Bregman projection:

OMD: $w_{t+1} = \underset{w\in\Omega}{\operatorname{argmin}} \langle \eta g_t, w\rangle + D_\psi(w, w_t)$.

can be rewritten as: $\nabla\psi(w_{t+1}') = \nabla\psi(w_t) - \eta g_t$.

$$w_{t+1}' = \underset{w\in\mathbb{R}^d}{\operatorname{argmin}} \langle \eta g_t, w\rangle + D_\psi(w, w_t)$$

$$w_{t+1} = \underset{w\in\Omega}{\operatorname{argmin}} D_\psi(w, w_{t+1}')$$

How to calculate $v = \underset{w\in\Omega}{\operatorname{argmin}} D_\psi(w, u)$?

① $\Omega = \{w : \|w\|_2 \le B\}$. $\psi(w) = \frac{1}{2}\|w\|_2^2$

$$v = \underset{w:\|w\|_2\le B}{\operatorname{argmin}} \|w-u\| = \begin{cases} u, & \|u\|_2 \le B \\ \frac{u}{\|u\|}\cdot B, & \|u\|_2 > B \end{cases}$$ (HW3. problem 3)

② $\Omega = \Delta^{d-1}$, $\psi(w) = \sum_i w_i \ln w_i$

$$v = \underset{w\in\Delta^{d-1}}{\operatorname{argmin}} \sum_i w_i \ln\frac{w_i}{u_i} = \frac{w}{\|w\|_1}$$ (exercise).

Goal: find w that approx minimizes

$$\frac{1}{m}\sum_{i=1}^m f(y_i\langle w, \phi(x_i)\rangle) + \frac{\lambda}{2}\|w\|_2^2.$$

$\in \mathbb{R}^N$

key idea:

keep track of coefficient of w_t. $\alpha_t \in \mathbb{R}^m$

maintain invariant that $w_t = \sum_{i=1}^m \alpha_t(i)\phi(x_i)$

Fact: $f: \mathbb{R} \to \mathbb{R}$ cvx.

$h(w) = f(\langle a, w\rangle + b)$

$z \in \partial f(\langle a, w\rangle + b)$

$z \cdot a \in \partial h(w)$

For $t = 1, 2, \cdots T$.

$\Omega = \mathbb{R}^d$.

sample $i_t \sim$ uniform $(\{1 \cdots m\})$.

$f_t(w) = \underbrace{f(y_{i_t}\langle w, \phi(x_{i_t})\rangle)}_{\ell_t(w)} + \frac{\lambda}{2}\|w\|^2$

calculating g_t:

— $v_t \in \partial \ell_t(w_t)$ ① find $z_t \in \partial f(y_{i_t}\langle w_t, \phi(x_{i_t})\rangle)$

— $g_t = v_t + \lambda w_t$ ② $v_t = z_t \cdot y_{i_t} \phi(x_{i_t})$

evaluated using kernel trick.

updating w_t:

$w_{t+1} \leftarrow w_t - \frac{1}{\lambda t}(\lambda w_t + v_t)$

$= (1 - \frac{1}{t}) w_t - \frac{1}{\lambda t} v_t$.

Q: can we modify the alg so that instead of keep w_t's, we keep α_t's?

Rewriting the recurrence of w_t:

$w_{t+1} = (1 - \frac{1}{t}) w_t - \frac{1}{\lambda t} \cdot z_t \cdot y_{i_t} \phi(x_{i_t})$.

$$\sum_{i=1}^m \alpha_{t+1}(i) \phi(x_i) = (1 - \frac{1}{t}) \sum_{i=1}^m \alpha_t(i) \phi(x_i) - \frac{1}{\lambda t} \cdot z_t \cdot y_{i_t} \phi(x_{i_t})$$

$$\alpha_{t+1}(i) = \begin{cases} (1 - \frac{1}{t}) \alpha_t(i) - \frac{1}{\lambda t} z_t y_{i_t}, & i = i_t \\ (1 - \frac{1}{t}) \alpha_t(i) & i \neq i_t \end{cases}$$

$$\Rightarrow \quad \alpha_{t+1} = -\frac{1}{\lambda t}\sum_{s=1}^{t} z_s\, y_{i_s}\, e_{i_s}, \qquad e_{i_s} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^m \quad \leftarrow i_s$$

In summary, developed a regularized loss minimization alg on the feature space w/ time complexity indept of feature dimension.

§ OCO for exp-concave fns.

Def: f is α-exp-concave if

$\exp(-\alpha f(x))$ is concave.

Lemma: If f is twice differentiable, then f α-exp-concave $\Leftrightarrow$ $\forall x$,

$$\nabla^2 f(x) \succeq \alpha \cdot \nabla f(x) \cdot \nabla f(x)^T \succeq 0.$$

$A \succeq B \Leftrightarrow$

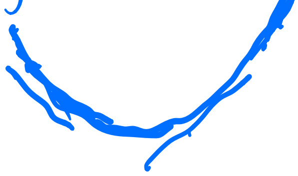
$A - B \succeq 0 \Leftrightarrow$

$A - B$ is p.s.d.

Fact: f is twice differentiable

f cvx $\Leftrightarrow \nabla^2 f(x) \succeq 0.$

f concave $\Leftrightarrow \qquad \preceq 0.$



(which also implies f is cvx).

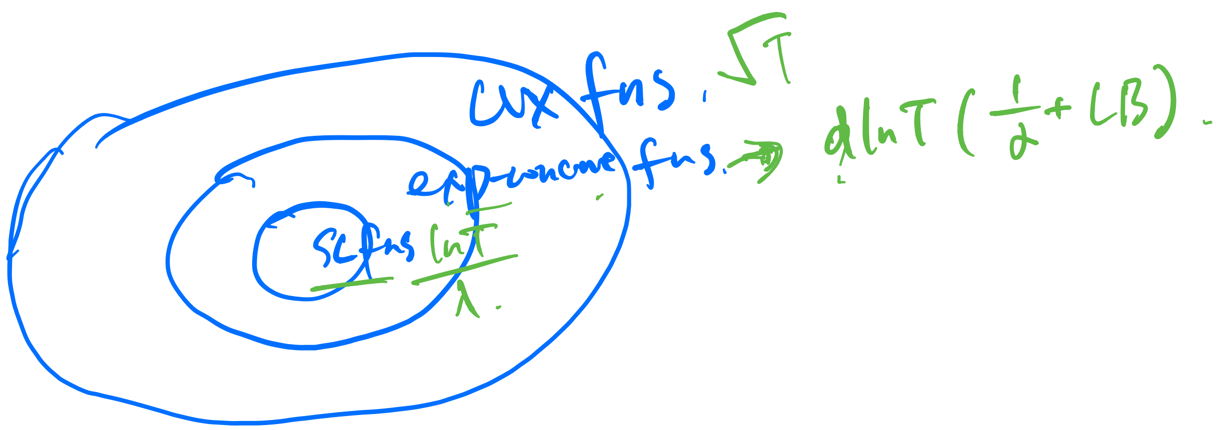
f λ-sc $\|\cdot\|_2$ $\Rightarrow$ f α-exp-concave ?✓

w/ sufficiently small α.

$\|\nabla f(x)\| \le G.$

$\Downarrow$

$\forall x,\ \nabla^2 f(x) \succeq \lambda \cdot I.$



Pf of Lemma,

f α-exp-concave

$\Leftrightarrow$ $g(w) = \exp(-\alpha f(w))$ is concave

$\Leftrightarrow$ $\forall w,\ \nabla^2 g(w) \preceq 0.$

$\nabla g(w) = (-\alpha)\cdot \exp(-\alpha f(w)) \nabla f(w)$

$\nabla^2 g(w) = (-\alpha)^2 \exp(-\alpha f(w))\, \nabla f(w) \cdot \nabla f(w)^T$

$+ (-\alpha) \exp(-\alpha f(w))$

$\nabla^2 f(w)$.

$$= \exp(-\alpha f(w)) \cdot \left(\alpha^2 \nabla f(w) \nabla f(w)^T - \alpha \nabla^2 f(w)\right)$$

$\boxtimes$.

2 Examples:

$$r_t(i) = \frac{\text{unit price of } i\text{th asset at the end of } t}{\text{------ } t-1.}$$

① . portfolio selection

$\exp(-f_t(w)) = \langle r_t, w \rangle$

$$\Omega = \Delta^{d-1}. \quad f_t(w) = -\ln(\langle r_t, w \rangle)$$

1-exp-concave.

$$R_T(w^*) = \underbrace{\sum_{t=1}^T f_t(w_t)}_{\text{negative log wealth of investor}} - \underbrace{\sum_{t=1}^T f_t(w^*)}_{\text{constant rebalanced portfolio (CRP)}}.$$

why is CRP an interesting benchmark?

$d=2$.

t	$r_t(1)$	$r_t(2)$
1	$1 \times \frac{1}{2}$	$\frac{1}{2} \times \frac{1}{2}$.
2	$1 \times \frac{1}{2}$.	$2 \times \frac{1}{2}$.
3	1	$\frac{1}{2}$

3

4 | 1 2

$w^* = (\frac{1}{2}, \frac{1}{2})$. initial wealth $= 1$ $\frac{9}{8}$.

$$\text{Wealth}_T = \prod_{t=1}^{T} \langle r_t, w^* \rangle \rightsquigarrow \begin{cases} \frac{3}{4} & t=\text{odd} \\ \frac{3}{2} & t=\text{even} \end{cases}$$

T even

$$= \left(\frac{9}{8}\right)^{\frac{T}{2}}.$$

2 $\Omega = \{ w \in \mathbb{R}^d : \|w\|_2 \le B \}$.

$$f_t(w) = \frac{1}{2}(\langle w, x_t \rangle - y_t)^2.$$

$$\|x_t\| \le R. \quad |y_t| \in Y.$$

Lemma $\Rightarrow$ $f_t(w)$ is $\frac{1}{(RB+Y)^2}$ - exp-concave.

Algorithms for online exp-concave optimization

Online newton step :

Ω. $\max_{u,v\in\Omega} \|u-v\|_2 \le B$.

$\{f_t\}_{t=1}^{T}$ are all α-exp-concave & L-Lip.

λ: parameter for the algorithm.

- initialize w_1 to be any pt in Ω:
- For $t = 1, 2 \ldots T$:
  - show w_t.
  - receive f_t.
  - update w_t:

$$w_{t+1} = \operatorname*{argmin}_{w\in\Omega} \langle w, g_t \rangle + D_{\varphi_t}(w, w_t), \quad g_t \in \partial f_t(w_t).$$

$$\varphi_t(w) = \frac{1}{2}\|w\|_{A_t}^2, \quad A_t = \lambda I + \tilde{\alpha} \sum_{s=1}^{t} g_s g_s^T$$

(Alternatively:

$$w'_{t+1} = w_t - A_t^{-1} \cdot g_t$$

$$w_{t+1} = \operatorname*{argmin}_{w\in\Omega} \|w - w'_{t+1}\|_{A_t}. \quad)$$

Analysis of online newton step:

Thm: $\lambda = \frac{1}{\tilde{\alpha} B^2}$, $\tilde{\alpha} = \min\left(\frac{1}{8 \cdot B \cdot L}, \frac{\alpha}{2}\right)$,

online newton step gives regret

$$R_T(\Omega) \leq O\left(\frac{1}{\tilde{\alpha}} d \ln T\right)$$

$$= O\left(\left(\frac{1}{\alpha} + L \cdot B\right) \cdot d \ln T\right).$$

Key Lemma: If f is α-exp-concave & L-Lip. then, for any $u, v \in \Omega$,

f_t w_t

$$f(u) \geq f(v) + \langle \nabla f(v), u-v \rangle + \frac{\tilde{\alpha}}{2} \cdot \left(u-v\right)^T \nabla f(v) \nabla f(v)^T (u-v).$$

λI

Pf of thm:

similar to "linearization" in OMD pf:

$$f_t(w_t) - f_t(u)$$

$$\leq \langle \nabla f_t(w_t), w_t - u \rangle - \frac{\tilde{\alpha}}{2}(w_t - u)^T \nabla f(w_t)$$

$$\nabla f(w_t)^T \cdot (w_t - u)$$

Next time:

complete the pf of thm.

stochastic multi-armed bandit