Linear Bandits.

Peter Bartlett

- Linear bandits.
  - Exponential weights with unbiased loss estimates.
  - Controlling loss estimates and their variance.
    - Barycentric spanner.
    - Uniform distribution.
    - John's distribution.
  - Lower bounds.
  - Stochastic mirror descent.
    - Full information.
    - Bandit information.

Linear bandits

At round t,

- Strategy chooses $a_t \in \mathcal{A} \subset \mathbb{R}^d$.
- Adversary chooses linear loss $\ell_t : \mathcal{A} \to [-1, 1]$.
- Strategy sees loss $\ell_t(a_t)$.

Loss is *linear* in action.

Aim to minimize regret:

$$\overline{R}_n = \mathbb{E} \sum_{t=1}^{n} \ell_t(a_t) - \inf_{a \in \mathcal{A}} \mathbb{E} \sum_{t=1}^{n} \ell_t(a).$$

Example: Packet routing

Consider the problem of packet-routing in a network (V, E). At round t,

- Strategy chooses a path $a_t \in \mathcal{A} \subset \{0,1\}^E$ from origin node to destination node.
- Adversary chooses delays $\ell_t \in \mathcal{L} = [0,1]^E$.
- See loss $\ell_t \cdot a_t$ (total delay).

Aim to minimize regret:

$$\overline{R}_n = \mathbb{E} \sum_{t=1}^{n} \ell_t \cdot a_t - \inf_{a \in \mathcal{A}} \mathbb{E} \sum_{t=1}^{n} \ell_t \cdot a.$$

Loss is *linear* in action.

Linear bandits vs k-armed bandits

This problem is closely related to the classical k-armed bandit problem:
At round t:

- Strategy chooses $a_t \in \mathcal{A} = \{1, \ldots, k\}$.
- Adversary chooses $\ell_t \in \mathcal{L} = [0,1]^{\mathcal{A}}$.
- See loss $\ell_t(a_t)$.

Aim to minimize regret:

$$\overline{R}_n = \mathbb{E} \sum_{t=1}^{n} \ell_t(a_t) - \inf_{a \in \mathcal{A}} \mathbb{E} \sum_{t=1}^{n} \ell_t(a).$$

Linear bandits vs k-armed bandits

This is unchanged (up to a constant factor) if we instead define

$$\mathcal{A} = \{e_1, \ldots, e_k\} \subset \mathbb{R}^k,$$
$$\mathcal{L} = \{\ell : \mathcal{A} \to [-1, 1] \text{ linear}\}.$$

And allowing the strategy to choose a in the convex hull of $\mathcal{A}$ does not change the regret

$$\overline{R}_n = \mathbb{E} \sum_{t=1}^{n} \ell_t(a_t) - \inf_{a \in \mathcal{A}} \mathbb{E} \sum_{t=1}^{n} \ell_t(a).$$

(But it might make the game easier for the strategy since it changes the information that the strategy sees.)

Finite covers

For a compact $\mathcal{A} \subseteq \mathbb{R}^d$, we can construct an ϵ-cover of size $O(1/\epsilon^d)$, for example, in the uniform metric

$$\rho(\hat{a}, a) = \|\hat{a} - a\|_\infty := \max_i |\hat{a}_i - a_i|.$$

Since we're aiming for $O(\sqrt{n})$ regret, we can think of $\mathcal{A}$ as having cardinality $|\mathcal{A}| = O(n^{d/2})$, so $\log |\mathcal{A}| = O(d \log n)$.

Exponential weights for linear bandits

Given $\mathcal{A}$, distribution μ on $\mathcal{A}$, mixing coefficient $\gamma > 0$, learning rate $\eta > 0$,

set q_1 uniform on $\mathcal{A}$.

for $t = 1, 2, \ldots, n$,

1. $p_t = (1 - \gamma)q_t + \gamma\mu$
2. choose $a_t \sim p_t$
3. observe $\ell_t^T a_t$
4. update $q_{t+1}(a) \propto q_t(a) \exp(-\eta\tilde{\ell}_t^T a)$,

where
$$\tilde{\ell}_t = \Sigma_t^{-1} a_t a_t^T \ell_t,$$
$$\Sigma_t = \mathbb{E}_{a\sim p_t} aa^T.$$

Unbiased loss estimates

- Assume $\mathrm{span}(\mathcal{A}) = \mathbb{R}^d$ (otherwise, we can project to a lower dimension) and that μ has support on a d-dimensional set. So $\mathbb{E}_{a\sim p_t} aa^T$ has rank d.
- Strategy observes $a_t^T \ell_t$ and a_t, so it can compute

$$\tilde{\ell}_t = \Sigma_t^{-1} a_t \left(a_t^T \ell_t\right).$$

- $\tilde{\ell}_t$ is unbiased:

$$\mathbb{E}\left[\tilde{\ell}_t | \mathcal{F}_{t-1}\right] = \left(\mathbb{E}_{a\sim p_t} aa^T\right)^{-1} \left(\mathbb{E}_{a_t\sim p_t} a_t a_t^T\right) \ell_t = \ell_t.$$

Regret bound

Theorem: For $\eta \sup_{a \in \mathcal{A}} \left| \tilde{\ell}_t^T a \right| \leq 1$,

$$\overline{R}_n \leq \gamma n + \frac{\log |\mathcal{A}|}{\eta} + (e-2)\eta \sum_{t=1}^{n} \mathbb{E}\mathbb{E}_{a \sim p_t} \left(\tilde{\ell}_t^T a \right)^2 .$$

So we need to control η times the magnitude of the loss estimates,

$$\eta \sup_{a \in \mathcal{A}} \left| \tilde{\ell}_t^T a \right|$$

and the variance term,

$$\mathbb{E}\mathbb{E}_{a \sim p_t} \left(\tilde{\ell}_t^T a \right)^2 .$$

Proof

The regret is

$$\mathbb{E}\left[\sum_{t=1}^{n}\left(\ell_t^T a_t - \ell_t^T a^*\right)\right].$$

We've seen that, given history $\mathcal{F}_{t-1}$,

$$\mathbb{E}\left[\tilde{\ell}_t|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\Sigma_t^{-1} a_t a_t^T \ell_t|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\ell_t|\mathcal{F}_{t-1}\right].$$

Lemma: Some unbiased estimates involving $\tilde{\ell}_t$:

$$\mathbb{E}\left[\ell_t^T a\right] = \mathbb{E}\left[\tilde{\ell}_t^T a\right],$$

$$\mathbb{E}\left[\ell_t^T a_t\right] = \mathbb{E}\left[\sum_{a\in\mathcal{A}} p_t(a)\mathbb{E}\left[\tilde{\ell}_t\middle|\mathcal{F}_{t-1}\right]^T a\right] = \mathbb{E}\left[\sum_{a\in\mathcal{A}} p_t(a)\tilde{\ell}_t^T a\right].$$

Proof

So we can write the strategy's expected cumulative loss as

$$\mathbb{E}\sum_{t=1}^{n} \ell_t^T a_t = \mathbb{E}\sum_{t=1}^{n}\sum_{a\in\mathcal{A}} p_t(a)\tilde{\ell}_t^T a.$$

We'll give up on the loss incurred in the exploration trials:

$$\begin{aligned}\sum_{t=1}^{n}\sum_{a\in\mathcal{A}} p_t(a)\tilde{\ell}_t^T a &= \sum_{t=1}^{n}\sum_{a\in\mathcal{A}} \left((1-\gamma)q_t(a) + \gamma\mu(a)\right)\tilde{\ell}_t^T a \\ &= (1-\gamma)\left(\sum_{t=1}^{n}\sum_{a\in\mathcal{A}} q_t(a)\tilde{\ell}_t^T a\right) + \underbrace{\gamma\sum_{t=1}^{n}\sum_{a\in\mathcal{A}} \mu(a)\tilde{\ell}_t^T a}_{\text{exploration}}.\end{aligned}$$

Proof

For q_t, we follow the standard analysis (see Adversarial Bandits), but instead of using non-negativity of the $\tilde{\ell}$s, we use a lower bound:

$$\begin{aligned}\log \mathbb{E} \exp\left(-\eta(X - \mathbb{E}X)\right) &\leq \mathbb{E}\left(\exp(-\eta X) - 1 + \eta X\right) \\ &\leq (e-2)\eta^2 \mathbb{E} X^2,\end{aligned}$$

where the last inequality uses $\exp(-x) \leq 1 - x + (e-2)x^2$ for $x \geq -1$.

So if $\eta \tilde{\ell}_t^T a \geq -1$ for all $a \in \mathcal{A}$, the previous analysis shows that, for any $a^* \in \mathcal{A}$, the first term above satisfies

$$\sum_{t=1}^{n} \sum_{a \in \mathcal{A}} q_t(a) \tilde{\ell}_t^T a \leq \sum_{t=1}^{n} \tilde{\ell}_t^T a^* + \frac{\log |\mathcal{A}|}{\eta} + (e-2)\eta \sum_{t=1}^{n} \sum_{a \in \mathcal{A}} q_t(a) \left(\tilde{\ell}_t^T a\right)^2.$$

Proof

Combining, and using the fact that $(1-\gamma)q_t(a) \leq p_t(a)$,

$$\sum_{t=1}^{n} \sum_{a \in \mathcal{A}} p_t(a) \tilde{\ell}_t^T a \leq \sum_{t=1}^{n} \tilde{\ell}_t^T a^*$$

$$+ \text{(exploration)} + \frac{\log |\mathcal{A}|}{\eta} + (e-2)\eta \sum_{t=1}^{n} \sum_{a \in \mathcal{A}} p_t(a) \left(\tilde{\ell}_t^T a \right)^2 .$$

The unbiasedness lemma gives

$$\overline{R}_n \leq \gamma n + \frac{\log |\mathcal{A}|}{\eta} + (e-2)\eta \sum_{t=1}^{n} \mathbb{E}_{a \sim p_t} \left(\tilde{\ell}_t^T a \right)^2 .$$

Controlling variance

Lemma: For $\mathcal{L} \subset [-1,1]^{\mathcal{A}}$, the variance term is bounded:

$$\mathbb{E}\mathbb{E}_{a\sim p_t}\left(\tilde{\ell}_t^T a\right)^2 \leq d.$$

$$\begin{aligned}
\mathbb{E}\left(\tilde{\ell}_t^T a\right)^2 &= a^T \mathbb{E}\left(\tilde{\ell}_t \tilde{\ell}_t^T\right) a \\
&= a^T \mathbb{E}\left(\left(\ell_t^T a_t\right)^2 \Sigma_t^{-1} a_t a_t^T \Sigma_t^{-1}\right) a \\
&\leq a^T \Sigma_t^{-1} \mathbb{E}\left(a_t a_t^T\right) \Sigma_t^{-1} a \\
&= a^T \Sigma_t^{-1} a.
\end{aligned}$$

$$\mathbb{E}_{a\sim p_t}\mathbb{E}\left(\tilde{\ell}_t^T a\right)^2 \leq \mathbb{E}\,\mathrm{tr}\left(a^T \Sigma_t^{-1} a\right) = \mathrm{tr}\left(\Sigma_t^{-1}\mathbb{E}\left(a a^T\right)\right) = \mathrm{tr}\left(I\right) = d.$$

Controlling the magnitude of the estimator

Lemma: For $\mathcal{L} \subset [-1,1]^{\mathcal{A}}$,

$$\left|\tilde{\ell}_t^T a\right| \leq \sup_{a,b\in\mathcal{A}} a^T \Sigma_t^{-1} b.$$

$$\begin{aligned} \left|\tilde{\ell}_t^T a\right| &= \left|a_t^T \ell_t \left(\Sigma_t^{-1} a_t\right)^T a\right| \\ &\leq \left|a_t^T \ell_t\right| \left|a_t^T \Sigma_t^{-1} a\right| \\ &\leq \sup_{a,b\in\mathcal{A}} a^T \Sigma_t^{-1} b. \end{aligned}$$

We'll see that typically $\sup_{a,b\in\mathcal{A}} a^T \Sigma_t^{-1} b \leq c_d/\gamma$.

Regret bound

Theorem: For $\mathcal{L} \subset [-1,1]^{\mathcal{A}}$, if

$$\sup_{a,b\in\mathcal{A}} a^T \Sigma_t^{-1} b \leq \frac{c_d}{\gamma},$$

$$\text{setting } \eta = \sqrt{\frac{\log|\mathcal{A}|}{n\left((e-2)d + c_d\right)}}$$

$$\gamma = c_d \eta$$

$$\text{gives } \overline{R}_n \leq 2\sqrt{n(d+c_d)\log|\mathcal{A}|}.$$

Exploration distributions

- (Dani, Hayes, Kakade, 2008):
  For μ uniform over *barycentric spanner*,
  $$\overline{R}_n = O\left(d\sqrt{n \log |\mathcal{A}|}\right) = \tilde{O}\left(d^{3/2}\sqrt{n}\right).$$
- (Cesa-Bianchi and Lugosi, 2009):
  For several combinatorial problems, $\mathcal{A} \subseteq \{0,1\}^d$, μ uniform over $\mathcal{A}$ gives
  $$\frac{\sup_{a\in\mathcal{A}} \|a\|_2^2}{\lambda_{\min}\left(\mathbb{E}_{a\sim\mu}[aa^T]\right)} = O(d),$$
  so
  $$\overline{R}_n = O\left(\sqrt{dn \log |\mathcal{A}|}\right) = \tilde{O}\left(d\sqrt{n}\right).$$
- (Bubeck, Cesa-Bianchi and Kakade, 2009): *John's Theorem*:
  $\tilde{O}\left(d\sqrt{n}\right)$.

Barycentric spanner

(Suppose that $\mathcal{A} \subseteq \mathbb{R}^d$ spans $\mathbb{R}^d$.)

A *barycentric spanner* of $\mathcal{A}$ is a set $\{b_1, \ldots, b_d\}$ that spans $\mathbb{R}^d$ and satisfies:
for all $a \in \mathcal{A}$ there is an $\alpha \in [-1, 1]^d$ such that $a = B\alpha$, where
$B = \begin{pmatrix} b_1 & \cdots & b_d \end{pmatrix}$.

- Every compact $\mathcal{A}$ has a barycentric spanner.
- If linear functions can be efficiently optimized over $\mathcal{A}$, then there is an efficient algorithm for finding an approximate barycentric spanner (that is, $|\alpha_i| \le 1 + \delta$; $O(d^2 \log d/\delta)$ linear optimizations).

Barycentric spanner

Lemma: If $\{b_1, \ldots, b_d\} \subset \mathcal{A}$ maximizes $\det(B)$, then it is a barycentric spanner.

Proof. For $a = B\alpha$,

$$\begin{aligned} |\det(B)| &\geq \left|\det\begin{pmatrix} a & b_2 & \cdots & b_d \end{pmatrix}\right| \\ &= \left|\sum_i \alpha_i \det\begin{pmatrix} b_i & b_2 & \cdots & b_d \end{pmatrix}\right| \\ &= |\alpha_1| \, |\det(B)| \, . \end{aligned}$$

$\square$

Barycentric spanner

Theorem: For $\mathcal{A} \subseteq [-1,1]^d$ and μ uniform on a barycentric spanner of $\mathcal{A}$,

$$\sup_{a,b\in\mathcal{A}} a^T \Sigma_t^{-1} b \leq \frac{d^2}{\gamma}$$

(that is, $c_d \leq d^2$). Hence,

$$\overline{R}_n \leq 2d\sqrt{2n \log |\mathcal{A}|}.$$

$$\Sigma_t = \frac{\gamma}{d} B B^T + \underbrace{(1-\gamma) \sum_{a\in\mathcal{A}} q_t(a) a a^T}_{M}.$$

Barycentric spanner: Proof

$$
\begin{aligned}
\sup_{a,b\in\mathcal{A}} a^T\Sigma_t^{-1}b &\le \sup_{\alpha,\beta\in[-1,1]^d} \alpha^T B^T \Sigma_t^{-1} B\beta \\
&\le \sup_{\|\alpha\|=\|\beta\|=\sqrt{d}} \alpha^T B^T \Sigma_t^{-1} B\beta \\
&= d\lambda_{\max}\left(B^T\Sigma_t^{-1}B\right) \\
&= d\lambda_{\max}\left(B^{-1}\Sigma_t B^{-T}\right)^{-1} \\
&= \frac{d}{\lambda_{\min}\left(B^{-1}\left(\frac{\gamma}{d}BB^T + M\right)B^{-T}\right)} \\
&\le \frac{d^2}{\gamma\lambda_{\min}\left(B^{-1}BB^TB^{-T}\right)} = \frac{d^2}{\gamma},
\end{aligned}
$$

where $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ denote the largest and smallest eigenvalues.

Other exploration distributions

Lemma:

$$\sup_{a,b\in\mathcal{A}} a^T \Sigma_t^{-1} b \le \frac{\sup_{a\in\mathcal{A}} \|a\|_2^2}{\gamma\lambda_{\min}\left(\mathbb{E}_{a\sim\mu}[aa^T]\right)}.$$

$$\begin{aligned}
\sup_{a,b\in\mathcal{A}} a^T \Sigma_t^{-1} b &\le \sup_{a\in\mathcal{A}} \|a\|_2^2 \lambda_{\max}\left(\Sigma_t^{-1}\right) \\
&= \frac{\sup_{a\in\mathcal{A}} \|a\|_2^2}{\lambda_{\min}\left(\Sigma_t\right)}. \\
\lambda_{\min}\left(\Sigma_t\right) &= \min_{\|v\|=1} \sum_{a\in\mathcal{A}} p_t(a) v^T a a^T v \\
&\ge \gamma \min_{\|v\|=1} \sum_{a\in\mathcal{A}} \mu(a) v^T a a^T v = \gamma \lambda_{\min}\left(\mathbb{E}_{a\sim\mu}[aa^T]\right).
\end{aligned}$$

John's distribution

Theorem: [John's Theorem] For any convex set $\mathcal{A} \subset \mathbb{R}^d$, denote the ellipsoid of minimal volume containing it as

$$E = \left\{x \in \mathbb{R}^d : (x-c)^T M(x-c) \leq 1\right\}.$$

Then there is a set $\{u_1, \ldots, u_m\} \subseteq E \cap \mathcal{A}$ of $m \leq d(d+1)/2+1$ contact points and a distribution p on this set such that any $x \in \mathbb{R}^d$ can be written

$$x = c + d \sum_{i=1}^{m} p_i \langle x-c, u_i - c\rangle (u_i - c),$$

where $\langle \cdot, \cdot \rangle$ is the inner product for which the minimal ellipsoid is the unit ball about its center c: $\langle x, y\rangle = x^T M y$.

John's distribution

This shows that

$$
\begin{aligned}
x - c &= d \sum_i p_i (u_i - c)(u_i - c)^T M (x - c) \\
\Leftrightarrow \qquad \tilde{x} &= d \sum_i p_i \tilde{u}_i \tilde{u}_i^T \tilde{x} \\
\Leftrightarrow \qquad \frac{1}{d} I &= \sum_i p_i \tilde{u}_i \tilde{u}_i^T,
\end{aligned}
$$

where $\tilde{u}_i = M^{1/2}(u_i - c)$, and similarly for $\tilde{x}$. Setting the exploration distribution μ to be the distribution p over the set of transformed contact points $\tilde{u}_i$, we see that, for $a, b \in \mathcal{A}$,

$$
\tilde{a}^T \mathbb{E}_{u \sim \mu} u u^T \tilde{b} = \frac{1}{d} \tilde{a}^T \tilde{b}.
$$

John's distribution

So if we shift the origin of the set $\mathcal{A}$ and of the u_i (and the corresponding introduction of a constant component in the losses), we have

$$\sup_{a,b\in\mathcal{A}} a^T \Sigma_t^{-1} b \le \frac{d}{\gamma},$$

that is, $c_d \le d$. Hence,

$$\overline{R}_n \le 2\sqrt{2nd\log|\mathcal{A}|}.$$

Exploration distributions

- (Dani, Hayes, Kakade, 2008):
  For μ uniform over *barycentric spanner*,
  $$\overline{R}_n = O\left(d\sqrt{n \log |\mathcal{A}|}\right) = \tilde{O}\left(d^{3/2}\sqrt{n}\right).$$
- (Cesa-Bianchi and Lugosi, 2009):
  For several combinatorial problems, $\mathcal{A} \subseteq \{0,1\}^d$, μ uniform over $\mathcal{A}$ gives
  $$\frac{\sup_{a \in \mathcal{A}} \|a\|_2^2}{\lambda_{\min}\left(\mathbb{E}_{a\sim\mu}[aa^T]\right)} = O(d),$$
  so
  $$\overline{R}_n = O\left(\sqrt{dn \log |\mathcal{A}|}\right) = \tilde{O}\left(d\sqrt{n}\right).$$
- (Bubeck, Cesa-Bianchi and Kakade, 2009): *John's Theorem*:
  $\tilde{O}\left(d\sqrt{n}\right)$.

Outline

- Linear bandits.
  - Exponential weights with unbiased loss estimates.
  - Controlling loss estimates and their variance.
    - Barycentric spanner.
    - Uniform distribution.
    - John's distribution.
  - Lower bounds.
  - Stochastic mirror descent.
    - Full information.
    - Bandit information.

Lower bounds

Lower bounds from the stochastic setting suffice.

Theorem: Consider $\mathcal{A} = \{\pm 1\}^d$, $\mathcal{L} \supseteq \{\pm e_i : 1 \leq i \leq d\}$. There is a constant c such that, for any strategy and any n, there is an i.i.d. adversary for which

$$\overline{R}_n \geq cd\sqrt{n}.$$

(Here, $\sqrt{nd \log |\mathcal{A}|} = O(d\sqrt{n})$.)

Lower bounds: proof

Probabilistic method: Fix $\epsilon \in (0, 1/2)$ and, for each $b \in \{\pm 1\}^d$, define P_b on $\mathcal{L}$ as

$$P_b(e_i) = \frac{1 - b_i \epsilon}{2d},$$
$$P_b(-e_i) = \frac{1 + b_i \epsilon}{2d}.$$

(so that the optimal $a^* = b$). We'll choose b uniformly, and show that the expected regret under this choice is large.

Lower bounds: proof

$$
\begin{aligned}
\overline{R}_n(P_b) &= \sum_{t=1}^{n} \sum_{i=1}^{d} \mathbb{E}\left[\ell_{t,i}\left(a_{t,i} - b_i\right)\right] \\
&= \sum_{t=1}^{n} \sum_{i=1}^{d} (a_{t,i} - b_i) \left(\frac{1 - 2b_i\epsilon}{2d} - \frac{1 + 2b_i\epsilon}{2d} \right) \\
&= \sum_{t=1}^{n} \sum_{i=1}^{d} (b_i - a_{t,i}) \frac{b_i\epsilon}{d} \\
&= \sum_{i=1}^{d} \underbrace{\frac{2\epsilon}{d} \sum_{t=1}^{n} 1[a_{t,i} \neq b_i]}_{\overline{R}_n^i(b_i)} .
\end{aligned}
$$

Lower bounds: proof

The regret of sub-game i, $\overline{R}_n^i(b_i)$, is at least the regret that would be incurred if the strategy knew that the adversary was using one of the P_b distributions, and also knew $\{b_j : j \neq i\}$. In that case, it would know

$$\theta := \mathbb{E} \sum_{j \neq i} l_{t,j} a_{t,j},$$

and so at each round, it would see a (± 1) Bernoulli random variable $\ell_t^T a_t$, with mean

$$\theta - b_i a_{t,i} \frac{\epsilon}{d}.$$

Notice that the $1/d$ here is crucial: because information about the ith component only arrives once every d rounds on average, the range of values of the unknown Bernoulli mean has shrunk. If the strategy saw the components of ℓ_i (even in the semi-bandit setting, with $\mathcal{A} = \{0, 1\}^d$ and feedback $(\ell_{t,1} a_{t,1}, \ldots, \ell_{t,d} a_{t,d})$), it would not suffer this disadvantage.

Lower bounds: proof

Using the same argument as we saw for the stochastic multi-armed bandit case (with a little extra work to show that θ is unlikely to be too close to 0 or 1, so that the variance of the Bernoulli is not too small), we see that

$$\mathbb{E}\overline{R}_n^i(b_i) \geq \frac{2\epsilon n}{d}\left(\frac{1}{2} - c\frac{\epsilon\sqrt{n}}{d}\right).$$

Choosing $\epsilon = d/(4c\sqrt{n})$ gives $\mathbb{E}\overline{R}_n^i(b_i) = \Omega(\sqrt{n})$, and so $\mathbb{E}\overline{R}_n(P_b) = \Omega(d\sqrt{n})$.

[NB: $\mathcal{A} = [-1, 1]^d$ $\mathcal{L} = \{\pm e_i\}$ has lower regret, because the strategy can use a_t to identify which direction $\pm e_i$ was played.]

[Open problem: when is $\Theta(d\sqrt{n})$ possible with an efficient strategy?]

Outline

- Linear bandits.
  - Exponential weights with unbiased loss estimates.
  - Controlling loss estimates and their variance.
    - Barycentric spanner.
    - Uniform distribution.
    - John's distribution.
  - Lower bounds.
  - Stochastic mirror descent.
    - Full information.
    - Bandit information.

Full information online prediction games

- Repeated game:

$$\text{Strategy plays } a_t \in \mathcal{A}$$
$$\text{Adversary reveals } \ell_t \in \mathcal{L}$$

- Aim to minimize regret:

$$R_n = \sum_{t=1}^{n} \ell_t(a_t) - \min_{a \in \mathcal{A}} \sum_{t=1}^{n} \ell_t(a).$$

Online Convex Optimization

- Choosing a_t to minimize past losses can fail.
- The strategy must avoid overfitting.
- First approach: gradient steps.
  Stay close to previous decisions, but move in a direction of improvement.

Online Convex Optimization

1. Gradient algorithm.
2. Regularized minimization
   - Bregman divergence
   - Regularized minimization $\Leftrightarrow$ minimizing latest loss and divergence from previous decision
   - Constrained minimization equivalent to unconstrained plus Bregman projection
   - Linearization
   - Mirror descent
3. Regret bound

Online Convex Optimization: Gradient Method

$$a_1 \in \mathcal{A},$$
$$a_{t+1} = \Pi_{\mathcal{A}} \left(a_t - \eta \nabla \ell_t(a_t) \right),$$

where $\Pi_{\mathcal{A}}$ is the Euclidean projection on $\mathcal{A}$,

$$\Pi_{\mathcal{A}}(x) = \arg\min_{a \in \mathcal{A}} \|x - a\|.$$

Theorem: For $G = \max_t \|\nabla \ell_t(a_t)\|$ and $D = \mathrm{diam}(\mathcal{A})$, the gradient strategy with $\eta = D/(G\sqrt{n})$ has regret satisfying

$$R_n \leq GD\sqrt{n}.$$

Online Convex Optimization: Gradient Method

Example: (2-ball, 2-ball)
$\mathcal{A} = \{a \in \mathbb{R}^d : \|a\| \leq 1\}$, $\mathcal{L} = \{a \mapsto v \cdot a : \|v\| \leq 1\}$. $D = 2$, $G \leq 1$.
Regret is no more than $2\sqrt{n}$.

(And $O(\sqrt{n})$ is optimal.)

Example: (1-ball, ∞-ball)
$\mathcal{A} = \Delta(k)$, $\mathcal{L} = \{a \mapsto v \cdot a : \|v\|_\infty \leq 1\}$.
$D = 2$, $G \leq \sqrt{k}$.
Regret is no more than $2\sqrt{kn}$.

Since competing with the whole simplex is equivalent to competing with the vertices (experts) for linear losses, this is worse than exponential weights ($\sqrt{k}$ versus $\log k$).

Gradient Method: Proof

Define
$$\tilde{a}_{t+1} = a_t - \eta \nabla \ell_t(a_t),$$
$$a_{t+1} = \Pi_{\mathcal{A}}(\tilde{a}_{t+1}).$$

Fix $a \in \mathcal{A}$ and consider the measure of progress $\|a_t - a\|$.

$$\begin{aligned} \|a_{t+1} - a\|^2 &\leq \|\tilde{a}_{t+1} - a\|^2 \\ &= \|a_t - a\|^2 + \eta^2 \|\nabla \ell_t(a_t)\|^2 - 2\eta \nabla_t(a_t) \cdot (a_t - a). \end{aligned}$$

By convexity,

$$\begin{aligned} \sum_{t=1}^{n} (\ell_t(a_t) - \ell_t(a)) &\leq \sum_{t=1}^{n} \nabla \ell_t(a_t) \cdot (a_t - a) \\ &\leq \frac{\|a_1 - a\|^2 - \|a_{n+1} - a\|^2}{2\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|\nabla \ell_t(a_t)\|^2 \end{aligned}$$

Online Convex Optimization

1. Gradient algorithm.
2. Regularized minimization
   - Bregman divergence
   - Regularized minimization $\Leftrightarrow$ minimizing latest loss and divergence from previous decision
   - Constrained minimization equivalent to unconstrained plus Bregman projection
   - Linearization
   - Mirror descent
3. Regret bound

Online Convex Optimization: A Regularization Viewpoint

- Suppose ℓ_t is linear: $\ell_t(a) = g_t \cdot a$.
- Suppose $\mathcal{A} = \mathbb{R}^d$.
- Then minimizing the regularized criterion

$$a_{t+1} = \arg\min_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \ell_s(a) + \frac{1}{2} \|a\|^2 \right)$$

corresponds to the gradient step

$$a_{t+1} = a_t - \eta \nabla \ell_t(a_t).$$

Online Convex Optimization: Regularization

Regularized minimization

Consider the family of strategies of the form:

$$a_{t+1} = \arg\min_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \ell_s(a) + R(a) \right).$$

The regularizer $R : \mathbb{R}^d \to \mathbb{R}$ is strictly convex and differentiable.

- R keeps the sequence of a_ts stable: it diminishes ℓ_t's influence.
- We can view the choice of a_{t+1} as trading off two competing forces: making $\ell_t(a_{t+1})$ small, and keeping a_{t+1} close to a_t.
- This is a perspective that motivated many algorithms in the literature.

Properties of Regularization Methods

In the unconstrained case ($\mathcal{A} = \mathbb{R}^d$), regularized minimization is equivalent to minimizing the latest loss and the distance to the previous decision. The appropriate notion of distance is the Bregman divergence $D_{\Phi_{t-1}}$:

Define

$$\Phi_0 = R,$$
$$\Phi_t = \Phi_{t-1} + \eta \ell_t,$$

so that

$$\begin{aligned} a_{t+1} &= \arg\min_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \ell_s(a) + R(a) \right) \\ &= \arg\min_{a \in \mathcal{A}} \Phi_t(a). \end{aligned}$$

Bregman Divergence

Definition: For a strictly convex, differentiable $\Phi : \mathbb{R}^d \to \mathbb{R}$, the Bregman divergence wrt Φ is defined, for $a, b \in \mathbb{R}^d$, as

$$D_\Phi(a, b) = \Phi(a) - (\Phi(b) + \nabla\Phi(b) \cdot (a - b)) \,.$$

$D_\Phi(a, b)$ is the difference between $\Phi(a)$ and the value at a of the linear approximation of Φ about b. (PICTURE)

Bregman Divergence

Example: For $a \in \mathbb{R}^d$, the squared euclidean norm, $\Phi(a) = \frac{1}{2}\|a\|^2$, has

$$D_\Phi(a,b) = \frac{1}{2}\|a\|^2 - \left(\frac{1}{2}\|b\|^2 + b \cdot (a-b)\right)$$
$$= \frac{1}{2}\|a-b\|^2,$$

the squared euclidean norm.

Bregman Divergence

Example: For $a \in [0, \infty)^d$, the unnormalized negative entropy, $\Phi(a) = \sum_{i=1}^{d} a_i (\ln a_i - 1)$, has

$$D_\Phi(a, b) = \sum_i (a_i(\ln a_i - 1) - b_i(\ln b_i - 1) - \ln b_i(a_i - b_i))$$

$$= \sum_i \left(a_i \ln \frac{a_i}{b_i} + b_i - a_i \right),$$

the unnormalized KL divergence.

Thus, for $a \in \Delta^d$, $\Phi(a) = \sum_i a_i \ln a_i$ has

$$D_\Phi(a, b) = \sum_i a_i \ln \frac{a_i}{b_i}.$$

Bregman Divergence

When the domain of Φ is $\mathcal{S} \subset \mathbb{R}^d$, in addition to differentiability and strict convexity, we make some more assumptions:

- $\mathcal{S}$ is closed, and its interior is convex.
- For a sequence approaching the boundary of $\mathcal{S}$, $\|\nabla\Phi(a_n)\| \to \infty$.

We say that such a Φ is a *Legendre function*.

Bregman Divergence Properties

1. $D_\Phi \geq 0$, $D_\Phi(a,a) = 0$.
2. $D_{A+B} = D_A + D_B$.
3. For ℓ linear, $D_{\Phi+\ell} = D_\Phi$.
4. *Bregman projection*, $\Pi^\Phi_{\mathcal{A}}(b) = \arg\min_{a\in\mathcal{A}} D_\Phi(a,b)$ is uniquely defined for closed, convex $\mathcal{A} \subset \mathcal{S}$ (that intersects the interior of $\mathcal{S}$).
5. *Generalized Pythagorus*: for closed, convex $\mathcal{A}$, $a^* = \Pi^\Phi_{\mathcal{A}}(b)$, $a \in \mathcal{A}$, $D_\Phi(a,b) \geq D_\Phi(a,a^*) + D_\Phi(a^*,b)$.
6. $\nabla_a D_\Phi(a,b) = \nabla\Phi(a) - \nabla\Phi(b)$.
7. For Φ^* the Legendre dual of Φ,

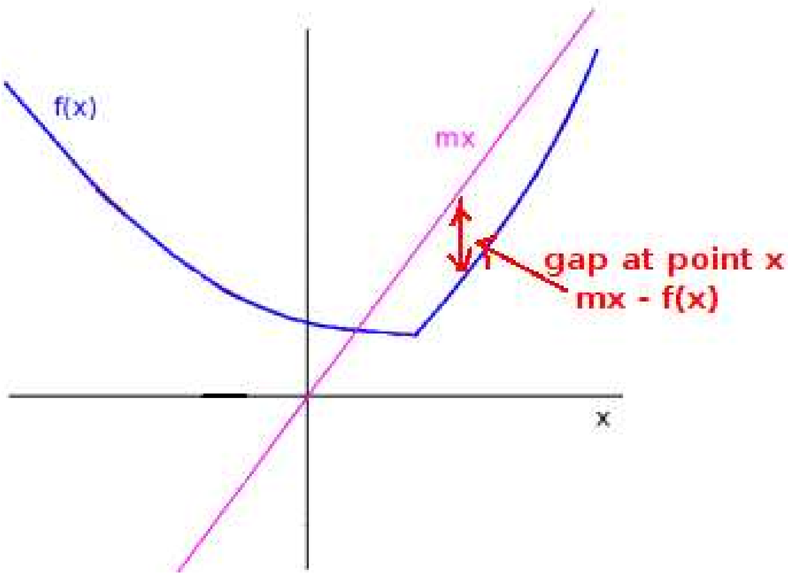
$$\nabla\Phi^* = (\nabla\Phi)^{-1},$$

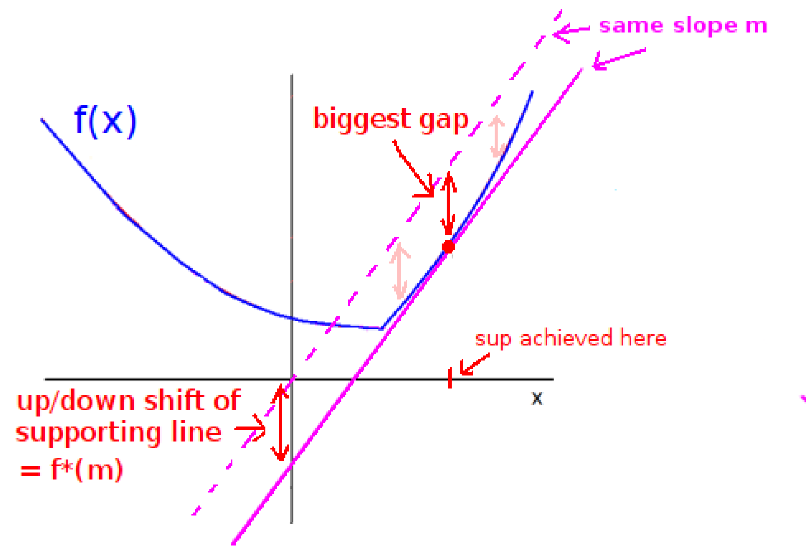
$$D_\Phi(a,b) = D_{\Phi^*}(\nabla\Phi(b), \nabla\Phi(a)).$$

Legendre Dual

Here, for a Legendre function $\Phi : \mathcal{S} \to \mathbb{R}$, we define the Legendre dual as

$$\Phi^*(u) = \sup_{v \in \mathcal{S}} \left(u \cdot v - \Phi(v) \right).$$





(http://maze5.net/)

Legendre Dual

Properties:

- Φ^* is Legendre.
- $\mathrm{dom}(\Phi^*) = \nabla\Phi(\text{int dom } \Phi)$.
- $\nabla\Phi^* = (\nabla\Phi)^{-1}$.
- $D_\Phi(a, b) = D_{\Phi^*}(\nabla\Phi(b), \nabla\Phi(a))$.
- $\Phi^{**} = \Phi$.

Examples:

- $\Phi = \frac{1}{2}\|\cdot\|_p^2$: $\Phi^* = \frac{1}{2}\|\cdot\|_q^2$, where $1/p + 1/q = 1$.
- $\Phi(a) = \sum_{i=1}^d e^{a_i}$: $\Phi^*(u) = \sum_i u_i(\ln u_i - 1)$.

Online Convex Optimization

1. Problem formulation
2. Empirical minimization fails.
3. Gradient algorithm.
4. Regularized minimization
   - Bregman divergence
   - Regularized minimization $\Leftrightarrow$ minimizing latest loss plus divergence from previous decision
   - Constrained minimization equivalent to unconstrained plus Bregman projection
   - Linearization
   - Mirror descent
5. Regret bounds

Properties of Regularization Methods

In the unconstrained case ($\mathcal{A} = \mathbb{R}^d$), regularized minimization is equivalent to minimizing the latest loss and the distance (Bregman divergence) to the previous decision.

Theorem: Define $\tilde{a}_1$ via $\nabla R(\tilde{a}_1) = 0$, and set

$$\tilde{a}_{t+1} = \arg\min_{a \in \mathbb{R}^d} \left(\eta \ell_t(a) + D_{\Phi_{t-1}}(a, \tilde{a}_t) \right).$$

Then

$$\tilde{a}_{t+1} = \arg\min_{a \in \mathbb{R}^d} \left(\eta \sum_{s=1}^{t} \ell_s(a) + R(a) \right).$$

Properties of Regularization Methods

Proof. By the definition of Φ_t,

$$\eta \ell_t(a) + D_{\Phi_{t-1}}(a, \tilde{a}_t) = \Phi_t(a) - \Phi_{t-1}(a) + D_{\Phi_{t-1}}(a, \tilde{a}_t).$$

The derivative wrt a is

$$\begin{aligned} &\nabla \Phi_t(a) - \nabla \Phi_{t-1}(a) + \nabla_a D_{\Phi_{t-1}}(a, \tilde{a}_t) \\ &= \nabla \Phi_t(a) - \nabla \Phi_{t-1}(a) + \nabla \Phi_{t-1}(a) - \nabla \Phi_{t-1}(\tilde{a}_t) \end{aligned}$$

Setting to zero shows that

$$\nabla \Phi_t(\tilde{a}_{t+1}) = \nabla \Phi_{t-1}(\tilde{a}_t) = \cdots = \nabla \Phi_0(\tilde{a}_1) = \nabla R(\tilde{a}_1) = 0,$$

So $\tilde{a}_{t+1}$ minimizes Φ_t. $\square$

Properties of Regularization Methods

Constrained minimization is equivalent to unconstrained minimization, followed by Bregman projection:

Theorem: For

$$a_{t+1} = \arg\min_{a \in \mathcal{A}} \Phi_t(a),$$
$$\tilde{a}_{t+1} = \arg\min_{a \in \mathbb{R}^d} \Phi_t(a),$$

we have

$$a_{t+1} = \Pi_{\mathcal{A}}^{\Phi_t}(\tilde{a}_{t+1}).$$

Properties of Regularization Methods

Proof. Let a'_{t+1} denote $\Pi_{\mathcal{A}}^{\Phi_t}(\tilde{a}_{t+1})$. First, by definition of a_{t+1},

$$\Phi_t(a_{t+1}) \leq \Phi_t(a'_{t+1}).$$

Conversely,

$$D_{\Phi_t}(a'_{t+1}, \tilde{a}_{t+1}) \leq D_{\Phi_t}(a_{t+1}, \tilde{a}_{t+1}).$$

But $\nabla\Phi_t(\tilde{a}_{t+1}) = 0$, so

$$D_{\Phi_t}(a, \tilde{a}_{t+1}) = \Phi_t(a) - \Phi_t(\tilde{a}_{t+1}).$$

Thus, $\Phi_t(a'_{t+1}) \leq \Phi_t(a_{t+1})$. $\square$

Properties of Regularization Methods

Example: For linear ℓ_t, regularized minimization is equivalent to minimizing the last loss plus the Bregman divergence wrt R to the previous decision:

$$
\begin{aligned}
&\arg\min_{a\in\mathcal{A}}\left(\eta\sum_{s=1}^{t}\ell_s(a)+R(a)\right)\\
&=\Pi_{\mathcal{A}}^{R}\left(\arg\min_{a\in\mathbb{R}^d}\left(\eta\ell_t(a)+D_R(a,\tilde{a}_t)\right)\right),
\end{aligned}
$$

because adding a linear function to Φ does not change D_Φ.

Linear Loss

We can replace ℓ_t by $\nabla \ell_t(a_t)$, and this leads to an upper bound on regret.

Thus, for convex losses, we can work with linear ℓ_t.

Regularization Methods: Mirror Descent

Regularized minimization for linear losses can be viewed as mirror descent—taking a gradient step in a dual space:

Theorem: The decisions

$$\tilde{a}_{t+1} = \arg\min_{a \in \mathbb{R}^d} \left(\eta \sum_{s=1}^{t} g_s \cdot a + R(a) \right)$$

can be written

$$\tilde{a}_{t+1} = (\nabla R)^{-1} \left(\nabla R(\tilde{a}_t) - \eta g_t \right).$$

This corresponds to first mapping from $\tilde{a}_t$ through ∇R, then taking a step in the direction $-g_t$, then mapping back through $(\nabla R)^{-1} = \nabla R^*$ to $\tilde{a}_{t+1}$.

Regularization Methods: Mirror Descent

Proof. For the unconstrained minimization, we have

$$\nabla R(\tilde{a}_{t+1}) = -\eta \sum_{s=1}^{t} g_s,$$

$$\nabla R(\tilde{a}_t) = -\eta \sum_{s=1}^{t-1} g_s,$$

so $\nabla R(\tilde{a}_{t+1}) = \nabla R(\tilde{a}_t) - \eta g_t$, which can be written

$$\tilde{a}_{t+1} = \nabla R^{-1} \left(\nabla R(\tilde{a}_t) - \eta g_t \right).$$

$\square$

Mirror Descent

Given:

compact, convex $\mathcal{A} \subseteq \mathbb{R}^d$, closed, convex $\mathcal{S} \supset \mathcal{A}$, $\eta > 0$, $\mathcal{S} \supset \mathcal{A}$,

Legendre $R : \mathcal{S} \to \mathbb{R}$. Set $a_1 \in \arg\min_{a \in \mathcal{A}} R(a)$.

For round t:

1. Play a_t; observe $\ell_t \in \mathbb{R}^d$.
2. $w_{t+1} = \nabla R^* \left(\nabla R(a_t) - \eta \nabla \ell_t(a_t)\right)$.
3. $a_{t+1} = \arg\min_{a \in \mathcal{A}} D_R(a, w_{t+1})$.
   [Always convex optimization.]

Exponential weights as mirror descent

For $\mathcal{A} = \Delta(k)$ and $R(a) = \sum_{i=1}^{k} (a_i \log a_i - a_i)$, this reduces to exponential weights:

$$
\begin{aligned}
\nabla R(u)_i &= \log a_i, \\
R^*(u) &= \sum_i e^{u_i}, \\
\nabla R^*(u)_i &= \exp(u_i), \\
\nabla R(w_{t+1})_i &= \log(w_{t+1,i}) = \log a_{t,i} - \eta \nabla \ell_t(a_t)_i, \\
w_{t+1,i} &= a_{t,i} \exp\left(-\eta \nabla \ell_t(a_t)_i\right), \\
D_R(a,b) &= \sum_i \left(a_i \log \frac{a_i}{b_i} + b_i - a_i \right), \\
a_{t+1,i} &\propto w_{t+1,i}.
\end{aligned}
$$

Mirror descent regret

Theorem: Suppose that, for all $a \in \mathcal{A} \cap \mathrm{int}(\mathcal{S})$, $\ell \in \mathcal{L}$, $\nabla R(a) - \eta \nabla \ell(a) \in \nabla R(\mathrm{int}(\mathcal{S}))$. For any $a \in \mathcal{A}$,

$$\sum_{t=1}^{n} (\ell_t(a_t) - \ell_t(a))$$
$$\leq \frac{1}{\eta} \left(R(a) - R(a_1) + \sum_{t=1}^{n} D_{R^*} \Big(\nabla R(a_t) - \eta \nabla \ell_t(a_t), \nabla R(a_t) \Big) \right).$$

Proof: Fix $a \in \mathcal{A}$. Since the ℓ_t are convex,

$$\sum_{t=1}^{n} (\ell_t(a_t) - \ell_t(a)) \leq \sum_{t=1}^{n} \nabla \ell_t(a_t)^T (a_t - a).$$

Mirror descent regret: proof

The choice of w_{t+1} and the fact that $\nabla R^{-1} = \nabla R^*$ show that

$$\nabla R(w_{t+1}) = \nabla R(a_t) - \eta \nabla \ell_t(a_t).$$

Hence,

$$\begin{aligned} \eta \nabla \ell_t(a_t)^T (a_t - a) &= (a - a_t)^T \left(\nabla R(w_{t+1}) - \nabla R(a_t) \right) \\ &= D_R(a, a_t) + D_R(a_t, w_{t+1}) - D_R(a, w_{t+1}). \end{aligned}$$

Generalized Pythagorus' inequality shows that the projection a_{t+1} satisfies

$$D_R(a, w_{t+1}) \geq D_R(a, a_{t+1}) + D_R(a_{t+1}, w_{t+1}).$$

Mirror descent regret: proof

$$\eta \sum_{t=1}^{n} \nabla \ell_t(a_t)^T (a_t - a)$$

$$\leq \sum_{t=1}^{n} \Big(D_R(a, a_t) + D_R(a_t, w_{t+1}) - D_R(a, w_{t+1})$$

$$- D_R(a, a_{t+1}) - D_R(a_{t+1}, w_{t+1}) \Big)$$

$$= D_R(a, a_1) - D_R(a, a_{n+1}) + \sum_{t=1}^{n} \left(D_R(a_t, w_{t+1}) - D_R(a_{t+1}, w_{t+1}) \right)$$

$$\leq D_R(a, a_1) + \sum_{t=1}^{n} D_R(a_t, w_{t+1}).$$

Mirror descent regret: proof

$$= D_R(a, a_1) + \sum_{t=1}^{n} D_{R^*}(\nabla R(w_{t+1}), \nabla R(a_t))$$

$$= D_R(a, a_1) + \sum_{t=1}^{n} D_{R^*}(\nabla R(a_t) - \eta \nabla \ell_t(a_t), \nabla R(a_t))$$

$$= R(a) - R(a_1) + \sum_{t=1}^{n} D_{R^*}(\nabla R(a_t) - \eta \nabla \ell_t(a_t), \nabla R(a_t)).$$

Linear bandit setting

- See only $\ell_t(a_t)$; $\nabla \ell_t(a_t)$ is unseen.
- Instead of a_t, strategy plays a noisy version, x_t.
- Strategy uses $\ell_t(x_t)$ to give an unbiased estimate of $\nabla \ell_t(a_t)$.

Stochastic mirror descent

Given:

compact, convex $\mathcal{A} \subseteq \mathbb{R}^d$, $\eta > 0$, $\mathcal{S} \supset \mathcal{A}$, Legendre $R : \mathcal{S} \to \mathbb{R}$.

Set $a_1 \in \arg\min_{a \in \mathcal{A}} R(a)$.

For round t:

1. Play noisy version x_t of a_t; observe $\ell_t(x_t)$.
2. Compute estimate $\tilde{g}_t$ of $\nabla \ell_t(a_t)$.
3. $w_{t+1} = \nabla R^* \left(\nabla R(a_t) - \eta \tilde{g}_t \right)$.
4. $a_{t+1} = \arg\min_{a \in \mathcal{A}} D_R(a, w_{t+1})$.

Regret of stochastic mirror descent

Theorem: Suppose that, for all $a \in \mathcal{A} \cap \mathrm{int}(\mathcal{S})$ and linear $\ell \in \mathcal{L}$, $\mathbb{E}[\tilde{g}_t | a_t] = \nabla \ell_t(a_t)$ and $\nabla R(a) - \eta \tilde{g}_t(a) \in \nabla R(\mathrm{int}(\mathcal{S}))$.
For any $a \in \mathcal{A}$,

$$\sum_{t=1}^{n} (\ell_t(a_t) - \ell_t(a))$$
$$\leq \frac{1}{\eta}\left(R(a) - R(a_1) + \sum_{t=1}^{n} \mathbb{E} D_{R^*}\left(\nabla R(a_t) - \eta \tilde{g}_t, \nabla R(a_t) \right) \right)$$
$$+ \sum_{t=1}^{n} \mathbb{E}\left[\| a_t - \mathbb{E}[x_t | a_t] \| \; \| \tilde{g}_t \|_* \right].$$

Regret: proof

$$
\begin{aligned}
&\mathbb{E}\sum_{t=1}^{n}\left(\ell_t(x_t)-\ell_t(a)\right)\\
&=\mathbb{E}\sum_{t=1}^{n}\left(\ell_t(x_t)-\ell_t(a_t)+\ell_t(a_t)-\ell_t(a)\right)\\
&=\mathbb{E}\sum_{t=1}^{n}\Big(\mathbb{E}\left[\,\ell_t^T(x_t-a_t)\middle|\,a_t\right]+\ell_t(a_t)-\ell_t(a)\Big)\\
&\le\mathbb{E}\sum_{t=1}^{n}\|a_t-\mathbb{E}[x_t|a_t]\|\;\;\|\tilde{g}_t\|_*+\mathbb{E}\sum_{t=1}^{n}\nabla\ell_t(a_t)^T(a_t-a)\\
&=\mathbb{E}\sum_{t=1}^{n}\|a_t-\mathbb{E}[x_t|a_t]\|\;\;\|\tilde{g}_t\|_*+\mathbb{E}\sum_{t=1}^{n}\tilde{g}_t^T(a_t-a).
\end{aligned}
$$

Regret: proof

Applying the regret bound for the (random) linear losses $a \mapsto \tilde{g}_t^T a$ gives

$$\leq \mathbb{E} \sum_{t=1}^{n} \|a_t - \mathbb{E}[x_t|a_t]\| \; \|\tilde{g}_t\|_*$$

$$+ \frac{1}{\eta}\left(R(a) - R(a_1) + \sum_{t=1}^{n} \mathbb{E} D_{R^*}\left(\nabla R(a_t) - \eta \tilde{g}_t, \nabla R(a_t)\right)\right).$$

Regret: Euclidean ball

Consider $B = \{a \in \mathbb{R}^d : \|a\| \leq 1\}$ (with the Euclidean norm).

Ingredients:

1. Distribution of x_t, given a_t:

$$x_t = \xi_t \frac{a_t}{\|a_t\|} + (1 - \xi_t)\epsilon_t e_{I_t},$$

   where ξ_t is Bernoulli($\|a_t\|$), ϵ_t is uniform ± 1, and I_t is uniform on $\{1, \ldots, d\}$, so $\mathbb{E}[x_t | a_t] = a_t$.

2. Estimate $\tilde{\ell}_t$ of loss ℓ_t:

$$\tilde{\ell}_t = d \frac{1 - \xi_t}{1 - \|a_t\|} x_t^T \ell_t x_t,$$

   so $\mathbb{E}[\tilde{\ell}_t | a_t] = \ell_t$.

Regret: Euclidean ball

Theorem: Consider stochastic mirror descent on $\mathcal{A} = (1-\gamma)B$, with these choices and $R(a) = -\log(1-\|a\|) - \|a\|$. Then for $\eta d \leq 1/2$,

$$\overline{R}_n \leq \gamma n + \frac{\log(1/\gamma)}{\eta} + \eta \sum_{t=1}^{n} \mathbb{E}\left[(1-\|a_t\|)\|\tilde{\ell}_t\|^2\right].$$

For $\gamma = 1/\sqrt{n}$ and $\eta = \sqrt{\log n/(2nd)}$,

$$\overline{R}_n \leq 3\sqrt{dn\log n}.$$

Proof: $\nabla R(a) = a/(1-\|a\|)$.

Linear bandits

Open question:

What geometric properties of $\mathcal{A}$ and $\mathcal{L}$ determine the regret?

$\mathcal{A}$	$\mathcal{L}$	$\overline{R}_n$
convex	$\ell : \mathcal{A} \to [-1, 1]$	$\tilde{O}(d\sqrt{n})$
$\\|\cdot\\|_2 \leq 1$	$\\|\cdot\\|_2 \leq 1$	$\tilde{O}(\sqrt{dn})$
Δ^{d-1}	$\\|\cdot\\|_\infty \leq 1$	$\tilde{O}(\sqrt{dn})$
$\\|\cdot\\|_\infty \leq 1$	$\{\pm e_i : 1 \leq i \leq d\}$	$\Omega(d\sqrt{n})$