s-1}]$ for all $t \leq T$. Using that τ_A is a stop time w.r.t. to the filtration $\mathcal{H}^t$, the Optional Stopping Time ensures $\mathbb{E}[M^{\tau_A}] = \mathbb{E}[M^1] = 0$, therefore:

$$\mathbb{E}\left[\sum_{t=1}^{\tau_A}\mathbb{E}_w\left[f(z(\lambda^t;\theta^t,w);\theta^*,w)|\mathcal{H}^{t-1}\right]\right] = \mathbb{E}\left[\sum_{t=1}^{\tau_A} f(z(\lambda^t;\theta^t,w^t);\theta^*,w^t)\right]$$

concluding the proof. □

Proposition B.4. *Run Algorithm 2 with a constant "step-size" rule $\eta_t \leftarrow \eta$ for all $t \geq 1$ where $\eta > 0$. Using δ_θ as in Definition 3.2, for each $t \in [T-1]$ it holds (here we use 0 to refer to the zero-vector $(0,\ldots,0)$ of dimension K):*

$$\mathbb{E}\left[V_h(0,\lambda^{t+1})\middle|\lambda^t,\theta^t\right] \leq \eta\left(\frac{2\eta}{\sigma_1}(\bar{C}^2+\bar{b}^2) + 2\bar{f} - \delta_{\theta^t}\|\lambda^t\|_1\right) + V_h(0,\lambda^t).$$

Proof. Let $\tilde{g}^t$ be the λ^t stochastic subgradient obtained in Step 6. of Algorithm 2. Here we abuse notation and use, *e.g.*, $\mathbb{E}[\tilde{g}^t|\lambda^t,\theta^t]$ to represent that $\tilde{g}^t$ is a random variable on w given a fixed pair $(\lambda^t,\theta^t) \in (\Lambda \times \Theta)$. The following bound holds

$$\mathbb{E}_{\mathcal{P}}[\|\tilde{g}^t\|_\infty^2] \leq \mathbb{E}[(\|c(z^t;\theta^t,w^t)\|_\infty + \|b\|_\infty)^2] \leq 2\mathbb{E}[\|c(z^t;\theta^t,w^t)\|_\infty^2 + \|b\|_\infty^2] \leq 2(\bar{C}^2+\bar{b}^2).$$

For any $\lambda \in \Lambda$ we have

$$\begin{aligned}
&\mathbb{E}[\tilde{g}^t|\lambda^t,\theta^t]^T(\lambda^t-\lambda)\\
=&\mathbb{E}[(\tilde{g}^t)^T(\lambda^t-\lambda)|\lambda^t,\theta^t]\\
\leq&\mathbb{E}\left[(\tilde{g}^t)^T(\lambda^t-\lambda^{t+1})+\frac{1}{\eta}V_h(\lambda,\lambda^t)-\frac{1}{\eta}V_h(\lambda,\lambda^{t+1})-\frac{1}{\eta}V_h(\lambda^{t+1},\lambda^t)\big|\lambda^t,\theta^t\right]\\
\leq&\mathbb{E}\left[(\tilde{g}^t)^T(\lambda^t-\lambda^{t+1})+\frac{1}{\eta}V_h(\lambda,\lambda^t)-\frac{1}{\eta}V_h(\lambda,\lambda^{t+1})-\frac{\sigma_1}{2\eta}\|\lambda^{t+1}-\lambda^t\|_1^2\big|\lambda^t,\theta^t\right]\\
\leq&\mathbb{E}\left[\frac{\eta}{\sigma_1}\|\tilde{g}^t\|_\infty^2+\frac{1}{\eta}V_h(\lambda,\lambda^t)-\frac{1}{\eta}V_h(\lambda,\lambda^{t+1})\big|\lambda^t,\theta^t\right]\\
\leq&\frac{2\eta}{\sigma_1}(\bar{C}^2+\bar{b}^2)+\frac{1}{\eta}V_h(\lambda,\lambda^t)-\mathbb{E}\left[\frac{1}{\eta}V_h(\lambda,\lambda^{t+1})\big|\lambda^t,\theta^t\right],
\end{aligned}$$

where we have used linearity of the expectation, the three point property, that $V_h(\cdot,\cdot)$ is σ_1 strongly convex on with respect to the $\|\cdot\|_1$ norm, Cauchy-Schwartz, and the bound for $\mathbb{E}[\|\tilde{g}^t\|_\infty^2]$ obtained before (same steps as in Step 1. and 2. of Proof B.3). Choosing $\lambda=(0,\dots,0)$ we get

$$\mathbb{E}\left[V_h(0,\lambda^{t+1})\big|\lambda^t,\theta^t\right]\leq\eta\left(\frac{2\eta}{\sigma_1}(\bar{C}^2+\bar{b}^2)-\mathbb{E}[\tilde{g}^t|\lambda^t,\theta^t]^T\lambda^t\right)+V_h(0,\lambda^t).$$

To finish the proof we now show that $\mathbb{E}[\tilde{g}^t|\lambda^t,\theta^t]^T\lambda^t\geq\|\lambda^t\|_1\delta_{\theta^t}-2\bar{f}$. Notice first that for any $(\lambda^t,\theta^t)\in(\Lambda\times\Theta)$ we have $\mathbb{E}[\tilde{g}^t(w)]^T\lambda^t=-\mathbb{E}[c(z(\lambda^t;\theta^t,w);\theta^t,w)]^T\lambda^t+p(\lambda^t)$ using that by definition $p(\lambda)=\sum_{k\in[K]}b_k([\lambda_k]_+-\alpha_k[-\lambda_k]_+)$. Let $\{z(w)\}_{w\in\mathcal{W}}$ be a series that satisfies $\delta_{\theta^t}=\mathbb{E}_{\mathcal{P}}[\min\{\|Tb_k-c_k(z(w);\theta^t,w)\|_\infty,\|c_k(z(w);\theta^t,w)-T\alpha_kb_k\|_\infty\}]$. Then,

$$\begin{aligned}
&\mathbb{E}[\tilde{g}^t|\lambda^t,\theta^t]^T\lambda^t\\
&=D(\lambda^t;\theta^t)-\mathbb{E}_{\mathcal{P}}[f(z(\lambda^t;\theta^t,w);\theta^t,w)]\\
&\geq\mathbb{E}_{\mathcal{P}}[\max_{z\in\mathcal{Z}}f(z;\theta^t,w)+\sum_{k\in[K]}\left([\lambda_k^t]_+(b_k-\mathbb{E}_{\mathcal{P}}[c_k(z;\theta^t,w)])+[-\lambda_k^t]_+(\mathbb{E}_{\mathcal{P}}[c_k(z;\theta^t,w)]-\alpha_kb_k)\right)]-\bar{f}\\
&\geq\mathbb{E}_{\mathcal{P}}[f(z(w);\theta^t,w)+\sum_{k\in[K]}\left([\lambda_k^t]_+(b_k-\mathbb{E}_{\mathcal{P}}[c_k(z(w);\theta^t,w)])+[-\lambda_k^t]_+(\mathbb{E}_{\mathcal{P}}[c_k(z(w);\theta^t,w)]-\alpha_kb_k)\right)]-\bar{f}\\
&\geq\mathbb{E}_{\mathcal{P}}[\sum_{k\in[K]}[\lambda_k^t]_+(b_k-\mathbb{E}_{\mathcal{P}}[c_k(z(w);\theta^t,w)])+[-\lambda_k^t]_+(\mathbb{E}_{\mathcal{P}}[c_k(z(w);\theta^t,w)]-\alpha_kb_k)]-2\bar{f}\\
&\geq\|\lambda^t\|_1\delta_{\theta^t}-2\bar{f},
\end{aligned}$$

where we have used the definition of $D(\lambda^t;\theta^t)$, $\bar{f}$, δ_{θ^t}, and the fact that $\|\lambda^t\|_1=\sum_{k\in[K]}\left([\lambda_k^t]_++[-\lambda_k^t]_+\right)$. □

B.4. Proof of Theorem 3.1

Proof. For any distribution $\mathcal{P}$ over $\mathcal{W}$ and for any $t'\in[T]$ we have

$$\begin{aligned}
OPT(\mathcal{P})&\leq\frac{t'}{T}OPT(\mathcal{P})+\frac{T-t'}{T}OPT(\mathcal{P})\\
&\leq t'D(\bar{\lambda}^{t'};\theta^*)+(T-t')\bar{f},
\end{aligned}$$

where we have used Proposition 2.1 and that a loose upper bound for $OPT(\mathcal{P})$ is $T\bar{f}$. Therefore,

$$\begin{aligned}
&Regret(A|\mathcal{P})\\
=&OPT(\mathcal{P}) - R(A|\mathcal{P})\\
\leq&\mathbb{E}\left[\tau_A D(\bar{\lambda}^{\tau_A};\theta^*) + (T-\tau_A)\bar{f} - \sum_{t=1}^{\tau_A} f(z^t;\theta^*,w^t)\right]\\
=&\mathbb{E}\left[\tau_A D(\bar{\lambda}^{\tau_A};\theta^*) - \sum_{t=1}^{\tau_A} f(z^t;\theta^*,w^t)\right] + \mathbb{E}[T-\tau_A]\bar{f}\\
\leq&\frac{2(\bar{C}^2+\bar{b}^2)}{\sigma_1}\eta\mathbb{E}[\tau_A] + \frac{1}{\eta}V_h(0,\lambda^1) + \frac{\bar{f}}{\underline{b}}\left(\bar{C} + \frac{C_h + \|\nabla h(\lambda^1)\|_\infty}{\eta}\right)\\
+&\mathbb{E}\left[\sum_{t=1}^{\tau_A}(c(z^t;\theta^*,w^t) - c(z^t;\theta^t,w^t))^T\lambda^t\right] + \frac{\bar{f}}{\underline{b}}\left\|\mathbb{E}\left[\sum_{t=1}^{\tau_A} c(z^t;\theta^*,w^t) - c(z^t;\theta^t,w^t)\right]\right\|_\infty,
\end{aligned}$$

where in the first inequality we have used the definition of $R(A|\mathcal{P})$ and the fact that Algorithm 2 runs for τ_A periods. The second inequality is obtained directly from Propositions B.2 and B.3. □

B.5. Proof of Proposition 3.1

Proof. A direct application of Proposition B.4 shows that whenever $\|\lambda^t\|_1 \geq C^\triangleright/\delta$ we have $\mathbb{E}[V_h(0,\lambda^{t+1})|(\lambda^t,\theta^t)] \leq V_h(0,\lambda^t)$. Then, for any $(\lambda^t,\theta^t) \in \Lambda\times\Theta$ we have

$$\begin{aligned}
&\mathbb{E}[V_h(0,\lambda^{t+1})|(\lambda^t,\theta^t)] \leq \max\left\{\max_{\|\lambda\|_1\leq\delta^{-1}C^\triangleright} V_h(0,\lambda) + \eta C^\triangleright,\, V_h(0,\lambda^1)\right\}\\
\Rightarrow&\mathbb{E}[V_h(0,\lambda^{t+1})] \leq \max\left\{\max_{\|\lambda\|_1\leq\delta^{-1}C^\triangleright} V_h(0,\lambda) + \eta C^\triangleright,\, V_h(0,\lambda^1)\right\}
\end{aligned}$$

Take now $h(\cdot) = \frac{1}{2}\|\cdot\|_2^2$, then for any $\lambda \in \Lambda$ we have $\nabla h(\lambda) = \lambda$ and $V_h(0,\lambda) = \frac{1}{2}\|\lambda\|_2^2$, therefore $\max_{\|\lambda\|_1\leq\delta^{-1}C^\triangleright} 0.5\|\lambda\|_2^2 = 0.5(C^\triangleright/\delta)^2$. Using Jensen inequality and previous results we get

$$\mathbb{E}[\|\lambda^{t+1}\|_2] \leq \max\left\{\sqrt{(C^\triangleright/\delta)^2 + 2\eta C^\triangleright}, \|\lambda^1\|_2\right\}$$

Finally, since $\|\lambda\|_\infty \leq \|\lambda\|_2$ for any $\lambda\in\Lambda$ is immediate that $\mathbb{E}[\|\lambda^t\|_\infty] \leq \max\left\{\sqrt{(C^\triangleright/\delta)^2 + 2\eta C^\triangleright}, \|\lambda^1\|_\infty\right\}$ for all $t\in[T]$ concluding the proof.

□

B.6. Proof of Proposition 3.2

Proof. Since $\alpha_k \neq -\infty$ by statement, Proposition B.1 shows $\dot{h}_k(\lambda^{t+1}) = \dot{h}_k(\lambda^t) - \eta\tilde{g}_k^t$ for any $t\in[T]$, which implies that $\dot{h}_k(\lambda^{\tau_A+1}) - \dot{h}_k(\lambda^1) = -\eta\sum_{t=1}^{\tau_A}\tilde{g}_k^t$ regardless of the τ_A value. Then, using the definition of $\tilde{g}^t$ we get

$$\begin{aligned}
&\sum_{t=1}^{\tau_A}\left(b_k(\mathbb{1}(\lambda_k\geq 0) + \alpha_k\mathbb{1}(\lambda_k<0)) - c_k(z^t;\theta^t,w^t)\right) = \frac{\dot{h}_k(\lambda^1) - \dot{h}_k(\lambda^{\tau_A+1})}{\eta}\\
\Rightarrow&\sum_{t=1}^{\tau_A}\left(b_k(\mathbb{1}(\lambda_k\geq 0) + \alpha_k\mathbb{1}(\lambda_k<0)) - c_k(z^t;\theta^*,w^t)\right) = \frac{\dot{h}_k(\lambda^1) - \dot{h}_k(\lambda^{\tau_A+1})}{\eta} + \sum_{t=1}^{\tau_A} c_k(z^t;\theta^t,w^t) - c_k(z^t;\theta^*,w^t).
\end{aligned}$$

Now, given that $(\mathbb{1}(\lambda'\geq 0) + \alpha_k\mathbb{1}(\lambda'<0)) \geq \alpha_k$ for any $\lambda'\in\mathbb{R}$ and that $\tau_A \leq T$ by definition, we have

$$\sum_{t=1}^{\tau_A}(b_k(\mathbb{1}(\lambda_k\geq 0) + \alpha_k\mathbb{1}(\lambda_k<0))) + (T-\tau_A)\alpha_k b_k \geq T\alpha_k b_k.$$

Combining the previous results and taking expectation we get

$$T\alpha_k b_k - \mathbb{E}[\sum_{t=1}^{\tau_A} c_k(z^t;\theta^*,w^t)] \leq \frac{\dot{h}_k(\lambda^1) - \mathbb{E}[\dot{h}_k(\lambda^{\tau_A+1})]}{\eta} + \mathbb{E}[T-\tau_A]\alpha_k b_k + \mathbb{E}\left[\sum_{t=1}^{\tau_A} c_k(z^t;\theta^t,w^t) - c_k(z^t;\theta^*,w^t)\right].$$

Finally, we conclude the proof by using Proposition B.2 and the definition of C_h. □

C. Extra Experimental Details and Results

C.1. Bidding Experiment

This experiment is based on data from Criteo (Diemert et al., 2017). Criteo is a Demand-Side Platform (DSP), which are entities which bid on behalf of hundreds or thousands of advertisers which set campaigns with them. The dataset from (Diemert et al., 2017) contains millions of bidding logs during one month of Criteo's operation. These bidding logs are all logs in which Criteo successfully acquired ad-space for its advertising clients through real-time second-price auctions (each log represents a different auction and ad-space). Each of these auctions occur when a user arrives to a website, app, etc., and each user is shown one ad few millisecond after its "arrival". Each bidding log contains. 1. Nine anonymized categorical columns containing characteristics of the ad-space and (possibly) about the user who has just "arrived". 2. The price Criteo paid for the ad-space, which corresponds to the second highest bid submitted to each auction. 3. The day of the auction and the advertiser whose ad was shown in the ad-space (the day is not included directly in the dataset, but appears in a Jupyter Notebook inside the compressed file that contains the dataset). 4. If a conversion occur after the ad was shown, *i.e.*, if the corresponding user performed an action of interest for the advertiser after watching the advertiser's ad. The dataset can be downloaded from https://ailab.criteo.com/criteo-attribution-modeling-bidding-dataset.

The experiment was performed as follows. We used the first 21 days of data as training, the next two days as validation, and the remaining seven days as test. The training data was used only to train a neural network to predict the probability of a conversion occurring. The model architecture was taken from (Pan et al., 2018) and uses as features the nine anonymized categorical columns, the day of the week, and an advertiser id to make a prediction if a conversion would occur or not. Parameters to be tuned for the neural network were the step-size for the Adam solver, embedding sizes, and other two specific network attributes (in total we tried 120 configurations). Once found the trained model with highest validation AUC (Area Under the Curve), we took this model predictions as if they were the real probabilities of a conversion occurring for unseen data. By having the advertiser id as an input on the model, we can get conversion probability estimates for all advertisers even when Criteo bid on behalf of only one advertiser per bidding log. The advertisers pay the DSP, in our context the bidder, each time the DSP bids on behalf of them. The payment corresponds to the probability of conversion times a known fixed value. The general simulator scheme for this experiment is shown in Algorithm 4.

Algorithm 4 Simulator Scheme

Input: Trained conversion prediction model σ, the set of all test bidding logs X_{test}, T the number of test bidding logs, $q \in \mathbb{R}_+^K$ the vector of payment per conversion values for the advertisers, $\{\mathrm{mp}^t\}_{t=1}^T$ the price Criteo paid for each ad spot in the test set in order.

for $t = 1, \ldots, T$ **do**

1. Read the t bidding test log and mp^t.
2. Use model σ to obtain estimated conversion probabilities $\mathrm{conv\_prob}$. Take $r_k^t = \mathrm{conv\_prob}_k \cdot q_k$ for all $k \in K$.
3. Using vector r^t and previous history, obtain (z^t, k^t) a pair of submitted bid and advertiser to bid on behalf of.
4. If $z^t \geq \mathrm{mp}^t$ then the auction is won, advertiser k^t pays $r_{k^t}^t$ to the bidder (the DSP), the bidder pays mp^t for the ad spot and obtains $r_{k^t}^t - \mathrm{mp}^t$ as profit.

end for

Algorithm 2 can be naturally incorporated in the simulator scheme by using the online optimization component of it to obtain (z^t, k^t) of Step 3. of the simulator. We only need the online optimization component of Algorithm 2, as we do not need to learn the distribution of the highest competing (mp) to solve Step 3. of Algorithm 2 (shown in Algorithm 3). We compare the performance of Algorithm 2 to using the Greedy Heuristic 5. When $\gamma = 1$, Algorithm 5 bids 'truthfully' on behalf of the advertiser with the highest valuation. This would be the optimal strategy if the advertisers had 'infinite' budgets and no lower bound requirements. Then, we can think of γ as a way to increase/decrease the bids in order to take the budgets into account. (For this example, we can think of Algorithm 2 as an online algorithm for obtaining γ variables per advertiser.)

Algorithm 5 Greedy Heuristic(γ)

Input: Vector $r \in \mathbb{R}_+^K$ and $\gamma > 0$.

Let $\mathcal{K}'$ be the set of advertisers with non depleted budgets. If $\mathcal{K}' = \emptyset$ do not bid, otherwise bid on behalf of $k^* \in \arg\max_{k \in \mathcal{K}'} r_k$ the amount γr_{k^*}.

Our test set contains 21073 iterations and 130 advertisers. (The original dataset had 700 advertisers but we removed all

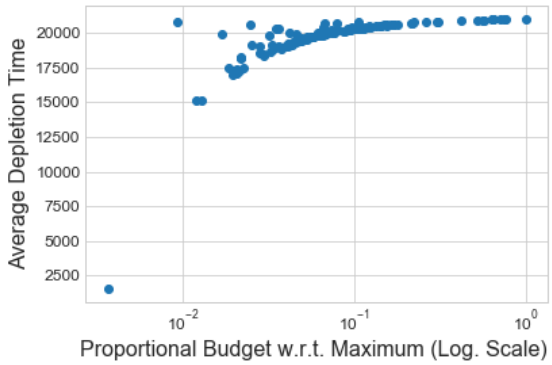

Figure 2. The x-axis in the figure shows the proportion of an advertiser budget w.r.t. the highest budget between all advertisers (shown on a logarithmic scale).

advertisers who appeared in less than 10,000 logs in either the training or validation plus test data.) Each iteration of the simulator scheme uses a batch of 128 test logs. The total budget of an advertiser is the total amount Criteo spent bidding on behalf of that advertiser in the test logs multiplied by 100. We run Algorithm 2 using traditional subgradient descent trying the fixed step sizes $\{1 * 10^{-i}\}_{i=0}^{3} \cup \{0.5 * 10^{-i}\}_{i=0}^{3}$ and $\{0.25 + 0.05 * i\}_{i=0}^{25}$ as γ parameters for the Greedy Heuristic 5. We run 100 simulations for each parameter and method pair. Each simulation is defined by the price advertisers would pay per conversion, which is the q vector in Algorithm 4. We sample q_k i.i.d. from $\text{Uniform}(0.5, 1.5)$ for all $k \in [K]$. We relaxed the ending condition of Algorithm 2 by allowing advertisers to overspend at most on one iteration. After that iteration, we consider an advertiser's budget as depleted and do not bid on behalf of it until the simulation's end. The final parameters chosen for Algorithms 2 and 5 were those that achieved the highest average profit.

An advertiser's budget depletion time correlates with its relative total maximum budget, fact that is shown in Figure 2. The x-axis is in logarithmic scale and shows the proportion of an advertiser budget w.r.t. the highest budget between all advertisers. Observe that as the relative budget increases, the average depletion time gets closer to the simulation end ($T = 21073$).

Finally, we run this experiment using a SLURM managed Linux cluster. We tried 120 parameters combinations for the conversion prediction architecture, running each parameter configuration for 25 epochs. Each parameter configuration took approximately 40 min to run using a Nvidia K80 GPU plus two Intel Xeon 4-core 3.0 Ghz (we used eight GPUs in parallel having a total run time of approximately 12 hours). For the experiment itself, we tried nine different step-sizes to run the subgradient descent step using Algorithm 2 and 26 γ values for 5, each configuration running 100 simulations. We used several cluster nodes each having 64GB of RAM and two Xeon 12-core Haswell with 2.3 Ghz per core. If we had used just one node it would have taken approximately 160 hours to run all required configurations.

C.2. Linear Contextual Bandits Experiment

We now describe in detail the methods used to implement Step 1. of Algorithm 2. First, let y^t be the variable that takes the value of one if an action is taken at period t and zero otherwise. Also, remember that $i(t) \in [d]$ is the action taken at period t (if any), and r^t the revenue observed at period t. We implemented Step 1. of Algorithm 2 in the following ways.

1. Gaussian Thompson Sampling as in Agrawal & Goyal (2013). Define $B(1) = I_d$ with I_d the identity matrix of size d, and $\hat{\theta}^1 = (1/\sqrt{d}, \ldots, 1/\sqrt{d})$. The Thompson Sampling procedure is composed of two steps which are updating a prior and sampling from a Gaussian posterior. We update the prior as follows. If $y^t = 1$, make $B(t+1) = I_d + \sum_{s \in [t]: y^s = 1} W^s_{i(s)} (W^s_{i(s)})^T$ and $\hat{\theta}^{t+1} = B(t+1)^{-1} (\sum_{s \in [t]: y^s = 1} W^s_{i(s)} r^t)$, otherwise $B(t+1) = B(t)$ and $\hat{\theta}^{t+1} = \hat{\theta}^t$. After the prior update, we sample θ^t from $\mathcal{N}(\hat{\theta}^t, \nu^2 B(t)^{-1})$ where $\mathcal{N}(\cdot, \cdot)$ represents a normal distribution defined by its mean and covariance matrix, and $\nu > 0$ a constant chosen as follows. When no randomness was added to the observed revenue term, we used $\nu = 0.1$ (remember that we could add randomness to both the matrices W^t and the observed revenue separately). When randomness was added to the observed revenue, we used $\nu = \frac{\text{rev\_err}}{10} * \sqrt{\log T * n}$ with $\text{rev\_err} = 0.1$ or 0.5 depending if a $\text{Uniform}(-0.1, 0.1)$ or $\text{Uniform}(-0.5, 0.5)$ is

added to the observed revenue term respectively. (The latter form of choosing ν was inspired on Agrawal & Goyal (2013) which uses $\nu = R\sqrt{9n \log T}$ to prove a regret bound for Thompson Sampling for linear contextual bandits without constraints.)

2. Least squares. Same as Thompson Sampling as described above, but Step 1. of Algorithm 2 uses $\theta^t = \hat{\theta}^t$. (This update is a core element of many learning approaches for linear contextual bandits (Agrawal & Goyal, 2013; Agrawal & Devanur, 2016) and can be understood as a Least Squares step).

3. Ridge regression. We use the Least Squares procedure as defined above for the first $\sqrt{T}/2$ actions, and then solve a ridge regression problem. We solve a ridge regression problem at Step 1. of iteration t using the set $\{W^s_{i(s)}, r^s\}_{s \in [t-1]: y^s = 1}$ with an ℓ_2 penalization parameter of $\alpha = 0.001$.

4. Ridge regression plus error. Same method as above but adds noise to the θ^t obtained from the ridge regression problem. We add an i.i.d. $\text{Uniform}(-0.3, 0.3)/\sqrt{\sum_{s=1}^t y^s}$ term to each coordinate of θ^t.

5. Known θ^*. Algorithm 2 using $\theta^t = \theta^*$ for all $t \in [T]$.

Figures 3 and 4 show how the different methods perform for $(d \times n)$ being $(5, 10)$ and $(50, 50)$ when $T = 10,000$, respectively. Each element of the x-axis represents a moving window composed of 250 iterations. The x-axis is composed of 9751 ticks . The y-axis shows the average relative revenue obtained in a window with respect to the proportional best revenue that could have been obtained ($\text{OPT}(\mathcal{P}) \cdot \frac{250}{10000}$). Importantly, the number of actions a method takes can vary between windows, which explains the following two facts. First, an initial revenue spike as many actions are taken when a simulation starts. The latter occurs as we took $\lambda^1 = 0$ which makes the cost component in Step 3. of Algorithm 2 zero. Second, a method may obtain a higher average revenue on a window than $\text{OPT}(\mathcal{P}) \cdot \frac{250}{10000}$ if more than 'average' actions are taken on that window.

Tables 2, 3, 4 show the average total relative revenue obtained for the different combinations of $d \times n$ and uncertainty used with respect to $\text{OPT}(\mathcal{P})$. In general, as long as the budget is spent properly, the revenue obtained by the 'Known θ^*' method when $W^t = W$ for all $t \in [T]$ should match $\text{OPT}(\mathcal{P})$. The latter as the best action to take is always the same. In the case when we still have $W^t = W$ for all $t \in [T]$, but the observed revenue has randomness, the 'Known θ^*' method may obtain a higher total revenue than $\text{OPT}(\mathcal{P})$.

Finally, we run this experiment using a SLURM managed Linux cluster. We used four nodes each having 64 GB of RAM and 20 cores of 2.5Ghz. We parallelized the code to run each combination of experiment setting and simulation number as a different run (the run-time was mostly spent on sampling from a Gaussian distribution for Thompson Sampling and solving Ridge Regression problems with thousands of points). The total running time was 12 hours. Processing the results was done in a local computer (Mac Book Pro 2015 version), spending around 30 minutes to aggregate the results obtained from the cluster.

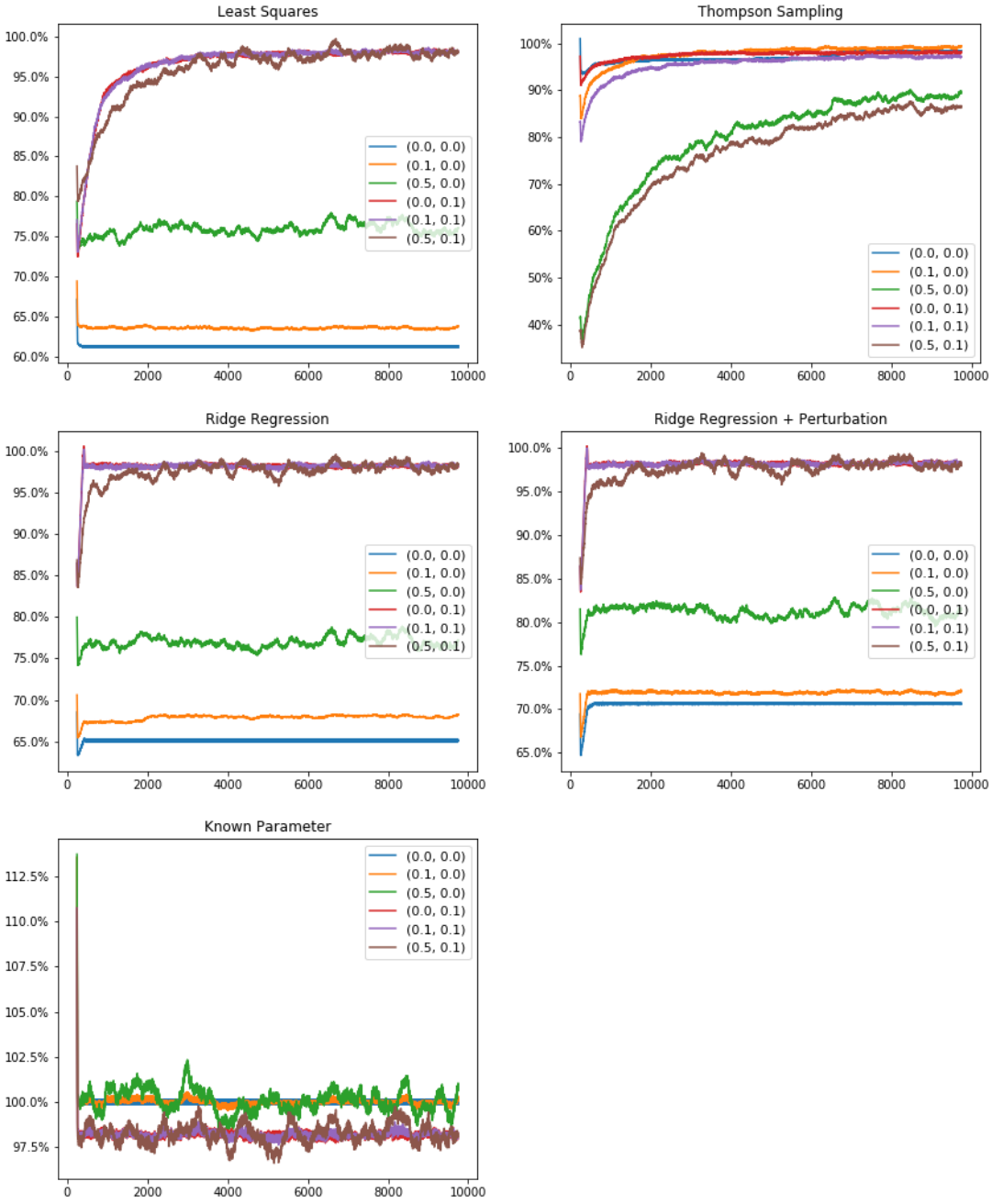

Figure 3. Moving average revenue for windows of 250 iterations against the proportional best average revenue possible using $d = 5$, $n = 10$.

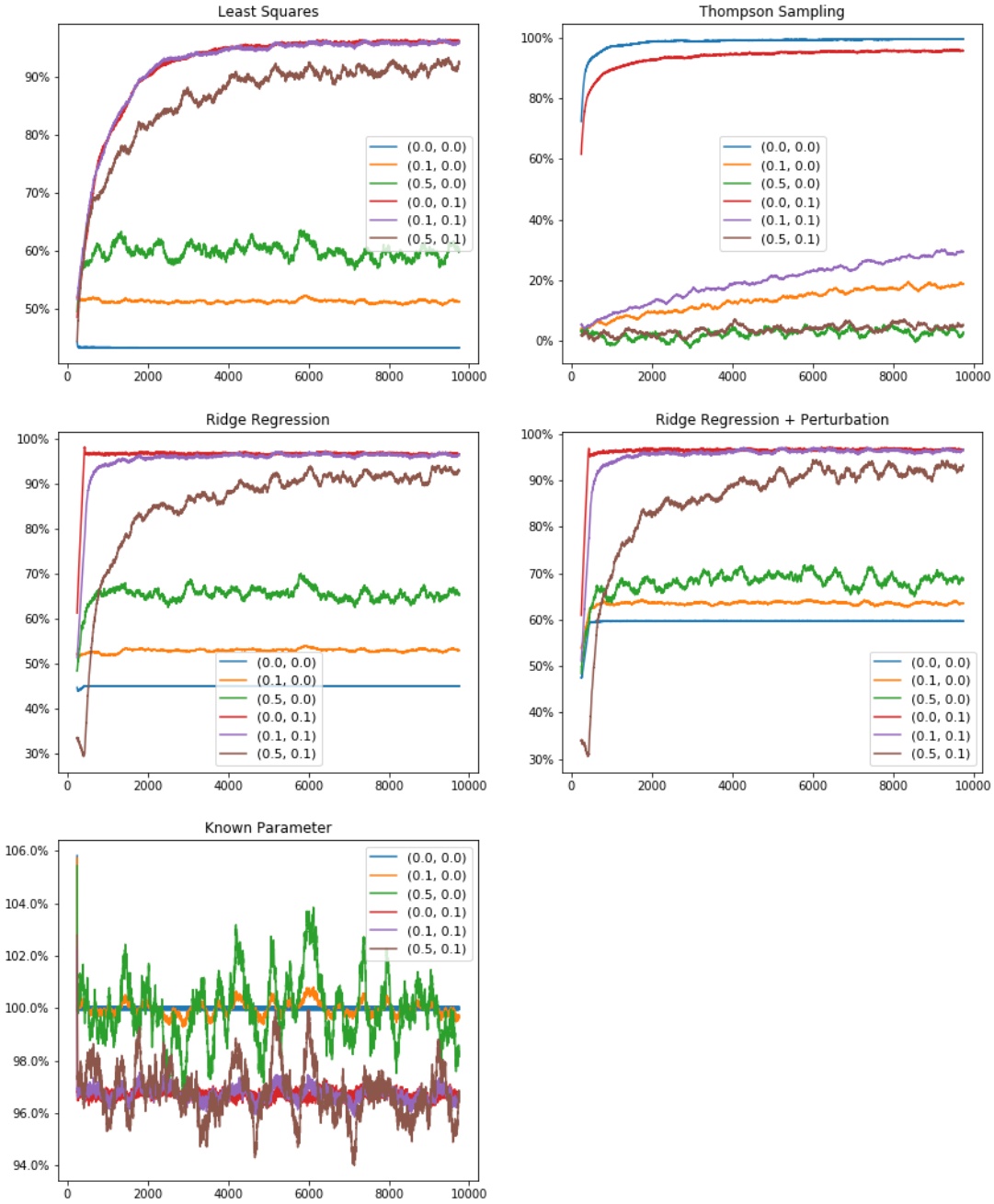

Figure 4. Moving average revenue for windows of 250 iterations against the proportional best average revenue possible using $d = 50$, $n = 50$.

$T = 1,000$	$d \times n$	$(0.0, 0.0)$	$(0.1, 0.0)$	$(0.5, 0.0)$	$(0.0, 0.1)$	$(0.1, 0.1)$	$(0.5, 0.1)$
Least Squares	5×5	76.2%	77.6%	84.2%	78.9%	79.4%	79.3%
Thompson Sampling	5×5	95.2%	74.2%	21.9%	85.4%	65.4%	19.0%
Ridge Regression	5×5	77.6%	79.0%	85.4%	90.4%	89.8%	83.5%
Ridge Reg. + Perturbation	5×5	80.8%	80.9%	86.0%	90.3%	89.6%	83.5%
Known Parameter	5×5	99.9%	100.1%	100.8%	92.4%	92.4%	92.2%
Least Squares	5×10	60.9%	63.2%	73.4%	80.1%	80.5%	82.6%
Thompson Sampling	5×10	94.2%	90.3%	51.2%	89.4%	85.9%	48.3%
Ridge Regression	5×10	64.5%	67.3%	76.5%	93.0%	92.8%	90.0%
Ridge Reg. + Perturbation	5×10	73.9%	73.9%	81.1%	92.8%	92.6%	90.2%
Known Parameter	5×10	100.0%	100.0%	99.9%	95.5%	95.5%	95.4%
Least Squares	10×5	70.9%	74.6%	78.1%	83.7%	84.0%	82.9%
Thompson Sampling	10×5	94.5%	91.0%	50.6%	89.0%	84.6%	47.6%
Ridge Regression	10×5	71.0%	75.2%	78.8%	92.5%	92.5%	89.0%
Ridge Reg. + Perturbation	10×5	82.0%	84.3%	84.7%	92.3%	92.3%	89.7%
Known Parameter	10×5	99.9%	99.9%	99.7%	94.5%	94.4%	94.2%
Least Squares	10×10	58.5%	62.5%	72.0%	75.7%	75.3%	76.7%
Thompson Sampling	10×10	92.2%	66.6%	14.7%	86.1%	62.4%	15.1%
Ridge Regression	10×10	59.0%	63.4%	72.4%	91.2%	90.4%	84.0%
Ridge Reg. + Perturbation	10×10	72.3%	73.6%	77.2%	90.9%	90.2%	84.1%
Known Parameter	10×10	100.0%	99.9%	99.7%	93.9%	93.9%	93.9%
Least Squares	25×25	44.0%	49.7%	54.0%	64.5%	66.0%	58.9%
Thompson Sampling	25×25	89.1%	5.4%	0.3%	74.4%	6.1%	0.7%
Ridge Regression	25×25	44.1%	50.6%	56.0%	86.1%	78.4%	46.8%
Ridge Reg. + Perturbation	25×25	69.5%	66.7%	61.4%	85.4%	78.0%	46.3%
Known Parameter	25×25	100.0%	100.0%	99.7%	90.7%	90.8%	91.3%
Least Squares	25×50	41.4%	48.1%	56.1%	64.7%	65.1%	68.1%
Thompson Sampling	25×50	89.0%	19.6%	3.3%	82.4%	20.5%	3.7%
Ridge Regression	25×50	43.3%	50.3%	62.7%	90.0%	85.8%	69.7%
Ridge Reg. + Perturbation	25×50	62.8%	64.0%	68.8%	89.5%	85.5%	69.1%
Known Parameter	25×50	100.0%	100.1%	100.3%	93.7%	93.8%	94.1%
Least Squares	50×25	49.1%	53.7%	59.1%	67.7%	68.1%	68.3%
Thompson Sampling	50×25	92.2%	18.3%	2.6%	82.7%	19.5%	2.8%
Ridge Regression	50×25	51.9%	55.9%	64.6%	89.4%	85.3%	67.7%
Ridge Reg. + Perturbation	50×25	70.8%	69.7%	71.8%	89.1%	85.2%	67.6%
Known Parameter	50×25	100.0%	100.0%	100.0%	92.9%	92.9%	92.6%
Least Squares	50×50	42.0%	52.2%	55.7%	62.3%	63.7%	58.7%
Thompson Sampling	50×50	87.5%	5.4%	1.5%	76.0%	6.7%	1.5%
Ridge Regression	50×50	43.6%	54.5%	62.1%	86.8%	76.8%	47.7%
Ridge Reg. + Perturbation	50×50	67.2%	68.8%	66.7%	86.0%	76.6%	47.2%
Known Parameter	50×50	100.0%	100.0%	100.0%	92.0%	91.9%	91.6%

Table 2. All percentages shown are the average revenue over 100 simulations divided by the best average revenue achievable ($\mathrm{OPT}(\mathcal{P})$).

$T = 5,000$	$d \times n$	$(0.0, 0.0)$	$(0.1, 0.0)$	$(0.5, 0.0)$	$(0.0, 0.1)$	$(0.1, 0.1)$	$(0.5, 0.1)$
Least Squares	5×5	76.7%	79.4%	87.1%	91.6%	91.5%	90.5%
Thompson Sampling	5×5	98.7%	88.6%	42.6%	93.2%	80.9%	36.7%
Ridge Regression	5×5	78.1%	79.4%	86.5%	95.1%	94.9%	92.4%
Ridge Reg. + Perturbation	5×5	80.0%	79.7%	87.2%	94.9%	94.8%	92.3%
Known Parameter	5×5	100.0%	100.0%	99.9%	95.9%	95.9%	96.0%
Least Squares	5×10	61.2%	63.5%	75.3%	93.1%	93.3%	92.6%
Thompson Sampling	5×10	97.3%	96.0%	71.7%	95.8%	93.0%	68.6%
Ridge Regression	5×10	64.9%	67.9%	79.6%	96.5%	96.5%	95.5%
Ridge Reg. + Perturbation	5×10	71.0%	71.9%	80.4%	96.4%	96.4%	95.3%
Known Parameter	5×10	100.0%	100.0%	100.0%	97.5%	97.5%	97.4%
Least Squares	10×5	71.3%	72.3%	80.9%	93.6%	93.4%	93.4%
Thompson Sampling	10×5	96.0%	96.4%	70.4%	95.2%	92.1%	67.1%
Ridge Regression	10×5	71.5%	73.7%	81.5%	96.3%	96.2%	95.5%
Ridge Reg. + Perturbation	10×5	77.0%	80.1%	83.0%	96.2%	96.1%	95.3%
Known Parameter	10×5	100.0%	100.0%	100.1%	97.0%	97.0%	97.0%
Least Squares	10×10	58.9%	63.3%	70.0%	91.0%	90.9%	91.3%
Thompson Sampling	10×10	96.2%	83.9%	29.5%	94.2%	80.7%	30.8%
Ridge Regression	10×10	59.4%	63.7%	70.4%	95.6%	95.4%	93.3%
Ridge Reg. + Perturbation	10×10	69.2%	69.8%	74.1%	95.5%	95.4%	93.1%
Known Parameter	10×10	100.0%	100.0%	100.1%	96.7%	96.6%	96.5%
Least Squares	25×25	44.6%	54.0%	58.6%	85.6%	85.6%	78.3%
Thompson Sampling	25×25	97.2%	12.6%	1.2%	88.6%	15.0%	1.9%
Ridge Regression	25×25	44.8%	54.7%	60.4%	93.4%	91.1%	76.4%
Ridge Reg. + Perturbation	25×25	64.9%	64.0%	66.2%	93.2%	90.9%	76.5%
Known Parameter	25×25	100.0%	100.1%	100.4%	95.0%	94.9%	94.7%
Least Squares	25×50	41.5%	48.1%	57.5%	87.7%	87.4%	84.4%
Thompson Sampling	25×50	94.6%	36.2%	7.3%	93.0%	39.9%	8.6%
Ridge Regression	25×50	43.5%	49.9%	68.0%	95.0%	94.2%	87.8%
Ridge Reg. + Perturbation	25×50	55.7%	58.0%	74.1%	94.9%	94.1%	87.0%
Known Parameter	25×50	100.0%	99.9%	99.6%	96.5%	96.5%	96.5%
Least Squares	50×25	49.3%	53.0%	57.8%	87.6%	87.9%	85.3%
Thompson Sampling	50×25	97.8%	34.3%	5.5%	92.3%	38.9%	7.1%
Ridge Regression	50×25	52.2%	55.3%	58.4%	94.6%	93.9%	86.8%
Ridge Reg. + Perturbation	50×25	66.0%	65.7%	67.8%	94.4%	93.7%	87.1%
Known Parameter	50×25	100.0%	100.0%	100.1%	96.0%	96.0%	96.0%
Least Squares	50×50	41.9%	52.7%	60.4%	85.8%	86.2%	79.6%
Thompson Sampling	50×50	96.4%	10.0%	1.8%	89.7%	14.3%	2.7%
Ridge Regression	50×50	43.6%	53.2%	68.2%	94.0%	91.5%	77.9%
Ridge Reg. + Perturbation	50×50	59.9%	61.3%	71.8%	93.7%	91.4%	77.8%
Known Parameter	50×50	100.0%	100.0%	100.2%	95.5%	95.5%	95.5%

Table 3. All percentages shown are the average revenue over 100 simulations divided by the best average revenue achievable ($\mathrm{OPT}(\mathcal{P})$).

$T = 10,000$	$d \times n$	$(0.0, 0.0)$	$(0.1, 0.0)$	$(0.5, 0.0)$	$(0.0, 0.1)$	$(0.1, 0.1)$	$(0.5, 0.1)$
Least Squares	5×5	76.8%	79.7%	85.4%	94.7%	94.6%	93.7%
Thompson Sampling	5×5	98.8%	92.4%	52.8%	95.4%	85.8%	47.0%
Ridge Regression	5×5	78.2%	79.7%	87.0%	96.5%	96.4%	95.0%
Ridge Reg. + Perturbation	5×5	80.1%	80.0%	88.6%	96.4%	96.4%	95.0%
Known Parameter	5×5	100.0%	100.0%	100.2%	97.0%	97.0%	97.1%
Least Squares	5×10	61.2%	63.5%	75.8%	95.9%	95.9%	95.4%
Thompson Sampling	5×10	96.8%	97.3%	79.0%	97.2%	95.1%	76.1%
Ridge Regression	5×10	65.0%	67.8%	76.8%	97.5%	97.5%	97.0%
Ridge Reg. + Perturbation	5×10	70.4%	71.7%	81.0%	97.5%	97.5%	97.0%
Known Parameter	5×10	100.0%	100.0%	100.1%	98.2%	98.2%	98.2%
Least Squares	10×5	71.4%	73.1%	81.7%	95.9%	95.9%	95.4%
Thompson Sampling	10×5	96.7%	97.7%	77.7%	96.8%	94.3%	74.6%
Ridge Regression	10×5	71.6%	75.0%	82.4%	97.3%	97.3%	96.8%
Ridge Reg. + Perturbation	10×5	76.4%	80.2%	83.3%	97.3%	97.3%	96.6%
Known Parameter	10×5	100.0%	100.0%	100.0%	97.8%	97.8%	97.8%
Least Squares	10×10	59.0%	64.5%	71.0%	94.5%	94.2%	93.5%
Thompson Sampling	10×10	96.4%	89.0%	38.8%	96.0%	86.3%	40.5%
Ridge Regression	10×10	59.4%	65.2%	71.8%	96.8%	96.7%	95.2%
Ridge Reg. + Perturbation	10×10	68.9%	70.4%	73.0%	96.7%	96.6%	95.0%
Known Parameter	10×10	100.0%	100.0%	100.1%	97.5%	97.5%	97.5%
Least Squares	25×25	44.5%	53.7%	67.1%	91.4%	91.2%	84.7%
Thompson Sampling	25×25	98.4%	18.5%	1.8%	92.3%	21.2%	2.7%
Ridge Regression	25×25	44.7%	54.7%	65.8%	95.3%	94.0%	83.4%
Ridge Reg. + Perturbation	25×25	65.8%	63.9%	69.6%	95.1%	94.0%	83.6%
Known Parameter	25×25	100.0%	100.0%	100.0%	96.2%	96.2%	95.9%
Least Squares	25×50	41.6%	48.0%	58.0%	92.7%	92.7%	90.4%
Thompson Sampling	25×50	97.8%	46.3%	10.4%	95.4%	50.8%	11.8%
Ridge Regression	25×50	43.6%	49.5%	67.1%	96.4%	96.0%	91.1%
Ridge Reg. + Perturbation	25×50	57.7%	59.2%	71.3%	96.3%	96.0%	91.2%
Known Parameter	25×50	100.0%	100.0%	100.0%	97.4%	97.4%	97.4%
Least Squares	50×25	49.3%	53.6%	58.8%	92.5%	92.8%	90.5%
Thompson Sampling	50×25	98.6%	44.8%	7.9%	94.8%	50.2%	10.3%
Ridge Regression	50×25	52.3%	55.1%	65.1%	96.1%	95.7%	91.3%
Ridge Reg. + Perturbation	50×25	63.9%	62.6%	69.9%	96.0%	95.7%	91.1%
Known Parameter	50×25	100.0%	100.0%	100.0%	97.0%	97.0%	97.1%
Least Squares	50×50	43.2%	51.2%	59.5%	91.4%	91.5%	85.8%
Thompson Sampling	50×50	98.1%	13.2%	2.3%	93.1%	19.7%	3.5%
Ridge Regression	50×50	44.9%	52.9%	65.0%	95.6%	94.5%	84.9%
Ridge Reg. + Perturbation	50×50	59.3%	63.2%	67.7%	95.5%	94.4%	85.2%
Known Parameter	50×50	100.0%	100.0%	99.9%	96.7%	96.7%	96.8%

Table 4. All percentages shown are the average revenue over 100 simulations divided by the best average revenue achievable ($\mathrm{OPT}(\mathcal{P})$).