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 $K^o$ has a nonempty interior, there exist $x'_0 \in -K^o$ and a ball $B'$ of radius $R'$ around $x'_0$ such that $B' \subseteq -K^o$. Since $B' \subseteq -K^o$, we have
\[
    c^T x' \geq 0 \quad \text{for all } c \in \Omega \subseteq K.
\]
We establish a one-to-one correspondence between points of $X$ under this mapping is a polyhedral set $X'$ defined by $A \left( \frac{R'}{R} (x' - x'_0) \right) \leq b$, which is equivalent to $A x' \leq b' = A x'_0 + \frac{R'}{R} b$. Note that, since $X' \subseteq B'$, condition (2) holds for $X'$.

Proof of Lemma 2. The result follows from Jensen’s inequality
\[
    h(E_P[\xi]) \geq E_P[h(\xi)] \quad (10)
\]
which holds for any concave function $h$ and probability measure $P$. Indeed, consider the function $h(x) = (c^T x)^\eta$, which is concave for $0 < \eta \leq 1$ as a composition of concave and linear functions. To apply (10), observe that $\Sigma(\xi, P)$ is the expected value of the random variable $\xi$ with respect to the probability measure $P$; that is, $\Sigma(\xi, P) = E_P[\xi]$, and the right-hand-side of (7) is the expected value of the random variable $h(\xi) = (c^T \xi)^\eta$ with respect to the probability measure $P$; that is:
\[
    \int_\Theta (c^T \xi(\theta))^\eta P(d\theta) = E_P[(c^T \xi)^\eta].
\]

Proof of Proposition 1. By Lemma 2,
\[
    (c_t^T x_t)^\eta = (c_t^T \Sigma(\xi_t, P_{t-1}))^\eta \geq \int_\Theta (c_t^T \xi_t(\theta))^\eta P_{t-1}(d\theta) = \frac{\int_\Theta (c_t^T \xi_t(\theta))^\eta P_{t-1}(d\theta)}{P_{t-1}(\Theta)} = \frac{P_t(\Theta)}{P_{t-1}(\Theta)}.
\]
By multiplying (11) for $t$ from 1 to $T$, we get $\prod_{t=1}^T (c_t^T x_t)^\eta \geq P_T(\Theta)$.

After observing that $\prod_{t=1}^T (c_t^T x_t)^\eta = (G_T(AA(\eta, P_0, \Sigma)))^\eta$, we get (8).

Lemma required in the proof of Theorem 1.

Lemma 3. Consider a learning rate $0 < \eta \leq 1$, a simplex $\bar{X}$, a probability measure $\bar{P}_0$ on $\bar{X}$, and a feasible solution $\bar{\delta} \in \bar{X}$. If there exists a constant $B'$ such that the inequality
\[
    \int_{\bar{X}} \prod_{t=1}^T (c_t^T \bar{x})^\eta \bar{P}_0(d\bar{x}) \geq B' \prod_{t=1}^T (c_t^T \bar{\delta})^\eta \quad (12)
\]
holds for all $c_t$, $t = 1, \ldots, T$ which are unit vectors, then it holds for all nonnegative $c_t$, $t = 1, \ldots, T$.

Proof. The proof is based on [17]. Notice that the function $\left( \int_{\bar{X}} (c^T \bar{x})^\eta \bar{P}_0(d\bar{x}) \right)^{1/\eta}$ is concave in $c$, 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 $c_T$’s which are unit vectors, then it holds for all nonnegative $c_T$’s. Set
\[
    f(\bar{x}) := \left( \frac{\prod_{t=1}^{T-1} (c_t^T \bar{x})^\eta}{\prod_{t=1}^{T-1} (c_t^T \bar{\delta})^\eta} \right)^{1/\eta}.
\]
We can rewrite (12) as \( \left\{ \int_X (\tilde{c}_i^T \tilde{x})_n \tilde{P}(d\tilde{x}) \right\}^{\frac{1}{\eta}} \geq (B')^{\frac{1}{\eta}} (\tilde{c}_i^T \tilde{\delta}) \), where the measure \( \tilde{P}(d\tilde{x}) \) is defined to be the product \( f(\tilde{x}) \tilde{P}_0(d\tilde{x}) \). The function on the left is concave in \( \tilde{c}_T \), and the function on the right is linear in \( \tilde{c}_T \). Therefore, it is sufficient to prove inequality (12) for \( \tilde{c}_T \)'s which define the extreme rays of a nonnegative orthant. Moreover, both sides are positive-homogeneous of degree 1 in \( \tilde{c}_T \). This implies that if (12) holds for all \( \tilde{c}_T \) which are unit vectors, then it holds for all nonnegative \( \tilde{c}_T \)'s. From now on \( \tilde{c}_T \) is assumed to be a unit vector.

Next, set
\[
 f(\tilde{x}) := \frac{\prod_{t \in \{1, \ldots, T-2, T\}} (\tilde{c}_i^T \tilde{x})^\eta}{\prod_{t \in \{1, \ldots, T-2, T\}} (\tilde{c}_i^T \tilde{\delta})^\eta},
\]
and rewrite (12) as
\[
 \left\{ \int_X (\tilde{c}_i^T \tilde{x})_n \tilde{P}(d\tilde{x}) \right\}^{\frac{1}{\eta}} \geq (B')^{\frac{1}{\eta}} (\tilde{c}_i^T \tilde{\delta}),
\]
where the measure \( \tilde{P}(d\tilde{x}) \) is defined above.

The same argument permits the assumption that \( \tilde{c}_{T-1} \) is a unit vector, and we can apply similar arguments to \( \tilde{c}_{T-2}, \ldots, \tilde{c}_1 \).

**Proof of Theorem 1.** It follows from Proposition 1 and the relation (6) that
\[
 (G_T(AA(\eta, P_0, \Sigma)))^T \geq \left( \int_X \prod_{t=1}^T (c_i^T x)^\eta P_0(dx) \right)^{\frac{1}{\eta}}.
\]
We also observe that \( (G_T(\tilde{\delta}))^T = \prod_{t=1}^T c_i^T \tilde{\delta} \). Therefore, to show (9), we need to prove
\[
 \left( \int_X \prod_{t=1}^T (c_i^T x)^\eta P_0(dx) \right)^{\frac{1}{\eta}} \geq (B')^{\frac{1}{\eta}} (\eta T + k - \frac{1}{\eta}) \prod_{t=1}^T c_i^T \tilde{\delta},
\]
where \( B' > 0 \) is a constant independent of \( T \) and \( \eta \).

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 \( \sim \) to represent the parameters on the simplex. Then,
\[
 \int_X \prod_{t=1}^T (c_i^T x)^\eta P_0(dx) \geq \frac{1}{Q} \int_{\tilde{X}} \prod_{t=1}^T (\tilde{c}_i^T \tilde{x})^\eta \tilde{P}_0(d\tilde{x}),
\]
where \( \tilde{P}_0 \) is a uniform distribution on \( \tilde{X} \), \( \tilde{c}_t \) is a vector of the (nonnegative) objective function values \( c_i^T x^t \) for every vertex \( x^t \) of \( X \), and \( Q \) is a constant that only depends on the polytope \( X \) (it does not depend on \( T \) and \( \eta \)). Note that for an arbitrary \( \tilde{\delta} \in \tilde{X} \), there is a corresponding \( \tilde{\delta} = \sum_{i=1}^k \tilde{\delta}_i x^i \in X \) (where \( k \) is the number of vertices of \( X \)), and \( \prod_{t=1}^T \tilde{c}_i^T \tilde{\delta} = \prod_{t=1}^T \tilde{c}_i^T \tilde{\delta} \).

By Lemma 3 in the Appendix, it is enough to consider sequences of \( \tilde{c}_t \)'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 \( \tilde{c}_{t,i} = 1 \) occurs \( T_i \) times out of \( T \), \( i = 1, \ldots, k \). Without loss of generality, we sort the indices \( i \) so that \( T_i \geq T_{i+1} \) for \( i = 1, \ldots, k' \), and \( T_i = 0 \) for \( i = k'+1, \ldots, k \) for some \( k' \leq k \). Then,
\[
 \left\{ \int_{\tilde{X}} \prod_{t=1}^T (\tilde{c}_i^T \tilde{x})_n \tilde{P}_0(d\tilde{x}) \right\}^{\frac{1}{\eta}} \geq \left\{ \int_{\tilde{X}} \prod_{t=1}^{k'} (\tilde{x}_i)_n \tilde{P}_0(d\tilde{x}) \right\}^{\frac{1}{\eta}}.
\]
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 \( z \geq 1 \): \( C_1 z^{-1/2} e^{-z} \leq \Gamma(z) \leq C_2 z^{-1/2} e^{-z} \). The function under the integral on the right-hand-side of (15) is an unnormalized Dirichlet \((T_i \eta + 1, \ldots, T_k \eta + 1)\) distribution on \( \tilde{X} \). Since its normalizing constant is equal to \( \frac{\prod_{i=1}^k \Gamma(T_i \eta + 1)}{\Gamma(T \eta + k)} \), and the inverse of the normalizing constant of the uniform distribution on \( \tilde{X} \) is \( \Gamma(k) \), we can continue equation (15) as follows

\[
\left\{ \frac{\int_{\tilde{X}} \prod_{i=1}^{k'} (\tilde{x}_i)^{T_i \eta} \tilde{p}_0(d\tilde{x})}{\prod_{i=1}^{k'} \left( \frac{T_i}{T} \right)^{T_i}} \right\}^{\frac{1}{\eta}} \geq \left\{ \frac{\prod_{i=1}^{k'} \Gamma(T_i \eta + 1)}{\Gamma(T \eta + k)} \frac{T^T}{\prod_{i=1}^{k'} T_i} \Gamma(k) \right\}^{\frac{1}{\eta}}
\]

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

\[
\left\{ \frac{\int_{\tilde{X}} \prod_{i=1}^{k'} (\tilde{x}_i)^{T_i \eta} \tilde{p}_0(d\tilde{x})}{\prod_{i=1}^{k'} \left( \frac{T_i}{T} \right)^{T_i}} \right\}^{\frac{1}{\eta}} \geq \left( \frac{\prod_{i=1}^{k'} C_1 e^{-1} (T_i \eta + 1)^{\frac{1}{2}} \Gamma(k)}{\prod_{i=1}^{k'} C_2 (T \eta + k)^{k^{-\frac{1}{2}}} \Gamma(k)} \right) \frac{1}{\eta} \geq \left( \frac{C_1 e^{-1} \Gamma(k)}{C_2 (\eta T + k)^{k^{-\frac{1}{2}}}} \right) \frac{1}{\eta}
\]

\[
\geq \left( \frac{B''}{(\eta T + k)^{k^{-\frac{1}{2}}}} \right)^{\frac{1}{\eta}},
\]

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

To obtain the optimal value of the learning rate \( \eta \), we need to minimize the expression

\[
\frac{k - \frac{1}{\eta} \ln(\eta T + k) + \frac{B}{\eta}}{\eta}
\]

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

The second term \( \frac{B}{\eta} \) is decreasing in \( \eta \). The first term is nonincreasing in \( \eta \) 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 \( \left( \frac{z}{z + k} - \ln(z + k) \right) \mid_{z = 0} \leq 0 \), 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 \( X(\epsilon) = \{ x \in X : c^T x \geq z(1 - \epsilon) \} \) for \( \epsilon \geq 0 \). Note that

\[
c^T x_\epsilon = \frac{\int_X (c^T x)(c^T x)^\eta \tilde{p}_0(dx)}{\int_X (c^T x)^\eta \tilde{p}_0(dx)} \geq z(1 - \epsilon) \frac{\int_X (c^T x)^\eta \tilde{p}_0(dx)}{\int_X (c^T x)^\eta \tilde{p}_0(dx)}.
\]
Therefore, to prove the statement it is sufficient to construct a sequence \( \{ \epsilon_t \} \) such that \( \epsilon_t \to 0 \) and

\[
\frac{\int_{X(\epsilon_t)} (c \mathbf{u} \cdot \mathbf{x})^t \mu \mathbb{P}_0(d\mathbf{x})}{\int_X (c \mathbf{u} \cdot \mathbf{x})^t \mu \mathbb{P}_0(d\mathbf{x})} \to 1 \text{ as } t \to \infty.
\]

The latter requirement is equivalent to

\[
\frac{\int_{X(\epsilon_t) \setminus X(\epsilon_t)} (c \mathbf{u} \cdot \mathbf{x})^t \mu \mathbb{P}_0(d\mathbf{x})}{\int_{X(\epsilon_t)} (c \mathbf{u} \cdot \mathbf{x})^t \mu \mathbb{P}_0(d\mathbf{x})} \to 0.
\]

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

\[
\int_{X(\epsilon_t)} (c \mathbf{u} \cdot \mathbf{x})^t \mu \mathbb{P}_0(d\mathbf{x}) \geq A \int_0^{\epsilon_t} (z(1 - \epsilon))^t \epsilon^{\tilde{n} - 1} d\epsilon \geq A(z(1 - \nu_t))^t \nu_t^{\tilde{n}} \int_0^{\nu_t} \epsilon^{\tilde{n} - 1} d\epsilon = \frac{A}{\tilde{n}} (z(1 - \nu_t))^t \nu_t^{\tilde{n}}.
\]

We have

\[
\frac{\int_{X(\epsilon_t) \setminus X(\epsilon_t)} (c \mathbf{u} \cdot \mathbf{x})^t \mu \mathbb{P}_0(d\mathbf{x})}{\int_{X(\epsilon_t)} (c \mathbf{u} \cdot \mathbf{x})^t \mu \mathbb{P}_0(d\mathbf{x})} \leq \frac{A}{\tilde{n}} (z(1 - \nu_t))^t \nu_t^{\tilde{n}}.
\]

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

\[
\frac{(1 - \epsilon_t)^t}{1 - \nu_t} \to 0.
\]

If \( \nu_t = \epsilon_t^2 \), the convergence to 0 occurs as long as \( (1 + \epsilon_t)^{\tilde{n}} \to \infty \) or as long as \( \eta t \ln(1 + \epsilon_t) + 2\tilde{n} \ln \epsilon_t \to \infty \).

We now see that \( \epsilon_t = \frac{1}{\tilde{n}} \) 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

\[
\eta t \ln \left( 1 + \frac{1}{t^q} \right) - 2\tilde{n}q \ln t = \eta t \left( \frac{1}{t^q} - \frac{1}{2t^{2q}} + \ldots \right) - 2\tilde{n}q \ln t = \eta \left( t^{1-q} - \frac{1}{2} t^{1-2q} + \ldots \right) - 2\tilde{n}q \ln t \to \infty.
\]