Online Appendix to "Linear Programming with Online Learning"
by T. Levina, Y. Levin, J. McGill and M. Nediak

Proof of Lemma 1. Since X is bounded, there exists a ball B of radius R around 0 such that $X \subseteq B$. Also, since K° has a nonempty interior, there exist $\mathbf{x}'_0 \in -K^\circ$ and a ball B' of radius R' around $\mathbf{x}'_0$ such that $B' \subseteq -K^\circ$. Since $B' \subseteq -K^\circ$, we have

$$\mathbf{c}^\top \mathbf{x}' \geq 0 \text{ for all } \mathbf{x}' \in B' \text{ and } \mathbf{c} \in \Omega \subseteq K.$$

We establish a one-to-one correspondence between points of B and B' as $\frac{\mathbf{x}}{R} = \frac{\mathbf{x}'-\mathbf{x}'_0}{R'}$. The image of X under this mapping is a polyhedral set X' defined by $A\left(\frac{R}{R'}(\mathbf{x}' - \mathbf{x}'_0)\right) \leq \mathbf{b}$, which is equivalent to $A\mathbf{x}' \leq \mathbf{b}' = A\mathbf{x}'_0 + \frac{R'}{R}\mathbf{b}$. Note that, since $X' \subseteq B'$, condition (2) holds for X'.

Proof of Lemma 2. The result follows from Jensen's inequality

$$h(E_{\mathsf{P}}[\boldsymbol{\xi}]) \geq E_{\mathsf{P}}[h(\boldsymbol{\xi})] \tag{10}$$

which holds for any concave function h and probability measure P. Indeed, consider the function $h(\mathbf{x}) = (\mathbf{c}^\top\mathbf{x})^\eta$, which is concave for $0 < \eta \leq 1$ as a composition of concave and linear functions. To apply (10), observe that $\Sigma(\boldsymbol{\xi}, \mathsf{P})$ is the expected value of the random variable $\boldsymbol{\xi}$ with respect to the probability measure P; that is, $\Sigma(\boldsymbol{\xi}, \mathsf{P}) = E_{\mathsf{P}}[\boldsymbol{\xi}]$, and the right-hand-side of (7) is the expected value of the random variable $h(\boldsymbol{\xi}) = (\mathbf{c}^\top\boldsymbol{\xi})^\eta$ with respect to the probability measure P; that is:

$$\int_\Theta (\mathbf{c}^\top\boldsymbol{\xi}(\theta))^\eta \mathsf{P}(d\theta) = E_{\mathsf{P}}[(\mathbf{c}^\top\boldsymbol{\xi})^\eta].$$

Proof of Proposition 1. By Lemma 2,

$$(\mathbf{c}_t^\top\mathbf{x}_t)^\eta = (\mathbf{c}_t^\top\Sigma(\boldsymbol{\xi}_t, \mathsf{P}^*_{t-1}))^\eta \geq \int_\Theta (\mathbf{c}_t^\top\boldsymbol{\xi}_t(\theta))^\eta \mathsf{P}^*_{t-1}(d\theta) = \frac{\int_\Theta (\mathbf{c}_t^\top\boldsymbol{\xi}_t(\theta))^\eta \mathsf{P}_{t-1}(d\theta)}{\mathsf{P}_{t-1}(\Theta)} = \frac{\mathsf{P}_t(\Theta)}{\mathsf{P}_{t-1}(\Theta)}. \tag{11}$$

By multiplying (11) for t from 1 to T, we get $\prod_{t=1}^{T}(\mathbf{c}_t^\top\mathbf{x}_t)^\eta \geq \mathsf{P}_T(\Theta)$.

After observing that $\prod_{t=1}^{T}(\mathbf{c}_t^\top\mathbf{x}_t)^\eta = (G_T(AA(\eta, \mathsf{P}_0, \Sigma)))^{\eta T}$ we get (8).

Lemma required in the proof of Theorem 1.

Lemma 3. *Consider a learning rate $0 < \eta \leq 1$, a simplex $\tilde{X}$, a probability measure $\tilde{\mathsf{P}}_0$ on $\tilde{X}$, and a feasible solution $\tilde{\boldsymbol{\delta}} \in \tilde{X}$. If there exists a constant B' such that the inequality*

$$\int_{\tilde{X}} \prod_{t=1}^{T}(\tilde{\mathbf{c}}_t^\top\tilde{\mathbf{x}})^\eta \tilde{\mathsf{P}}_0(d\tilde{\mathbf{x}}) \geq B' \prod_{t=1}^{T}(\tilde{\mathbf{c}}_t^\top\tilde{\boldsymbol{\delta}})^\eta \tag{12}$$

holds for all $\tilde{\mathbf{c}}_t$, $t = 1, \ldots, T$ which are unit vectors, then it holds for all nonnegative $\tilde{\mathbf{c}}_t$, $t = 1, \ldots, T$.

Proof. The proof is based on [17]. Notice that the function $\left\{\int_{\tilde{X}} (\tilde{\mathbf{c}}^\top\tilde{\mathbf{x}})^\eta \tilde{\mathsf{P}}_0(d\tilde{\mathbf{x}})\right\}^{\frac{1}{\eta}}$ is concave in $\tilde{\mathbf{c}}$, if $0 < \eta \leq 1$. This follows from Minkowski's inequality for integrals [11, Theorem 198].

Next we move in steps starting with $t = T$, first showing that if (12) holds for all $\tilde{\mathbf{c}}_T$'s which are unit vectors, then it holds for all nonnegative $\tilde{\mathbf{c}}_T$'s. Set

$$f(\tilde{\mathbf{x}}) := \frac{\prod_{t=1}^{T-1}(\tilde{\mathbf{c}}_t^\top\tilde{\mathbf{x}})^\eta}{\prod_{t=1}^{T-1}(\tilde{\mathbf{c}}_t^\top\tilde{\boldsymbol{\delta}})^\eta}.$$

We can rewrite (12) as $\left\{\int_{\tilde{X}}(\tilde{\mathbf{c}}_T^\top\tilde{\mathbf{x}})^\eta\tilde{\mathsf{P}}(d\tilde{\mathbf{x}})\right\}^{\frac{1}{\eta}} \geq (B')^{\frac{1}{\eta}}\,(\tilde{\mathbf{c}}_T^\top\tilde{\boldsymbol{\delta}})$, where the measure $\tilde{\mathsf{P}}(d\tilde{\mathbf{x}})$ is defined to be the product $f(\tilde{\mathbf{x}})\,\tilde{\mathsf{P}}_0(d\tilde{\mathbf{x}})$. The function on the left is concave in $\tilde{\mathbf{c}}_T$, and the function on the right is linear in $\tilde{\mathbf{c}}_T$. Therefore, it is sufficient to prove inequality (12) for $\tilde{\mathbf{c}}_T$'s which define the extreme rays of a nonnegative orthant. Moreover, both sides are positive-homogeneous of degree 1 in $\tilde{\mathbf{c}}_T$. This implies that if (12) holds for all $\tilde{\mathbf{c}}_T$ which are unit vectors, then it holds for all nonnegative $\tilde{\mathbf{c}}_T$'s. From now on $\tilde{\mathbf{c}}_T$ is assumed to be a unit vector.

Next, set

$$f(\tilde{\mathbf{x}}) := \frac{\prod_{t\in\{1,\ldots,T-2,T\}}(\tilde{\mathbf{c}}_t^\top\tilde{\mathbf{x}})^\eta}{\prod_{\in\{1,\ldots,T-2,T\}}(\tilde{\mathbf{c}}_t^\top\tilde{\boldsymbol{\delta}})^\eta},$$

and rewrite (12) as $\left\{\int_{\tilde{X}}(\tilde{\mathbf{c}}_{T-1}^\top\tilde{\mathbf{x}})^\eta\tilde{\mathsf{P}}(d\tilde{\mathbf{x}})\right\}^{\frac{1}{\eta}} \geq (B')^{\frac{1}{\eta}}\,(\tilde{\mathbf{c}}_{T-1}^\top\tilde{\boldsymbol{\delta}})$, where the measure $\tilde{\mathsf{P}}(d\tilde{\mathbf{x}})$ is defined above. The same argument permits the assumption that $\tilde{\mathbf{c}}_{T-1}$ is a unit vector, and we can apply similar arguments to $\tilde{\mathbf{c}}_{T-2},\ldots,\tilde{\mathbf{c}}_1$.

Proof of Theorem 1. It follows from Proposition 1 and the relation (6) that

$$(G_T(AA(\eta,\mathsf{P}_0,\Sigma)))^T \geq \left(\int_X \prod_{t=1}^T(\mathbf{c}_t^\top\mathbf{x})^\eta\mathsf{P}_0(d\mathbf{x})\right)^{\frac{1}{\eta}}.$$

We also observe that $(G_T(\boldsymbol{\delta}))^T = \prod\limits_{t=1}^T \mathbf{c}_t^\top\boldsymbol{\delta}$. Therefore, to show (9), we need to prove

$$\left(\int_X \prod_{t=1}^T(\mathbf{c}_t^\top\mathbf{x})^\eta\mathsf{P}_0(d\mathbf{x})\right)^{\frac{1}{\eta}} \geq (B')^{\frac{1}{\eta}}(\eta T+k)^{-\frac{k-1/2}{\eta}}\prod_{t=1}^T\mathbf{c}_t^\top\boldsymbol{\delta}, \tag{13}$$

where $B'>0$ is a constant independent of T and η.

Map the set X onto a $(k-1)$-dimensional simplex $\tilde{X}$ in k-dimensional space so that the vertices of $\tilde{X}$ correspond to the vertices of X. Note: we use $\tilde{}$ to represent the parameters on the simplex. Then,

$$\int_X \prod_{t=1}^T(\mathbf{c}_t^\top\mathbf{x})^\eta\mathsf{P}_0(d\mathbf{x}) \geq \frac{1}{Q}\int_{\tilde{X}}\prod_{t=1}^T(\tilde{\mathbf{c}}_t^\top\tilde{\mathbf{x}})^\eta\tilde{\mathsf{P}}_0(d\tilde{\mathbf{x}}), \tag{14}$$

where $\tilde{\mathsf{P}}_0$ is a uniform distribution on $\tilde{X}$, $\tilde{\mathbf{c}}_t$ is a vector of the (nonnegative) objective function values $\mathbf{c}_t^\top\mathbf{x}^i$ for every vertex $\mathbf{x}^i$ of X, and Q is a constant that only depends on the polytope X (it does not depend on T and η). Note that for an arbitrary $\tilde{\boldsymbol{\delta}}\in\tilde{X}$, there is a corresponding $\boldsymbol{\delta}=\sum_{i=1}^k\tilde{\delta}_i\mathbf{x}^i\in X$ (where k is the number of vertices of X), and $\prod_{t=1}^T\mathbf{c}_t^\top\boldsymbol{\delta}=\prod_{t=1}^T\tilde{\mathbf{c}}_t^\top\tilde{\boldsymbol{\delta}}$.

By Lemma 3 in the Appendix, it is enough to consider sequences of $\tilde{\mathbf{c}}_t$'s which are unit vectors. These sequences are merely a mathematical construction used in the proof and may not necessarily correspond to sequences of objective vectors which can be observed in the problem.

Suppose that $\tilde{\mathbf{c}}_{t,i}=1$ occurs T_i times out of T, $i=1,\ldots,k$. Without loss of generality, we sort the indices i so that $T_i\geq 1$ for $i=1,\ldots,k'$, and $T_i=0$ for $i=k'+1,\ldots,k$ for some $k'\leq k$. Then,

$$\frac{\left\{\int_{\tilde{X}}\prod_{t=1}^T(\tilde{\mathbf{c}}_t^\top\tilde{\mathbf{x}})^\eta\tilde{\mathsf{P}}_0(d\tilde{\mathbf{x}})\right\}^{\frac{1}{\eta}}}{\prod_{t=1}^T\tilde{\mathbf{c}}_t^\top\tilde{\boldsymbol{\delta}}} = \frac{\left\{\int_{\tilde{X}}\prod_{i=1}^{k'}(\tilde{\mathbf{x}}_i)^{T_i\eta}\tilde{\mathsf{P}}_0(d\tilde{\mathbf{x}})\right\}^{\frac{1}{\eta}}}{\prod_{i=1}^{k'}\tilde{\delta}_i^{\,T_i}} \geq \frac{\left\{\int_{\tilde{X}}\prod_{i=1}^{k'}(\tilde{\mathbf{x}}_i)^{T_i\eta}\tilde{\mathsf{P}}_0(d\tilde{\mathbf{x}})\right\}^{\frac{1}{\eta}}}{\prod_{i=1}^{k'}(\frac{T_i}{T})^{T_i}}. \tag{15}$$

The last inequality holds because the optimal solution to the problem

$$\max\left\{\prod_{i=1}^{k'}\tilde{\delta}_i^{\,T_i} : \sum_{i=1}^{k}\tilde{\delta}_i=1,\ \tilde{\delta}_i\geq 0,\ i=1,\ldots,k\right\}$$

is given by $\tilde{\delta}_i = \frac{T_i}{T}$, $i = 1, \ldots, k$.

By a simple corollary of Stirling's approximation to the gamma function $\Gamma(z)$ (see [1, Article 6.1.37]), there exist constants C_1 and C_2 such that for all $z \geq 1$: $C_1 z^{z-1/2} e^{-z} \leq \Gamma(z) \leq C_2 z^{z-1/2} e^{-z}$. The function under the integral on the right-hand-side of (15) is an unnormalized Dirichlet$(T_1\eta + 1, \ldots, T_k\eta + 1)$ distribution on $\tilde{X}$. Since its normalizing constant is equal to $\dfrac{\prod_{i=1}^{k} \Gamma(T_i\eta + 1)}{\Gamma(T\eta + k)}$, and the inverse of the normalizing constant of the uniform distribution on $\tilde{X}$ is $\Gamma(k)$, we can continue equation (15) as follows

$$
\begin{aligned}
\frac{\left\{\int_{\tilde{X}} \prod_{i=1}^{k'} (\tilde{\mathbf{x}}_i)^{T_i\eta} \tilde{\mathsf{P}}_0(d\tilde{\mathbf{x}})\right\}^{\frac{1}{\eta}}}{\prod_{i=1}^{k'} \left(\frac{T_i}{T}\right)^{T_i}} &= \left\{\frac{\prod_{i=1}^{k'} \Gamma(T_i\eta + 1)}{\Gamma(T\eta + k)} \Gamma(k)\right\}^{\frac{1}{\eta}} \frac{T^T}{\prod_{i=1}^{k'} T_i^{T_i}} \\
&\geq \left\{\prod_{i=1}^{k'} \left[\frac{C_1 (T_i\eta + 1)^{T_i\eta + 1/2} e^{-T_i\eta - 1}}{(T_i\eta)^{T_i\eta}}\right] \cdot \frac{(T\eta)^{T\eta} \Gamma(k)}{C_2 (T\eta + k)^{T\eta + k - 1/2} e^{-T\eta - k}}\right\}^{\frac{1}{\eta}} \\
&= \left\{\prod_{i=1}^{k'} \left[C_1 \left(1 + \frac{1}{T_i\eta}\right)^{T_i\eta} (T_i\eta + 1)^{1/2} e^{-1}\right] \cdot \frac{\Gamma(k)}{C_2 \left(1 + \frac{k}{T\eta}\right)^{T\eta} (T\eta + k)^{k - 1/2} e^{-k}}\right\}^{\frac{1}{\eta}}.
\end{aligned} \tag{16}
$$

Moreover, since $\left(1 + \frac{1}{T_i\eta}\right)^{T_i\eta} \geq 1$ and $\left(1 + \frac{k}{T\eta}\right)^{T\eta} \leq e^k$, inequality (16) implies

$$
\begin{aligned}
\frac{\left\{\int_{\tilde{X}} \prod_{i=1}^{k'} (\tilde{\mathbf{x}}_i)^{T_i\eta} \tilde{\mathsf{P}}_0(d\tilde{\mathbf{x}})\right\}^{\frac{1}{\eta}}}{\prod_{i=1}^{k'} \left(\frac{T_i}{T}\right)^{T_i}} &\geq \left(\frac{\prod_{i=1}^{k'} [C_1 e^{-1} (T_i\eta + 1)^{\frac{1}{2}}] \Gamma(k)}{C_2 (T\eta + k)^{k - \frac{1}{2}}}\right)^{\frac{1}{\eta}} \geq \left(\frac{(C_1 e^{-1})^{k'} \Gamma(k)}{C_2 (\eta T + k)^{k - \frac{1}{2}}}\right)^{\frac{1}{\eta}} \\
&\geq \left(\frac{B''}{(\eta T + k)^{k - \frac{1}{2}}}\right)^{\frac{1}{\eta}},
\end{aligned} \tag{17}
$$

where $B'' > 0$ is a constant which does not depend on T and η. Combining (14), (15), and (17), we get (13).

To obtain the optimal value of the learning rate η, we need to minimize the expression

$$\frac{k - \frac{1}{2}}{\eta} \ln(\eta T + k) + \frac{B}{\eta}$$

with respect to η. We show next that this expression is decreasing in η, and, therefore, the optimal $\eta = 1$.

The second term $\frac{B}{\eta}$ is decreasing in η. The first term is nonincreasing in η since

$$\frac{d}{d\eta}\left(\frac{1}{\eta} \ln(\eta T + k)\right) = \frac{1}{\eta^2}\left(\frac{\eta T}{\eta T + k} - \ln(\eta T + k)\right) \leq 0.$$

Indeed, let $z = \eta T$, and observe that $\left(\frac{z}{z+k} - \ln(z + k)\right)\Big|_{z=0} \leq 0$, and $\frac{d}{dz}\left(\frac{z}{z+k} - \ln(z + k)\right) = -\frac{z}{(z+k)^2} < 0$, for $z > 0$. Thus, $\frac{z}{z+k} - \ln(z + k) \leq 0$ for all $z \geq 0$.

Proof of Proposition 2. Let $X(\epsilon) = \{\mathbf{x} \in X : \mathbf{c}^\top \mathbf{x} \geq z(1 - \epsilon)\}$ for $\epsilon \geq 0$. Note that

$$\mathbf{c}^\top \mathbf{x}_t = \frac{\int_X (\mathbf{c}^\top \mathbf{x})(\mathbf{c}^\top \mathbf{x})^{\eta t} \mathsf{P}_0(d\mathbf{x})}{\int_X (\mathbf{c}^\top \mathbf{x})^{\eta t} \mathsf{P}_0(d\mathbf{x})} \geq z(1 - \epsilon) \frac{\int_{X(\epsilon)} (\mathbf{c}^\top \mathbf{x})^{\eta t} \mathsf{P}_0(d\mathbf{x})}{\int_X (\mathbf{c}^\top \mathbf{x})^{\eta t} \mathsf{P}_0(d\mathbf{x})}.$$

Therefore, to prove the statement it is sufficient to construct a sequence $\{\epsilon_t\}$ such that $\epsilon_t \to 0$ and

$$\frac{\int_{X(\epsilon_t)} (\mathbf{c}^\top \mathbf{x})^{\eta t} \mathsf{P}_0(d\mathbf{x})}{\int_X (\mathbf{c}^\top \mathbf{x})^{\eta t} \mathsf{P}_0(d\mathbf{x})} \to 1 \text{ as } t \to \infty.$$

The latter requirement is equivalent to

$$\frac{\int_{X\setminus X(\epsilon_t)} (\mathbf{c}^\top \mathbf{x})^{\eta t} \mathsf{P}_0(d\mathbf{x})}{\int_{X(\epsilon_t)} (\mathbf{c}^\top \mathbf{x})^{\eta t} \mathsf{P}_0(d\mathbf{x})} \to 0.$$

Note that $\int_{X\setminus X(\epsilon_t)} (\mathbf{c}^\top \mathbf{x})^{\eta t} \mathsf{P}_0(d\mathbf{x}) \le (z(1-\epsilon_t))^{\eta t} \mathsf{P}_0(X)$. Also, for all sufficiently small ϵ, the set $X(\epsilon)$ contains a pyramidal (conic shaped) subset $K(\epsilon)$ with a vertex in $X(0)$ and the base in the plane $\{\mathbf{x} : \mathbf{c}^\top \mathbf{x} = z(1-\epsilon)\}$. The hyperarea of the base of $K(\epsilon)$ is directly proportional to $\epsilon^{\bar{n}-1}$ times a constant (here, $\bar{n}$ is the relative dimension of X). Let $\{\nu_t\}$ be any nonnegative sequence such that $\nu_t \le \epsilon_t$. Then,

$$\int_{X(\epsilon_t)} (\mathbf{c}^\top \mathbf{x})^{\eta t} \mathsf{P}_0(d\mathbf{x}) \ge A \int_0^{\epsilon_t} (z(1-\epsilon))^{\eta t} \epsilon^{\bar{n}-1} d\epsilon \ge A(z(1-\nu_t))^{\eta t} \int_0^{\nu_t} \epsilon^{\bar{n}-1} d\epsilon = \frac{A}{\bar{n}} (z(1-\nu_t))^{\eta t} \nu_t^{\bar{n}}.$$

We have

$$\frac{\int_{X\setminus X(\epsilon_t)} (\mathbf{c}^\top \mathbf{x})^{\eta t} \mathsf{P}_0(d\mathbf{x})}{\int_{X(\epsilon_t)} (\mathbf{c}^\top \mathbf{x})^{\eta t} \mathsf{P}_0(d\mathbf{x})} \le \frac{(z(1-\epsilon_t))^{\eta t} \mathsf{P}_0(X)}{\frac{A}{\bar{n}} (z(1-\nu_t))^{\eta t} \nu_t^{\bar{n}}}.$$

The right-hand-side of this inequality converges to 0 as $t \to \infty$ as long as

$$\left(\frac{1-\epsilon_t}{1-\nu_t}\right)^{\eta t} \frac{1}{\nu_t^{\bar{n}}} \to 0.$$

If $\nu_t = \epsilon_t^2$ the convergence to 0 occurs as long as $(1+\epsilon_t)^{\eta t} \epsilon_t^{2\bar{n}} \to \infty$ or as long as $\eta t \ln(1+\epsilon_t) + 2\bar{n} \ln \epsilon_t \to \infty$. We now see that $\epsilon_t = \frac{1}{t^q}$ can be used for any $q \in [\frac{1}{2}, 1)$. Indeed, from the Taylor series expansion of $\ln(1+\frac{1}{t^q})$, we get

$$\eta t \ln\left(1+\frac{1}{t^q}\right) - 2\bar{n} q \ln t = \eta t \left(\frac{1}{t^q} - \frac{1}{2t^{2q}} + \dots\right) - 2\bar{n} q \ln t = \eta \left(t^{1-q} - \frac{1}{2} t^{1-2q} + \dots\right) - 2\bar{n} q \ln t \to \infty.$$