E-companion

EC.1. Appendix to Introduction: Exponential Limit for Uniform MLE

For $X_1, ..., X_n \sim U[0,\theta]$, $\theta > 0$, the joint density at $(x_1, ...x_n)$ is

$$\prod_{i=1}^{n} \frac{1}{\theta} \mathbb{I}\Big[0 \le x_i \le \theta\Big] = \frac{1}{\theta^n} \mathbb{I}\Big[0 \le \min_{i=1,...,n} x_i \le \max_{i=1,...,n} x_i \le \theta\Big]$$

From the above, we see that the maximum likelihood is given by $\hat{\theta}_n = \max_{i=1,...,n} X_i$. For a normal approximation, we would typically analyse $\sqrt{n}(\hat{\theta}_n - \theta)$. However, instead we analyse a more concentrated asymptotic $n(\theta - \hat{\theta}_n)$. For this observe

$$\mathbb{P}(n(\theta - \hat{\theta}_n) \ge z) = \mathbb{P}\Big(\max_{i=1,...,n} X_i \le \theta - \frac{z}{n}\Big) = \Big(1 - \frac{z}{n\theta}\Big)^n \xrightarrow[n\to\infty]{} e^{-\frac{z}{\theta}}$$

From the above, we see that

$$n(\theta - \hat{\theta}_n) \xrightarrow[n\to\infty]{\mathcal{D}} \exp(\theta^{-1})$$

So the limit here is not normally distributed under a $\sqrt{n}$ normalization but is order n and is exponentially distributed.

EC.2. Appendix to Section 3: Main Results

EC.2.1. Appendix to Section 3.3

EC.2.1.1. Sub-exponential noise (D2) implies Condition (C1)

LEMMA EC.1. *For a Lispchitz continuous function f, if Condition (D2) holds, that is*

$$\sup_{t\ge 0} \mathbb{E}\big[e^{\lambda \|c_t\|} \big| \mathcal{F}_t\big] < \infty$$

then Condition (C2) holds that is

$$\Big[|f(\boldsymbol{x}_{t+1}) - f(\boldsymbol{x}_t)| \Big| \mathcal{F}_t\Big] \le \alpha_t Y\,, \quad \textit{with} \quad \mathbb{E}[e^{\eta Y}] < \infty \tag{EC.1}$$

for some $\eta > 0$.

Proof. Since $f(x)$ is Lipschitz

$$|f(\boldsymbol{x}_{t+1}) - f(\boldsymbol{x}_t)| \le K\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\| \le \alpha_t K ||\boldsymbol{c}_t||\,.$$

Since Condition (D2) holds, we take $M \ge \sup_t \mathbb{E}\big[e^{\lambda\|\boldsymbol{c}_t\|}\big|\mathcal{F}_t\big]$. We let Y be the random variable with CCDF: $\mathbb{P}(Y \ge y) = 1 \wedge (Me^{-\frac{\lambda}{K}y})$. Thus for $y \in \mathbb{R}_+$

$$\begin{aligned}\mathbb{P}\left(|f(\boldsymbol{x}_{t+1}) - f(\boldsymbol{x}_t)| \ge \alpha_t y \middle| \mathcal{F}_t\right) &\le \mathbb{P}\left(\|\boldsymbol{c}_t\| \ge y/K \middle| \mathcal{F}_t\right) \\ &\le \min\big\{1, e^{-(\lambda/K)y}\mathbb{E}\big[e^{\lambda\|\boldsymbol{c}_t\|}\big|\mathcal{F}_t\big]\big\} \le \mathbb{P}(Y \ge y)\end{aligned}$$

Above, we apply a Chernoff bound. From this inequality above, we see that Condition (C2) follows from Condition (D2). □

EC.2.1.2. Proof of Lemma 1: sharpness is equivalent to non-vanishing gradient for convex functions. We now prove Lemma 1.

LEMMA 1. *If the function $l(\boldsymbol{x})$ is absolutely continuous then the gradient condition (D1) implies the function is sharp, (D1′). Moreover, if the function $l(\boldsymbol{x})$ is convex, then the gradient condition (D1) is equivalent to the function being sharp (D1′).*

Proof. First let's assume condition (D1) holds. Let $\boldsymbol{x}(t) = \boldsymbol{x}^\star + (1-t)(\boldsymbol{x} - \boldsymbol{x}^\star)$. Thus we have

$$\begin{aligned} l(\boldsymbol{x}) - l(\boldsymbol{x}^\star) &= \int_0^1 \frac{dl(\boldsymbol{x}(t))}{dt} dt \\ &= \int_0^1 \nabla l(\boldsymbol{x}(t))(\boldsymbol{x}(t) - \boldsymbol{x}^\star) dt \\ &\geq \int_0^1 \kappa \|\boldsymbol{x}(t) - \boldsymbol{x}^\star\| dt \\ &= \int_0^1 (1-t)\kappa \|\boldsymbol{x} - \boldsymbol{x}^\star\| dt \\ &= \frac{\kappa}{2} \|\boldsymbol{x} - \boldsymbol{x}^\star\| \end{aligned}$$

The first equality follows since absolute continuity implies the fundamental theorem of calculus holds. The second equality holds by the chain rule. The third equality follows by the gradient condition (D1). We then apply the definition of $\boldsymbol{x}(t)$ and integrate. Thus as required, we see that condition (D1) implies (D1′).

If we also suppose that the function $l(\boldsymbol{x})$ is convex and that (D1′) holds then

$$l(\boldsymbol{x}^\star) - l(\boldsymbol{x}) \geq \nabla l(\boldsymbol{x})(\boldsymbol{x}^\star - \boldsymbol{x}) \,.$$

So

$$\nabla l(\boldsymbol{x})(\boldsymbol{x} - \boldsymbol{x}^\star) \geq l(\boldsymbol{x}) - l(\boldsymbol{x}^\star) \geq \kappa' \|\boldsymbol{x} - \boldsymbol{x}^\star\|$$

The first inequality rearranges the convexity definition above. The second inequality applies the Sharp Condition (D1′). So we see, as required, for a convex function, the Sharp Condition (D1′) implies the gradient condition (D1). □

EC.2.1.3. Proof of Lemma 2: with d or more active constraints, SGD is not normally distributed. Duchi and Ruan (2021) is designed for smooth problems with fewer active constraints than the problem dimension. Once the number of active constraints exceeds the dimension of the problem, then the normal approximation no longer holds. With the Lemma below, we can say that the limiting distribution has an exponential concentration for PSGD. Calculations can speculate the form of the asymptotic optimality; however, the general theory of asymptotic optimality of stochastic optimization is incomplete, particularly in settings where the normal approximation is invalid. As we indicate, it requires a better understanding of asymptotic optimality in the presence of Sharpness.

LEMMA 2. *Suppose that at the optimum* $-\nabla l(\boldsymbol{x}^\star) \in \text{relint}\, \mathcal{N}_{\mathcal{X}}(\boldsymbol{x}^\star)$, *where* $\mathcal{N}_{\mathcal{X}}(\boldsymbol{x}^\star) := \{\boldsymbol{v} : \boldsymbol{v}^\top(\boldsymbol{x}^\star - \boldsymbol{y}) \le 0, \forall \boldsymbol{y} \in \mathcal{X}\}$ *and* $\nabla l(\boldsymbol{x}^\star) \neq 0$ *and that there are at least* d *active constraints at* $\boldsymbol{x}^\star$ *(w.l.o.g.* $i = 1, ..., d$*) and*

$$\{\nabla l_i(\boldsymbol{x}^\star) : i = 1, ..., d\} \text{ are linearly independent} \tag{D1''}$$

then the function f *is sharp at* $\boldsymbol{x}^\star$ *and Assumption* (D1) *holds.*

Proof. Let x_∞ be such that $\|\boldsymbol{x}\|_\infty < x_\infty$ for all $x \in \mathcal{X}$. Let $\boldsymbol{c} = \nabla l(\boldsymbol{x}^\star)$. Let $\boldsymbol{c}_i = \nabla l_i(\boldsymbol{x}^\star)$ and $b_i = \boldsymbol{c}_i^\top \boldsymbol{x}^\star$. Let $\mathcal{P} = \{\boldsymbol{x} : \boldsymbol{c}_i^\top \boldsymbol{x} \ge b_i,\, i = 1, ..., d, \|\boldsymbol{x}\|_\infty \le x_\infty\}$. Notice that by convexity

$$\mathcal{X} \subseteq \mathcal{P}. \tag{EC.2}$$

Notice by linear independence $\boldsymbol{x}^\star$ is the unique point such that $\boldsymbol{c}_i^\top \boldsymbol{x}^\star = b_i$, $i = 1, ..., d$. That is $\boldsymbol{x}^\star$ is an extreme point of the polytope $\mathcal{P}$. Since $-\nabla l(\boldsymbol{x}^\star) \in \text{relint}\, \mathcal{N}_{\mathcal{X}}(\boldsymbol{x}^\star)$, $\boldsymbol{x}^\star$ miniminizes $\boldsymbol{c}^\top \boldsymbol{x}$ over $\boldsymbol{x} \in \mathcal{P}$. Thus by Lemma EC.8

$$\boldsymbol{c}^\top(\boldsymbol{x} - \boldsymbol{x}^\star) \ge K\|\boldsymbol{c}\|\|\boldsymbol{x} - \boldsymbol{x}^\star\|, \qquad \forall \boldsymbol{x} \in \mathcal{P}\,. \tag{EC.3}$$

By convexity

$$l(\boldsymbol{x}) - l(\boldsymbol{x}^\star) \ge \boldsymbol{c}^\top(\boldsymbol{x} - \boldsymbol{x}^\star) \tag{EC.4}$$

combining (EC.2), (EC.3) and (EC.4) we see that

$$l(\boldsymbol{x}) - l(\boldsymbol{x}^\star) \ge K\|\boldsymbol{c}\|\|\boldsymbol{x} - \boldsymbol{x}^\star\|\,, \qquad \forall \boldsymbol{x} \in \mathcal{X}\,.$$

Thus we see that the function $l(\boldsymbol{x})$ is sharp on $\mathcal{X}$. Condition (D1) then follows by Lemma 1 since the function $l(\boldsymbol{x})$ is convex. □

EC.2.2. Finite number of Projections for Interior Optimum.

We say a projection step is trivial if $x \in \mathcal{X}$ and thus $\Pi_{\mathcal{X}}(x) = x$. Otherwise, we say the projection at x is non-trivial. We can show that in instances where the optimum is in the interior only a finite number of non-trivial projections are required.

PROPOSITION EC.1. *Under the assumptions of Theorem 2, if* $\mathcal{X}^\star$ *belongs to the interior of* $\mathcal{X}$, *then the number of (non-trivial) projection steps required by Projected Stochastic Gradient Descent is finite and bounded in expectation.*

Proof. Let the random variable N denote the number of non-trivial projections. By Theorem 2 we have that

$$\mathbb{P}\left(\min_{\boldsymbol{x}^\star \in \mathcal{X}^\star} \|\boldsymbol{x}_{t+1} - \boldsymbol{x}^\star\| \ge z\right) \le Je^{-It^\gamma z}\,.$$

We let $\tilde{z}$ be the distance from the set of optima to the boundary of $\mathcal{X}$, that is,

$$\tilde{z} = \min_{\boldsymbol{x}\in\mathcal{X}^\star, \boldsymbol{y}\notin\mathcal{X}} \|\boldsymbol{x}^\star - \boldsymbol{y}\|\,.$$

Since $\mathcal{X}^\star$ belongs to the interior of $\mathcal{X}$, we have that $\tilde{z} > 0$. Note that if a projection is non-trivial then $\min_{\boldsymbol{x}^\star\in\mathcal{X}^\star} ||\boldsymbol{x} - \boldsymbol{x}^\star|| \geq \tilde{z}$. Thus

$$N \leq \sum_{t=0}^{\infty} \mathbb{I}\Big[\min_{\boldsymbol{x}^\star\in\mathcal{X}^\star} \|\boldsymbol{x}_{t+1} - \boldsymbol{x}^\star\| \geq \tilde{z}\Big]$$

and so, as required,

$$\mathbb{E}[N] \leq \sum_{t=0}^{\infty} \mathbb{P}\left(\min_{\boldsymbol{x}^\star\in\mathcal{X}^\star} \|\boldsymbol{x}_{t+1} - \boldsymbol{x}^\star\| \geq \tilde{z}\right) \leq \sum_{t=0}^{\infty} Je^{-It^\gamma z} < \infty\,.$$

□

EC.2.3. Kiefer-Wolfowitz: Proof of Theorem 3

We now restate and prove Theorem 3.

THEOREM 3. *If Conditions (D1), (D2), (D3) hold and if*

$$\nu \leq \left(\frac{\kappa}{3cd^{\frac{1}{2}}}\right)^{\frac{1}{2}}$$

then the Kiefer-Wolfowitz algorithm satisfies

$$\mathbb{P}\left(\min_{\boldsymbol{x}\in\mathcal{X}^\star} \|\boldsymbol{x}_{t+1} - \boldsymbol{x}\| \geq z\right) \leq \hat{J}e^{-\frac{\hat{I}}{\alpha_t}z}, \quad \mathbb{E}\left[\min_{\boldsymbol{x}\in\mathcal{X}^\star} \|\boldsymbol{x}_{t+1} - \boldsymbol{x}\|\right] \leq \hat{K}\alpha_t, \quad \mathbb{E}\left[l(\boldsymbol{x}_{t+1}) - \min_{\boldsymbol{x}\in\mathcal{X}} l(\boldsymbol{x})\right] \leq \hat{L}\alpha_t$$

where above $\hat{J}, \hat{I}, \hat{K}, \hat{L}$ are positive constants.

Proof. The proof here combines the proof ideas for Kiefer-Wolfowitz, see Fabian (1967) (or, more recently, Broadie et al. (2011)), with the proof in Theorem 2. As with the proof of Theorem 2 our goal is to verify Conditions C1 and C2, so that we can apply Theorem 1.

We can write the KW recursion as

$$\boldsymbol{y}_{t+1} = \boldsymbol{x}_t - \alpha_t \nabla l(\boldsymbol{x}_t) + \alpha_t \boldsymbol{\delta}_t + \alpha_t \boldsymbol{\epsilon}_t \tag{EC.5}$$

$$\boldsymbol{x}_{t+1} = \Pi_{\mathcal{X}}(\boldsymbol{y}_{t+1}) \tag{EC.6}$$

where

$$\boldsymbol{\delta}_t = \nabla l(\boldsymbol{x}) - \frac{\boldsymbol{l}(\boldsymbol{x}_t + \boldsymbol{\nu}_t) - \boldsymbol{l}(\boldsymbol{x}_t - \boldsymbol{\nu}_t)}{2\nu_t}$$

$$\boldsymbol{\epsilon}_t = \frac{\boldsymbol{l}(\boldsymbol{x}_t + \boldsymbol{\nu}_t) - \boldsymbol{l}(\boldsymbol{x}_t - \boldsymbol{\nu}_t)}{2\nu_t} - \frac{\boldsymbol{l}(\boldsymbol{x}_t + \boldsymbol{\nu}_t, \hat{w}_t^+) - \boldsymbol{l}(\boldsymbol{x}_t - \boldsymbol{\nu}_t, \hat{w}_t^-)}{2\nu_t}\,.$$

Letting $\boldsymbol{x}_t^\star$ be the projection of $\boldsymbol{x}_t$ onto $\mathcal{X}^\star$, then

$$\begin{aligned}\|\boldsymbol{x}_{t+1}-\boldsymbol{x}_{t+1}^\star\|^2 &\le \|\boldsymbol{x}_{t+1}-\boldsymbol{x}_t^\star\|^2\\ &\le \|\boldsymbol{y}_{t+1}-\boldsymbol{x}_t^\star\|^2 = \|\boldsymbol{y}_{t+1}-\boldsymbol{x}_t+\boldsymbol{x}_t-\boldsymbol{x}_t^\star\|^2\\ &= \|\boldsymbol{x}_t-\boldsymbol{x}_t^\star\|^2 - 2\alpha_t\nabla l(\boldsymbol{x}_t)^\top(\boldsymbol{x}_t-\boldsymbol{x}_t^\star)+2\alpha_t\boldsymbol{\delta}_t^\top(\boldsymbol{x}_t-\boldsymbol{x}_t^\star)+2\alpha_t\boldsymbol{\epsilon}_t^\top(\boldsymbol{x}_t-\boldsymbol{x}_t^\star)+\|\boldsymbol{y}_{t+1}-\boldsymbol{x}_t\|^2.\end{aligned}$$

The first inequality above follows since $\boldsymbol{x}_{t+1}^\star$ is a projection. The second follows since $\boldsymbol{x}_{t+1}$ is a projection. We then expand.

We let E_t be the positive number defined below in (EC.16). On the event $\{\|\boldsymbol{x}_t-\boldsymbol{x}_t^\star\|\ge E_t\}$, we have

$$\begin{aligned}&\|\boldsymbol{x}_{t+1}-\boldsymbol{x}_{t+1}^\star\|\\ \le&\|\boldsymbol{x}_t-\boldsymbol{x}_t^\star\|\sqrt{1-2\alpha_t\nabla l(\boldsymbol{x}_t)^\top\frac{(\boldsymbol{x}_t-\boldsymbol{x}_t^\star)}{\|\boldsymbol{x}_t-\boldsymbol{x}_t^\star\|^2}+2\alpha_t\boldsymbol{\delta}_t^\top\frac{(\boldsymbol{x}_t-\boldsymbol{x}_t^\star)}{\|\boldsymbol{x}_t-\boldsymbol{x}_t^\star\|^2}+2\alpha_t\boldsymbol{\epsilon}_t^\top\frac{(\boldsymbol{x}_t-\boldsymbol{x}_t^\star)}{\|\boldsymbol{x}_t-\boldsymbol{x}_t^\star\|^2}+\frac{\|\boldsymbol{y}_{t+1}-\boldsymbol{x}_t\|^2}{\|\boldsymbol{x}_t-\boldsymbol{x}_t^\star\|^2}}\\ \le&\|\boldsymbol{x}_t-\boldsymbol{x}_t^\star\|\\ &-\alpha_t\nabla l(\boldsymbol{x}_t)^\top\frac{(\boldsymbol{x}_t-\boldsymbol{x}_t^\star)}{\|\boldsymbol{x}_t-\boldsymbol{x}_t^\star\|} &\text{(EC.7)}\\ &+\alpha_t\boldsymbol{\delta}_t^\top\frac{(\boldsymbol{x}_t-\boldsymbol{x}_t^\star)}{\|\boldsymbol{x}_t-\boldsymbol{x}_t^\star\|} &\text{(EC.8)}\\ &+\alpha_t\boldsymbol{\epsilon}_t^\top\frac{(\boldsymbol{x}_t-\boldsymbol{x}_t^\star)}{\|\boldsymbol{x}_t-\boldsymbol{x}_t^\star\|} &\text{(EC.9)}\\ &+\frac{\|\boldsymbol{y}_{t+1}-\boldsymbol{x}_t\|^2}{2\|\boldsymbol{x}_t-\boldsymbol{x}_t^\star\|}. &\text{(EC.10)}\end{aligned}$$

We now analyse the conditional expectation of the four terms above. Term (EC.7) is bounded using to the sharpness condition (D1)

$$-\nabla l(\boldsymbol{x}_t)^\top\frac{(\boldsymbol{x}_t-\boldsymbol{x}_t^\star)}{\|\boldsymbol{x}-\boldsymbol{x}^\star\|}\le-\kappa. \tag{EC.11}$$

Term (EC.8) is bounded by the Taylor approximation condition (D3). Specifically

$$\boldsymbol{\delta}_t^\top\frac{(\boldsymbol{x}_t-\boldsymbol{x}_t^\star)}{\|\boldsymbol{x}_t-\boldsymbol{x}_t^\star\|}\le\|\boldsymbol{\delta}_t\|=\left\|\nabla l(\boldsymbol{x})-\frac{\boldsymbol{l}(\boldsymbol{x}_t+\boldsymbol{\nu})-\boldsymbol{l}(\boldsymbol{x}_t-\boldsymbol{\nu})}{2\nu}\right\|\le cd^{\frac{1}{2}}\nu^2. \tag{EC.12}$$

Term (EC.9) has zero mean

$$\mathbb{E}\left[\boldsymbol{\epsilon}_t^\top\frac{(\boldsymbol{x}_t-\boldsymbol{x}_t^\star)}{\|\boldsymbol{x}_t-\boldsymbol{x}_t^\star\|}\middle|\mathcal{F}_t\right]=0. \tag{EC.13}$$

For Term (EC.10), $\boldsymbol{y}_{t+1}=\boldsymbol{x}_t-\alpha_t\boldsymbol{c}_t$

$$\mathbb{E}\left[\frac{\|\boldsymbol{y}_{t+1}-\boldsymbol{x}_t\|^2}{\|\boldsymbol{x}_t-\boldsymbol{x}_t^\star\|}\middle|\mathcal{F}_t\right]\le\frac{\alpha_t^2}{E_t}\mathbb{E}\left[||\boldsymbol{c}_t||^2\middle|\mathcal{F}_t\right]\le\frac{\alpha_t^2\sigma_l^2}{2E_t\nu^2} \tag{EC.14}$$

Since the variance of $l(\boldsymbol{x}, \hat{w})$ is bounded (by σ_l^2), the variance of $||\boldsymbol{c}_t||$ is bounded. Above, we let σ_l^2/ν^2 define this upper bound. Applying bounds (EC.11), (EC.12), (EC.13) and (EC.14) respectively to the terms (EC.7), (EC.8), (EC.9) and (EC.10) gives

$$\mathbb{E}\Big[\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_{t+1}^\star\| \Big| \mathcal{F}_t\Big] \leq \|\boldsymbol{x}_t - \boldsymbol{x}_t^\star\| - \alpha_t\kappa + \alpha_t c\nu^2 + \frac{\alpha_t^2\sigma_l^2}{2E_t\nu^2}\,. \tag{EC.15}$$

Notice if we choose

$$\nu \leq \sqrt{\frac{\kappa}{3cd^{\frac{1}{2}}}} \qquad \text{and} \qquad E_t = \frac{3\sigma_l^2}{4\nu^2\kappa}\alpha_t\,, \tag{EC.16}$$

then application to (EC.15) gives

$$\mathbb{E}\Big[\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_{t+1}^\star\| \Big| \mathcal{F}_t\Big] \leq \|\boldsymbol{x}_t - \boldsymbol{x}_t^\star\| - \alpha_t\frac{\kappa}{3} \qquad \text{on the event} \qquad \Big\{\|\boldsymbol{x}_t - \boldsymbol{x}_t^\star\| \geq \frac{3\sigma_l^2}{\nu^4\kappa}\alpha_t\Big\}\,.$$

This verifies that Condition (C1) of Theorem 1 holds.

We must also verify Condition (C2). (The argument that follows is more-or-less identical to the verification of (C2) in Theorem 2.) For this notice that

$$\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_{t+1}^\star\| \leq \|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t^\star\| \leq \|\boldsymbol{y}_{t+1} - \boldsymbol{x}_t^\star\| \leq \|\boldsymbol{y}_{t+1} - \boldsymbol{x}_t\| + \|\boldsymbol{x}_t - \boldsymbol{x}_t^\star\| = \alpha_t\|\boldsymbol{c}_t\| + \|\boldsymbol{x}_t - \boldsymbol{x}_t^\star\|\,.$$

and

$$\begin{aligned}\|\boldsymbol{x}_t - \boldsymbol{x}_t^\star\| &\leq \|\boldsymbol{x}_t - \boldsymbol{x}_{t+1}^\star\| \leq \|\boldsymbol{x}_t - \boldsymbol{x}_{t+1}\| + \|\boldsymbol{x}_{t+1} - \boldsymbol{x}_{t+1}^\star\| \\ &\leq \|\boldsymbol{y}_{t+1} - \boldsymbol{x}_t\| + \|\boldsymbol{x}_{t+1} - \boldsymbol{x}_{t+1}^\star\| = \alpha_t\|\boldsymbol{c}_t\| + \|\boldsymbol{x}_{t+1} - \boldsymbol{x}_{t+1}^\star\|\,.\end{aligned}$$

Thus

$$\Big|\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_{t+1}^\star\| - \|\boldsymbol{x}_t - \boldsymbol{x}_t^\star\|\Big| \leq \alpha_t\|\boldsymbol{c}_t\| \tag{EC.17}$$

Since Condition (D2) holds, we take $M \geq \sup_t \mathbb{E}\big[e^{\lambda\|\boldsymbol{c}_t\|}\big|\mathcal{F}_t\big]$. We let Y be the random variable with CCDF: $\mathbb{P}(Y \geq y) = 1 \wedge (Me^{-\lambda y})$. Thus for $y \in \mathbb{R}_+$

$$\begin{aligned}\mathbb{P}\left(|f(\boldsymbol{x}_{t+1}) - f(\boldsymbol{x}_t)| \geq \alpha_t y \middle| \mathcal{F}_t\right) &\leq \mathbb{P}\left(\|\boldsymbol{c}_t\| \geq y \middle| \mathcal{F}_t\right) \\ &\leq \min\left\{1, e^{-\lambda y}\mathbb{E}\big[e^{\lambda\|\boldsymbol{c}_t\|}\big|\mathcal{F}_t\big]\right\} \leq \mathbb{P}(Y \geq y)\end{aligned}$$

Above, we apply (EC.17) and a Chernoff bound. From this inequality above, we see that Condition (C2) follows from Condition (D2).

We can now apply Theorem 1 which gives:

$$\mathbb{P}\Big(\min_{\boldsymbol{x}\in\mathcal{X}^\star}\|\boldsymbol{x}_{t+1} - \boldsymbol{x}\| \geq z\Big) \leq \hat{I}e^{-\frac{\hat{J}}{\alpha_t}z} \quad \text{and} \quad \mathbb{E}\left[\min_{\boldsymbol{x}\in\mathcal{X}^\star}\|\boldsymbol{x}_{t+1} - \boldsymbol{x}\|\right] \leq \hat{K}\alpha_t$$

for constants $\hat{I}$, $\hat{J}$ and $\hat{K}$. Since we also assume in addition that $l : \mathcal{X} \to \mathbb{R}$ is Lispchitz continuous (with Lipschitz constant $\hat{L}/\hat{K}$) we have, as required,

$$\mathbb{E}\left[l(\boldsymbol{x}_{t+1}) - \min_{\boldsymbol{x}\in\mathcal{X}} l(\boldsymbol{x})\right] \leq \frac{\hat{L}}{\hat{K}}\mathbb{E}\left[\min_{\boldsymbol{x}\in\mathcal{X}^\star}\|\boldsymbol{x}_{t+1} - \boldsymbol{x}\|\right] \leq \hat{L}\alpha_t\,.$$

□

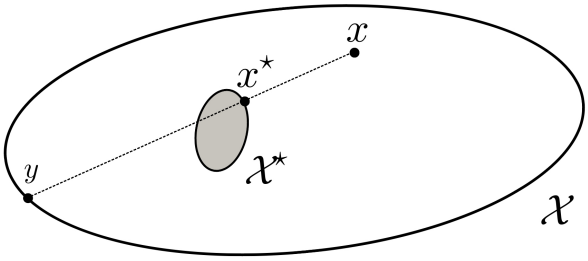


Figure EC.1 Here we give the $\boldsymbol{x}$, $\boldsymbol{x}^\star$ and $\boldsymbol{y}$ terms from Lemma EC.2. Here we take $\boldsymbol{x}$ project onto $\mathcal{X}^\star$ to give $\boldsymbol{x}^\star$, then $\boldsymbol{y}$ is the boundary value on the line passing from $\boldsymbol{x}$ to $\boldsymbol{x}^\star$

EC.2.4. Stochastic Frank-Wolfe: Proof of Theorem 4

The main aim of this section is to prove Theorem 4. We also show that the distance function satisfies the conditions of our main theorem. This suggests that if the objective function behaves linearly rather than quadratically near the optimum, we should anticipate faster convergence. We also discuss how linear convergence can hold for Stochastic Frank-Wolfe in the same manner that we proved for Projected Stochastic Gradient Descent.

Before proceeding with the proof of Theorem 4 we require a couple of lemmas. Lemma EC.2 is used to show that there is sufficient negative drift in the Frank-Wolfe algorithm.

LEMMA EC.2. *If Condition (D1) and Condition (E2) hold then there exists a $\hat{\kappa} > 0$ such that for all $\boldsymbol{x} \in \mathcal{X} \backslash \mathcal{X}^\star$ there exists $\boldsymbol{y} \in \mathcal{X}$ such that*

$$(\boldsymbol{y} - \boldsymbol{x})^\top \nabla l(\boldsymbol{x}) \leq -\hat{\kappa}\,.$$

Proof. The idea of the proof is as follows. The derivative from $\boldsymbol{x}$ and $\boldsymbol{x}^\star$ at $\boldsymbol{x}$ bounded above by $-\kappa$, by Assumption (D1). Since $\boldsymbol{x}^\star$ is in the interior by (E2), we can increase the directional derivative further by replacing $\boldsymbol{x}^\star$ with $\boldsymbol{y}$, where $\boldsymbol{y}$ is the point on the boundary of $\mathcal{X}$ on the line between $\boldsymbol{x}$ and $\boldsymbol{x}^\star$. See Figure EC.1. We now proceed with the formal argument.

By Condition (E2), there exists a constant $d > 0$ such that

$$\min_{\boldsymbol{x}^\star \in \mathcal{X}^\star, \boldsymbol{y} \in \partial\mathcal{X}} \|\boldsymbol{y} - \boldsymbol{x}^\star\| \geq d\,. \tag{EC.18}$$

(Here $\partial\mathcal{X} := \bar{\mathcal{X}} \backslash \mathcal{X}^\circ$ is the boundary of $\mathcal{X}$.) By Condition (D1), for all $\boldsymbol{x} \notin \mathcal{X}^\star$ there exists $\boldsymbol{x}^\star \in \mathcal{X}^\star$

$$\frac{(\boldsymbol{x}^\star - \boldsymbol{x})^\top}{\|\boldsymbol{x}^\star - \boldsymbol{x}\|} \nabla l(\boldsymbol{x}) \leq -\kappa\,. \tag{EC.19}$$

We let $\boldsymbol{y}(t) = \boldsymbol{x} + t(\boldsymbol{x}^\star - \boldsymbol{x})$ for $t \in \mathbb{R}$. Notice that

$$\frac{(\boldsymbol{y}(t) - \boldsymbol{x})^\top}{\|\boldsymbol{y}(t) - \boldsymbol{x}\|} = \frac{(\boldsymbol{x}^\star - \boldsymbol{x})^\top}{\|\boldsymbol{x}^\star - \boldsymbol{x}\|} \tag{EC.20}$$

Letting $t^\star = \max\{t : \boldsymbol{y}(t) \in \mathcal{X}\}$, we see that

$$\boldsymbol{y} := \boldsymbol{y}(t^\star) \in \delta\mathcal{X} \tag{EC.21}$$

Combining (EC.18-EC.21), we see that

$$(\boldsymbol{y}-\boldsymbol{x})^\top \nabla l(\boldsymbol{x}) = \|\boldsymbol{y}-\boldsymbol{x}\| \frac{(\boldsymbol{x}^\star-\boldsymbol{x})^\top}{\|\boldsymbol{x}^\star-\boldsymbol{x}\|}\nabla l(\boldsymbol{x}) \leq -d\kappa =: -\hat{\kappa}$$

as required. □

We now restate and prove Theorem 4

THEOREM 4. *For learning rates of the form $\alpha_t = a/(u+t)^\gamma$ with $a,u>0$ and $\gamma \in [0,1]$, if Conditions (D1), (D2), (E1) and (E2) hold and if $m_t \geq (3\sigma/\kappa\alpha_t)^2$ then the stochastic Frank-Wolfe algorithm satisfies*

$$\mathbb{P}\left(l(\boldsymbol{x}_{t+1}) - \min_{\boldsymbol{x}\in\mathcal{X}} l(\boldsymbol{x}) \geq z\right) \leq Ie^{-\frac{J}{\alpha_t}z},$$

for constants I,J.

Proof of Theorem 4. The proof here combines ideas from Theorem 2 with the adjustments for stochastic effects for the Frank-Wolfe algorithm given in Theorem 3 from Hazan and Luo (2016).

In the proof we define $D := \max_{\boldsymbol{x},\boldsymbol{v}} \|\boldsymbol{x}-\boldsymbol{v}\|$ and we let $\epsilon(\boldsymbol{x}_t) = l(\boldsymbol{x}_t) - l(\boldsymbol{x}^\star)$ and we define σ such that

$$\mathbb{E}\left[\|\nabla l(\boldsymbol{x}_t) - \boldsymbol{c}_t^i\|^2\right] \leq \sigma^2 \qquad\qquad \forall i,t.$$

(Note that σ is finite by the moment generating function condition (D2))

By Condition (E1)

$$\begin{aligned}\frac{\epsilon(\boldsymbol{x}_{t+1})^2}{2} - \frac{\epsilon(\boldsymbol{x}_t)^2}{2} &\leq \epsilon(\boldsymbol{x}_t)\left(\boldsymbol{x}_{t+1}-\boldsymbol{x}_t\right)^\top \nabla\epsilon(\boldsymbol{x}_t) + \frac{K}{2}\|\boldsymbol{x}_{t+1}-\boldsymbol{x}_t\|^2 \\ &= \alpha_t\epsilon(\boldsymbol{x}_t)(\boldsymbol{v}_t-\boldsymbol{x}_t)^\top\boldsymbol{c}_t + \alpha_t\epsilon(\boldsymbol{x}_t)(\boldsymbol{v}_t-\boldsymbol{x}_t)^\top\left[\nabla\epsilon(\boldsymbol{x}_t)-\boldsymbol{c}_t\right] + \frac{K}{2}\alpha_t^2\|\boldsymbol{v}_t-\boldsymbol{x}_t\|^2.\end{aligned} \tag{EC.22}$$

We now consider the event where the following bound holds

$$\left\{\epsilon(\boldsymbol{x}_t) \geq \frac{3\alpha_t K D^2}{2\kappa}\right\}. \tag{EC.23}$$

Thus

$$\begin{aligned}&\epsilon(\boldsymbol{x}_{t+1}) \\ &\leq \sqrt{\epsilon(\boldsymbol{x}_t)^2 + 2\alpha_t\epsilon(\boldsymbol{x}_t)(\boldsymbol{v}_t-\boldsymbol{x}_t)^\top\boldsymbol{c}_t + 2\alpha_t\epsilon(\boldsymbol{x}_t)(\boldsymbol{v}_t-\boldsymbol{x}_t)^\top\left[\nabla\epsilon(\boldsymbol{x}_t)-\boldsymbol{c}_t\right] + K\alpha_t^2\|\boldsymbol{v}_t-\boldsymbol{x}_t\|^2} \\ &= \epsilon(\boldsymbol{x}_t)\sqrt{1 + 2\frac{\alpha_t}{\epsilon(\boldsymbol{x}_t)}(\boldsymbol{v}_t-\boldsymbol{x}_t)^\top\boldsymbol{c}_t + 2\frac{\alpha_t}{\epsilon(\boldsymbol{x}_t)}(\boldsymbol{v}_t-\boldsymbol{x}_t)^\top\left[\nabla\epsilon(\boldsymbol{x}_t)-\boldsymbol{c}_t\right] + K\frac{\alpha_t^2}{\epsilon(\boldsymbol{x}_t)^2}\|\boldsymbol{v}_t-\boldsymbol{x}_t\|^2} \\ &\leq \epsilon(\boldsymbol{x}_t) + \alpha_t(\boldsymbol{v}_t-\boldsymbol{x}_t)^\top\boldsymbol{c}_t + \alpha_t(\boldsymbol{v}_t-\boldsymbol{x}_t)^\top\left[\nabla\epsilon(\boldsymbol{x}_t)-\boldsymbol{c}_t\right] + \frac{K}{2}\frac{\alpha_t^2}{\epsilon(\boldsymbol{x}_t)}\|\boldsymbol{v}_t-\boldsymbol{x}_t\|^2 \\ &\leq \epsilon(\boldsymbol{x}_t) + \alpha_t(\boldsymbol{y}_t-\boldsymbol{x}_t)^\top\boldsymbol{c}_t + \alpha_t\|\boldsymbol{v}_t-\boldsymbol{x}_t\|\|\nabla\epsilon(\boldsymbol{x}_t)-\boldsymbol{c}_t\| + \frac{\alpha_t\kappa}{3} \\ &\leq \epsilon(\boldsymbol{x}_t) + \alpha_t(\boldsymbol{y}_t-\boldsymbol{x}_t)^\top\boldsymbol{c}_t + \alpha_t D\|\nabla\epsilon(\boldsymbol{x}_t)-\boldsymbol{c}_t\| + \frac{\alpha_t\kappa}{3}\end{aligned} \tag{EC.24}$$

In the first inequality, above we rearrange the expression (EC.22). In the second inequality, we apply the inequality $\sqrt{1+z} \leq 1 + z/2$. In the third equality, we note that $\boldsymbol{v}_t^\top \boldsymbol{c}_t \leq \boldsymbol{y}_t^\top \boldsymbol{c}_t$, by the definition of $\boldsymbol{v}_t$ (11b). Here we let $\boldsymbol{y}_t \in \mathcal{X}$ be as defined in Lemma EC.2. Also, we apply the Cauchy-Schwarz Inequality and the bound (EC.23). In the final inequality we note that $\|\boldsymbol{v}_t - \boldsymbol{x}_t\| \geq D$.

Taking the conditional expectation of (EC.24), we see that, on the event (EC.23), the following holds

$$\begin{aligned}\mathbb{E}\left[l(\boldsymbol{x}_{t+1}) - l(\boldsymbol{x}_t)|\mathcal{F}_t\right] &= \mathbb{E}\left[\epsilon(\boldsymbol{x}_{t+1}) - \epsilon(\boldsymbol{x}_t)|\mathcal{F}_t\right] \\ &\leq \alpha_t(\boldsymbol{y}_t - \boldsymbol{x}_t)^\top \mathbb{E}[\boldsymbol{c}_t|\mathcal{F}_t] + \alpha_t D\mathbb{E}\left[\|\nabla\epsilon(\boldsymbol{x}_t) - \boldsymbol{c}_t\||\mathcal{F}_t\right] + \frac{\alpha_t\kappa}{3} \\ &\leq -\alpha_t\hat{\kappa} + \alpha_t D\mathbb{E}\left[\|\nabla l(\boldsymbol{x}_t) - \boldsymbol{c}_t\||\mathcal{F}_t\right] + \frac{\alpha_t\kappa}{3}. \end{aligned} \tag{EC.25}$$

Notice that, since $m_t \geq (3\sigma D/\hat{\kappa}\alpha_t)^2$,

$$\mathbb{E}\left[\|\nabla l(\boldsymbol{x}_t) - \boldsymbol{c}_t\|\,|\mathcal{F}_t\right] \leq \sqrt{\mathbb{E}\left[\|\nabla l(\boldsymbol{x}_t) - \boldsymbol{c}_t\|^2\,|\mathcal{F}_t\right]} \leq \frac{\sigma}{\sqrt{m_t}} \leq \frac{\hat{\kappa}}{3D}.$$

Now applying this inequality to (EC.25) gives

$$\mathbb{E}\left[l(\boldsymbol{x}_{t+1}) - l(\boldsymbol{x}_t)|\mathcal{F}_t\right] \leq -\alpha_t\frac{\hat{\kappa}}{3}$$

on the event $l(\boldsymbol{x}_t) - l(\boldsymbol{x}^\star) \geq 3KD/\alpha\hat{\kappa}$. Thus Condition (C1) is met.

For Condition (C2), since l is Lipschitz continuous and the set $\mathcal{X}$ is bounded we have

$$\|l(\boldsymbol{x}_{t+1}) - l(\boldsymbol{x}_t)\| \leq L\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\| \leq \alpha_t L\|\boldsymbol{v}_t - \boldsymbol{x}_t\| \leq 2\alpha_t L\max_{\boldsymbol{x}\in\mathcal{X}}\|\boldsymbol{x}\|.$$

Thus we see that Condition (C2) holds with a constant upperbound $Y = 2L\max_{\boldsymbol{x}\in\mathcal{X}}\|\boldsymbol{x}\|$.

Here we see that the conditions of Theorem 1 are met, and thus we have that

$$\mathbb{P}\left(l(\boldsymbol{x}_{t+1}) - l(\boldsymbol{x}_t) \geq z\right) \leq Ie^{-\frac{J}{\alpha_t}z},$$

as required. □

EC.2.4.1. Cones satisfy Condition (E1) Below we recall that we define the matrix norm $\|\cdot\|_S$ for a positive semi-definite matrix S by

$$\|\boldsymbol{x}\|_S := \sqrt{\boldsymbol{x}^\top S\boldsymbol{x}}$$

LEMMA EC.3. *For a symmetric positive definite matrix S, the distance function*

$$d_{\mathcal{X}^\star}(\boldsymbol{x}) = \min_{x^\star\in\mathcal{X}^\star}\|\boldsymbol{x} - \boldsymbol{x}^\star\|_S$$

satisfies Condition (E1).

Proof. We must show that the function

$$d_{\mathcal{X}^\star}(\boldsymbol{x})^2 = \min_{x^\star \in \mathcal{X}^\star} \|\boldsymbol{x} - \boldsymbol{x}^\star\|_S^2$$

is strongly convex.

Given $\boldsymbol{x}$, we let $\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}^\star \in \mathcal{X}^\star} \|\boldsymbol{x} - \boldsymbol{x}^\star\|_S$. By the Envelope Theorem,

$$\nabla d_{\mathcal{X}}(\boldsymbol{x})^2 = 2S(\boldsymbol{x} - \boldsymbol{x}^\star) \tag{EC.26}$$

Also

$$\|\boldsymbol{x} - \boldsymbol{y}\|_S^2 \leq \lambda_{\max}(S)\|\boldsymbol{x} - \boldsymbol{y}\|^2 \tag{EC.27}$$

where $\lambda_{\max}(S)$ is the maximum eigenvalue of S.

Now for any $\boldsymbol{y}$ and $\boldsymbol{x}$,

$$\begin{aligned} d_{\mathcal{X}^\star}(\boldsymbol{y})^2 = \min_{\boldsymbol{y}^\star \in \mathcal{X}^\star} \|\boldsymbol{y} - \boldsymbol{y}^\star\|_S^2 &\leq \|\boldsymbol{y} - \boldsymbol{x}^\star\|^2 \\ &= \|\boldsymbol{y} - \boldsymbol{x} + \boldsymbol{x} - \boldsymbol{x}^\star\|_S^2 \\ &= \|\boldsymbol{x} - \boldsymbol{x}^\star\|_S^2 + 2(\boldsymbol{y} - \boldsymbol{x})^\top S(\boldsymbol{x} - \boldsymbol{x}^\star) + \|\boldsymbol{y} - \boldsymbol{x}\|_S^2 \\ &\leq d_{\mathcal{X}^\star}(\boldsymbol{x})^2 + (\boldsymbol{y} - \boldsymbol{x})\nabla d_{\mathcal{X}^\star}^2(\boldsymbol{x}) + \lambda_{\max}(S)\|\boldsymbol{x} - \boldsymbol{y}\|^2 \end{aligned}$$

In the first inequality, we apply the sub-optimality of $\boldsymbol{x}^\star$ with respect to the point $\boldsymbol{y}$. In the second inequality, we apply (EC.26) and (EC.27). Thus from the above inequality we see that $d_{\mathcal{X}^\star}(x)^2$ is a $\lambda_{\max}(S)$–smoothly convex function, as required. □

EC.2.5. Stochastic Frank-Wolfe Boundary case

PROPOSITION EC.2. *For learning rates of the form $\alpha_t = a/(u+t)^\gamma$ with $a, u > 0$ and $\gamma \in [0,1)$, if Conditions (D1), (D2), (E1) and (E2) hold and if $m_t \geq (2\sigma E/KD\alpha_t)^2$ then the stochastic Frank-Wolfe algorithm satisfies*

$$\limsup_{t\to\infty} \frac{1}{\sqrt{\alpha_t}} \mathbb{E}\left[l(\boldsymbol{x}_{t+1}) - l(\boldsymbol{x}^\star)\right] < \infty,$$

Proof. The proof here combines ideas from Theorem 2 with the adjustments for stochastic effects for the Frank-Wolfe algorithm given in Theorem 3 from Hazan and Luo (2016).

In the proof we let $\epsilon(\boldsymbol{x}_t) = l(\boldsymbol{x}_t) - l(\boldsymbol{x}^\star)$ and we define σ such that

$$\mathbb{E}\left[\|\nabla l(\boldsymbol{x}_t) - \boldsymbol{c}_t^i\|^2\right] \leq \sigma^2 \qquad \forall i, t.$$

(Note that σ is finite by the moment generating function condition (D2)). We define $D := \max_{\boldsymbol{x},\boldsymbol{v}} \|\boldsymbol{x} - \boldsymbol{v}\|$ and $E := \max_{\boldsymbol{x}} \epsilon(\boldsymbol{x})$.

By Condition (E1)

$$\begin{aligned}
&\frac{\epsilon(\boldsymbol{x}_{t+1})^2}{2} - \frac{\epsilon(\boldsymbol{x}_t)^2}{2} \\
&\leq \epsilon(\boldsymbol{x}_t)\big(\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\big)^\top \nabla\epsilon(\boldsymbol{x}_t) + \frac{K}{2}\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\|^2 \\
&= \alpha_t \epsilon(\boldsymbol{x}_t)(\boldsymbol{v}_t - \boldsymbol{x}_t)^\top \boldsymbol{c}_t + \alpha_t \epsilon(\boldsymbol{x}_t)(\boldsymbol{v}_t - \boldsymbol{x}_t)^\top \left[\nabla\epsilon(\boldsymbol{x}_t) - \boldsymbol{c}_t\right] + \frac{K}{2}\alpha_t^2\|\boldsymbol{v}_t - \boldsymbol{x}_t\|^2 \\
&\leq \alpha_t \epsilon(\boldsymbol{x}_t)(\boldsymbol{x}^\star - \boldsymbol{x}_t)^\top \boldsymbol{c}_t + \alpha_t \epsilon(\boldsymbol{x}_t)\|\boldsymbol{v}_t - \boldsymbol{x}_t\|\|\nabla\epsilon(\boldsymbol{x}_t) - \boldsymbol{c}_t\| + \frac{K}{2}\alpha_t^2\|\boldsymbol{v}_t - \boldsymbol{x}_t\|^2 \\
&\leq \alpha_t \epsilon(\boldsymbol{x}_t)(\boldsymbol{x}^\star - \boldsymbol{x}_t)^\top \boldsymbol{c}_t + \alpha_t ED\|\nabla\epsilon(\boldsymbol{x}_t) - \boldsymbol{c}_t\| + \alpha_t^2\frac{KD^2}{2}
\end{aligned} \tag{EC.28}$$

Notice that, since $m_t \geq (2\sigma E/KD\alpha_t)^2$,

$$\mathbb{E}\big[\|\nabla l(\boldsymbol{x}_t) - \boldsymbol{c}_t\| \,|\mathcal{F}_t\big] \leq \sqrt{\mathbb{E}\left[\|\nabla l(\boldsymbol{x}_t) - \boldsymbol{c}_t\|^2 \,|\mathcal{F}_t\right]} \leq \frac{\sigma}{\sqrt{m_t}} \leq \alpha_t \frac{KD}{2E}. \tag{EC.29}$$

Taking expectations in (EC.28) gives

$$\begin{aligned}
&\mathbb{E}\left[\frac{\epsilon(\boldsymbol{x}_{t+1})^2}{2}\right] - \mathbb{E}\left[\frac{\epsilon(\boldsymbol{x}_t)^2}{2}\right] \\
&\leq \alpha_t \mathbb{E}\left[\epsilon(\boldsymbol{x}_t)\boldsymbol{c}_t^\top(\boldsymbol{x}^\star - \boldsymbol{x}_t)\right] + \alpha_t ED\mathbb{E}\left[\|\nabla\epsilon(\boldsymbol{x}_t) - \boldsymbol{c}_t\|\right] + \alpha_t^2\frac{KD^2}{2} \\
&\leq \alpha_t \mathbb{E}\left[\epsilon(\boldsymbol{x}_t)\nabla\epsilon(\boldsymbol{x}_t)^\top(\boldsymbol{x}^\star - \boldsymbol{x}_t)\right] + \alpha_t ED\mathbb{E}\left[\|\nabla\epsilon(\boldsymbol{x}_t) - \boldsymbol{c}_t\|\right] + \alpha_t^2\frac{KD^2}{2} \\
&\leq \alpha_t \mathbb{E}\left[\left(\frac{\nabla\epsilon(\boldsymbol{x}_t)^2}{2}\right)^\top(\boldsymbol{x}^\star - \boldsymbol{x}_t)\right] + \alpha_t^2 KD^2 \\
&\leq -\alpha_t\frac{\epsilon(\boldsymbol{x}_t)^2}{2} + \alpha_t^2 KD^2
\end{aligned}$$

In the third equality above, we apply (EC.29). In the final equality, we apply the convexity of $\epsilon(\boldsymbol{x})^2$. Thus, we see that

$$\mathbb{E}\left[\frac{\epsilon(\boldsymbol{x}_{t+1})^2}{2}\right] \leq (1-\alpha_t)\mathbb{E}\left[\frac{\epsilon(\boldsymbol{x}_t)^2}{2}\right] + \alpha_t^2 KD^2.$$

Consequently, by Lemma EC.5 (and Lemma EC.4) given below

$$\limsup_{t\to\infty} \frac{1}{\alpha_t}\mathbb{E}[\epsilon(\boldsymbol{x}_{t+1})^2/2] < \infty$$

Thus

$$\limsup_{t\to\infty} \frac{\mathbb{E}[\epsilon(\boldsymbol{x}_{t+1})]}{\sqrt{\alpha_t}} \leq \left(\frac{\mathbb{E}[\epsilon(\boldsymbol{x}_{t+1})^2]}{\alpha_t}\right)^{1/2} < \infty.$$

□

We require the following techincal lemma which we then extend.

LEMMA EC.4. *If ξ_n is a positive sequence such that*

$$\xi_{n+1} \le \xi_n\Big(1 - A\alpha_n\Big) + \alpha_n B$$

and

$$\sum_{n=1}^{\infty} \alpha_n = \infty, \qquad \limsup_{n\to\infty} \alpha_n \le 0$$

then

$$\limsup_{n\to\infty} \xi_n \le \frac{B}{A}\,.$$

Proof. Rearranging gives

$$\xi_{n+1} - \xi_n \le -\alpha_n(A\xi_n - B).$$

If $\xi_n > B/A + \epsilon$ for some $\epsilon > 0$ then

$$\xi_{n+1} - \xi_n \le -\alpha_n(A\xi_n - B) \le -\alpha_n(A[B/A + \epsilon] - B) = -\alpha_n A\epsilon\,.$$

So ξ_n is decreasing when $\xi_n > B/A + \epsilon$ holds and, since $\sum_n \alpha_n = \infty$, there exists N s.t. $\xi_N \le B/A + \epsilon$. Let N_0 be the first value of N where $\xi_N \le B/A + \epsilon$ occurs.

Notice, ξ_n can only increase when $\xi_n \le B/A + \epsilon$, and since ξ_n is a positive then

$$\xi_{n+1} \le \xi_n + \alpha_n B\,.$$

Thus, we see that

$$\xi_n \le \frac{B}{A} + \epsilon + \alpha_n A\epsilon, \qquad \forall n \ge N_0\,.$$

Therefore

$$\limsup_{n\to\infty} \xi_n \le \frac{B}{A} + \epsilon + \limsup_{n\to\infty} \alpha_n B \le \frac{B}{A} + \epsilon\,.$$

Since ϵ is arbitrary the results holds. □

The following is an extension of the above lemma. Note for $\beta_n = \alpha_n = a/(u+t)^\gamma$ below and , for $\gamma < 1$, we can take $C = 0$ below. (We can consider the case $\gamma = 1$, but we require to take a sufficiently small.)

LEMMA EC.5. *If ξ_n is a positive sequence such that*

$$\xi_{n+1} \le \xi_n\Big(1 - A\alpha_n\Big) + \alpha_n\beta_n B$$

and

$$\sum_{n=1}^{\infty} \alpha_n = \infty, \qquad \lim_{n\to\infty} \alpha_n = 0, \quad \frac{\beta_n}{\beta_{n+1}} \le (1 + C\alpha_n)$$

with $A > C$ then

$$\limsup_{n\to\infty} \frac{\xi_n}{\beta_n} \le \frac{A - C}{B}.$$

Proof. Since $\lim_{n\to\infty}\alpha_n = 0$, take N such that $\alpha_n < \delta$ for all $n \geq N$.

Now defining $\xi'_n = \xi_n/\beta_n$ for $n \geq N$ gives

$$\begin{aligned}
\xi'_{n+1} = &\frac{\xi_{n+1}}{\beta_{n+1}} \leq \frac{\beta_n}{\beta_{n+1}}\Big(1 - A\alpha_n\Big)\xi'_n + \alpha_n \frac{\beta_n}{\beta_{n+1}} B \\
&\leq (1 + C\alpha_n)\,(1 - A\alpha_n)\,\xi'_n + \alpha_n\,(1 + C\alpha_n)\,B \\
&\leq (1 - (A - C + \delta)\alpha_n)\,\xi'_n + \alpha_n(1 + C\delta)B \\
&= (1 - A'\alpha_n)\xi'_n + \alpha_n B'
\end{aligned}$$

where we define $A' = A - C + \delta$ and $B' = (1 + C\delta)B$. Applying Lemma EC.4 gives

$$\limsup_{n\to\infty} \xi'_n \leq \frac{A'}{B'},$$

which recalling the definitions of ξ'_n, A',B' and recalling that δ is arbitrary gives the result. □

EC.2.6. Appendix to Section 3.6 : Linear Convergence Proofs

As discussed, our proof follows the main argument of Theorem 3.2 of Davis et al. (2019). We divide the procedure into S stages. We consider PSGD with constant step size within each stage, as defined in (12). The task of each stage is to half the error with the optimum. We apply our bound Lemma EC.6, which is a stronger concentration bound than Theorem 4.1, used in Davis et al. (2019). This leads to some improvements in the bounds found there.

EC.2.6.1. Exponential Concentration for constant step-size and unbounded state-space. Below, we state an exponential concentration bound for constant step sizes. We do not require the function $f(x)$ or the set $\mathcal{X}$ to be bounded (or constrained) for this result to hold.

LEMMA EC.6. *For constant step sizes* α

$$\mathbb{P}(f(\boldsymbol{x}_{t+1}) - f(\boldsymbol{x}^\star) \geq z | \mathcal{F}_0) \leq e^{-\frac{Q}{\alpha}z}\left\{ e^{\frac{Q}{\alpha}(f(\boldsymbol{x}_0) - f(\boldsymbol{x}^\star))} e^{-t\frac{Q}{\alpha}\kappa/2} + D\frac{e^{Q\kappa/2}}{1 - e^{-Q\kappa/2}} e^{QB} \right\}.$$

Proof. There are no boundedness assumptions placed in Lemma 4. We restate the conclusion of that result here:

$$\mathbb{E}[e^{\eta L_{t+1}} | \mathcal{F}_{T_0}] \leq \mathbb{E}[e^{\eta L_{T_1}} | \mathcal{F}_{T_0}] \prod_{k=T_1}^{t} \rho_t + D \sum_{\tau=T_1+1}^{t+1} \prod_{k=\tau}^{t} \rho_k, \tag{EC.30}$$

for $t \geq T_1 \geq T_0$. If we consider the above terms for constant step sizes $\alpha = \alpha_t$ then

$$\begin{aligned}
L_{t+1} &= f(\boldsymbol{x}_{t+1}) - f(\boldsymbol{x}^\star) - \alpha B \\
\rho_t = \rho &:= e^{-\alpha\eta\kappa + \alpha^2\eta^2 E} \leq e^{-\alpha\eta\frac{\kappa}{2}} \qquad \text{for} \qquad \alpha\eta \leq Q \\
T_0 &= \min\{t : \frac{\alpha_t - \alpha_{t+1}}{\alpha_t} \leq \frac{\kappa}{2B}\} = 0 \\
T_1 &= 0,
\end{aligned}$$

also

$$\prod_{k=0}^{t}\rho_k=\rho^{t+1}\le 1 \quad \text{and} \quad \sum_{l=1}^{t+1}\prod_{k=l}^{t}\rho_k=\sum_{l=1}^{t+1}\rho^{t-l}\le\sum_{l=-1}^{\infty}\rho^{l}=\frac{\rho^{-1}}{1-\rho}.$$

with these terms the above expression (EC.30) gives the requied bound

$$\mathbb{E}[e^{\eta(f(\boldsymbol{x}_{t+1})-f(\boldsymbol{x}^\star))}|\mathcal{F}_0]\le e^{\eta(f(\boldsymbol{x}_0)-f(\boldsymbol{x}^\star))}\rho^{t+1}+D\frac{\rho^{-1}}{1-\rho}e^{\alpha\eta B}.$$

Applying Markov's inequality gives

$$\mathbb{P}\left(f(\boldsymbol{x}_{t+1})-f(\boldsymbol{x}^\star)\ge z\right)\le e^{-\eta z}\left\{e^{\eta(f(\boldsymbol{x}_0)-f(\boldsymbol{x}^\star))}\rho^{t+1}+D\frac{\rho^{-1}}{1-\rho}e^{\alpha\eta B}\right\}.$$

Taking $\eta=Q/\alpha$ gives the required bound. □

EC.2.6.2. Linear Convergence under Exponential Concentration We note that while we generally assume that the set $\mathcal{X}$ is bounded, the above lemma and the linear convergence results of this section apply to unbounded constraint sets. We now prove Theorem 5.

THEOREM 5. *We assume that $\mathcal{X}$ is a convex set that may be unbounded. Assume Conditions (C1) and (C2) hold for a stochastic approximation procedure with rates given in (12):*
a) If, for $\hat{\epsilon}>0$ and $\hat{\delta}\in(0,1)$, we set

$$S=\log\left(\frac{F}{\hat{\epsilon}}\right),\quad \hat{\alpha}_s=\frac{2^{-s}F\kappa}{E\log\left(\frac{RS}{\hat{\delta}}\right)},\quad \text{and}\quad t_s=\left\lceil\frac{2}{\kappa^2}\log\left(\frac{RS}{\hat{\delta}}\right)\right\rceil$$

then with probability greater than $1-\hat{\delta}$ it holds that $\min_{\boldsymbol{x}\in\mathcal{X}^\star}\|\hat{\boldsymbol{x}}_S-\boldsymbol{x}\|\le\hat{\epsilon}$. Moreover, the number of iterations (12) required to achieve this bound is

$$\left\lceil\log_2\left(\frac{F}{\hat{\epsilon}}\right)\right\rceil\left\lceil\frac{2}{\kappa^2}\log\left(\frac{R}{\hat{\delta}}\right)+\log\left(\left\lceil\log_2\left(\frac{F}{\hat{\epsilon}}\right)\right\rceil\right)\right\rceil.$$

(Above $F=\min_{x^\star\in\mathcal{X}^\star}\|\boldsymbol{x}_0-\boldsymbol{x}^\star\|$ and R and E are time-independent constants that depend on the constants given in Conditions (C1) and (C2).)

b) For $\hat{\alpha}_s=\frac{a}{2^s\log(s+1)}$ and $t_s=\log^2(s+1)$, there exists positive constants A and M such that $\forall\hat{\delta}\in(0,1)$ if $a\ge A/\hat{\delta}$ then

$$\mathbb{P}\left(\min_{\boldsymbol{x}\in\mathcal{X}^\star}\|\hat{\boldsymbol{x}}_s-\boldsymbol{x}\|\le 2^{-s}M,\quad\forall s\in\mathbb{N}\right)\ge 1-\hat{\delta}.$$

Proof of Theorem 5. First, we recall some notation: $f(\hat{\boldsymbol{x}}_s):=\min_{\boldsymbol{x}\in\mathcal{X}^\star}\|\hat{\boldsymbol{x}}_s-\boldsymbol{x}\|$ and $F=f(\boldsymbol{x}_0)-\min_{x\in\mathcal{X}}f(x)$. The constants D and E are the moment generating function constants as defined in Lemma 3 and Lemma 4, respectively.

We define the event $\mathcal{E}_s:=\{f(\hat{\boldsymbol{x}}_s)\le 2^{-s}F\}$. So $\mathbb{P}(\mathcal{E}_0)=1$. We inductively analyze $\mathbb{P}(\mathcal{E}_s)$. Notice

$$\mathbb{P}(\mathcal{E}_s)\ge\mathbb{P}(\mathcal{E}_s|\mathcal{E}_{s-1})\mathbb{P}(\mathcal{E}_{s-1})=(1-\mathbb{P}(\mathcal{E}_s^c|\mathcal{E}_{s-1}))\,\mathbb{P}(\mathcal{E}_{s-1})\ge\mathbb{P}(\mathcal{E}_{s-1})-\mathbb{P}(\mathcal{E}_s^c|\mathcal{E}_{s-1}). \tag{EC.31}$$

By Lemma EC.6, for $\hat{x}_{s-1}$ such that $f(\hat{\boldsymbol{x}}_{s-1}) \le 2^{-s+1}F$ we have

$$\mathbb{P}\left(\mathcal{E}_s^c|\hat{\boldsymbol{x}}_{s-1}\right) = \mathbb{P}\left(f(\hat{\boldsymbol{x}}_s) \ge 2^{-s}F|\hat{\boldsymbol{x}}_{s-1}\right) \le e^{-\frac{Q}{\hat{\alpha}_s}z}\left[e^{\frac{Q}{\hat{\alpha}_s}(f(\hat{\boldsymbol{x}}_{s-1})-f(\boldsymbol{x}^\star))}e^{-t\frac{Q}{\hat{\alpha}_s}\kappa/2} + D\frac{e^{Q\kappa/2}}{1-e^{-Q\kappa/2}}e^{QB}\right]$$
$$\le e^{-\frac{Q}{\hat{\alpha}_s}z}\left[\exp\left\{\frac{Q}{\hat{\alpha}_s}2^{-s+1}F - t\frac{Q}{\hat{\alpha}_s}\kappa/2\right\} + D\frac{e^{Q\kappa/2}}{1-e^{-Q\kappa/2}}e^{QB}\right].$$

(Here we apply Lemma EC.6 for times $t = T_{s-1}, ..., T_s - 1$ with expectation $\mathbb{E}[\cdot]$ given by $\mathbb{E}[\cdot|\hat{x}_{s-1}]$.)

Notice that the term in curly brackets above is negative iff $t_s \ge 2^{-s+1}F/\kappa\hat{\alpha}_s$. If this holds then

$$\mathbb{P}\left(\mathcal{E}_s^c|\mathcal{E}_{s-1}\right) \le Re^{-2^{-s}F\kappa/2E\hat{\alpha}} \qquad \text{where} \quad R := 1 + D\frac{e^{Q\kappa/2}}{1-e^{-Q\kappa/2}}e^{QB}.$$

Applying this to (EC.31), $\mathbb{P}(\mathcal{E}_s) \ge \mathbb{P}(\mathcal{E}_{s-1}) - \mathbb{P}(\mathcal{E}_s^c|\mathcal{E}_{s-1}) \ge \mathbb{P}(\mathcal{E}_{s-1}) - Re^{-2^{-s}F\kappa/2E\hat{\alpha}_s}$. So we have

$$\mathbb{P}(\mathcal{E}_S) \ge 1 - \sum_{s=1}^{S} Re^{-2^{-s}F\kappa/2E\hat{\alpha}_s}. \tag{EC.32}$$

The total number of computations/samples required is $\sum_{s=1}^{S} t_s \ge \sum_{s=1}^{S} 2^{-s}F/\kappa\hat{\alpha}_s$.

We now prove part a). Given the bounds above, we can optimize the number of samples to achieve a probability $1-\hat{\delta}$. That is we solve

$$\text{minimize} \quad \sum_{s=1}^{S}\frac{2^{-s+1}F}{\kappa\hat{\alpha}_s} \quad \text{such that} \quad \sum_{s=1}^{S} Re^{-2^{-s+1}F\kappa/2E\hat{\alpha}_s} \le \hat{\delta} \quad \text{over} \quad \hat{\alpha}_s > 0.$$

A short calculation shows that this is minimized by $\hat{\alpha}_s = 2^{-s}F\kappa/E\log(RS/\hat{\delta})$ and thus since $t_s \ge 2^{-s+1}F/\kappa\hat{\alpha}_s$ we define $t_s = \left\lceil \frac{2}{\kappa^2}\log\left(\frac{RS}{\hat{\delta}}\right)\right\rceil$ and the number of samples required here is $S \times t_s$ which equals $S\left\lceil \frac{2}{\kappa^2}\log\left(\frac{RS}{\hat{\delta}}\right)\right\rceil$. Since for an $\hat{\epsilon}$ approximation, we require S to be such that $\hat{\epsilon} \ge 2^{-S}F$, we take $S = \lceil \log_2(F/\hat{\epsilon})\rceil$. Thus we see that an $\hat{\epsilon}$ approximation can be achieved with a probability greater than $1-\hat{\delta}$ in a number of samples given by

$$\left\lceil \log_2\left(\frac{F}{\hat{\epsilon}}\right)\right\rceil\left\lceil \frac{2}{\kappa^2}\log\left(\frac{R}{\hat{\delta}}\right) + \log\left(\left\lceil \log_2\left(\frac{F}{\hat{\epsilon}}\right)\right\rceil\right)\right\rceil.$$

This gives the part a) of Theorem 5.

Notice, we can make the sum (EC.32) finite for $S = \infty$. Specifically if we take $\hat{\alpha}_s = a/2^s\log(s+1)$ and $t_s = (\log(s+1))^2$ then the Condition (C1) holds $\forall s \ge s_0$ for $s_0 = \lceil e^{2F/\kappa}\rceil + 1$ and thus

$$\sum_{s=s_0}^{\infty} Re^{-2^{-s+1}F\kappa/2E\hat{\alpha}_s} = \sum_{s=s_0}^{\infty} R\frac{1}{(s+1)^{aF\kappa/2E}} \le R\int_{s_0}^{\infty}\frac{1}{s^{aF\kappa/2E}}ds \le \frac{2RE}{aF\kappa}\cdot\frac{1}{s_0^{aF\kappa/2E}} \le \frac{2RE}{aF\kappa}.$$

The above sum is less than $\hat{\delta}$ for $a \ge RE/2\hat{\delta}F\kappa$. Letting $A = 2RE/F\kappa$ and $M = 2^{s_0}F$, we see that for $a \ge A/\hat{\kappa}$ gives

$$\mathbb{P}\left(\exists s \in \mathbb{N} \text{ s.t. } \min_{x\in\mathcal{X}^\star}\|x_s - x\| \ge 2^{-s}F\right) \le \sum_{s=1}^{\infty}\mathbb{P}\left(\mathcal{E}_s^c \cup \left(\cap_{s'\le s}\mathcal{E}_{s'}\right)\right) \le \sum_{s=1}^{\infty}\mathbb{P}\left(\mathcal{E}_s^c|\mathcal{E}_{s-1}\right) \le \sum_{s=1}^{\infty} Re^{-\frac{2^s F\kappa}{2E\hat{\alpha}}} \le \hat{\delta}.$$

Thus for learning rates $\hat{\alpha}_s = a/2^s\log(s+1)$ with $a \ge M/\hat{\delta}$ if it holds that $\mathbb{P}(\forall s, \min_{x\in\mathcal{X}^\star}\|x_s - x\| \le 2^{-s}M) \ge 1-\hat{\delta}$. This gives the 2nd part of Theorem 5. □

EC.2.6.3. Application to Specific Stochastic Approximation Algorithms. The following is the equivalent linear convergence result for Kiefer-Wolfowotiz

COROLLARY EC.1 **(Linear Convergence in Projected Stochastic Gradient Descent).** *We assume that $\mathcal{X}$ is a convex set that may be unbounded. Assume Conditions (D1) and (D2) hold for PSGD with rates given in (12). If, for $\hat{\epsilon} > 0$ and $\hat{\delta} \in (0,1)$, we set*

$$S = \log\left(\frac{F}{\hat{\epsilon}}\right), \quad \hat{\alpha}_s = \frac{2^{-s} F\kappa}{E \log\left(\frac{GS}{\hat{\delta}}\right)}, \quad \text{and} \quad t_s = \left\lceil \frac{2}{\kappa^2} \log\left(\frac{GS}{\hat{\delta}}\right) \right\rceil$$

then with probability greater than $1-\hat{\delta}$ it holds that $\min_{\boldsymbol{x}\in\mathcal{X}^\star} \|\hat{\boldsymbol{x}}_S - \boldsymbol{x}\| \le \hat{\epsilon}$. Moreover, the number of iterations (12) required to achieve this bound is

$$\left\lceil \log_2\left(\frac{F}{\hat{\epsilon}}\right) \right\rceil \left\lceil \frac{2}{\kappa^2} \log\left(\frac{G}{\hat{\delta}}\right) + \log\left(\left\lceil \log_2\left(\frac{F}{\hat{\epsilon}}\right)\right\rceil\right) \right\rceil .$$

COROLLARY EC.2 **(Linear Convergence of Kiefer-Wolfowitz).** *Assume Conditions (D1), (D2), (D3) hold. For the KW algorithm, (9), with step-sizes given in (12): If, for $\hat{\epsilon} > 0$ and $\hat{\delta} \in (0,1)$, we set*

$$S = \log\left(\frac{F}{\hat{\epsilon}}\right), \quad \hat{\alpha}_s = \frac{2^{-s} F\kappa}{E \log\left(\frac{GS}{\hat{\delta}}\right)}, \quad \nu_s = \sqrt{\frac{\kappa}{3c}} \quad \text{and} \quad t_s = \left\lceil \frac{2}{\kappa^2} \log\left(\frac{GS}{\hat{\delta}}\right) \right\rceil$$

then with probability greater than $1-\hat{\delta}$ it holds that $\min_{\boldsymbol{x}\in\mathcal{X}^\star} \|\hat{\boldsymbol{x}}_S - \boldsymbol{x}\| \le \hat{\epsilon}$. Moreover, the number of iterations (12) required to achieve this bound is

$$\left\lceil \log_2\left(\frac{F}{\hat{\epsilon}}\right) \right\rceil \left\lceil \frac{2}{\kappa^2} \log\left(\frac{G}{\hat{\delta}}\right) + \log\left(\left\lceil \log_2\left(\frac{F}{\hat{\epsilon}}\right)\right\rceil\right) \right\rceil .$$

The proof of Corollary EC.2 is identical to the proof of Theorem 5

COROLLARY EC.3 **(Linear Convergence of Stochastic Frank-Wolfe).** *We assume that $\mathcal{X}$ is a convex set that may be unbounded. Also, Assume Condition D1, D2, E1 and E2 hold. For the SFW algorithm with step-sizes given by (12). If we set*

$$S = \log\left(\frac{F}{\hat{\epsilon}}\right), \quad \hat{\alpha}_s = \frac{2^{-s} F\kappa}{E \log\left(\frac{GS}{\hat{\delta}}\right)}, \quad m_s := \left\lceil \left(\frac{3\sigma}{\kappa\alpha}\right)^2 \right\rceil, \quad \text{and} \quad t_s = \left\lceil \frac{2}{\kappa^2} \log\left(\frac{GS}{\hat{\delta}}\right) \right\rceil$$

then with probability greater than $1-\hat{\delta}$ it holds that $\min_{\boldsymbol{x}\in\mathcal{X}^\star} \|\hat{\boldsymbol{x}}_S - \boldsymbol{x}\| \le \hat{\epsilon}$. Moreover, the number of iterations (12) required to achieve this bound is

$$\left\lceil \log_2\left(\frac{F}{\hat{\epsilon}}\right) \right\rceil \left\lceil \frac{2}{\kappa^2} \log\left(\frac{G}{\hat{\delta}}\right) + \log\left(\left\lceil \log_2\left(\frac{F}{\hat{\epsilon}}\right)\right\rceil\right) \right\rceil .$$

EC.3. Appendix to Theoretical Results

EC.3.1. List of Notations for Theorem 1

There are several time-independent constants (usually denoted with a capital letter) in Theorem 1. We list these here.

B	Bound where the drift condition holds. See (C1).
$C = \dfrac{(1+H)}{2QG^n} + B$	Constant in Proposition 1
$D = \mathbb{E}[e^{\lambda Z}] = \mathbb{E}[e^{\lambda[Y+\kappa/2]}]$	Moment Generating Function, see (18).
$E = \mathbb{E}\left[\dfrac{e^{\lambda Z} - 1 - \lambda Z}{\lambda^2}\right]$	See Lemma 4
$F = \max_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x}) - \min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})$	
$G = \dfrac{1}{4^\gamma}, \quad \text{for } \alpha_t = \dfrac{a}{(u+t)^\gamma}$	See Lemma 5 and Proof in Section 4.1.2
$H = De^{\frac{\kappa QG^n}{2}}/(1 - e^{-\frac{\kappa QG^n}{2}})$	Constant in Proposition 1 Defined after (26).
$I = (1+H)e^{QG/(F/\alpha_{T_2} - B)}$	Constant in Theorem 1
$J = QG^n$	Exponent in Theorem 1
$K = \dfrac{I}{J}$	Constant in Theorem 1
$n = \begin{cases} 1 & \text{for } \gamma < 1 \\ \lceil \frac{\alpha_0 B + F}{a \log 2} \rceil & \text{for } \gamma = 1 \end{cases}$	
$Q = \lambda \wedge (\kappa/2E)$	
$T_0 = \min\left\{t \geq 0 : \dfrac{(\alpha_s - \alpha_{s+1})}{\alpha_s} < \kappa/2B, \forall s \geq t\right\}$	See Lemma 3 and its proof.
$T_1 = \begin{cases} \frac{\alpha_0 B + F}{\frac{a}{1-\gamma}\left[1 - \frac{1}{2^{1-\gamma}}\right]} & \text{for } \gamma < 1 \\ u2^n & \text{for } \gamma = 1 \end{cases}$	
$T_2 = T_0 \vee T_1$	
Y	Sub-exponential Random Variable defined in C2
$Z = Y + \dfrac{\kappa}{2}$	Random Variable defined in Lemma 3

EC.3.2. Technical Lemmas for the Proof of Proposition 1

LEMMA 3. *Given Conditions (C1) and (C2) hold, there exists a deterministic constant T_0 such that the sequence of random variables $(L_t : t \geq T_0)$ satisfies*

$$\mathbb{E}\big[L_{t+1} - L_t \big| \mathcal{F}_t\big]\mathbb{I}[L_t \geq 0] < -\alpha_t \kappa, \tag{17}$$

and

$$[|L_{t+1} - L_t||\mathcal{F}_t] \leq \alpha_t Z \quad \textit{where} \quad D := \mathbb{E}[e^{\lambda Z}] < \infty\,. \tag{18}$$

Proof of Lemma 3. Applying the definition of L_t and the drift Condition (C1) gives

$$\begin{aligned}\mathbb{E}[L_{t+1}-L_t|\mathcal{F}_t] &= \mathbb{E}[f(\boldsymbol{x}_{t+1})-f(\boldsymbol{x}_t)|\mathcal{F}_t]+(\alpha_t-\alpha_{t+1})B\\ &\le -2\alpha_t\kappa+(\alpha_t-\alpha_{t+1})B\\ &\le -2\alpha_t\kappa\left[1-(\alpha_t-\alpha_{t+1})B/\alpha_t\kappa\right]\end{aligned}$$

Since $(\alpha_t-\alpha_{t+1})/\alpha_t\to 0$, there exists a constant T_0 such that $(\alpha_t-\alpha_{t+1})/\alpha_t<\kappa/2B$ for all $t\ge T_0$. Specifically we can take $T_0=\min\{t\ge 0:(\alpha_s-\alpha_{s+1})/\alpha_s<\kappa/2B,\forall s\ge t\}$. This gives the first drift condition (17).

For the second condition, for $t\ge T_0$ with T_0 as just defined:

$$\begin{aligned}\left[|L_{t+1}-L_t|\middle|\mathcal{F}_t\right] &\le \left[|f(\boldsymbol{x}_{t+1})-f(\boldsymbol{x}_t)|\middle|\mathcal{F}_t\right]+|\alpha_{t+1}-\alpha_t|B\\ &\le \alpha_t Y+\alpha_t\frac{(\alpha_t-\alpha_{t+1})}{\alpha_t}B\\ &\le \alpha_t(Y+\kappa/2)\,.\end{aligned}$$

Taking $Z=Y+\kappa/2$, it is clear that condition (18) holds for Z as an immediate consequence of the boundedness condition on Y in (C2). □

Lemma 5. *If α_t, $t\in\mathbb{Z}_+$, is a decreasing positive sequence, then*

$$\min_{s=\hat{t},\dots,t}\left\{\frac{\sum_{k=s}^t\alpha_k}{\sum_{k=s}^t\alpha_k^2}\right\}=\frac{\sum_{k=\hat{t}}^t\alpha_k}{\sum_{k=\hat{t}}^t\alpha_k^2}\,. \tag{21}$$

Moreover, if α_t, $t\in\mathbb{Z}_+$ satisfies the learning rate condition (13) then

$$\frac{1}{\alpha_{\lfloor t/2^n\rfloor}}\ge\frac{G^n}{\alpha_t}\qquad\text{and}\qquad\min_{s=\lfloor t/2^n\rfloor,\dots,t}\left\{\frac{\sum_{k=s}^t\alpha_k}{\sum_{k=s}^t\alpha_k^2}\right\}\ge\frac{G^n}{\alpha_t} \tag{22}$$

for some constant $G\in(0,1]$ and for $n\in\mathbb{N}$ such that $t/2^n>1$.

Proof of Lemma 5. It is straight-forward to show that for $a,a',A,A'>0$

$$\frac{a}{a'}\le\frac{A}{A'}\quad\text{if and only if}\quad\frac{a+A}{a'+A'}\le\frac{A}{A'}\,. \tag{EC.33}$$

[Note that both expressions above are equivalent to $AA'+aA'\le AA'+a'A$.]

Take positive numbers a_s,a'_s, $s=1,\dots,t$. If

$$\frac{a_s}{a'_s}\le\frac{a_k}{a'_k}$$

for $k=s+1,\dots,t$, then

$$\sum_{k=s+1}^t a'_k a_s\le\sum_{k=s+1}^t a_k a'_s. \tag{EC.34}$$

Thus,

$$\frac{a_s}{a'_s} \leq \frac{\sum_{k=s+1}^{t} a_k}{\sum_{k=s+1}^{t} a'_k}.$$

Thus applying (EC.33) with $A = \sum_{k=s+1}^{t} a_k$ and $A' = \sum_{k=s+1}^{t} a'_k$ gives

$$\frac{\sum_{k=s}^{t} a_k}{\sum_{k=s}^{t} a'_k} \leq \frac{\sum_{k=s+1}^{t} a_k}{\sum_{k=s+1}^{t} a'_k}. \tag{EC.35}$$

Finally, taking $a_k = \alpha_k$ and $a'_k = \alpha_k^2$, we see that (EC.34) holds since α_t is decreasing. Thus, from (EC.35), we see that the result (21) holds.

If the condition (13) holds then $\liminf_{t\to\infty} \alpha_{2t}/\alpha_t > 0$ implies

$$\frac{\alpha_t}{\alpha_{\lfloor t/2 \rfloor}} > \sqrt{G} \tag{EC.36}$$

for some $1 \geq G > 0$. Thus

$$\frac{\alpha_t}{\alpha_{\lfloor t/2^n \rfloor}} = \frac{\alpha_t}{\alpha_{\lfloor t/2 \rfloor}} \times \ldots \times \frac{\alpha_{\lfloor t/2^{n-1} \rfloor}}{\alpha_{\lfloor t/2^n \rfloor}} \geq G^{\frac{n}{2}} \geq G^n. \tag{EC.37}$$

Since the sequence is decreasing and (EC.36) holds, we have that

$$\frac{\sum_{k=\lfloor t/2^n \rfloor}^{t} \alpha_k}{\sum_{k=\lfloor t/2^n \rfloor}^{t} \alpha_k^2} \geq \frac{(t - \lfloor t/2^n \rfloor)\alpha_t}{(t - \lfloor t/2^n \rfloor)\alpha_{\lfloor t/2^n \rfloor}^2} = \frac{\alpha_t^2}{\alpha_{\lfloor t/2^n \rfloor}^2} \frac{1}{\alpha_t} = \frac{\alpha_t^2}{\alpha_{\lfloor t/2 \rfloor}^2} \times \ldots \times \frac{\alpha_{\lfloor t/2^{n-1} \rfloor}^2}{\alpha_{\lfloor t/2^n \rfloor}^2} \frac{1}{\alpha_t} \geq \frac{G^n}{\alpha_t}.$$

Applying this to (21) with $s = \lfloor t/2^n \rfloor$ gives

$$\min_{s=\lfloor t/2^n \rfloor, \ldots, t} \left\{ \frac{\sum_{k=s}^{t} \alpha_k}{\sum_{k=s}^{t} \alpha_k^2} \right\} \geq \frac{G^n}{\alpha_t}.$$

Thus the above along with (EC.37) proves that (22) holds as required. □

LEMMA EC.7. *For $\alpha_t = a/(u+t)^\gamma$ with $0 \leq \gamma < 1$ Taking*

$$n = \begin{cases} 1, & \text{for } \gamma < 1, \\ 1 + \left\lceil \frac{\alpha_0 B + F}{a \log 2} \right\rceil, & \text{for } \gamma = 1, \end{cases} \qquad \text{and} \qquad T_1 = \begin{cases} u + \frac{2^{1+\gamma}}{au^{-\gamma}}[\alpha_0 B + F], & \text{for } \gamma < 1, \\ u 2^n, & \text{for } \gamma = 1, \end{cases}$$

it holds that

$$\sum_{s=\lfloor t/2^n \rfloor}^{t} \alpha_s \geq \alpha_0 B + F, \qquad \forall t \geq T_1. \tag{EC.38}$$

Proof. We consider the case of $\gamma < 1$ separately from the case where $\gamma = 1$.

First we take $\gamma < 1$ and $n = 1$. In the following expression, we take $t = xu$ with $x \geq 1$,

$$\sum_{s=\lfloor t/2 \rfloor}^{t} \alpha_s \geq \frac{t}{2}\alpha_t = \frac{a}{2}\frac{t}{(u+t)^\gamma} = \frac{a}{2}u^{1-\gamma}\frac{x}{(1+x)^\gamma} \geq \frac{au^{1-\gamma}}{2^{1+\gamma}}x = a\frac{u^{-\gamma}}{2^{1+\gamma}}t \tag{EC.39}$$

Thus, $t = ux$ with $x \geq 1$ and right-hand side of (EC.39) is greater than $\alpha_0 B + F$ for

$$T_1 = \frac{2^{1+\gamma}}{au^{-\gamma}}[\alpha_0 B + F] + u,$$

and for any t such that $t \geq T_1$. This completes the proof for $\gamma < 1$

Second, we take $\gamma = 1$. We assume that $t \geq T_1 := 2^n u$ and we will take $n = 1 + \left\lceil \frac{\alpha_0 B + F}{a \log 2} \right\rceil$.

$$\sum_{s=\lfloor t/2^n \rfloor}^{t} \alpha_s \geq \int_{t/2^n}^{t} \frac{a}{u+s} ds = a \log\left(\frac{u+t}{u+t2^{-n}}\right) = an \log 2 + a \log\left(\frac{u+t}{u2^n + tk}\right) \geq an \log 2 + a \log \frac{1}{2}.$$

The last inequality above holds since $t \geq T_1 := 2^n u$. Notice that

$$an \log 2 + a \log \frac{1}{2} = a(n-1) \log 2 \geq \alpha_0 B + F, \qquad \text{for } n = 1 + \left\lceil \frac{\alpha_0 B + F}{a \log 2} \right\rceil.$$

Thus the required bound (EC.38) holds for n and T_1 as specified for $\gamma = 1$. □

EC.4. Appendix to Applications and Numerical Examples

This section aims to provide a simple application of the main results of Theorem 1 and Theorem 2. Given the importance of Linear Programming (LP) and Markov Decision Processes (MDP) in operations research, we briefly explore these problem settings. However, we emphasize that linear objectives are a special case of the results proven in Theorem 1 and Theorem 2. The results are proved under conditions that apply to non-smooth, non-convex objectives and general convex constraints. We refer to Birge and Louveaux (2011) and Shapiro et al. (2021) as standard texts on stochastic linear programming. For the linear programming formulation of MDPs, we refer to Schweitzer and Seidmann (1985).

EC.4.1. Linear Programming

Here we consider a linear program in which the cost function that we wish to minimize must be sampled and where the optimization constraints are deterministic. We are interested in solving a linear program of the form

$$\text{minimize } \bar{\boldsymbol{c}}^\top \boldsymbol{x} \quad \text{subject to } H\boldsymbol{x} \leq \boldsymbol{b} \text{ over } \boldsymbol{x} \in \mathbb{R}^d, \tag{EC.40}$$

where $\bar{\boldsymbol{c}} \in \mathbb{R}^d \backslash \{\boldsymbol{0}\}$, $H \in \mathbb{R}^{p \times d}$ and $\boldsymbol{b} \in \mathbb{R}^p$. We assume $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^d : H\boldsymbol{x} \leq \boldsymbol{b}\}$ is a bounded polytope.

We suppose that the constraint set $\mathcal{X}$ is deterministic and known, however, the cost vector $\bar{\boldsymbol{c}}$ is unknown but can be sampled.

Specifically, we let $\boldsymbol{c}_t$, $t \in \mathbb{Z}_+$, be an independent, mean $\bar{\boldsymbol{c}}$, sub-exponential random vectors in $\mathbb{R}^d$. That is

$$\mathbb{E}[\boldsymbol{c}_t | \mathcal{F}_t] = \bar{\boldsymbol{c}} \qquad \text{and} \qquad \sup_{t \in \mathbb{Z}_+} \mathbb{E}\left[e^{\lambda \|\boldsymbol{c}_t\|} \middle| \mathcal{F}_t\right] < \infty \tag{EC.41}$$

for some $\lambda > 0$. We then apply projected stochastic gradient descent (6). Notice that Condition (D1) is satisfied by (EC.41). Further Condition (D2) holds for any linear program. This is a consequence of the following technical lemma.

LEMMA EC.8. *If $\mathcal{X}$ is a bounded polytope and $\mathcal{X}^\star = argmin_{\boldsymbol{x}\in\mathcal{X}} \bar{\boldsymbol{c}}^\top \boldsymbol{x}$., then there exists a constant $K > 0$ such that*

$$\frac{\bar{\boldsymbol{c}}^\top(\boldsymbol{x}-\boldsymbol{x}^\star)}{||\bar{\boldsymbol{c}}||||\boldsymbol{x}-\boldsymbol{x}^\star||} \geq K,$$

for $\boldsymbol{x}^\star$ the projection of $\boldsymbol{x}$ onto $\mathcal{X}^\star$. Thus Condition (D2) holds for PSGD applied to the LP (EC.40).

The proof of Lemma EC.8 requires some careful bounding between optimal solutions and sub-optimal extreme points. The proof is given below. The result bounds the angle between optimal and sub-optimal points for a polytope.

Proof of Lemma EC.8. We assume without loss of generality that $\bar{\boldsymbol{c}}^\top \boldsymbol{x}^\star = 0$ and $||\bar{\boldsymbol{c}}|| = 1$. Let $\mathcal{E}$ be the extreme points of $\mathcal{X}$. Let $\mathcal{E}^\star$ be the extreme points in $\mathcal{X}^\star$. Then let $\mathcal{E}' := \mathcal{E} \setminus \mathcal{E}^\star$ and $\mathcal{X}'$ is the convex closure of $\mathcal{E}'$. Let $a := \min_{\boldsymbol{x}\in\mathcal{X}'} \bar{\boldsymbol{c}}^\top \boldsymbol{x}$ and $D := \max_{\boldsymbol{x}^\star\in\mathcal{X}^\star, \boldsymbol{x}'\in\mathcal{X}'} ||\boldsymbol{x}^\star - \boldsymbol{x}'||$. We will show we can take $K := a/D$.

For all $\boldsymbol{x} \in \mathcal{X} \setminus \mathcal{X}^\star$, $\boldsymbol{x}$ must be a convex combination of a point in $\mathcal{X}^\star$ and a point in $\mathcal{X}'$. Specifically,

$$\boldsymbol{x} = (1-p)\boldsymbol{x}_0 + p\boldsymbol{x}_1, \tag{EC.42}$$

for $\boldsymbol{x}_0 \in \mathcal{X}^\star$ and $\boldsymbol{x}_1 \in \mathcal{X}'$ and $p \in (0,1]$. Then, as required,

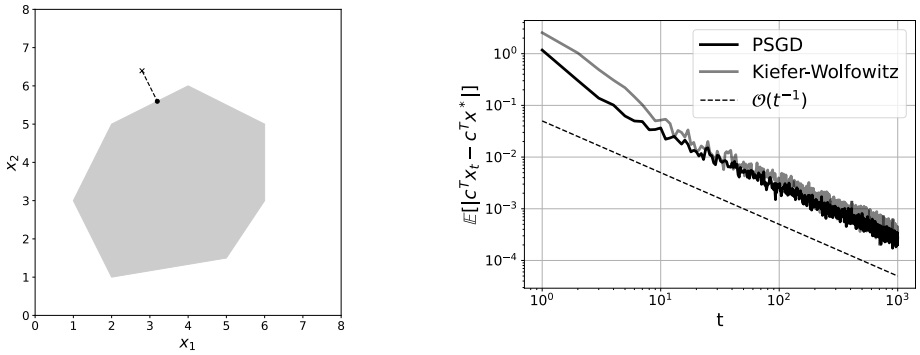
$$\frac{\bar{\boldsymbol{c}}^\top \boldsymbol{x}}{||\boldsymbol{x}-\boldsymbol{x}^\star||} \geq \frac{\bar{\boldsymbol{c}}^\top(\boldsymbol{x}-\boldsymbol{x}_0)}{||\boldsymbol{x}-\boldsymbol{x}_0||} = \frac{\bar{\boldsymbol{c}}^\top(\boldsymbol{x}_1-\boldsymbol{x}_0)}{||\boldsymbol{x}_1-\boldsymbol{x}_0||} \geq \frac{a}{D} = K > 0\,.$$

The first inequality above uses the fact that $\boldsymbol{x}^\star$ is closest to $\boldsymbol{x}$. The equality applies (EC.42). Then finally, we apply the definitions of a, D and K. □

Thus, we see that both Theorem 2 and Theorem 5 hold in the context of linear programming problems with an unknown objective function.

EC.4.1.1. Polytope Example We consider the problem with two variables with the constraints being the polytope in Figure EC.2(a). We assume that the cost vector $\bar{\boldsymbol{c}} = [4,6]^\top$ is unknown but can be sampled from a joint Gaussian distribution of independent random variables with mean vector $\bar{\boldsymbol{c}}$ and variance 1. This problem is analytically tractable. Given the costs, we can calculate the reference solution to be $x^\star = [2,1]^T$.

The convergence rate for the PSGD and Kiefer-Wolfowitz should be $O(1/t^\gamma)$ in expectation when the error is measured by the L^1-norm. Evidence for the convergence rate is shown in Figure EC.2(b). Increasing the batch size above 50 substantially reduces the noise of sampling costs, and the algorithm may perform better than $O(1/t)$. In this case, the algorithm converges, reaching the optimum solution after 7 iterations. This occurs because the chance of observing any sample perturbing the stochastic gradient descent algorithm away from the optimal point is a rare event. However, when there is a non-negligible probability of an iteration leaving the optimal point, then the $O(1/t)$ is found as anticipated.



(a) Polytope (b) Convergence for batch size $B = 5$

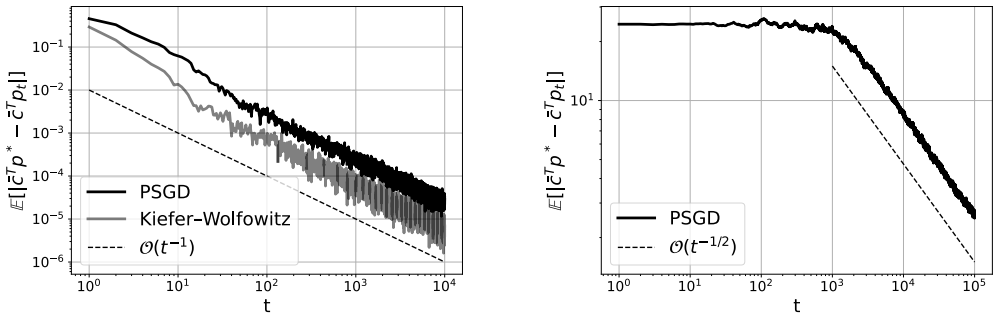
Figure EC.2 Polytope of the two variables linear programming problem and convergence of projected stochastic gradient descent on the two variables linear programming example. In Figure EC.2(a), the shaded area is the bounded polytope and the cross is one of the points of iterations, and the black point is the corresponding projection. In Figure EC.2(b), expectation is computed over 1000 realizations. The parameters of step size are chosen as $a = 1, u = 1$ and $\gamma = 1$ such that $\alpha_t = 1/(1+t)$. The costs $\boldsymbol{c}_t$ are computed with batch size $B = 5$. The parameter $v = 1$ is chosen for Kiefer-Wolfowitz. The fitted slope is -1.02 and -1.05 for PSGD and Kiefer-Wolfowitz.

EC.4.1.2. Probability Simplex This section considers a higher dimension for the optimization over the probability simplex as an example. There are simple, efficient algorithms for projection onto the probability simplex (Duchi et al. 2008). The problem that we solve is formulated as follows

$$\text{minimize} \quad p_1\bar{c}_1 + p_2\bar{c}_2 + ... + p_n\bar{c}_n = \bar{\boldsymbol{c}}^{\boldsymbol{T}}\boldsymbol{p} \quad \text{subject to} \quad \sum_{i=1}^{n} p_i = 1 \quad \text{over} \quad p_i \geq 0, \ \forall i = 1, ..., n,$$

where $\bar{c}_1 < \bar{c}_2 < ... < \bar{c}_n$ and $n = 50$. We label the polytope due to the constraint as $\mathcal{P}$ and suppose that the cost vector $\bar{\mathbf{c}}$ is unknown but can be sampled from a normal distribution with a certain mean vector and covariance matrix. In particular, for $t \in \mathbb{Z}_+$, we apply the stochastic gradient descent iteration: $\boldsymbol{p}_{t+1} = \Pi_{\mathcal{P}}(\boldsymbol{p}_t - \alpha_t \boldsymbol{c}_t)$, where $\mathbf{c_t} \sim \mathcal{N}(\bar{\mathbf{c}}, \mathbf{1})$ and $\alpha_t = a/t$ with $a > 0$. According to the special settings above, the minimum of this problem is $\boldsymbol{p}^* = (1, 0, ..., 0)$. We expect that $\mathbb{E}\left[|\bar{\boldsymbol{c}}^{\boldsymbol{T}}\boldsymbol{p_t} - \bar{\boldsymbol{c}}^{\boldsymbol{T}}\boldsymbol{p}^*|\right] = O\left(1/t\right)$. Figure EC.3(a) confirms that the PSGD and Kiefer-Wolfowitz converge with an order of -1.

However, a itltohught falls somewhat outside the scope of this paper's results, it is also possible to consider the multi-armed bandit variation of this problem. Here, the natural generalization of the projected gradient descent algorithm applies importance sampling. Here we sample an index i_t according to the distribution $\boldsymbol{p}_t$ and apply the updated $p_{i,t+1} = p_{i,t} - \alpha_t \frac{c_{i,t}}{p_{i,t}} \mathbb{I}[i = i_t]$ for $i = 1, ..., n$. Simulations suggest a rate of convergence of the order of $O(1/\sqrt{t})$, see Figure EC.3(b).



(a) Probability simplex example (b) Multi-arm bandit problem

Figure EC.3 Convergence of projected stochastic gradient descent on the probability simplex example and multi-arm bandit problem. Expectations are computed over 100 realizations. The parameter $v = 1$ is chosen for Kiefer-Wolfowitz. The step size parameter is chosen as $a = 1$ such that $\alpha_t = 1/t$ for both. In Figure EC.3(a), the fitted slope is -1.01 and -1.01 for PSGD and Kiefer-Wolfowitz. In Figure EC.3(b), the fitted slope is -0.475.

EC.4.2. Markov Decision Processes

We now optimize a discounted Markov Decision Process (MDP) using the results from the last section. Here we use a linear programming approach to give the convergence of a simple policy gradient algorithm for an MDP in which the system dynamics are known but the costs are unknown.

An MDP can be formulated as a linear program, where the primal form of this linear program solves for the optimal value function, and the dual form finds the optimal occupancy measure. In this linear programming formulation, the dual problem takes the form:

$$\begin{aligned} \text{minimize} \quad & \sum_{s\in\mathcal{S}}\sum_{a\in\mathcal{A}} \bar{c}(s,a)x(s,a) && \text{(Dual)} \\ \text{subject to} \quad & \sum_{a\in\mathcal{A}} x(s',a) = \xi(s') + \beta \sum_{s\in\mathcal{S}}\sum_{a\in\mathcal{A}} x(s,a)P(s'|s,a), \qquad \forall s' \in \mathcal{S} \\ \text{over} \quad & (x(s,a) : s\in\mathcal{S}, a\in\mathcal{A}) \in \mathbb{R}_+^{\mathcal{S}\times\mathcal{A}}. \end{aligned}$$

Here $(\xi(s) : s \in \mathcal{S})$ is a positive vector. We assume that the dynamics as given by $(P(s'|s,a) : a \in \mathcal{A},\ s, s' \in \mathcal{S})$ are known but costs $(\bar{c}(s,a) : a \in \mathcal{A}, s \in \mathcal{S})$ are unknown and must be sampled, then above we have a linear program with an unknown objective and known constraints. For this reason, we can apply the analysis developed in the last section.

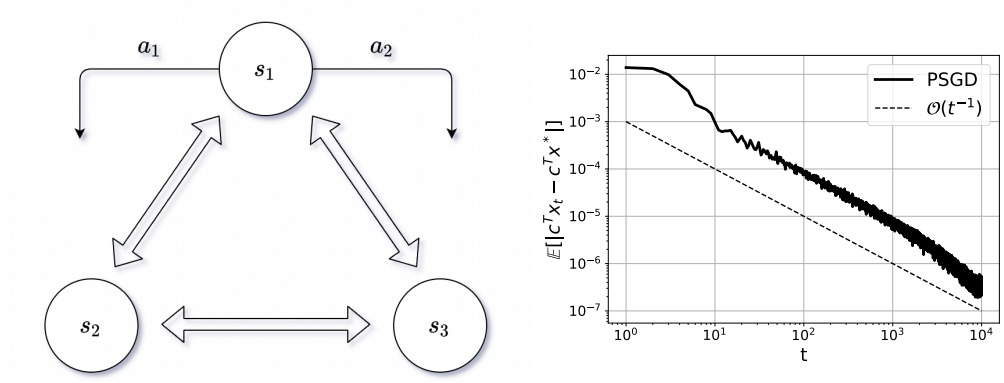
Here we assume that we can sample costs $\hat{\boldsymbol{c}} = (\hat{c}(s,a) : s \in \mathcal{S}, a \in \mathcal{A})$ where the states and actions are distributed according to some predetermined probability distribution $\boldsymbol{\pi} = (\pi(s,a) : a \in \mathcal{A}, s \in \mathcal{S})$. There are several ways of sampling the cost vector $\boldsymbol{c}_t$ for each t. The most straightforward one is as follows. For each t, the cost $\boldsymbol{c}_t = (c_t(s,a) : s \in \mathcal{S},\ a \in \mathcal{A})$ is sampled by first taking IID sample

(s_t, a_t), with distribution $\boldsymbol{\pi} = (\pi(s,a) : s \in \mathcal{S}, a \in \mathcal{A})$ where $\pi(s,a) > 0$ for all $s \in \mathcal{S}$ and $a \in \mathcal{A}$, and then defining

$$c_t(s,a) = \frac{\hat{c}(s_t,a_t)}{\pi(s_t,a_t)} \mathbb{I}[(s_t,a_t) = (s,a)]\,. \tag{EC.43}$$

We allow for the possibility of averaging batches of costs of the form (EC.43). We then consider the projected gradient descent algorithm $\boldsymbol{x}_{t+1} = \Pi_{\mathcal{X}}(\boldsymbol{x}_t - \alpha_t \boldsymbol{c}_t)$. The projection above is onto the constraint set of the dual problem (Dual). Our above observation on Linear Programs holds here. Specifically, Theorem 2 and Theorem 5 hold for this PSGD algorithm.

EC.4.2.1. Three-state two-action Markov decision process We now consider the first reinforcement learning application of our results, a relatively simple MDP. We consider an MDP with three states $\mathcal{S} = \{s_1, s_2, s_3\}$. In each state, there are two actions $\mathcal{A} = \{a_1, a_2\}$, corresponding to move anticlockwise (a_1) and clockwise (a_2). Figure EC.4(a) shows the states and actions. When we choose to take an action, the probability of going to the desired state is 2/3; otherwise, one of the states uniformly at random. We assume that the costs $c(s,a)$ are independent normally distributed with $c(s_i,a_j) \sim N(i,1)$, for $i = 1,2,3$. The states and actions are sampled according to the predetermined probability distribution $\boldsymbol{\pi} = (\pi(s,a) = 1/6 : s \in \mathcal{S}, a \in \mathcal{A})$. Figure EC.4(b) demonstrates the correct convergence rate as predicted.



(a) Three-state two-action MDP graph (b) Three-state two-action MDP example

Figure EC.4 The three-state two-action MDP graph and convergence of projected stochastic gradient descent on the three-state two-action MDP example. In Figure EC.4(b), the expectation is computed over 20 realizations. The costs $c_t(s,a)$ are computed with batch size $B = 200$. The parameters of step size are chosen as $a = 0.1, u = 1$ and $\gamma = 1$ such that $\alpha_t = 0.1/(1+t)$. The fitted slope is -1.29.

EC.4.2.2. Blackjack We now consider a larger tabular reinforcement learning problem for the game of Blackjack. Blackjack is a simple card game where a player is initially dealt two cards. The player is dealt cards sequentially before deciding to stop. The player must attempt to reach a total that is more than the dealer but not more than 21. The problem is described in more detail in Sutton and Barto (2018). The states of the problem depend on three factors which are: the player's current points (4–22); usable ace (with or without); dealer's showing card (1–10), which gives 290 states in total.

We label the states in sequence starting with s_1 being no usable ace, the player's current points 4 and dealer's showing card 1, and ending with s_{290} being usable ace, the player's current points 21 and dealer's showing card 10. The actions simply consist of hitting (a_1) and sticking (a_0). Denote the collection of states $\mathcal{S} = \{s_i : i = 1, ..., 290\}$ and the collection of actions $\mathcal{A} = \{a_i : i = 0, 1\}$.

We assume that the reward $\bar{\boldsymbol{r}} = (\bar{r}(s,a) : s \in \mathcal{S}, a \in \mathcal{A})$ can be sampled for each iteration of the projected stochastic gradient descent by carrying on the following procedure. We first simulate IID samples (s_t^i, a_t^i), $i = 1, ..., B$ from the distribution $\boldsymbol{\pi} = (\pi(s,a) = 1/580 : s \in \mathcal{A}, a \in \mathcal{A})$ and then define the cost similar as Equation (EC.43) with $\bar{c}(s_t^i, a_t^i) = -\bar{r}(s_t^i, a_t^i)$. In addition, according to the rules, it is reasonable to set the discount factor $\beta = 1$. Applying the PSGD with the learning rate of the form $\alpha_t = a/(b+t)^\gamma$, for $a, b > 0$ and $\gamma \in [0, 1]$, the projected stochastic gradient descent converges with a rate of $O(1/t^\gamma)$ in expectation. The rate $O(1/t)$ with $\gamma = 1$ is shown in Figure EC.5.

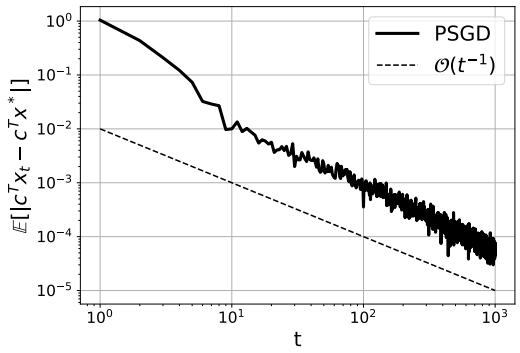


Figure EC.5 Convergence of projected stochastic gradient descent on the Blackjack example. The expectation is computed over 10 realizations. The costs $c_t(s,a)$ are computed with batch size $B = 200$. The parameters of step size are chosen as $a = 0.1, u = 1$ and $\gamma = 1$ such that $\alpha_t = 0.1/(1+t)$. The fitted slope is -1.23.