بسم الله الرحمن الرحیم

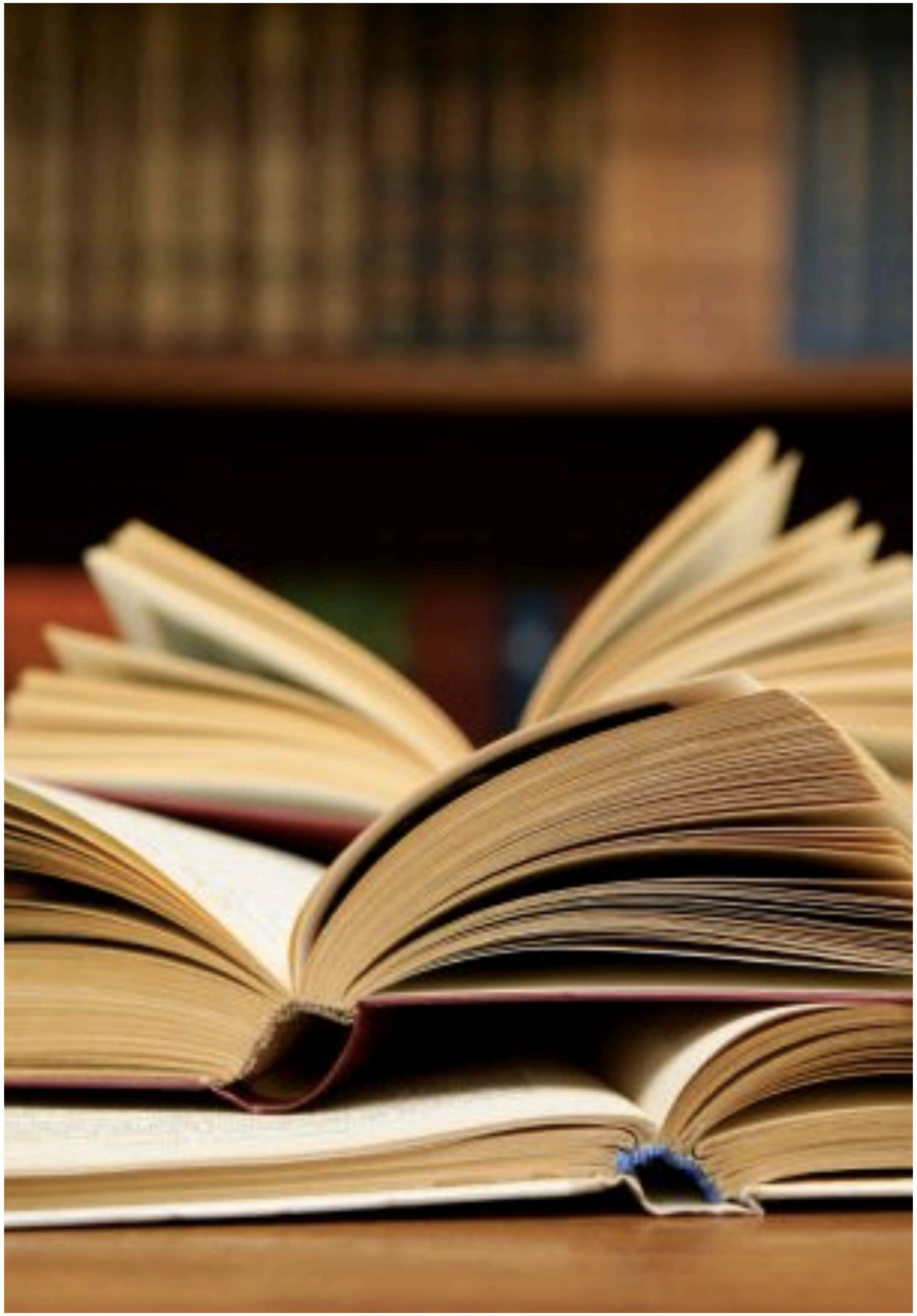
یادگیری برخط

جلسه هجدهم:
پیروی از پیش‌روی منظم شده
و کاهش آینه‌ای (۱)

فهرست

- بهینه‌سازی برخط خطی
- الگوریتم‌های جدید بهینه‌سازی برخط
- بندیت‌سازی
- حل چند مسئله، مثلا بندیت خطی

صورت مسئله



بهینه‌سازی خطی برخط

$\mathcal{L} \subset \mathbb{R}^d$ $\qquad$ $\mathcal{A} \subset \mathbb{R}^d$

۱ـ انتخاب عمل $a_t \in A$

۲ـ دریافت $y_t \in L$

۳ـ هزینه: $\langle a_t, y_t \rangle$

تفاوت با بندیت

بهینه‌سازی برخط غیر خطی $l_t(x)$

بهینه‌سازی خطی برخط

$\mathcal{L} \subset \mathbb{R}^d$ $\mathcal{A} \subset \mathbb{R}^d$

۱ــ انتخاب عمل $a_t \in A$

۲ــ دریافت $y_t \in L$

۳ــ هزینه: $\langle a_t, y_t \rangle$

تفاوت با بندیت

بهینه‌سازی برخط غیر خطی $l_t(x)$

$$R_n(a) = \sum_{t=1}^{n} \langle a_t - a, y_t \rangle$$

$$R_n = \max_{a \in \mathcal{A}} R_n(a)$$

الگوریتم بهینه‌سازی برخط

کاهش گرادیان برخط

کاهش گرادیان

$$a_{t+1} = a_t - \eta y_t$$

$$a_{t+1} = \Pi(a_t - \eta y_t)$$

کاهش گرادیان تصویر شده

کاهش گرادیان برخط

کاهش گرادیان

$$a_{t+1} = a_t - \eta y_t$$

کاهش گرادیان تصویر شده

$$a_{t+1} = \Pi(a_t - \eta y_t)$$

$$a_{t+1} = \arg\min_{a \in A} \|a - (a_t - \eta y_t)\|^2$$

کاهش گرادیان برخط

کاهش گرادیان

$$a_{t+1} = a_t - \eta y_t$$

$$a_{t+1} = \Pi(a_t - \eta y_t)$$

کاهش گرادیان تصویر شده

$$a_{t+1} = \arg\min_{a \in A} \|a - (a_t - \eta y_t)\|^2$$

$$\arg\min_{a \in A} \|a - (a_t - \eta y_t)\|^2$$

کاهش گرادیان برخط

کاهش گرادیان

$$a_{t+1} = a_t - \eta y_t$$

کاهش گرادیان تصویر شده

$$a_{t+1} = \Pi(a_t - \eta y_t)$$

$$a_{t+1} = \arg\min_{a \in A} \|a - (a_t - \eta y_t)\|^2$$

$$\arg\min_{a \in A} \|a - (a_t - \eta y_t)\|^2 = \arg\min_{a \in A}(a - (a_t - \eta y_t))^\top (a - (a_t - \eta y_t))$$

کاهش گرادیان برخط

کاهش گرادیان

$$a_{t+1} = a_t - \eta y_t$$

کاهش گرادیان تصویر شده

$$a_{t+1} = \Pi(a_t - \eta y_t)$$

$$a_{t+1} = \arg\min_{a \in A} \|a - (a_t - \eta y_t)\|^2$$

$$\begin{aligned} \arg\min_{a \in A} \|a - (a_t - \eta y_t)\|^2 &= \arg\min_{a \in A} (a - (a_t - \eta y_t))^\top (a - (a_t - \eta y_t)) \\ &= \arg\min_{a \in A} \|a\|^2 - 2a^\top (a_t - \eta y_t) + \|a_t - \eta y_t\|^2 \end{aligned}$$

کاهش گرادیان برخط

کاهش گرادیان

$$a_{t+1} = a_t - \eta y_t$$

کاهش گرادیان تصویر شده

$$a_{t+1} = \Pi(a_t - \eta y_t)$$

$$a_{t+1} = \arg\min_{a \in A} \|a - (a_t - \eta y_t)\|^2$$

$$\begin{aligned}
\arg\min_{a \in A} \|a - (a_t - \eta y_t)\|^2 &= \arg\min_{a \in A} (a - (a_t - \eta y_t))^\top (a - (a_t - \eta y_t)) \\
&= \arg\min_{a \in A} \|a\|^2 - 2a^\top (a_t - \eta y_t) + \|a_t - \eta y_t\|^2 \\
&= \arg\min_{a \in A} \|a\|^2 + 2\eta a^\top y_t - 2a^\top a_t + \|a_t - \eta y_t\|^2
\end{aligned}$$

کاهش گرادیان برخط

کاهش گرادیان

$$a_{t+1} = a_t - \eta y_t$$

کاهش گرادیان تصویر شده

$$a_{t+1} = \Pi(a_t - \eta y_t)$$

$$a_{t+1} = \arg\min_{a \in A} \|a - (a_t - \eta y_t)\|^2$$

$$\begin{aligned}
\arg\min_{a \in A} \|a - (a_t - \eta y_t)\|^2 &= \arg\min_{a \in A} (a - (a_t - \eta y_t))^\top (a - (a_t - \eta y_t)) \\
&= \arg\min_{a \in A} \|a\|^2 - 2a^\top (a_t - \eta y_t) + \|a_t - \eta y_t\|^2 \\
&= \arg\min_{a \in A} \|a\|^2 + 2\eta a^\top y_t - 2a^\top a_t + \|a_t - \eta y_t\|^2 \\
&= \arg\min_{a \in A} 2\eta a^\top y_t + \|a - a_t\|^2
\end{aligned}$$

کاهش گرادیان برخط

کاهش گرادیان

$$a_{t+1} = a_t - \eta y_t$$

کاهش گرادیان تصویر شده

$$a_{t+1} = \Pi(a_t - \eta y_t)$$

$$a_{t+1} = \arg\min_{a \in A} \|a - (a_t - \eta y_t)\|^2$$

$$\begin{aligned}
\arg\min_{a \in A} \|a - (a_t - \eta y_t)\|^2 &= \arg\min_{a \in A} (a - (a_t - \eta y_t))^\top (a - (a_t - \eta y_t)) \\
&= \arg\min_{a \in A} \|a\|^2 - 2a^\top (a_t - \eta y_t) + \|a_t - \eta y_t\|^2 \\
&= \arg\min_{a \in A} \|a\|^2 + 2\eta a^\top y_t - 2a^\top a_t + \|a_t - \eta y_t\|^2 \\
&= \arg\min_{a \in A} 2\eta a^\top y_t + \|a - a_t\|^2
\end{aligned}$$

$$a_{t+1} = \arg\min_{a \in A} \eta \langle a, y_t \rangle + D_F(a, a_t)$$

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} (\eta \langle a, y_t \rangle + D_F(a, a_t))$$

روش‌های پیروی از پیش‌رو

$$a_{t+1} = \arg\min_{a \in A} \sum_{s} \langle a, y_s \rangle$$

روش‌های پیروی از پیش‌رو

$$a_{t+1} = \arg\min_{a \in A} \sum_s \langle a, y_s \rangle$$

$$a_{t+1} = \arg\min_{a \in A} \eta \sum_s \langle a, y_s \rangle + F(a)$$

پیروی از پیش‌روی منظم شده

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$$

تابع لژاندر

لژاندر: مانند بشقاب ته گود

(a) C is non-empty;

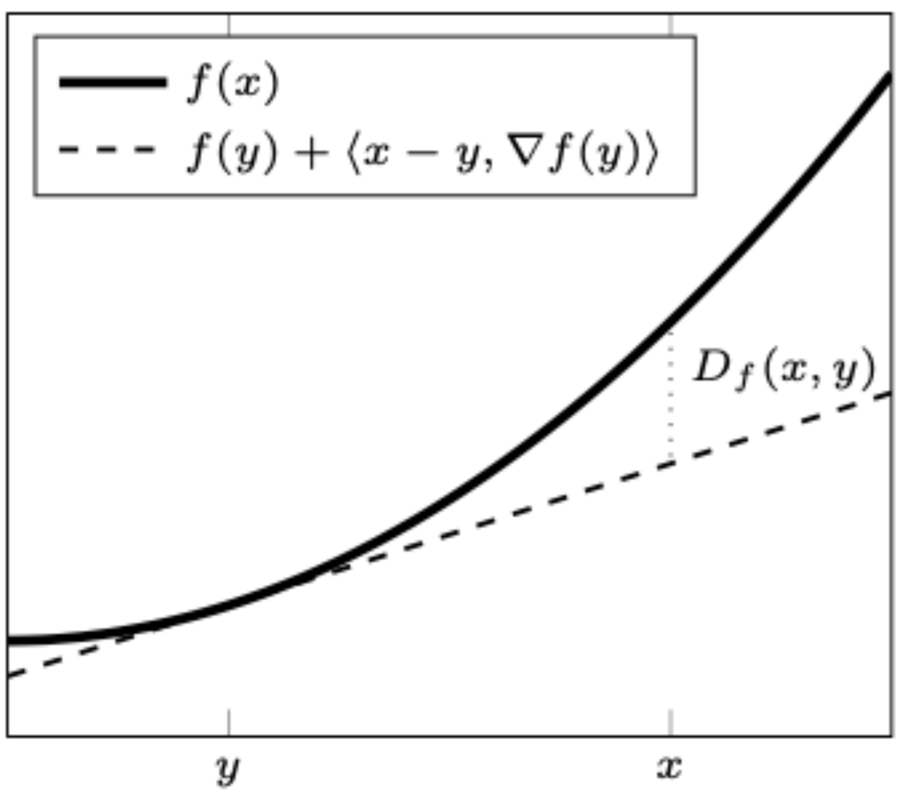
(b) f is differentiable and strictly convex on C; and

(c) $\lim_{n\to\infty} \|\nabla f(x_n)\|_2 = \infty$ for any sequence $(x_n)_n$ with $x_n \in C$ for all n and $\lim_{n\to\infty} x_n = x$ and some $x \in \partial C$.

فاصله D_f

لژاندر: مانند بشقاب ته گود

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y) \rangle$$



چند تابع خوب لژاندر

$f(x) = \frac{1}{2}\|x\|_2^2$ $\qquad$ $\mathrm{dom}(f) = \mathbb{R}^d$

چند تابع خوب لژاندر

$f(x) = \frac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y) \rangle$$

چند تابع خوب لژاندر

$f(x) = \frac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y) \rangle$$

$$= \frac{1}{2}\|x\|_2^2 - \frac{1}{2}\|y\|_2^2 - \langle x - y, y \rangle$$

چند تابع خوب لژاندر

$f(x) = \frac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$

$$\begin{aligned} D_f(x,y) &= f(x) - f(y) - \langle x - y, \nabla f(y) \rangle \\ &= \frac{1}{2}\|x\|_2^2 - \frac{1}{2}\|y\|_2^2 - \langle x - y, y \rangle \\ &= \frac{1}{2}(\|x\|_2^2 + \|y\|_2^2 - 2\langle x, y \rangle) \end{aligned}$$

چند تابع خوب لژاندر

$f(x) = \frac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$

$$D_f(x, y) = f(x) - f(y) - \langle x - y, \nabla f(y)\rangle$$

$$= \frac{1}{2}\|x\|_2^2 - \frac{1}{2}\|y\|_2^2 - \langle x - y, y\rangle$$

$$= \frac{1}{2}(\|x\|_2^2 + \|y\|_2^2 - 2\langle x, y\rangle) \quad = \frac{1}{2}(\|x - y\|_2^2)$$

چند تابع خوب لژاندر

$f(x) = \frac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y)\rangle$$

$$= \frac{1}{2}\|x\|_2^2 - \frac{1}{2}\|y\|_2^2 - \langle x - y, y\rangle$$

$$= \frac{1}{2}(\|x\|_2^2 + \|y\|_2^2 - 2\langle x, y\rangle) \quad = \frac{1}{2}(\|x - y\|_2^2)$$

$f(x) \;=\; \sum_i x_i \log(x_i) \;-\; x_i \qquad \mathrm{dom}(f) \;=\; [0,\infty)^d$

برای توزیع x و y

چند تابع خوب لژاندر

$$f(x) = \tfrac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$$

$$\begin{aligned} D_f(x,y) &= f(x) - f(y) - \langle x - y, \nabla f(y)\rangle \\ &= \frac{1}{2}\|x\|_2^2 - \frac{1}{2}\|y\|_2^2 - \langle x - y, y\rangle \\ &= \frac{1}{2}(\|x\|_2^2 + \|y\|_2^2 - 2\langle x, y\rangle) \quad = \frac{1}{2}(\|x - y\|_2^2) \end{aligned}$$

$$f(x) \;=\; \textstyle\sum_i x_i \log(x_i) \;-\; x_i \qquad \mathrm{dom}(f) \;=\; [0,\infty)^d$$

$$\nabla f(x) \;=\; \log(x)$$

برای توزیع x و y

چند تابع خوب لژاندر

$$f(x) = \tfrac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$$

$$\begin{aligned} D_f(x,y) &= f(x) - f(y) - \langle x - y, \nabla f(y)\rangle \\ &= \frac{1}{2}\|x\|_2^2 - \frac{1}{2}\|y\|_2^2 - \langle x - y, y\rangle \\ &= \frac{1}{2}(\|x\|_2^2 + \|y\|_2^2 - 2\langle x, y\rangle) \quad = \frac{1}{2}(\|x - y\|_2^2) \end{aligned}$$

$$f(x) \;=\; \textstyle\sum_i x_i \log(x_i) \;-\; x_i \qquad \mathrm{dom}(f) \;=\; [0,\infty)^d$$

$$\nabla f(x) \;=\; \log(x)$$

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y)\rangle$$

برای توزیع x و y

چند تابع خوب لژاندر

$$f(x) = \tfrac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$$

$$D_f(x,y) = f(x) - f(y) - \langle x-y, \nabla f(y)\rangle$$

$$= \frac{1}{2}\|x\|_2^2 - \frac{1}{2}\|y\|_2^2 - \langle x-y, y\rangle$$

$$= \frac{1}{2}(\|x\|_2^2 + \|y\|_2^2 - 2\langle x, y\rangle) \quad = \frac{1}{2}(\|x-y\|_2^2)$$

$$f(x) \;=\; \textstyle\sum_i x_i \log(x_i) \;-\; x_i \qquad \mathrm{dom}(f) \;=\; [0,\infty)^d$$

برای توزیع x و y

$$\nabla f(x) \;=\; \log(x)$$

$$D_f(x,y) = f(x) - f(y) - \langle x-y, \nabla f(y)\rangle$$

$$= \langle x, \log(x)\rangle - \langle x,1\rangle - \langle y, \log(y)\rangle + \langle y,1\rangle - \langle x, \log(y)\rangle + \langle y, \log(y)\rangle$$

چند تابع خوب لژاندر

$$f(x) = \tfrac{1}{2}\|x\|_2^2 \qquad \mathrm{dom}(f) = \mathbb{R}^d$$

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y)\rangle$$

$$= \frac{1}{2}\|x\|_2^2 - \frac{1}{2}\|y\|_2^2 - \langle x - y, y\rangle$$

$$= \frac{1}{2}(\|x\|_2^2 + \|y\|_2^2 - 2\langle x, y\rangle) \quad = \frac{1}{2}(\|x - y\|_2^2)$$

$$f(x) \;=\; \textstyle\sum_i x_i \log(x_i) \;-\; x_i \qquad \mathrm{dom}(f) \;=\; [0,\infty)^d$$

برای توزیع x و y

$$\nabla f(x) \;=\; \log(x)$$

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y)\rangle$$

$$= \langle x, \log(x)\rangle - \langle x, 1\rangle - \langle y, \log(y)\rangle + \langle y, 1\rangle - \langle x, \log(y)\rangle + \langle y, \log(y)\rangle$$

$$= \sum x_i \log(x_i/y_i)$$

دو الگوریتم

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

پیروی از پیش‌روی منظم شده

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a)\right)$$

مثال: کاهش آینه‌ای

کاهش آینه‌ای

$$a_1 = \mathrm{argmin}_{a\in\mathcal{A}} F(a)$$

$$a_{t+1} = \mathrm{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$A = \mathbb{R}^d$ $\quad F(a) = \frac{1}{2}\|a\|_2^2$ $\quad \nabla F(a) = a$ $\quad D(a, a_t) = \frac{1}{2}\|a - a_t\|_2^2$

$$a_{t+1} = \mathrm{argmin}_{a\in\mathbb{R}^d} \eta\langle a, y_t\rangle + \frac{1}{2}\|a - a_t\|_2^2$$

مثال: کاهش آینه‌ای

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

$A = \mathbb{R}^d$ $\quad F(a) = \frac{1}{2}\|a\|_2^2 \quad \nabla F(a) = a \quad D(a, a_t) = \frac{1}{2}\|a - a_t\|_2^2$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathbb{R}^d} \eta \langle a, y_t \rangle + \frac{1}{2}\|a - a_t\|_2^2$$

$$\nabla(\ldots) = \eta y_t + a - a_t = 0$$

مثال: کاهش آینه‌ای

کاهش آینه‌ای

$$a_1 = \text{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \text{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

$A = \mathbb{R}^d$ $\quad F(a) = \frac{1}{2}\|a\|_2^2 \quad \nabla F(a) = a \quad D(a, a_t) = \frac{1}{2}\|a - a_t\|_2^2$

$$a_{t+1} = \text{argmin}_{a \in \mathbb{R}^d} \eta \langle a, y_t \rangle + \frac{1}{2}\|a - a_t\|_2^2$$

$$\nabla(\ldots) = \eta y_t + a - a_t = 0$$

کاهش گرادیان برخط

مثال: کاهش آینه‌ای

کاهش آینه‌ای

$$a_1 = \text{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \text{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

$A = \mathbb{R}^d$ $\quad F(a) = \frac{1}{2}\|a\|_2^2 \quad \nabla F(a) = a \quad D(a, a_t) = \frac{1}{2}\|a - a_t\|_2^2$

$$a_{t+1} = \text{argmin}_{a \in \mathbb{R}^d} \eta \langle a, y_t \rangle + \frac{1}{2}\|a - a_t\|_2^2$$

$$\nabla(\ldots) = \eta y_t + a - a_t = 0$$

$$a_{t+1} = a_t - \eta y_t$$

کاهش گرادیان برخط

مثال: کاهش آینه‌ای

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

$A =$ محدب $\quad F(a) = \frac{1}{2}\|a\|_2^2 \quad \nabla F(a) = a \quad D(a, a_t) = \frac{1}{2}\|a - a_t\|_2^2$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathbb{R}^d} \eta \langle a, y_t \rangle + \frac{1}{2}\|a - a_t\|_2^2$$

مثال: کاهش آینه‌ای

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a\in\mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$A =$ محدب $\quad F(a) = \frac{1}{2}\|a\|_2^2 \quad \nabla F(a) = a \quad D(a, a_t) = \frac{1}{2}\|a - a_t\|_2^2$

$$a_{t+1} = \operatorname{argmin}_{a\in\mathbb{R}^d} \eta\langle a, y_t\rangle + \frac{1}{2}\|a - a_t\|_2^2$$

$$a_{t+1} = \Pi(a_t - \eta y_t)$$

کاهش گرادیان برخط
تصویر شده

مثال: کاهش آینه‌ای

$$A = \mathbb{R}^d \qquad F(a) = \frac{1}{2}\|a\|_2^2 \qquad \nabla F(a) = a \qquad D(a, a_t) = \frac{1}{2}\|a - a_t\|_2^2$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathbb{R}^d} \eta \langle a, y_t \rangle + \frac{1}{2}\|a - a_t\|_2^2$$

$$\nabla(\ldots) = \eta y_t + a - a_t = 0$$

$$a_{t+1} = a_t - \eta y_t$$

کاهش گرادیان برخط

$A =$ محدب

کاهش گرادیان برخط تصویرشده

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \eta \langle a, y_t \rangle + \frac{1}{2}\|a - a_t\|_2^2 \;= \Pi(a_t - \eta y_t)$$

مثال: پیروی از پیش‌روی منظم شده

پیروی از پیش‌روی منظم شده

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$$

$A = \mathbb{R}^d$ $\quad F(a) = \frac{1}{2}\|a\|_2^2$ $\quad \nabla F(a) = a$ $\quad D(a, a_t) = \frac{1}{2}\|a - a_t\|_2^2$

مثال: پیروی از پیش‌روی منظم شده

پیروی از پیش‌روی منظم شده

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$$

$A = \mathbb{R}^d$ $\quad F(a) = \frac{1}{2}\|a\|_2^2$ $\quad \nabla F(a) = a$ $\quad D(a, a_t) = \frac{1}{2}\|a - a_t\|_2^2$

$$a_{t+1} = \arg\min_a \eta \sum_s \langle a, y_s \rangle + \frac{1}{2}\|a\|^2$$

مثال: پیروی از پیش‌روی منظم شده

پیروی از پیش‌روی
منظم شده

$$a_1 = \text{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \text{argmin}_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$$

$$A = \mathbb{R}^d \qquad F(a) = \tfrac{1}{2}\|a\|_2^2 \qquad \nabla F(a) = a \qquad D(a, a_t) = \tfrac{1}{2}\|a - a_t\|_2^2$$

$$a_{t+1} = \arg\min_{a} \eta \sum_{s} \langle a, y_s \rangle + \frac{1}{2}\|a\|^2$$

$$\nabla(\ldots) = \eta \sum y_s + a = 0$$

مثال: پیروی از پیش‌روی منظم شده

پیروی از پیش‌روی منظم شده

$$a_1 = \text{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \text{argmin}_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$$

$A = \mathbb{R}^d$ $\quad F(a) = \frac{1}{2}\|a\|_2^2 \quad \nabla F(a) = a \quad D(a, a_t) = \frac{1}{2}\|a - a_t\|_2^2$

$$a_{t+1} = \arg\min_a \eta \sum_s \langle a, y_s \rangle + \frac{1}{2}\|a\|^2$$

$$\nabla(\ldots) = \eta \sum y_s + a = 0$$

$$a_{t+1} = -\eta \sum_{s=1}^{t} y_t$$

مثال: پیروی از پیش‌روی منظم شده

پیروی از پیش‌روی منظم شده

$$a_1 = \mathrm{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$$

$A = \mathbb{R}^d$ $\quad F(a) = \frac{1}{2}\|a\|_2^2$ $\quad \nabla F(a) = a$ $\quad D(a, a_t) = \frac{1}{2}\|a - a_t\|_2^2$

$$a_{t+1} = \arg\min_a \eta \sum_s \langle a, y_s \rangle + \frac{1}{2}\|a\|^2$$

$$\nabla(\ldots) = \eta \sum y_s + a = 0$$

$$a_{t+1} = -\eta \sum_{s=1}^{t} y_t \quad = a_t - \eta y_t$$

مثال: پیروی از پیش‌روی منظم شده

پیروی از پیش‌روی منظم شده

$$a_1 = \mathrm{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$$

$A = \mathbb{R}^d \qquad F(a) = \frac{1}{2}\|a\|_2^2 \qquad \nabla F(a) = a \qquad D(a, a_t) = \frac{1}{2}\|a - a_t\|_2^2$

$$a_{t+1} = \arg\min_a \eta \sum_s \langle a, y_s \rangle + \frac{1}{2}\|a\|^2$$

$$\nabla(\ldots) = \eta \sum y_s + a = 0$$

$$a_{t+1} = -\eta \sum_{s=1}^{t} y_t \quad = a_t - \eta y_t$$

کاهش گرادیان برخط

مثال: پیروی از پیش‌روی منظم شده

پیروی از پیش‌روی
منظم شده

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$$

$A = \mathscr{P}_{d-1} \qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \operatorname{dom}(f) = [0, \infty)^d$

مثال: پیروی از پیش‌روی منظم شده

پیروی از پیش‌روی منظم شده

$$a_1 = \text{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \text{argmin}_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$$

$A = \mathscr{P}_{d-1}$ $\quad f(x) = \sum_i x_i \log(x_i) - x_i \quad \text{dom}(f) = [0, \infty)^d$

$$a_{t+1} = \text{argmin}_{a \in \mathcal{A}} \eta \sum_{s=1}^{t} \langle a, y_s \rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

قضیه: در شرایط خوب با تابع لژاندر f

$$y = \mathrm{argmin}_{z \in A} f(z)$$

$$\boxed{\tilde{y} = \mathrm{argmin}_{z \in \mathbb{R}^d} f(z)}$$

$$y = \mathrm{argmin}_{z \in A} D_f(z, \tilde{y}).$$

$$A = \mathscr{P}_{d-1} \qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \mathrm{dom}(f) = [0, \infty)^d$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}} \, \eta \sum_{s=1}^{t} \langle a, y_s \rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$\boxed{\tilde{y} = \mathrm{argmin}_{z \in \mathbb{R}^d} \, f(z)}$$

$$y = \mathrm{argmin}_{z \in A} \, D_f(z, \tilde{y}).$$

$A = \mathscr{P}_{d-1}$ $\qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \mathrm{dom}(f) = [0, \infty)^d$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}} \, \eta \sum_{s=1}^{t} \langle a, y_s \rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$\boxed{\tilde{y} = \mathrm{argmin}_{z \in \mathbb{R}^d} f(z)}$$

$$y = \mathrm{argmin}_{z \in A} D_f(z, \tilde{y}).$$

$$\tilde{y} = \underset{a \in \mathbb{R}^d}{\arg\min} \;\; \eta \sum_{s=1}^{t} \langle a, y_s \rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$A = \mathscr{P}_{d-1} \qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \mathrm{dom}(f) = [0, \infty)^d$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}} \, \eta \sum_{s=1}^{t} \langle a, y_s \rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$\tilde{y} = \mathrm{argmin}_{z \in \mathbb{R}^d} f(z)$$

$$y = \mathrm{argmin}_{z \in A} D_f(z, \tilde{y}).$$

$$\tilde{y} = \arg\min_{a \in \mathbb{R}^d} \; \eta \sum_{s=1}^{t} \langle a, y_s \rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$\nabla(\ldots) = \eta \sum_s y_s + \log(a) = 0$$

$$A = \mathscr{P}_{d-1} \qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \mathrm{dom}(f) = [0,\infty)^d$$

$$a_{t+1} = \mathrm{argmin}_{a\in\mathcal{A}}\, \eta \sum_{s=1}^{t} \langle a, y_s\rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$\tilde{y} = \mathrm{argmin}_{z\in\mathbb{R}^d} f(z)$$

$$y = \mathrm{argmin}_{z\in A} D_f(z, \tilde{y}).$$

$$\tilde{y} = \arg\min_{a\in\mathbb{R}^d}\; \eta \sum_{s=1}^{t} \langle a, y_s\rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$\nabla(\ldots) = \eta \sum_s y_s + \log(a) = 0 \qquad \tilde{y} = \exp(-\eta \sum_s y_s)$$

$$A = \mathscr{P}_{d-1} \qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \mathrm{dom}(f) = [0, \infty)^d$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}}\, \eta \sum_{s=1}^{t} \langle a, y_s \rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$\tilde{y} = \mathrm{argmin}_{z \in \mathbb{R}^d} f(z)$$

$$y = \mathrm{argmin}_{z \in A} D_f(z, \tilde{y}).$$

$$\tilde{y} = \arg\min_{a \in \mathbb{R}^d}\ \eta \sum_{s=1}^{t} \langle a, y_s \rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$\nabla(\ldots) = \eta \sum_s y_s + \log(a) = 0 \qquad \tilde{y} = \exp(-\eta \sum_s y_s)$$

$$a_{t+1} = \arg\min_{a \in \mathscr{P}_{d-1}} D_f(a, \tilde{y})$$

$A = \mathscr{P}_{d-1} \qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \mathrm{dom}(f) = [0, \infty)^d$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}} \eta \sum_{s=1}^{t} \langle a, y_s \rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$\tilde{y} = \exp(-\eta \sum_s y_s)$$

$$\boxed{\tilde{y} = \mathrm{argmin}_{z \in \mathbb{R}^d} f(z)}$$

$$y = \mathrm{argmin}_{z \in A} D_f(z, \tilde{y}).$$

$$A = \mathscr{P}_{d-1} \qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \mathrm{dom}(f) = [0,\infty)^d$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}}\, \eta \sum_{s=1}^{t} \langle a, y_s \rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$\tilde{y} = \exp(-\eta \sum_s y_s)$$

$$\boxed{\tilde{y} = \mathrm{argmin}_{z \in \mathbb{R}^d} f(z)}$$

$$y = \mathrm{argmin}_{z \in A} D_f(z, \tilde{y}).$$

$$\begin{aligned} D_f(x,y) &= f(x) - f(y) - \langle x - y, \nabla f(y) \rangle \\ &= \langle x, \log(x) \rangle - \langle x, 1 \rangle - \langle y, \log(y) \rangle + \langle y, 1 \rangle - \langle x, \log(y) \rangle + \langle y, \log(y) \rangle \\ &= \sum x_i \log(x_i / y_i) + C \end{aligned}$$

$$A = \mathscr{P}_{d-1} \qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \mathrm{dom}(f) = [0,\infty)^d$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}}\, \eta \sum_{s=1}^{t} \langle a, y_s \rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$\tilde{y} = \exp(-\eta \sum_s y_s)$$

$$\tilde{y} = \mathrm{argmin}_{z \in \mathbb{R}^d}\, f(z)$$

$$y = \mathrm{argmin}_{z \in A}\, D_f(z, \tilde{y}).$$

$$\begin{aligned} D_f(x,y) &= f(x) - f(y) - \langle x - y, \nabla f(y) \rangle \\ &= \langle x, \log(x) \rangle - \langle x, 1 \rangle - \langle y, \log(y) \rangle + \langle y, 1 \rangle - \langle x, \log(y) \rangle + \langle y, \log(y) \rangle \\ &= \sum x_i \log(x_i / y_i) + C \end{aligned}$$

$$a_{t+1} = \arg\min_{a \in \mathscr{P}_{d-1}} \sum_i x_i \log(x_i / \tilde{y}_i)$$

$$A = \mathscr{P}_{d-1} \qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \mathrm{dom}(f) = [0,\infty)^d$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}}\, \eta \sum_{s=1}^{t} \langle a, y_s \rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$\tilde{y} = \exp(-\eta \sum_s y_s)$$

$$\boxed{\tilde{y} = \mathrm{argmin}_{z \in \mathbb{R}^d} f(z)}$$

$$y = \mathrm{argmin}_{z \in A} D_f(z, \tilde{y}).$$

$$\begin{aligned} D_f(x,y) &= f(x) - f(y) - \langle x - y, \nabla f(y) \rangle \\ &= \langle x, \log(x) \rangle - \langle x, 1 \rangle - \langle y, \log(y) \rangle + \langle y, 1 \rangle - \langle x, \log(y) \rangle + \langle y, \log(y) \rangle \\ &= \sum x_i \log(x_i / y_i) + C \end{aligned}$$

$$a_{t+1} = \arg\min_{a \in \mathscr{P}_{d-1}} \sum_i x_i \log(x_i / \tilde{y}_i)$$

$$\nabla \sum_i x_i \log(x_i / \tilde{y}_i) = \alpha 1$$

$$A = \mathscr{P}_{d-1} \qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \mathrm{dom}(f) = [0,\infty)^d$$

$$a_{t+1} = \mathrm{argmin}_{a\in\mathcal{A}}\, \eta \sum_{s=1}^{t} \langle a, y_s\rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$\tilde{y} = \exp(-\eta \sum_s y_s)$$

$$\tilde{y} = \mathrm{argmin}_{z\in\mathbb{R}^d}\, f(z)$$

$$y = \mathrm{argmin}_{z\in A}\, D_f(z, \tilde{y}).$$

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y)\rangle$$
$$= \langle x, \log(x)\rangle - \langle x, 1\rangle - \langle y, \log(y)\rangle + \langle y, 1\rangle - \langle x, \log(y)\rangle + \langle y, \log(y)\rangle$$
$$= \sum x_i \log(x_i/y_i) + C$$

$$a_{t+1} = \arg\min_{a\in\mathscr{P}_{d-1}} \sum_i x_i \log(x_i/\tilde{y}_i)$$

$$\nabla \sum_i x_i \log(x_i/\tilde{y}_i) = \alpha 1 \qquad \log(x_i) + 1 - \log(\tilde{y}_i) = \alpha$$

$$A = \mathscr{P}_{d-1} \qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \mathrm{dom}(f) = [0,\infty)^d$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}}\, \eta \sum_{s=1}^{t} \langle a, y_s \rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$\tilde{y} = \exp(-\eta \sum_s y_s)$$

$$\tilde{y} = \mathrm{argmin}_{z \in \mathbb{R}^d} f(z)$$

$$y = \mathrm{argmin}_{z \in A} D_f(z, \tilde{y}).$$

$$\begin{aligned} D_f(x,y) &= f(x) - f(y) - \langle x - y, \nabla f(y) \rangle \\ &= \langle x, \log(x) \rangle - \langle x, 1 \rangle - \langle y, \log(y) \rangle + \langle y, 1 \rangle - \langle x, \log(y) \rangle + \langle y, \log(y) \rangle \\ &= \sum x_i \log(x_i / y_i) + C \end{aligned}$$

$$a_{t+1} = \arg\min_{a \in \mathscr{P}_{d-1}} \sum_i x_i \log(x_i / \tilde{y}_i)$$

$$\nabla \sum_i x_i \log(x_i / \tilde{y}_i) = \alpha 1 \qquad \log(x_i) + 1 - \log(\tilde{y}_i) = \alpha$$

$$\log(x_i / \tilde{y}_i) = \alpha - 1$$

$$A = \mathscr{P}_{d-1} \qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \mathrm{dom}(f) = [0,\infty)^d$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}}\, \eta \sum_{s=1}^{t} \langle a, y_s \rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$\tilde{y} = \exp(-\eta \sum_s y_s)$$

$$\boxed{\tilde{y} = \mathrm{argmin}_{z \in \mathbb{R}^d} f(z)}$$

$$y = \mathrm{argmin}_{z \in A} D_f(z, \tilde{y}).$$

$$\begin{aligned} D_f(x,y) &= f(x) - f(y) - \langle x - y, \nabla f(y) \rangle \\ &= \langle x, \log(x) \rangle - \langle x, 1 \rangle - \langle y, \log(y) \rangle + \langle y, 1 \rangle - \langle x, \log(y) \rangle + \langle y, \log(y) \rangle \\ &= \sum x_i \log(x_i / y_i) + C \end{aligned}$$

$$a_{t+1} = \arg\min_{a \in \mathscr{P}_{d-1}} \sum_i x_i \log(x_i / \tilde{y}_i)$$

$$\nabla \sum_i x_i \log(x_i/\tilde{y}_i) = \alpha 1 \qquad \log(x_i) + 1 - \log(\tilde{y}_i) = \alpha$$

$$\log(x_i/\tilde{y}_i) = \alpha - 1 \qquad x_i \propto \tilde{y}_i$$

$$A = \mathscr{P}_{d-1} \qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \mathrm{dom}(f) = [0,\infty)^d$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}}\, \eta \sum_{s=1}^{t} \langle a, y_s \rangle + \sum_{i=1}^{d} a_i \log(a_i) - a_i$$

$$\tilde{y} = \exp(-\eta \sum_s y_s)$$

$$\boxed{\tilde{y} = \mathrm{argmin}_{z \in \mathbb{R}^d} f(z)}$$

$$y = \mathrm{argmin}_{z \in A} D_f(z, \tilde{y}).$$

$$\begin{aligned} D_f(x,y) &= f(x) - f(y) - \langle x - y, \nabla f(y) \rangle \\ &= \langle x, \log(x) \rangle - \langle x, 1 \rangle - \langle y, \log(y) \rangle + \langle y, 1 \rangle - \langle x, \log(y) \rangle + \langle y, \log(y) \rangle \\ &= \sum x_i \log(x_i / y_i) + C \end{aligned}$$

$$a_{t+1} = \arg\min_{a \in \mathscr{P}_{d-1}} \sum_i x_i \log(x_i / \tilde{y}_i)$$

$$\nabla \sum_i x_i \log(x_i / \tilde{y}_i) = \alpha 1 \qquad \log(x_i) + 1 - \log(\tilde{y}_i) = \alpha$$

$$\log(x_i / \tilde{y}_i) = \alpha - 1 \qquad x_i \propto \tilde{y}_i \qquad x_i = \tilde{y}_i / \|\tilde{y}\|$$

مثال: پیروی از پیش‌روی منظم شده

پیروی از پیش‌روی
منظم شده

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$$

$$A = \mathscr{P}_{d-1} \qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \operatorname{dom}(f) = [0, \infty)^d$$

$$a_{t+1,i} = \frac{\exp\left(-\eta \sum_{s=1}^{t} y_{si}\right)}{\sum_{j=1}^{d} \exp\left(-\eta \sum_{s=1}^{t} y_{sj}\right)}.$$

عملیات محاسباتی کاهش آینه‌ای

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

عملیات محاسباتی کاهش آینه‌ای

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

عملیات محاسباتی کاهش آینه‌ای

$$a_{t+1} = \operatorname{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$$\begin{aligned}
\tilde{a}_{t+1} &= \operatorname{argmin}_{a\in\mathcal{D}} \eta\langle a, y_t\rangle + D_F(a, a_t)\\
a_{t+1} &= \operatorname{argmin}_{a\in\mathcal{A}} D_F(a, \tilde{a}_{t+1}).
\end{aligned}$$

بسم الله الرحمن الرحیم

یادگیری برخط

جلسه نوزدهم:
پیروی از پیش‌روی منظم شده
و کاهش آینه‌ای (۲)

مرور

- بهینه‌سازی برخط خطی
- الگوریتم‌های جدید بهینه‌سازی برخط
- بندیت‌سازی
- حل چند مسئله، مثلا بندیت خطی

تابع لژاندر

لژاندر: مانند بشقاب ته‌گود

(a) C is non-empty;

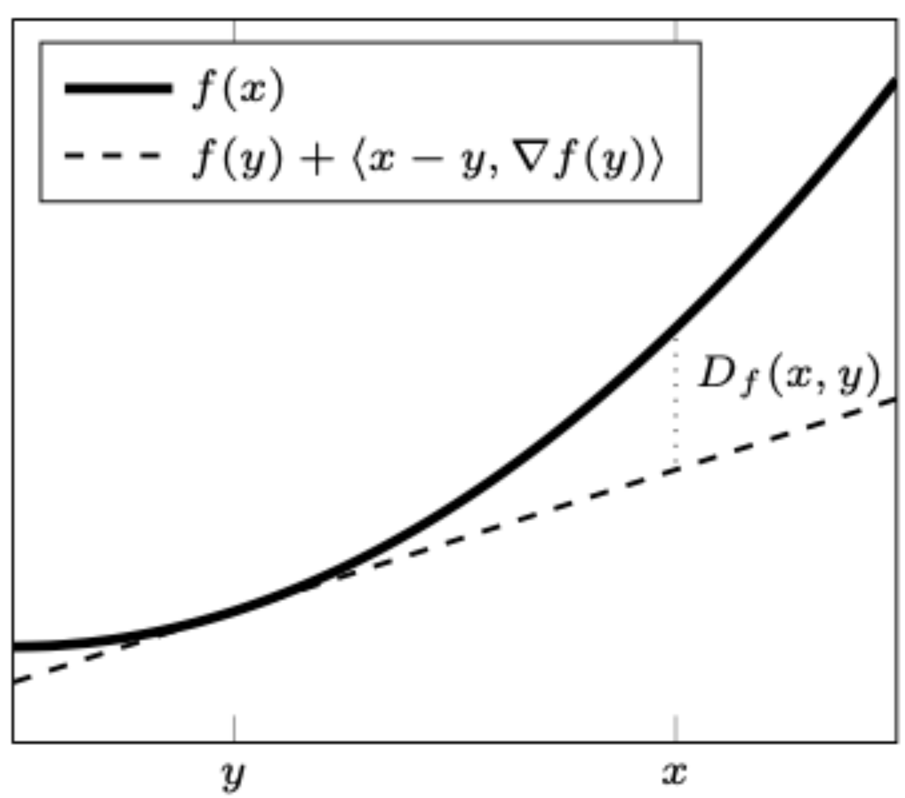
(b) f is differentiable and strictly convex on C; and

(c) $\lim_{n\to\infty} \|\nabla f(x_n)\|_2 = \infty$ for any sequence $(x_n)_n$ with $x_n \in C$ for all n and $\lim_{n\to\infty} x_n = x$ and some $x \in \partial C$.

فاصله $\mathbf{D_f}$

لژاندر: مانند بشقاب ته گود

$$D_f(x, y) = f(x) - f(y) - \langle x - y, \nabla f(y) \rangle$$



دو الگوریتم

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a\in\mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

پیروی از پیش‌روی منظم شده

$$a_1 = \operatorname{argmin}_{a\in\mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a\in\mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s\rangle + F(a)\right)$$

مثال: کاهش آینه‌ای

$A = \mathbb{R}^d$ $\quad F(a) = \frac{1}{2}\|a\|_2^2$ $\quad \nabla F(a) = a$ $\quad D(a, a_t) = \frac{1}{2}\|a - a_t\|_2^2$

$$a_{t+1} = \text{argmin}_{a \in \mathbb{R}^d}\, \eta\langle a, y_t\rangle + \frac{1}{2}\|a - a_t\|_2^2$$

$$\nabla(\ldots) = \eta y_t + a - a_t = 0$$

$$a_{t+1} = a_t - \eta y_t$$

کاهش گرادیان برخط

$A =$ محدب

کاهش گرادیان برخط تصویرشده

$$a_{t+1} = \text{argmin}_{a \in \mathcal{A}}\, \eta\langle a, y_t\rangle + \frac{1}{2}\|a - a_t\|_2^2 \;= \Pi(a_t - \eta y_t)$$

مثال: پیروی از پیش‌روی منظم شده

پیروی از پیش‌روی منظم شده

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$$

$$A = \mathscr{P}_{d-1} \qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \operatorname{dom}(f) = [0, \infty)^d$$

$$a_{t+1,i} = \frac{\exp\left(-\eta \sum_{s=1}^{t} y_{si}\right)}{\sum_{j=1}^{d} \exp\left(-\eta \sum_{s=1}^{t} y_{sj}\right)}.$$

قضیه: در شرایط خوب با تابع لژاندر f

$$y = \operatorname{argmin}_{z \in A} f(z)$$

$$\boxed{\tilde{y} = \operatorname{argmin}_{z \in \mathbb{R}^d} f(z)}$$

$$y = \operatorname{argmin}_{z \in A} D_f(z, \tilde{y}).$$

تحلیل پشیمانی

کاهش آینه‌ای

پیروی از پیش‌روی منظم شده

تحلیل پیشمانی کاهش آینه‌ای

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

قضیه:

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$R_n(a) \leq \frac{1}{\eta} \left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1}) \right)$$

$$\tilde{a}_{t+1} = \operatorname{argmin}_{a \in \mathcal{D}} \eta \langle a, y_t \rangle + D_F(a, a_t)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} D_F(a, \tilde{a}_{t+1}).$$

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle$$

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle$$

$$\langle a - a_{t+1}, \eta y_t + \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \geq 0$$

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle$$

$$\langle a - a_{t+1}, \eta y_t + \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \geq 0$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle$$

$$\langle a - a_{t+1}, \eta y_t + \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \geq 0$$

$$a_{t+1} = \text{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

$$\langle a_{t+1} - a, y_t \rangle \leq \frac{1}{\eta} \langle a - a_{t+1}, \nabla F(a_{t+1}) - \nabla F(a_t) \rangle$$

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle$$

$$\langle a - a_{t+1}, \eta y_t + \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \geq 0$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

$$\begin{aligned} \langle a_{t+1} - a, y_t \rangle &\leq \frac{1}{\eta} \langle a - a_{t+1}, \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \\ &= \frac{1}{\eta} \left(D(a, a_t) - D(a, a_{t+1}) - D(a_{t+1}, a_t) \right) \end{aligned}$$

$$\begin{array}{ccc} F(a) & -F(a) & -F(a_{t+1}) \\ -F(a_t) & +F(a_{t+1}) & +F(a_t) \\ -\langle a - a_t, \nabla F(a_t) \rangle & +\langle a - a_{t+1}, \nabla F(a_{t+1}) \rangle & +\langle a_{t+1} - a_t, \nabla F(a_t) \rangle \end{array}$$

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle$$

$$\langle a - a_{t+1}, \eta y_t + \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \geq 0$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

$$\begin{aligned} \langle a_{t+1} - a, y_t \rangle &\leq \frac{1}{\eta} \langle a - a_{t+1}, \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \\ &= \frac{1}{\eta} \left(D(a, a_t) - D(a, a_{t+1}) - D(a_{t+1}, a_t) \right) \end{aligned}$$

$$\begin{array}{ccc} \cancel{F(a)} & -\cancel{F(a)} & -F(a_{t+1}) \\ -F(a_t) & +F(a_{t+1}) & +F(a_t) \\ -\langle a - a_t, \nabla F(a_t) \rangle & +\langle a - a_{t+1}, \nabla F(a_{t+1}) \rangle & +\langle a_{t+1} - a_t, \nabla F(a_t) \rangle \end{array}$$

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle$$

$$\langle a - a_{t+1}, \eta y_t + \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \geq 0$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

$$\begin{aligned}
\langle a_{t+1} - a, y_t \rangle &\leq \frac{1}{\eta} \langle a - a_{t+1}, \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \\
&= \frac{1}{\eta} \left(D(a, a_t) - D(a, a_{t+1}) - D(a_{t+1}, a_t) \right)
\end{aligned}$$

$$\begin{array}{ccc}
\cancel{F(a)} & -\cancel{F(a)} & -\cancel{F(a_{t+1})} \\
-F(a_t) & +\cancel{F(a_{t+1})} & +F(a_t) \\
-\langle a - a_t, \nabla F(a_t) \rangle & +\langle a - a_{t+1}, \nabla F(a_{t+1}) \rangle & +\langle a_{t+1} - a_t, \nabla F(a_t) \rangle
\end{array}$$

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle$$

$$\langle a - a_{t+1}, \eta y_t + \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \geq 0$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

$$\begin{aligned} \langle a_{t+1} - a, y_t \rangle &\leq \frac{1}{\eta} \langle a - a_{t+1}, \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \\ &= \frac{1}{\eta} \left(D(a, a_t) - D(a, a_{t+1}) - D(a_{t+1}, a_t) \right) \end{aligned}$$

$$\begin{array}{ccc} \cancel{F}(a) & -\cancel{F}(a) & -\cancel{F}(a_{t+1}) \\ -\cancel{F}(a_t) & +\cancel{F}(a_{t+1}) & +\cancel{F}(a_t) \\ -\langle a - a_t, \nabla F(a_t) \rangle & +\langle a - a_{t+1}, \nabla F(a_{t+1}) \rangle & +\langle a_{t+1} - a_t, \nabla F(a_t) \rangle \end{array}$$

$$\langle a_t - a, y_t \rangle = \langle a_t - a_{t+1}, y_t \rangle + \langle a_{t+1} - a, y_t \rangle$$

$$\langle a - a_{t+1}, \eta y_t + \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \geq 0$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t) \right)$$

$$\begin{aligned} \langle a_{t+1} - a, y_t \rangle &\leq \frac{1}{\eta} \langle a - a_{t+1}, \nabla F(a_{t+1}) - \nabla F(a_t) \rangle \\ &= \frac{1}{\eta} \left(D(a, a_t) - D(a, a_{t+1}) - D(a_{t+1}, a_t) \right) \end{aligned}$$

$$\begin{array}{ccc} \cancel{F(a)} & -\cancel{F(a)} & -\cancel{F(a_{t+1})} \\ -\cancel{F(a_t)} & +\cancel{F(a_{t+1})} & +\cancel{F(a_t)} \\ -\langle a - \cancel{a_t}, \nabla F(a_t) \rangle & +\langle a - a_{t+1}, \nabla F(a_{t+1}) \rangle & +\langle a_{t+1} - \cancel{a_t}, \nabla F(a_t) \rangle \end{array}$$

$$
\begin{aligned}
R_n &= \sum_{t=1}^{n} \langle a_t - a, y_t \rangle \\
&\leq \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle + \frac{1}{\eta} \sum_{t=1}^{n} \left(D(a, a_t) - D(a, a_{t+1}) - D(a_{t+1}, a_t) \right)
\end{aligned}
$$

$$
\begin{aligned}
R_n &= \sum_{t=1}^{n} \langle a_t - a, y_t \rangle \\
&\leq \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle + \frac{1}{\eta} \sum_{t=1}^{n} \left(D(a, a_t) - D(a, a_{t+1}) - D(a_{t+1}, a_t) \right) \\
&= \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle + \frac{1}{\eta} \left(D(a, a_1) - D(a, a_{n+1}) - \sum_{t=1}^{n} D(a_{t+1}, a_t) \right)
\end{aligned}
$$

$$
\begin{aligned}
R_n &= \sum_{t=1}^{n} \langle a_t - a, y_t \rangle \\
&\leq \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle + \frac{1}{\eta} \sum_{t=1}^{n} \left(D(a, a_t) - D(a, a_{t+1}) - D(a_{t+1}, a_t) \right) \\
&= \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle + \frac{1}{\eta} \left(D(a, a_1) - D(a, a_{n+1}) - \sum_{t=1}^{n} D(a_{t+1}, a_t) \right) \\
&\leq \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle + \frac{F(a) - F(a_1)}{\eta} - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t)
\end{aligned}
$$

$$\eta y_t = \nabla F(a_t) - \nabla F(a_{t+1})$$

$$\langle a_t - a_{t+1}, y_t \rangle = \frac{1}{\eta} \langle a_t - a_{t+1}, \nabla F(a_t) - \nabla F(\tilde{a}_{t+1}) \rangle$$

$$\eta y_t = \nabla F(a_t) - \nabla F(a_{t+1})$$

$$\langle a_t - a_{t+1}, y_t \rangle = \frac{1}{\eta} \langle a_t - a_{t+1}, \nabla F(a_t) - \nabla F(\tilde{a}_{t+1}) \rangle$$

$$= \frac{1}{\eta} \left(D(a_{t+1}, a_t) + D(a_t, \tilde{a}_{t+1}) - D(a_{t+1}, \tilde{a}_{t+1}) \right)$$

$$\eta y_t = \nabla F(a_t) - \nabla F(a_{t+1})$$

$$
\begin{aligned}
\langle a_t - a_{t+1}, y_t \rangle &= \frac{1}{\eta} \langle a_t - a_{t+1}, \nabla F(a_t) - \nabla F(\tilde{a}_{t+1}) \rangle \\
&= \frac{1}{\eta} \left(D(a_{t+1}, a_t) + D(a_t, \tilde{a}_{t+1}) - D(a_{t+1}, \tilde{a}_{t+1}) \right) \\
&\leq \frac{1}{\eta} \left(D(a_{t+1}, a_t) + D(a_t, \tilde{a}_{t+1}) \right)
\end{aligned}
$$

تحلیل پشیمانی پیروی از پیش‌روی منظم شده

پیروی از پیش‌روی منظم شده

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$$

قضیه:

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

مثال (گوی واحد)

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $\eta = \sqrt{1/n} \qquad F(a) = \frac{1}{2}\|a\|_2^2$

آنگاه $R_n \leq \sqrt{n}.$

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

مثال (گوی واحد)

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $\quad F(a) = \frac{1}{2}\|a\|_2^2 \qquad \eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} (\eta \langle a, y_t \rangle + D_F(a, a_t))$$

کاهش گرادیان تصویر شده

مثال (گوی واحد)

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $F(a) = \frac{1}{2}\|a\|_2^2$ $\eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}$.

قضیه:

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$R_n(a) \leq \frac{1}{\eta} \left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1}) \right)$$

$$\tilde{a}_{t+1} = \operatorname{argmin}_{a \in \mathcal{D}} \eta \langle a, y_t \rangle + D_F(a, a_t)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} D_F(a, \tilde{a}_{t+1}).$$

مثال (گوی واحد)

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $F(a) = \frac{1}{2}\|a\|_2^2$ $\eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

مثال (گوی واحد)

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $F(a) = \frac{1}{2}\|a\|_2^2$ $\eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

مثال (گوی واحد)

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $F(a) = \frac{1}{2}\|a\|_2^2$ $\quad \eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2$$

مثال (گوی واحد)

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $F(a) = \frac{1}{2}\|a\|_2^2$ $\quad \eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2 = \frac{\eta^2}{2}\|y_t\|_2^2$$

مثال (گوی واحد)

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $\eta = \sqrt{1/n} \qquad F(a) = \frac{1}{2}\|a\|_2^2$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2 = \frac{\eta^2}{2}\|y_t\|_2^2$$

$$R_n \leq \frac{\mathrm{diam}_F(\mathcal{A})}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n}\|y_t\|_2^2$$

مثال (گوی واحد)

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $\quad F(a) = \frac{1}{2}\|a\|_2^2 \qquad \eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2 = \frac{\eta^2}{2}\|y_t\|_2^2$$

$$R_n \leq \frac{\operatorname{diam}_F(\mathcal{A})}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n}\|y_t\|_2^2 \leq \frac{1}{2\eta} + \frac{\eta n}{2}$$

مثال (گوی واحد)

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $\eta = \sqrt{1/n}$ $F(a) = \frac{1}{2}\|a\|_2^2$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2 = \frac{\eta^2}{2}\|y_t\|_2^2$$

$$R_n \leq \frac{\mathrm{diam}_F(\mathcal{A})}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n}\|y_t\|_2^2 \leq \frac{1}{2\eta} + \frac{\eta n}{2} = \sqrt{n}.$$

مثال (گوی واحد) ==> حالت کلی

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $\quad F(a) = \frac{1}{2}\|a\|_2^2 \quad \eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n \leq O(\sqrt{nDG})$$

مثال (گوی واحد) ==> حالت کلی

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \le 1\}$$

الگوریتم کاهش آینه‌ای $\quad F(a) = \frac{1}{2}\|a\|_2^2 \qquad \eta = \sqrt{1/n}$

آنگاه $R_n \le \sqrt{n}.$

$$R_n(a) \le \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$R_n \le O(\sqrt{nD}G)$$

مثال (گوی واحد) ==> حالت کلی

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $F(a) = \frac{1}{2}\|a\|_2^2$ $\eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2$$

$$R_n \leq O(\sqrt{nDG})$$

مثال (گوی واحد) ==> حالت کلی

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $F(a) = \frac{1}{2}\|a\|_2^2 \qquad \eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2 = \frac{\eta^2}{2}\|y_t\|_2^2$$

$$R_n \leq O(\sqrt{nDG})$$

مثال (گوی واحد) ==> حالت کلی

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \le 1\}$$

الگوریتم کاهش آینه‌ای $\quad F(a) = \frac{1}{2}\|a\|_2^2 \qquad \eta = \sqrt{1/n}$

آنگاه $\; R_n \le \sqrt{n}.$

$$R_n(a) \le \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2 = \frac{\eta^2}{2}\|y_t\|_2^2$$

$$R_n \le O(\sqrt{nDG})$$

$$R_n \le \frac{\operatorname{diam}_F(\mathcal{A})}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n}\|y_t\|_2^2$$

مثال (گوی واحد) ==> حالت کلی

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \le 1\}$$

الگوریتم کاهش آینه‌ای $\eta = \sqrt{1/n}$ $\quad F(a) = \frac{1}{2}\|a\|_2^2$

آنگاه $R_n \le \sqrt{n}.$

$$R_n(a) \le \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2 = \frac{\eta^2}{2}\|y_t\|_2^2$$

$$R_n \le O(\sqrt{nD}G)$$

$$R_n \le \frac{\operatorname{diam}_F(\mathcal{A})}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n}\|y_t\|_2^2 \le \frac{1}{2\eta} + \frac{\eta n}{2}$$

مثال (گوی واحد) ==> حالت کلی

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $\eta = \sqrt{1/n} \qquad F(a) = \frac{1}{2}\|a\|_2^2$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2 = \frac{\eta^2}{2}\|y_t\|_2^2$$

$$R_n \leq O(\sqrt{nDG})$$

$$R_n \leq \frac{\mathrm{diam}_F(\mathcal{A})}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n}\|y_t\|_2^2 \leq \frac{1}{2\eta} + \frac{\eta n}{2} = \sqrt{n}.$$

مثال: کاهش آینه‌ای روی سادک

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a\in\mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$A = \mathscr{P}_{d-1}$ $\quad f(x) = \sum_i x_i \log(x_i) - x_i \quad \mathrm{dom}(f) = [0,\infty)^d$

$y_t \in \mathcal{L} = [0,1]^d$

مثال: کاهش آینه‌ای روی سادک

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a\in\mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$A = \mathscr{P}_{d-1}$ $\quad f(x) = \sum_i x_i \log(x_i) - x_i \quad \mathrm{dom}(f) = [0,\infty)^d$

$D_f(x, y) =$ $\quad y_t \in \mathcal{L} = [0,1]^d$

مثال: کاهش آینه‌ای روی سادک

کاهش آینه‌ای

$$a_1 = \mathrm{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

$A = \mathscr{P}_{d-1}$ $\quad f(x) = \sum_i x_i \log(x_i) - x_i \quad \mathrm{dom}(f) = [0, \infty)^d$

$D_f(x, y) = \sum x_i \log(x_i / y_i) + \langle y, 1 \rangle - \langle x, 1 \rangle \quad y_t \in \mathcal{L} = [0, 1]^d$

مثال: کاهش آینه‌ای روی سادک

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a\in\mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$A = \mathscr{P}_{d-1}$ $\quad f(x) = \sum_i x_i \log(x_i) - x_i \quad \mathrm{dom}(f) = [0,\infty)^d$

$$D_f(x,y) = \sum x_i \log(x_i/y_i) + \langle y, 1\rangle - \langle x, 1\rangle \qquad y_t \in \mathcal{L} = [0,1]^d$$

$$a_1 = \frac{1}{d}1$$

مثال: کاهش آینه‌ای روی سادک

کاهش آینه‌ای

$$a_1 = \mathrm{argmin}_{a\in\mathcal{A}}\, F(a)$$

$$a_{t+1} = \mathrm{argmin}_{a\in\mathcal{A}}\left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$$A = \mathscr{P}_{d-1} \qquad f(x) = \textstyle\sum_i x_i \log(x_i) - x_i \qquad \mathrm{dom}(f) = [0,\infty)^d$$

$$D_f(x,y) = \sum x_i \log(x_i/y_i) + \langle y, 1\rangle - \langle x, 1\rangle \qquad y_t \in \mathcal{L} = [0,1]^d$$

$$a_1 = \frac{1}{d}1$$

$$a_{t+1} = \arg\min_a \langle a, \eta y_t\rangle + \langle a, \log(a/a_t) - 1\rangle$$

مثال: کاهش آینه‌ای روی سادک

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

$$A = \mathscr{P}_{d-1} \qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \operatorname{dom}(f) = [0, \infty)^d$$

$$D_f(x, y) = \sum x_i \log(x_i / y_i) + \langle y, 1 \rangle - \langle x, 1 \rangle \qquad y_t \in \mathcal{L} = [0,1]^d$$

$$a_1 = \frac{1}{d} 1$$

$$a_{t+1} = \arg\min_a \langle a, \eta y_t \rangle + \langle a, \log(a / a_t) - 1 \rangle$$

مثال: کاهش آینه‌ای روی سادک

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a\in\mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$$A = \mathscr{P}_{d-1} \qquad f(x) \;=\; \textstyle\sum_i x_i \log(x_i) \;-\; x_i \qquad \operatorname{dom}(f) \;=\; [0,\infty)^d$$

$$D_f(x, y) = \sum x_i \log(x_i/y_i) + \langle y, 1\rangle - \langle x, 1\rangle \qquad y_t \in \mathcal{L} = [0,1]^d$$

$$a_1 = \frac{1}{d}1$$

$$a_{t+1} = \arg\min_a \langle a, \eta y_t\rangle + \langle a, \log(a/a_t) - 1\rangle \quad \Rightarrow \quad \tilde{a}_{t+1} = a_t \exp(-\eta y_t)$$

مثال: کاهش آینه‌ای روی سادک

کاهش آینه‌ای

$$a_1 = \text{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \text{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

$$A = \mathscr{P}_{d-1} \qquad f(x) = \sum_i x_i \log(x_i) - x_i \qquad \text{dom}(f) = [0, \infty)^d$$

$$D_f(x, y) = \sum x_i \log(x_i / y_i) + \langle y, 1 \rangle - \langle x, 1 \rangle \qquad y_t \in \mathcal{L} = [0, 1]^d$$

$$a_1 = \frac{1}{d}1 \qquad \tilde{a}_{t+1} = a_t \exp(-\eta y_t)$$

$$a_{t+1} = \arg\min_a \langle a, \eta y_t \rangle + \langle a, \log(a / a_t) - 1 \rangle \Rightarrow a_{t+1} = \frac{a_t \exp(-\eta y_t)}{\|a_t \exp(-\eta y_t)\|_1}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\leq \quad \boxed{\frac{\log(d)}{\eta}}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \le \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\le \quad \boxed{\frac{\log(d)}{\eta}} \quad \boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P, Q) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\leq \quad \boxed{\frac{\log(d)}{\eta}} \quad \boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty} \quad \boxed{-\frac{1}{\eta} \sum_{t=1}^{n} \frac{1}{2} \|a_t - a_{t+1}\|_1^2}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a)-F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta}\sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P,Q) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\leq \quad \boxed{\frac{\log(d)}{\eta}} \quad \boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty} \quad \boxed{-\frac{1}{\eta}\sum_{t=1}^{n} \frac{1}{2}\|a_t - a_{t+1}\|_1^2}$$

$$\boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty - \frac{1}{\eta}\frac{1}{2}\|a_t - a_{t+1}\|_1^2}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P, Q) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\leq \quad \boxed{\frac{\log(d)}{\eta}} \quad \boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty} \quad \boxed{-\frac{1}{\eta} \sum_{t=1}^{n} \frac{1}{2} \|a_t - a_{t+1}\|_1^2}$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty - \frac{1}{\eta} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \le \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P, Q) \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\le \quad \frac{\log(d)}{\eta} \qquad \sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty \qquad -\frac{1}{\eta} \sum_{t=1}^{n} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty - \frac{1}{\eta} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\le \frac{\log(d)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|y_t\|_\infty^2$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \le \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P, Q) \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\le \quad \boxed{\frac{\log(d)}{\eta}} \quad \boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty} \quad \boxed{-\frac{1}{\eta} \sum_{t=1}^{n} \frac{1}{2} \|a_t - a_{t+1}\|_1^2}$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty - \frac{1}{\eta} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\le \frac{\log(d)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|y_t\|_\infty^2 \quad \le \frac{\log(d)}{\eta} + \frac{\eta n}{2}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a)-F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P,Q) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\leq \quad \boxed{\frac{\log(d)}{\eta}} \quad \boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty} \quad \boxed{-\frac{1}{\eta} \sum_{t=1}^{n} \frac{1}{2} \|a_t - a_{t+1}\|_1^2}$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty - \frac{1}{\eta} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\leq \frac{\log(d)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|y_t\|_\infty^2 \quad \leq \frac{\log(d)}{\eta} + \frac{\eta n}{2} \quad = \sqrt{2n \log(d)}$$

مثال: کاهش آینه‌ای روی سادک

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

$A = \mathscr{P}_{d-1}$ $\quad F(a) = \frac{1}{2}\|a\|_2^2$ $\quad D_f(x, y) = \frac{1}{2}\|a_{t+1} - a_t\|_2^2$ $\quad y_t \in \mathcal{L} = [0, 1]^d$

مثال: کاهش آینه‌ای روی سادک

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

$A = \mathscr{P}_{d-1}$ $\quad F(a) = \frac{1}{2}\|a\|_2^2$ $\quad D_f(x,y) = \frac{1}{2}\|a_{t+1} - a_t\|_2^2$ $\quad y_t \in \mathcal{L} = [0,1]^d$

$$a_1 = \frac{1}{d}1$$

مثال: کاهش آینه‌ای روی سادک

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a\in\mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

$A = \mathscr{P}_{d-1}$ $\quad F(a) = \frac{1}{2}\|a\|_2^2$ $\quad D_f(x, y) = \frac{1}{2}\|a_{t+1} - a_t\|_2^2$ $\quad y_t \in \mathcal{L} = [0,1]^d$

$$a_1 = \frac{1}{d}1$$

$$a_{t+1} = \arg\min_a \langle a, \eta y_t\rangle + \frac{1}{2}\|a - a_t\|^2$$

مثال: کاهش آینه‌ای روی سادک

کاهش آینه‌ای

$$a_1 = \text{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \text{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

$A = \mathscr{P}_{d-1}$ $\quad F(a) = \frac{1}{2}\|a\|_2^2$ $\quad D_f(x, y) = \frac{1}{2}\|a_{t+1} - a_t\|_2^2$ $\quad y_t \in \mathcal{L} = [0, 1]^d$

$$a_1 = \frac{1}{d}1$$

$$a_{t+1} = \arg\min_a \langle a, \eta y_t \rangle + \frac{1}{2}\|a - a_t\|^2$$

مثال: کاهش آینه‌ای روی سادک

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

$A = \mathscr{P}_{d-1}$ $\quad F(a) = \frac{1}{2}\|a\|_2^2$ $\quad D_f(x, y) = \frac{1}{2}\|a_{t+1} - a_t\|_2^2$ $\quad y_t \in \mathcal{L} = [0, 1]^d$

$$a_1 = \frac{1}{d}1$$

$$a_{t+1} = \arg\min_a \langle a, \eta y_t \rangle + \frac{1}{2}\|a - a_t\|^2$$

$$\tilde{a}_{t+1} = a_t - \eta y_t$$

مثال: کاهش آینه‌ای روی سادک

کاهش آینه‌ای

$$a_1 = \text{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \text{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

$A = \mathscr{P}_{d-1}$ $\quad F(a) = \frac{1}{2}\|a\|_2^2$ $\quad D_f(x, y) = \frac{1}{2}\|a_{t+1} - a_t\|_2^2$ $\quad y_t \in \mathcal{L} = [0,1]^d$

$$a_1 = \frac{1}{d}1$$

$$a_{t+1} = \arg\min_a \langle a, \eta y_t \rangle + \frac{1}{2}\|a - a_t\|^2$$

$$\tilde{a}_{t+1} = a_t - \eta y_t$$

$$a_{t+1} = \Pi(\tilde{a}_{t+1})$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P, Q) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\leq \quad \boxed{\frac{\log(d)}{\eta}} \quad \boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty} \quad \boxed{-\frac{1}{\eta} \sum_{t=1}^{n} \frac{1}{2} \|a_t - a_{t+1}\|_1^2}$$

$$\boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty - \frac{1}{\eta} \frac{1}{2} \|a_t - a_{t+1}\|_1^2}$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\leq \frac{\log(d)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|y_t\|_\infty^2 \quad \leq \frac{\log(d)}{\eta} + \frac{\eta n}{2} \quad = \sqrt{2n \log(d)}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P, Q) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\leq \quad \frac{۱}{\eta} \quad \sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty \quad - \frac{1}{\eta} \sum_{t=1}^{n} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty - \frac{1}{\eta} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\leq \frac{\log(d)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|y_t\|_\infty^2 \quad \leq \frac{\log(d)}{\eta} + \frac{\eta n}{2} \quad = \sqrt{2n \log(d)}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \le \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P, Q) \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\le \quad \frac{\text{١}}{\eta} \qquad \sum_{t=1}^{n} \|a_t - a_{t+1}\|_{\text{٢}} \|y_t\|_\infty \qquad - \frac{1}{\eta} \sum_{t=1}^{n} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty - \frac{1}{\eta} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\le \frac{\log(d)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|y_t\|_\infty^2 \quad \le \frac{\log(d)}{\eta} + \frac{\eta n}{2} \quad = \sqrt{2n \log(d)}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P,Q) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\leq \quad \frac{۱}{\eta} \qquad \sum_{t=1}^{n} \|a_t - a_{t+1}\|_{۲} \|y_t\|_{۲} \qquad -\frac{1}{\eta} \sum_{t=1}^{n} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty - \frac{1}{\eta} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\leq \frac{\log(d)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|y_t\|_\infty^2 \quad \leq \frac{\log(d)}{\eta} + \frac{\eta n}{2} \quad = \sqrt{2n \log(d)}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P, Q) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\leq \quad \frac{\text{۱}}{\eta} \qquad \sum_{t=1}^{n} \|a_t - a_{t+1}\|_{\text{۲}} \|y_t\|_{\text{۲}} \qquad - \frac{1}{\eta} \sum_{t=1}^{n} \cancel{\frac{1}{2}} \|a_t - a_{t+1}\|_1^2$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty - \frac{1}{\eta} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\leq \frac{\log(d)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|y_t\|_\infty^2 \quad \leq \frac{\log(d)}{\eta} + \frac{\eta n}{2} \quad = \sqrt{2n \log(d)}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P, Q) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\leq \quad \frac{\text{۱}}{\eta} \quad \sum_{t=1}^{n} \|a_t - a_{t+1}\|_{\text{۲}} \|y_t\|_{\text{۲}} \quad - \frac{1}{\eta} \sum_{t=1}^{n} \cancel{\frac{1}{2}} \|a_t - a_{t+1}\|_{\text{۲}}^2$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty - \frac{1}{\eta} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\leq \frac{\log(d)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|y_t\|_\infty^2 \quad \leq \frac{\log(d)}{\eta} + \frac{\eta n}{2} \quad = \sqrt{2n \log(d)}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \le \frac{F(a)-F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t\rangle - \frac{1}{\eta}\sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P,Q) \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P\|Q)},$$

$$\le \quad \frac{\text{۱}}{\eta} \qquad \sum_{t=1}^{n} \|a_t - a_{t+1}\|_{\text{۲}} \|y_t\|_{\text{۲}} \qquad -\frac{1}{\eta}\sum_{t=1}^{n} \cancel{\frac{1}{2}} \|a_t - a_{t+1}\|_{\text{۲}}^2$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_{\text{۲}} \|y_t\|_{\text{۲}} - \frac{1}{\eta} \cancel{\frac{1}{2}} \|a_t - a_{t+1}\|_{\text{۲}}^2$$

$$\max_{x\in\mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\le \frac{\log(d)}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n} \|y_t\|_\infty^2 \quad \le \frac{\log(d)}{\eta} + \frac{\eta n}{2} \quad = \sqrt{2n\log(d)}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \le \frac{F(a)-F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta}\sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P,Q) \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P\|Q)},$$

$$\le \quad \frac{\text{۱}}{\eta} \quad \sum_{t=1}^{n} \|a_t - a_{t+1}\|_{\text{۲}} \|y_t\|_{\text{۲}} \quad - \frac{1}{\eta}\sum_{t=1}^{n} \cancel{\frac{1}{2}} \|a_t - a_{t+1}\|_{\text{۲}}^2$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_{\text{۲}} \|y_t\|_{\text{۲}} - \frac{1}{\eta} \cancel{\frac{1}{2}} \|a_t - a_{t+1}\|_{\text{۲}}^2$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\le \frac{\text{۱}}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n} \|y_t\|_\infty^2 \quad \le \frac{\log(d)}{\eta} + \frac{\eta n}{2} \quad = \sqrt{2n\log(d)}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \le \frac{F(a)-F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta}\sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P,Q) \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P\|Q)},$$

$$\le \quad \frac{\text{۱}}{\eta} \qquad \sum_{t=1}^{n} \|a_t - a_{t+1}\|_{\text{۲}} \|y_t\|_{\text{۲}} \qquad -\frac{1}{\eta}\sum_{t=1}^{n} \cancel{\frac{1}{2}} \|a_t - a_{t+1}\|_{\text{۲}}^2$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_{\text{۲}} \|y_t\|_{\text{۲}} - \frac{1}{\eta} \cancel{\frac{1}{2}} \|a_t - a_{t+1}\|_{\text{۲}}^2$$

$$\max_{x\in\mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\le \frac{\text{۱}}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n} \|y_t\|_{\text{۲}}^2 \quad \le \frac{\log(d)}{\eta} + \frac{\eta n}{2} \quad = \sqrt{2n\log(d)}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \le \frac{F(a)-F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P,Q) \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P\|Q)},$$

$$\le \quad \boxed{\frac{\text{۱}}{\eta}} \quad \boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_{\text{۲}} \|y_t\|_{\text{۲}}} \quad \boxed{-\frac{1}{\eta} \sum_{t=1}^{n} \cancel{\frac{1}{2}} \|a_t - a_{t+1}\|_{\text{۲}}^2}$$

$$\boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_{\text{۲}} \|y_t\|_{\text{۲}} - \frac{1}{\eta} \cancel{\frac{1}{2}} \|a_t - a_{t+1}\|_{\text{۲}}^2}$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\le \frac{\text{۱}}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|y_t\|_{\text{۲}}^2 \quad \le \frac{\text{۱}}{\eta} + \frac{\eta n}{2} \quad = \sqrt{2n \log(d)}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P, Q) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\leq \quad \frac{\text{۱}}{\eta} \quad \sum_{t=1}^{n} \|a_t - a_{t+1}\|_{\text{۲}} \|y_t\|_{\text{۲}} \quad - \frac{1}{\eta} \sum_{t=1}^{n} \cancel{\frac{1}{2}} \|a_t - a_{t+1}\|_{\text{۲}}^2$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_{\text{۲}} \|y_t\|_{\text{۲}} - \frac{1}{\eta} \cancel{\frac{1}{2}} \|a_t - a_{t+1}\|_{\text{۲}}^2$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\leq \frac{\text{۱}}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|y_t\|_{\text{۲}}^2 \quad \leq \frac{\text{۱}}{\eta} + \frac{\eta n}{2} \quad d = \sqrt{2n \log(d)}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P,Q) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P\|Q)},$$

$$\leq \quad \frac{\text{۱}}{\eta} \qquad \sum_{t=1}^{n} \|a_t - a_{t+1}\|_{\text{۲}} \|y_t\|_{\text{۲}} \qquad -\frac{1}{\eta} \sum_{t=1}^{n} \cancel{\frac{1}{2}} \|a_t - a_{t+1}\|_{\text{۲}}^2$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_{\text{۲}} \|y_t\|_{\text{۲}} - \frac{1}{\eta} \cancel{\frac{1}{2}} \|a_t - a_{t+1}\|_{\text{۲}}^2$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\leq \frac{\text{۱}}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|y_t\|_{\text{۲}}^2 \quad \leq \frac{\text{۱}}{\eta} + \frac{\eta n}{2} d = \cancel{\sqrt{2n\log(d)}} \; O(\sqrt{nd})$$

بسم الله الرحمن الرحیم

یادگیری برخط

جلسه بیستم:
پیروی از پیش‌روی منظم شده
و کاهش آینه‌ای (۳)

مرور

- بهینه‌سازی برخط خطی
- الگوریتم‌های جدید بهینه‌سازی برخط
- بندیت‌سازی
- حل چند مسئله، مثلا بندیت خطی

تابع لژاندر

لژاندر: مانند بشقاب ته‌گود

(a) C is non-empty;

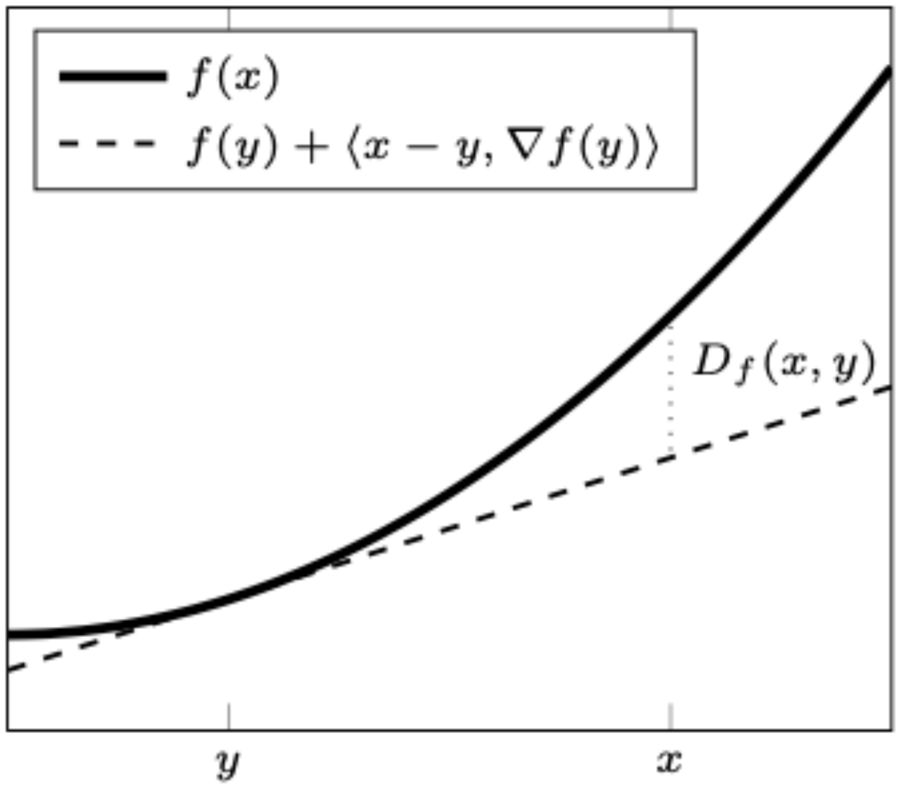
(b) f is differentiable and strictly convex on C; and

(c) $\lim_{n\to\infty} \|\nabla f(x_n)\|_2 = \infty$ for any sequence $(x_n)_n$ with $x_n \in C$ for all n and $\lim_{n\to\infty} x_n = x$ and some $x \in \partial C$.

فاصله D_f

لژاندر: مانند بشقاب ته گود

$$D_f(x,y) = f(x) - f(y) - \langle x - y, \nabla f(y) \rangle$$



دو الگوریتم

کاهش آینه‌ای

$$a_1 = \mathrm{argmin}_{a\in\mathcal{A}} F(a)$$

$$a_{t+1} = \mathrm{argmin}_{a\in\mathcal{A}} \left(\eta\langle a, y_t\rangle + D_F(a, a_t)\right)$$

پیروی از پیش‌روی
منظم شده

$$a_1 = \mathrm{argmin}_{a\in\mathcal{A}} F(a)$$

$$a_{t+1} = \mathrm{argmin}_{a\in\mathcal{A}} \left(\eta \sum_{s=1}^{t}\langle a, y_s\rangle + F(a)\right)$$

مثال: کاهش آینه‌ای

$$A = \mathbb{R}^d \qquad F(a) = \tfrac{1}{2}\|a\|_2^2 \qquad \nabla F(a) = a \qquad D(a, a_t) = \tfrac{1}{2}\|a - a_t\|_2^2$$

$$a_{t+1} = \mathrm{argmin}_{a \in \mathbb{R}^d}\, \eta\langle a, y_t\rangle + \frac{1}{2}\|a - a_t\|_2^2$$

$$\nabla(\dots) = \eta y_t + a - a_t = 0$$

$$a_{t+1} = a_t - \eta y_t$$

کاهش گرادیان برخط

$A =$ محدب

کاهش گرادیان برخط تصویرشده

$$a_{t+1} = \mathrm{argmin}_{a \in \mathcal{A}}\, \eta\langle a, y_t\rangle + \frac{1}{2}\|a - a_t\|_2^2 \;\; = \Pi(a_t - \eta y_t)$$

مثال: پیروی از پیش‌روی منظم شده

پیروی از پیش‌روی منظم شده

$$a_1 = \mathrm{argmin}_{a\in\mathcal{A}} F(a)$$

$$a_{t+1} = \mathrm{argmin}_{a\in\mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s\rangle + F(a)\right)$$

$A = \mathscr{P}_{d-1}$ $\quad f(x) = \sum_i x_i \log(x_i) - x_i \quad \mathrm{dom}(f) = [0,\infty)^d$

$$a_{t+1,i} = \frac{\exp\left(-\eta \sum_{s=1}^{t} y_{si}\right)}{\sum_{j=1}^{d} \exp\left(-\eta \sum_{s=1}^{t} y_{sj}\right)}.$$

قضیه: در شرایط خوب با تابع لژاندر f

$$y = \operatorname{argmin}_{z \in A} f(z)$$

$$\boxed{\tilde{y} = \operatorname{argmin}_{z \in \mathbb{R}^d} f(z)}$$

$$y = \operatorname{argmin}_{z \in A} D_f(z, \tilde{y}).$$

تحلیل پیشمانی کاهش آینه‌ای

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

قضیه:

$$R_n(a) \le \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$R_n(a) \le \frac{1}{\eta} \left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1}) \right)$$

$$\tilde{a}_{t+1} = \operatorname{argmin}_{a \in \mathcal{D}} \eta \langle a, y_t \rangle + D_F(a, a_t)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} D_F(a, \tilde{a}_{t+1}).$$

تحلیل پشیمانی پیروی از پیش‌روی منظم شده

پیروی از پیش‌روی منظم شده

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$$

قضیه:

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

مثال (گوی واحد)

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $\quad F(a) = \frac{1}{2}\|a\|_2^2 \quad \eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

مثال (گوی واحد)

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $\quad F(a) = \frac{1}{2}\|a\|_2^2 \quad \eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

مثال (گوی واحد)

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $F(a) = \frac{1}{2}\|a\|_2^2$ $\eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2$$

مثال (گوی واحد)

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $F(a) = \frac{1}{2}\|a\|_2^2$ $\quad \eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2 = \frac{\eta^2}{2}\|y_t\|_2^2$$

مثال (گوی واحد)

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $F(a) = \frac{1}{2}\|a\|_2^2$ $\eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2 = \frac{\eta^2}{2}\|y_t\|_2^2$$

$$R_n \leq \frac{\mathrm{diam}_F(\mathcal{A})}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n}\|y_t\|_2^2$$

مثال (گوی واحد)

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $F(a) = \frac{1}{2}\|a\|_2^2$ $\eta = \sqrt{1/n}$

آنگاه $R_n \leq \sqrt{n}$.

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2 = \frac{\eta^2}{2}\|y_t\|_2^2$$

$$R_n \leq \frac{\mathrm{diam}_F(\mathcal{A})}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n}\|y_t\|_2^2 \leq \frac{1}{2\eta} + \frac{\eta n}{2}$$

مثال (گوی واحد)

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

الگوریتم کاهش آینه‌ای $\eta = \sqrt{1/n}$ $\quad F(a) = \frac{1}{2}\|a\|_2^2$

آنگاه $R_n \leq \sqrt{n}.$

$$R_n(a) \leq \frac{1}{\eta}\left(F(a) - F(a_1) + \sum_{t=1}^{n} D(a_t, \tilde{a}_{t+1})\right)$$

$$D(a_t, \tilde{a}_{t+1}) = \frac{1}{2}\|\tilde{a}_{t+1} - a_t\|_2^2 = \frac{\eta^2}{2}\|y_t\|_2^2$$

$$R_n \leq \frac{\operatorname{diam}_F(\mathcal{A})}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n}\|y_t\|_2^2 \leq \frac{1}{2\eta} + \frac{\eta n}{2} = \sqrt{n}.$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\leq \quad \boxed{\frac{\log(d)}{\eta}}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \le \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\le \quad \boxed{\frac{\log(d)}{\eta}} \quad \boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P,Q) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\leq \boxed{\frac{\log(d)}{\eta}} \quad \boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty} \quad \boxed{-\frac{1}{\eta} \sum_{t=1}^{n} \frac{1}{2} \|a_t - a_{t+1}\|_1^2}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \le \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P,Q) \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\le \quad \boxed{\frac{\log(d)}{\eta}} \quad \boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty} \quad \boxed{-\frac{1}{\eta} \sum_{t=1}^{n} \frac{1}{2} \|a_t - a_{t+1}\|_1^2}$$

$$\boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty - \frac{1}{\eta} \frac{1}{2} \|a_t - a_{t+1}\|_1^2}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P,Q) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\leq \quad \boxed{\frac{\log(d)}{\eta}} \quad \boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty} \quad \boxed{-\frac{1}{\eta}\sum_{t=1}^{n} \frac{1}{2} \|a_t - a_{t+1}\|_1^2}$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty - \frac{1}{\eta} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P, Q) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\leq \quad \boxed{\frac{\log(d)}{\eta}} \quad \boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty} \quad \boxed{-\frac{1}{\eta} \sum_{t=1}^{n} \frac{1}{2} \|a_t - a_{t+1}\|_1^2}$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty - \frac{1}{\eta} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\leq \frac{\log(d)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|y_t\|_\infty^2$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P, Q) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\leq \qquad \frac{\log(d)}{\eta} \qquad \sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty \qquad -\frac{1}{\eta} \sum_{t=1}^{n} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty - \frac{1}{\eta} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\leq \frac{\log(d)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|y_t\|_\infty^2 \quad \leq \frac{\log(d)}{\eta} + \frac{\eta n}{2}$$

مثال کاهش آینه‌ای روی سادک

$$R_n(a) \le \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

$$\delta(P, Q) \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \| Q)},$$

$$\le \quad \boxed{\frac{\log(d)}{\eta}} \quad \boxed{\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty} \quad \boxed{-\frac{1}{\eta} \sum_{t=1}^{n} \frac{1}{2} \|a_t - a_{t+1}\|_1^2}$$

$$\sum_{t=1}^{n} \|a_t - a_{t+1}\|_1 \|y_t\|_\infty - \frac{1}{\eta} \frac{1}{2} \|a_t - a_{t+1}\|_1^2$$

$$\max_{x \in \mathbb{R}} ax - bx^2/2 = a^2/(2b).$$

$$\le \frac{\log(d)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \|y_t\|_\infty^2 \quad \le \frac{\log(d)}{\eta} + \frac{\eta n}{2} \quad = \sqrt{2n \log(d)}$$

نسخه بندیتی

بهینه‌سازی خطی برخط

$\mathcal{L} \subset \mathbb{R}^d \quad \mathcal{A} \subset \mathbb{R}^d$

تفاوت با یادگیری برخط
$y_t \in L$

۱ـ انتخاب عمل $A_t \in A$
۲ـ دریافت $\langle A_t, y_t \rangle$

$$R_n(a) = \mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - a, y_t \rangle\right]$$

$$R_n = \max_{a \in \mathcal{A}} R_n(a)$$

چگونه تغییر بدهیم؟

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

چگونه تغییر بدهیم؟

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

y_t؟

چگونه تغییر بدهیم؟

کاهش آینه‌ای

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \langle a, y_t \rangle + D_F(a, a_t)\right)$$

تصادف؟

y_t؟

Input Legendre potential F, action set $\mathcal{A}$ and learning rate $\eta > 0$

Choose $\bar{A}_1 = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} F(a)$

for $t = 1, \ldots, n$ **do**

Choose measure P_t on $\mathcal{A}$ with mean $\bar{A}_t$

Sample action A_t from P_t and observe $\langle A_t, y_t \rangle$

Compute estimate $\hat{Y}_t$ of the loss vector y_t

Update:

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \langle a, \hat{Y}_t \rangle + D_F(a, \bar{A}_t) \qquad \text{(Mirror descent)}$$

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \sum_{s=1}^{t} \langle a, \hat{Y}_s \rangle + F(a) \quad \text{(follow-the-regularised-leader)}$$

end for

Input Legendre potential F, action set $\mathcal{A}$ and learning rate $\eta > 0$

Choose $\bar{A}_1 = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} F(a)$

for $t = 1, \ldots, n$ **do**

  Choose measure P_t on $\mathcal{A}$ with mean $\bar{A}_t$

  Sample action A_t from P_t and observe $\langle A_t, y_t \rangle$

  Compute estimate $\hat{Y}_t$ of the loss vector y_t

  Update:

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \langle a, \hat{Y}_t \rangle + D_F(a, \bar{A}_t) \qquad \text{(Mirror descent)}$$

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \sum_{s=1}^{t} \langle a, \hat{Y}_s \rangle + F(a) \qquad \text{(follow-the-regularised-leader)}$$

end for

Input Legendre potential F, action set $\mathcal{A}$ and learning rate $\eta > 0$

Choose $\bar{A}_1 = \operatorname{argmin}_{a \in \mathcal{A} \cap \mathrm{dom}(F)} F(a)$

for $t = 1, \ldots, n$ **do**

Choose measure P_t on $\mathcal{A}$ with mean $\bar{A}_t$

Sample action A_t from P_t and observe $\langle A_t, y_t \rangle$

Compute estimate $\hat{Y}_t$ of the loss vector y_t

Update:

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \mathrm{dom}(F)} \eta \langle a, \hat{Y}_t \rangle + D_F(a, \bar{A}_t) \qquad \text{(Mirror descent)}$$

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \mathrm{dom}(F)} \eta \sum_{s=1}^{t} \langle a, \hat{Y}_s \rangle + F(a) \quad \text{(follow-the-regularised-leader)}$$

end for

تحلیل ـ ازکران پشیمانی کاهش آینه‌ای

قضیه: پشیمانی کاهش آینه‌ای

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

تحلیل ـ ازکران پشیمانی کاهش آینه‌ای

قضیه: پشیمانی کاهش آینه‌ای

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

امید؟

تحلیل ـ از کران پشیمانی کاهش آینه‌ای

قضیه: پشیمانی کاهش آینه‌ای

$$R_n(a) \le \frac{F(a)-F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta}\sum_{t=1}^{n} D(a_{t+1}, a_t).$$

امید؟

a_t؟

تحلیل ـ از کران پشیمانی کاهش آینه‌ای

قضیه: پشیمانی کاهش آینه‌ای

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n} \langle a_t - a_{t+1}, y_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(a_{t+1}, a_t).$$

امید؟

a_t؟

$E[\langle a_t, y_t \rangle] = E[\langle \bar{A}_t, y_t \rangle]$؟

تحلیل

نااریبی

$$\mathbb{E}\left[\langle A_t, y_t\rangle\right] = \mathbb{E}\left[\langle \bar{A}_t, y_t\rangle\right] = \mathbb{E}\left[\mathbb{E}\left[\langle \bar{A}_t, y_t\rangle \mid \bar{A}_t\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\langle \bar{A}_t, \hat{Y}_t\rangle \,\middle|\, \bar{A}_t\right]\right]$$

تحلیل

ناااریبی

$$\mathbb{E}\left[\langle A_t, y_t\rangle\right] = \mathbb{E}\left[\langle \bar{A}_t, y_t\rangle\right] = \mathbb{E}\left[\mathbb{E}\left[\langle \bar{A}_t, y_t\rangle \mid \bar{A}_t\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\langle \bar{A}_t, \hat{Y}_t\rangle \,\middle|\, \bar{A}_t\right]\right]$$

$$R_n(a) = \mathbb{E}\left[\sum_{t=1}^{n}\langle A_t, y_t\rangle - \langle a, y_t\rangle\right] = \mathbb{E}\left[\sum_{t=1}^{n}\langle \bar{A}_t - a, \hat{Y}_t\rangle\right]$$

تحلیل

ناااریبی

$$\mathbb{E}\left[\langle A_t, y_t\rangle\right] = \mathbb{E}\left[\langle \bar{A}_t, y_t\rangle\right] = \mathbb{E}\left[\mathbb{E}\left[\langle \bar{A}_t, y_t\rangle \,|\, \bar{A}_t\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\langle \bar{A}_t, \hat{Y}_t\rangle \,\middle|\, \bar{A}_t\right]\right]$$

$$R_n(a) = \mathbb{E}\left[\sum_{t=1}^{n}\langle A_t, y_t\rangle - \langle a, y_t\rangle\right] = \mathbb{E}\left[\sum_{t=1}^{n}\langle \bar{A}_t - a, \hat{Y}_t\rangle\right]$$

اگر بردارهای a_t داشته باشیم که با الگوریتم کاهش آینه‌ای براساس y_t تغییر کنند، داریم:

قضیه: پشیمانی کاهش آینه‌ای

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n}\langle a_t - a_{t+1}, y_t\rangle - \frac{1}{\eta}\sum_{t=1}^{n} D(a_{t+1}, a_t).$$

تحلیل

ناااریبی

$$\mathbb{E}\left[\langle A_t, y_t\rangle\right] = \mathbb{E}\left[\langle \bar{A}_t, y_t\rangle\right] = \mathbb{E}\left[\mathbb{E}\left[\langle \bar{A}_t, y_t\rangle \mid \bar{A}_t\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\langle \bar{A}_t, \hat{Y}_t\rangle \,\middle|\, \bar{A}_t\right]\right]$$

$$R_n(a) = \mathbb{E}\left[\sum_{t=1}^{n}\langle A_t, y_t\rangle - \langle a, y_t\rangle\right] = \mathbb{E}\left[\sum_{t=1}^{n}\langle \bar{A}_t - a, \hat{Y}_t\rangle\right]$$

$\tilde{Y}_t$ $\bar{A}_t$

اگر بردارهای a_t داشته باشیم که با الگوریتم کاهش آینه‌ای براساس y_t تغییر کنند، داریم:

قضیه: پشیمانی کاهش آینه‌ای

$$R_n(a) \leq \frac{F(a) - F(a_1)}{\eta} + \sum_{t=1}^{n}\langle a_t - a_{t+1}, y_t\rangle - \frac{1}{\eta}\sum_{t=1}^{n} D(a_{t+1}, a_t).$$

تحلیل پشیمانی دو الگوریتم بندیت

$$R_n(a) \leq \mathbb{E}\left[\frac{F(a) - F(\bar{A}_1)}{\eta} + \sum_{t=1}^{n} \langle \bar{A}_t - \bar{A}_{t+1}, \hat{Y}_t \rangle - \frac{1}{\eta} \sum_{t=1}^{n} D(\bar{A}_{t+1}, \bar{A}_t)\right]$$

تحلیل پشیمانی دو الگوریتم بندیت

$$R_n(a) \leq \mathbb{E}\left[\frac{F(a) - F(\bar{A}_1)}{\eta} + \sum_{t=1}^{n} \langle \bar{A}_t - \bar{A}_{t+1}, \hat{Y}_t \rangle - \frac{1}{\eta}\sum_{t=1}^{n} D(\bar{A}_{t+1}, \bar{A}_t)\right]$$

$$R_n \leq \frac{\mathrm{diam}_F(\mathcal{A})}{\eta} + \frac{1}{\eta}\sum_{t=1}^{n} \mathbb{E}\left[D(\bar{A}_t, \tilde{A}_{t+1})\right]$$

$$\tilde{A}_{t+1} = \mathrm{argmin}_{a \in \mathrm{dom}(F)}\, \eta\langle a, \hat{Y}_t \rangle + D_F(a, \bar{A}_t)$$

مثال: بندیت خطی روی گوی واحد

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

مثال: بندیت خطی روی گوی واحد

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

$$P_t = (1-\gamma)\tilde{P}_t + \gamma\pi,$$

$$\tilde{P}_t(B) = \frac{\int_B \exp\left(-\eta \sum_{s=1}^{t-1} \hat{Y}_s(a)\right) da}{\int_{\mathcal{A}} \exp\left(-\eta \sum_{s=1}^{t-1} \hat{Y}_s(a)\right) da}.$$

مثال: بندیت خطی روی گوی واحد

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$$

$$P_t = (1-\gamma)\tilde{P}_t + \gamma\pi,$$

$$\tilde{P}_t(B) = \frac{\int_B \exp\left(-\eta \sum_{s=1}^{t-1} \hat{Y}_s(a)\right) da}{\int_{\mathcal{A}} \exp\left(-\eta \sum_{s=1}^{t-1} \hat{Y}_s(a)\right) da}.$$

$$R_n \leq 2d\sqrt{3n(1+\log_+(2n/d))}.$$

مثال: بندیت خطی روی گوی واحد

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|_2 \le 1\}$$

Input Legendre potential F, action set $\mathcal{A}$ and learning rate $\eta > 0$
Choose $\bar{A}_1 = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} F(a)$
for $t = 1, \ldots, n$ **do**
  Choose measure P_t on $\mathcal{A}$ with mean $\bar{A}_t$
  Sample action A_t from P_t and observe $\langle A_t, y_t \rangle$
  Compute estimate $\hat{Y}_t$ of the loss vector y_t
  Update:

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \langle a, \hat{Y}_t \rangle + D_F(a, \bar{A}_t) \quad \text{(Mirror descent)}$$

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \sum_{s=1}^{t} \langle a, \hat{Y}_s \rangle + F(a) \quad \text{(follow-the-regularised-leader)}$$

end for

مثال: بندیت خطی روی گوی واحد

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|$$

$$E_t \in \{0,1\}$$

$$\mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$$

$$U_t \sim \{\pm e_1, \dots, \pm e_d\}$$

$$A_t = E_t U_t + \frac{(1-E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

Input Legendre potential F, action set $\mathcal{A}$ and learning

Choose $\bar{A}_1 = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} F(a)$

for $t = 1, \dots, n$ **do**

  Choose measure P_t on $\mathcal{A}$ with mean $\bar{A}_t$

  Sample action A_t from P_t and observe $\langle A_t, y_t \rangle$

  Compute estimate $\hat{Y}_t$ of the loss vector y_t

  Update:

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \langle a, \hat{Y}_t \rangle + D_F(a, \bar{A}_t) \qquad \text{(Mirror descent)}$$

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \sum_{s=1}^{t} \langle a, \hat{Y}_s \rangle + F(a) \qquad \text{(follow-the-regularised-leader)}$$

end for

مثال: بندیت خطی روی گوی واحد

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|$$

Input Legendre potential F, action set $\mathcal{A}$ and learning

Choose $\bar{A}_1 = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} F(a)$

for $t = 1, \dots, n$ **do**

Choose measure P_t on $\mathcal{A}$ with mean $\bar{A}_t$

Sample action A_t from P_t and observe $\langle A_t, y_t \rangle$

Compute estimate $\hat{Y}_t$ of the loss vector y_t

Update:

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \langle a, \hat{Y}_t \rangle + D_F(a, \bar{A}_t)$$

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \sum_{s=1}^{t} \langle a, \hat{Y}_s \rangle + F(a) \quad \text{(follow-the-regularised-leader)}$$

end for

$$E_t \in \{0, 1\}$$

$$\mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$$

$$U_t \sim \{\pm e_1, \dots, \pm e_d\}$$

$$A_t = E_t U_t + \frac{(1 - E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

$$\hat{Y}_t = \frac{d\, E_t\, A_t \langle A_t, y_t \rangle}{1 - \|\bar{A}_t\|}$$

$$\mathbb{E}_t[A_t] = \bar{A}_t$$

$$E_t \in \{0, 1\}$$

$$\mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$$

$$U_t \sim \{\pm e_1, \dots, \pm e_d\}$$

$$A_t = E_t U_t + \frac{(1 - E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

$$\hat{Y}_t = \frac{d\, E_t\, A_t \langle A_t, y_t \rangle}{1 - \|\bar{A}_t\|}$$

$$\mathbb{E}_t[A_t] = \bar{A}_t$$

$$E_t \in \{0, 1\}$$

$$\mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$$

$$U_t \sim \{\pm e_1, \ldots, \pm e_d\}$$

$$A_t = E_t U_t + \frac{(1 - E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

$$E[\hat{Y}_t]$$

$$\hat{Y}_t = \frac{d\, E_t\, A_t \langle A_t, y_t \rangle}{1 - \|\bar{A}_t\|}$$

$$\mathbb{E}_t[A_t] = \bar{A}_t$$

$$E_t \in \{0,1\}$$

$$\mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$$

$$U_t \sim \{\pm e_1, \ldots, \pm e_d\}$$

$$A_t = E_t U_t + \frac{(1-E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

$$E[\hat{Y}_t] = d \cdot E[A_t \langle A_t, y_t \rangle]$$

$$\hat{Y}_t = \frac{d\, E_t\, A_t \langle A_t, y_t \rangle}{1 - \|\bar{A}_t\|}$$

$$\mathbb{E}_t[A_t] = \bar{A}_t$$

$$E_t \in \{0, 1\}$$

$$\mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$$

$$U_t \sim \{\pm e_1, \ldots, \pm e_d\}$$

$$A_t = E_t U_t + \frac{(1 - E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

$$\begin{aligned} E[\hat{Y}_t] &= d \cdot E[A_t \langle A_t, y_t \rangle] \\ &= d \cdot \frac{1}{2d} \sum_i e_i \langle e_i, y_t \rangle - e_i \langle -e_i, y_t \rangle \end{aligned}$$

$$\hat{Y}_t = \frac{d\, E_t\, A_t \langle A_t, y_t \rangle}{1 - \|\bar{A}_t\|}$$

$$\mathbb{E}_t[A_t] = \bar{A}_t$$

$$E_t \in \{0,1\}$$

$$\mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$$

$$U_t \sim \{\pm e_1, \ldots, \pm e_d\}$$

$$A_t = E_t U_t + \frac{(1-E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

$$\hat{Y}_t = \frac{d\, E_t\, A_t \langle A_t, y_t \rangle}{1 - \|\bar{A}_t\|}$$

$$\begin{aligned} E[\hat{Y}_t] &= d \cdot E[A_t \langle A_t, y_t \rangle] \\ &= d \cdot \frac{1}{2d} \sum_i e_i \langle e_i, y_t \rangle - e_i \langle -e_i, y_t \rangle \\ &= d \cdot \frac{1}{2d} 2y_t \end{aligned}$$

مثال: بندیت خطی روی گوی واحد

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|$$

$$\tilde{A} = \{x \in \mathbb{R}^d : \|x\|_2 \leq r\}$$

$$E_t \in \{0,1\}$$

$$\mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$$

$$U_t \sim \{\pm e_1, \ldots, \pm e_d\}$$

$$A_t = E_t U_t + \frac{(1-E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

$$\hat{Y}_t = \frac{d\, E_t\, A_t \langle A_t, y_t \rangle}{1 - \|\bar{A}_t\|}$$

Input Legendre potential F, action set $\mathcal{A}$ and learning

Choose $\bar{A}_1 = \operatorname{argmin}_{a \in \mathcal{A} \cap \mathrm{dom}(F)} F(a)$

for $t = 1, \ldots, n$ **do**

Choose measure P_t on $\mathcal{A}$ with mean $\bar{A}_t$

Sample action A_t from P_t and observe $\langle A_t, y_t \rangle$

Compute estimate $\hat{Y}_t$ of the loss vector y_t

Update:

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \mathrm{dom}(F)} \eta \langle a, \hat{Y}_t \rangle + D_F(a, \bar{A}_t)$$

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \mathrm{dom}(F)} \eta \sum_{s=1}^{t} \langle a, \hat{Y}_s \rangle + F(a) \quad \text{(follow-the-regularised-leader)}$$

end for

مثال: بندیت خطی روی گوی واحد

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|$$

$$\tilde{\mathcal{A}} = \{x \in \mathbb{R}^d : \|x\|_2 \leq r\}$$

$$E_t \in \{0, 1\}$$

$$\mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$$

$$U_t \sim \{\pm e_1, \dots, \pm e_d\}$$

$$A_t = E_t U_t + \frac{(1 - E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

Input Legendre potential F, action set $\mathcal{A}$ and learning

Choose $\bar{A}_1 = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} F(a)$

for $t = 1, \dots, n$ **do**

  Choose measure P_t on $\mathcal{A}$ with mean $\bar{A}_t$

  Sample action A_t from P_t and observe $\langle A_t, y_t \rangle$

  Compute estimate $\hat{Y}_t$ of the loss vector y_t

  Update:

$$\hat{Y}_t = \frac{d\, E_t\, A_t \langle A_t, y_t \rangle}{1 - \|\bar{A}_t\|}$$

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \langle a, \hat{Y}_t \rangle + D_F(a, \bar{A}_t)$$ (Mirror descent)

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \sum_{s=1}^{t} \langle a, \hat{Y}_s \rangle + F(a)$$ (follow-the-regularised-leader)

end for

مثال: بندیت خطی روی گوی واحد

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|$$

$$\tilde{\mathcal{A}} = \{x \in \mathbb{R}^d : \|x\|_2 \leq r\}$$

$$E_t \in \{0,1\}$$

$$\mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$$

$$U_t \sim \{\pm e_1, \ldots, \pm e_d\}$$

$$A_t = E_t U_t + \frac{(1-E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

$$\hat{Y}_t = \frac{d\, E_t\, A_t \langle A_t, y_t\rangle}{1 - \|\bar{A}_t\|}$$

Input Legendre potential F, action set $\mathcal{A}$ and learning

Choose $\bar{A}_1 = \mathrm{argmin}_{a \in \mathcal{A} \cap \mathrm{dom}(F)} F(a)$

for $t = 1, \ldots, n$ **do**

Choose measure P_t on $\mathcal{A}$ with mean $\bar{A}_t$

Sample action A_t from P_t and observe $\langle A_t, y_t\rangle$

Compute estimate $\hat{Y}_t$ of the loss vector y_t

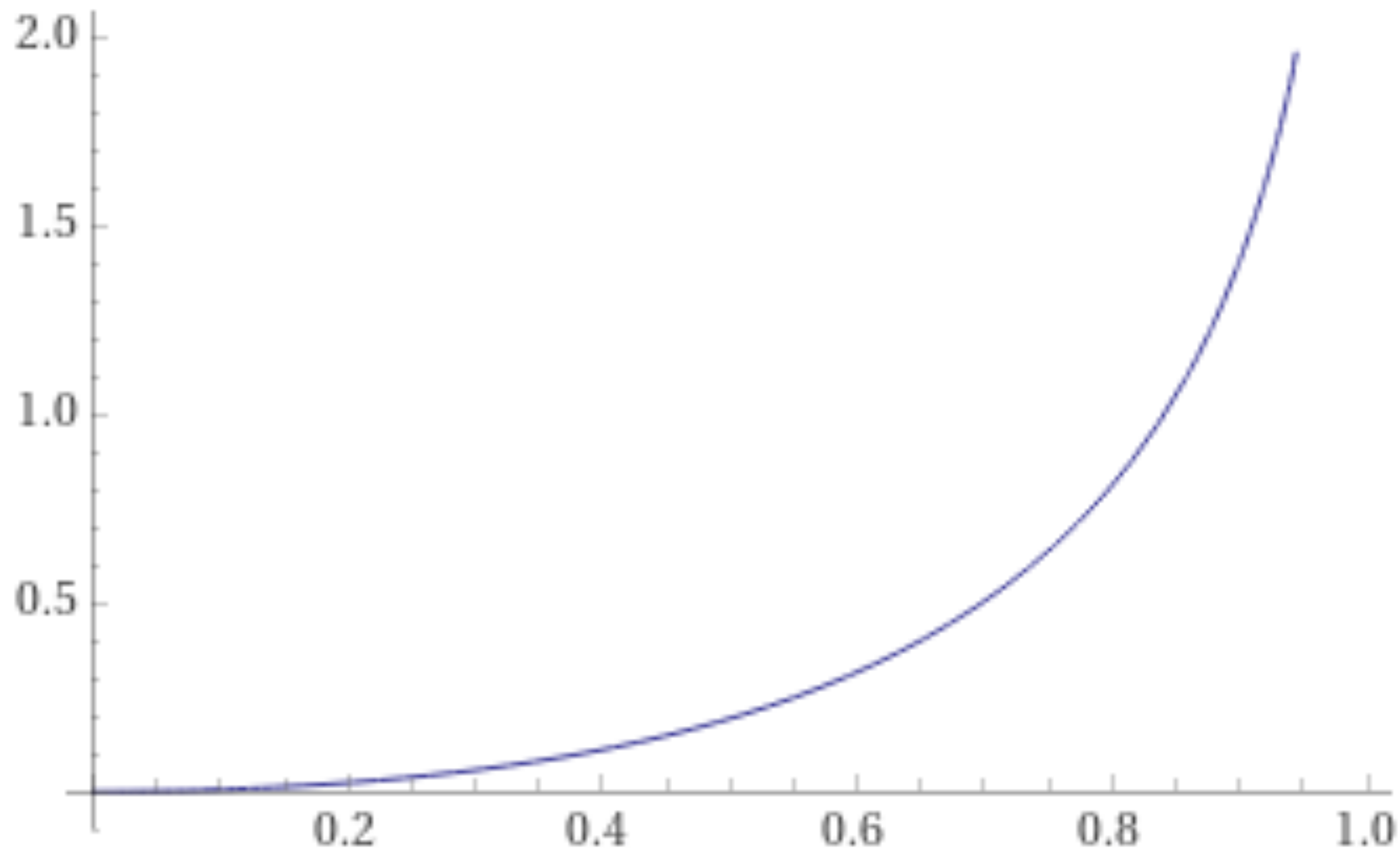
Update:

$$\bar{A}_{t+1} = \mathrm{argmin}_{a \in \mathcal{A} \cap \mathrm{dom}(F)}\ \eta\langle a, \hat{Y}_t\rangle + D_F(a, \bar{A}_t)$$

$$\bar{A}_{t+1} = \mathrm{argmin}_{a \in \mathcal{A} \cap \mathrm{dom}(F)}\ \eta \sum_{s=1}^{t} \langle a, \hat{Y}_s\rangle + F(a) \quad \text{(follow-the-regularised-leader)}$$

$$F(a) = -\log(1 - \|a\|) - \|a\|$$

end for



$$F(a) = -\log(1 - \|a\|) - \|a\|$$

پیروی از پیش‌روی منظم شده بندیتی روی گوی واحد

پیروی از پیش‌روی منظم شده

$$F(a) = -\log(1 - \|a\|) - \|a\|$$

$$a_1 = \operatorname{argmin}_{a \in \mathcal{A}} F(a)$$

$$a_{t+1} = \operatorname{argmin}_{a \in \mathcal{A}} \left(\eta \sum_{s=1}^{t} \langle a, y_s \rangle + F(a) \right)$$

$$\arg\min_{x \in \tilde{A}} \langle a, \eta \sum \hat{Y}_s \rangle - \log(1 - \|a\|) - \|a\|$$

$$\arg\min_{x\in\tilde{A}} \langle a, \eta \sum \hat{Y}_s \rangle - \log(1 - \|a\|) - \|a\|$$

شاید نقطه بهینه در قیود صدق کند:

$$\arg\min_{x\in\tilde{A}} \langle a, \eta \sum \hat{Y}_s \rangle - \log(1 - \|a\|) - \|a\|$$

شاید نقطه بهینه در قیود صدق کند:

$$\nabla(\ldots) = \eta \sum \hat{Y}_s - \frac{1}{1 - \|a\|} \times \frac{-a}{\|a\|} - \frac{a}{\|a\|}$$

$$\arg\min_{x\in\tilde{A}}\langle a, \eta \sum \hat{Y}_s\rangle - \log(1 - \|a\|) - \|a\|$$

شاید نقطه بهینه در قیود صدق کند:

$$\begin{aligned}\nabla(\ldots) &= \eta \sum \hat{Y}_s - \frac{1}{1 - \|a\|} \times \frac{-a}{\|a\|} - \frac{a}{\|a\|} \\ &= \eta \sum \hat{Y}_s + \frac{a}{\|a\|}(\frac{1}{1 - \|a\|} - 1)\end{aligned}$$

$$\arg\min_{x\in\tilde{A}} \langle a, \eta \sum \hat{Y}_s \rangle - \log(1 - \|a\|) - \|a\|$$

شاید نقطه بهینه در قیود صدق کند:

$$\nabla(\ldots) = \eta \sum \hat{Y}_s - \frac{1}{1-\|a\|} \times \frac{-a}{\|a\|} - \frac{a}{\|a\|}$$

$$= \eta \sum \hat{Y}_s + \frac{a}{\|a\|}\left(\frac{1}{1-\|a\|} - 1\right) = \eta \sum \hat{Y}_s + \frac{a}{1-\|a\|}$$

$$\arg\min_{x\in\tilde{A}} \langle a, \eta \sum \hat{Y}_s \rangle - \log(1 - \|a\|) - \|a\|$$

شاید نقطه بهینه در قیود صدق کند:

$$\nabla F(a) = \frac{a}{1-\|a\|}$$

$$\nabla(\ldots) = \eta \sum \hat{Y}_s - \frac{1}{1-\|a\|} \times \frac{-a}{\|a\|} - \frac{a}{\|a\|}$$

$$= \eta \sum \hat{Y}_s + \frac{a}{\|a\|}\left(\frac{1}{1-\|a\|} - 1\right) = \eta \sum \hat{Y}_s + \frac{a}{1-\|a\|}$$

$$\arg\min_{x\in\tilde{A}} \langle a, \eta \sum \hat{Y}_s \rangle - \log(1 - \|a\|) - \|a\|$$

شاید نقطه بهینه در قیود صدق کند:

$$\nabla F(a) = \frac{a}{1 - \|a\|}$$

$$\nabla(\ldots) = \eta \sum \hat{Y}_s - \frac{1}{1-\|a\|} \times \frac{-a}{\|a\|} - \frac{a}{\|a\|}$$

$$= \eta \sum \hat{Y}_s + \frac{a}{\|a\|}\left(\frac{1}{1-\|a\|} - 1\right) = \eta \sum \hat{Y}_s + \frac{a}{1-\|a\|}$$

$$\nabla(\ldots) = 0$$

$$\arg\min_{x\in\tilde{A}} \langle a, \eta \sum \hat{Y}_s \rangle - \log(1 - \|a\|) - \|a\|$$

شاید نقطه بهینه در قیود صدق کند:

$$\nabla F(a) = \frac{a}{1 - \|a\|}$$

$$\nabla(\ldots) = \eta \sum \hat{Y}_s - \frac{1}{1 - \|a\|} \times \frac{-a}{\|a\|} - \frac{a}{\|a\|}$$

$$= \eta \sum \hat{Y}_s + \frac{a}{\|a\|}\left(\frac{1}{1 - \|a\|} - 1\right) = \eta \sum \hat{Y}_s + \frac{a}{1 - \|a\|}$$

$$\nabla(\ldots) = 0 \Rightarrow \frac{a}{1 - \|a\|} = -\eta \sum \hat{Y}_s$$

$$\arg\min_{x\in\tilde{A}} \langle a, \eta \sum \hat{Y}_s \rangle - \log(1 - \|a\|) - \|a\|$$

شاید نقطه بهینه در قیود صدق کند:

$$\nabla F(a) = \frac{a}{1 - \|a\|}$$

$$\nabla(\ldots) = \eta \sum \hat{Y}_s - \frac{1}{1 - \|a\|} \times \frac{-a}{\|a\|} - \frac{a}{\|a\|}$$

$$= \eta \sum \hat{Y}_s + \frac{a}{\|a\|}\left(\frac{1}{1 - \|a\|} - 1\right) = \eta \sum \hat{Y}_s + \frac{a}{1 - \|a\|}$$

$$\nabla(\ldots) = 0 \Rightarrow \frac{a}{1 - \|a\|} = -\eta \sum \hat{Y}_s$$

a موازی $-\eta \sum \hat{Y}_s$
با اندازه‌ای مناسب

$$\arg\min_{x\in\tilde{A}} \langle a, \eta \sum \hat{Y}_s \rangle - \log(1-\|a\|) - \|a\|$$

شاید نقطه بهینه در قیود صدق کند:

$$\nabla F(a) = \frac{a}{1-\|a\|}$$

$$\nabla(\ldots) = \eta \sum \hat{Y}_s - \frac{1}{1-\|a\|} \times \frac{-a}{\|a\|} - \frac{a}{\|a\|}$$

$$= \eta \sum \hat{Y}_s + \frac{a}{\|a\|}\left(\frac{1}{1-\|a\|} - 1\right) = \eta \sum \hat{Y}_s + \frac{a}{1-\|a\|}$$

$$\nabla(\ldots) = 0 \Rightarrow \frac{a}{1-\|a\|} = -\eta \sum \hat{Y}_s$$

a موازی $-\eta \sum \hat{Y}_s$

با اندازه‌ای مناسب

$$\frac{\|a\|}{1-\|a\|} = \eta \|\sum \hat{Y}_s\|$$

$$\arg\min_{x\in\tilde{A}} \langle a, \eta \sum \hat{Y}_s \rangle - \log(1 - \|a\|) - \|a\|$$

شاید نقطه بهینه در قیود صدق کند:

$$\nabla(\ldots) = \eta \sum \hat{Y}_s - \frac{1}{1-\|a\|} \times \frac{-a}{\|a\|} - \frac{a}{\|a\|}$$

$$\nabla F(a) = \frac{a}{1-\|a\|}$$

$$= \eta \sum \hat{Y}_s + \frac{a}{\|a\|}\left(\frac{1}{1-\|a\|} - 1\right) = \eta \sum \hat{Y}_s + \frac{a}{1-\|a\|}$$

$$\nabla(\ldots) = 0 \Rightarrow \frac{a}{1-\|a\|} = -\eta \sum \hat{Y}_s$$

a موازی $-\eta \sum \hat{Y}_s$
با اندازه‌ای مناسب

$$\frac{\|a\|}{1-\|a\|} = \eta\| \sum \hat{Y}_s\| \Rightarrow \frac{\|a\|}{1-\|a\|} = \eta\| \sum \hat{Y}_s\|$$

$$\arg\min_{x\in\tilde{A}} \langle a, \eta \sum \hat{Y}_s \rangle - \log(1 - \|a\|) - \|a\|$$

شاید نقطه بهینه در قیود صدق کند:

$$\nabla F(a) = \frac{a}{1 - \|a\|}$$

$$\nabla(\ldots) = \eta \sum \hat{Y}_s - \frac{1}{1 - \|a\|} \times \frac{-a}{\|a\|} - \frac{a}{\|a\|}$$

$$= \eta \sum \hat{Y}_s + \frac{a}{\|a\|}\left(\frac{1}{1 - \|a\|} - 1\right) = \eta \sum \hat{Y}_s + \frac{a}{1 - \|a\|}$$

$$\nabla(\ldots) = 0 \Rightarrow \frac{a}{1 - \|a\|} = -\eta \sum \hat{Y}_s$$

a موازی $-\eta \sum \hat{Y}_s$

با اندازه‌ای مناسب

$$\frac{\|a\|}{1 - \|a\|} = \eta \|\sum \hat{Y}_s\| \Rightarrow \frac{\|a\|}{1 - \|a\|} = \eta \|\sum \hat{Y}_s\| \Rightarrow \frac{\|a\|}{1} = \frac{\eta \|\sum \hat{Y}_s\|}{1 + \eta \|\sum \hat{Y}_s\|}$$

$$\arg\min_{x\in\tilde{A}} \langle a, \eta \sum \hat{Y}_s \rangle - \log(1 - \|a\|) - \|a\|$$

شاید نقطه بهینه در قیود صدق کند:

$$\nabla F(a) = \frac{a}{1 - \|a\|}$$

$$\nabla(\ldots) = \eta \sum \hat{Y}_s - \frac{1}{1 - \|a\|} \times \frac{-a}{\|a\|} - \frac{a}{\|a\|}$$

$$= \eta \sum \hat{Y}_s + \frac{a}{\|a\|}\left(\frac{1}{1 - \|a\|} - 1\right) = \eta \sum \hat{Y}_s + \frac{a}{1 - \|a\|}$$

$$\nabla(\ldots) = 0 \Rightarrow \frac{a}{1 - \|a\|} = -\eta \sum \hat{Y}_s$$

a موازی $-\eta \sum \hat{Y}_s$
با اندازه‌ای مناسب

$$\frac{\|a\|}{1 - \|a\|} = \eta \| \sum \hat{Y}_s \| \Rightarrow \frac{\|a\|}{1 - \|a\|} = \eta \| \sum \hat{Y}_s \| \Rightarrow \frac{\|a\|}{1} = \frac{\eta \| \sum \hat{Y}_s \|}{1 + \eta \| \sum \hat{Y}_s \|}$$

$$a = \frac{-\eta \sum \hat{Y}_s}{1 + \eta \| \sum \hat{Y}_s \|}$$

$$\arg\min_{x\in\tilde{A}}\langle a, \eta \sum \hat{Y}_s\rangle - \log(1-\|a\|) - \|a\|$$

شاید نقطه بهینه در قیود صدق کند:

$$\nabla F(a) = \frac{a}{1-\|a\|}$$

$$\nabla(\ldots) = \eta\sum \hat{Y}_s - \frac{1}{1-\|a\|}\times\frac{-a}{\|a\|} - \frac{a}{\|a\|}$$

$$= \eta\sum \hat{Y}_s + \frac{a}{\|a\|}\left(\frac{1}{1-\|a\|} - 1\right) = \eta\sum \hat{Y}_s + \frac{a}{1-\|a\|}$$

$$\nabla(\ldots) = 0 \Rightarrow \frac{a}{1-\|a\|} = -\eta\sum \hat{Y}_s$$

a موازی $-\eta\sum \hat{Y}_s$

با اندازه‌ای مناسب

$$\frac{\|a\|}{1-\|a\|} = \eta\|\sum \hat{Y}_s\| \Rightarrow \frac{\|a\|}{1-\|a\|} = \eta\|\sum \hat{Y}_s\| \Rightarrow \frac{\|a\|}{1} = \frac{\eta\|\sum \hat{Y}_s\|}{1+\eta\|\sum \hat{Y}_s\|}$$

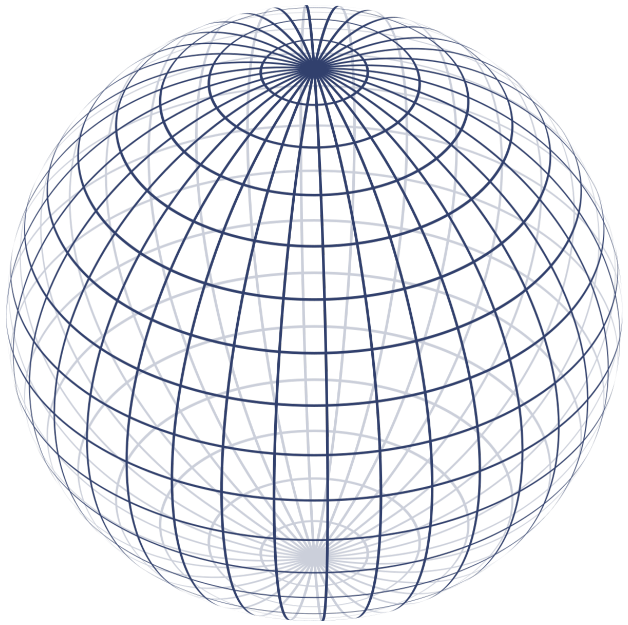
$$a = \frac{-\eta\sum \hat{Y}_s}{1+\eta\|\sum \hat{Y}_s\|} \qquad a = \frac{-\eta\hat{L}_{t-1}}{1+\eta\|\hat{L}_{t-1}\|}$$

شاید نقطه بهینه در قیود صدق کند:

$$a = \frac{-\eta \hat{L}_{t-1}}{1 + \eta \|\hat{L}_{t-1}\|}$$

شاید نقطه بهینه در قیود صدق کند:

$$a = \frac{-\eta \hat{L}_{t-1}}{1 + \eta \|\hat{L}_{t-1}\|}$$

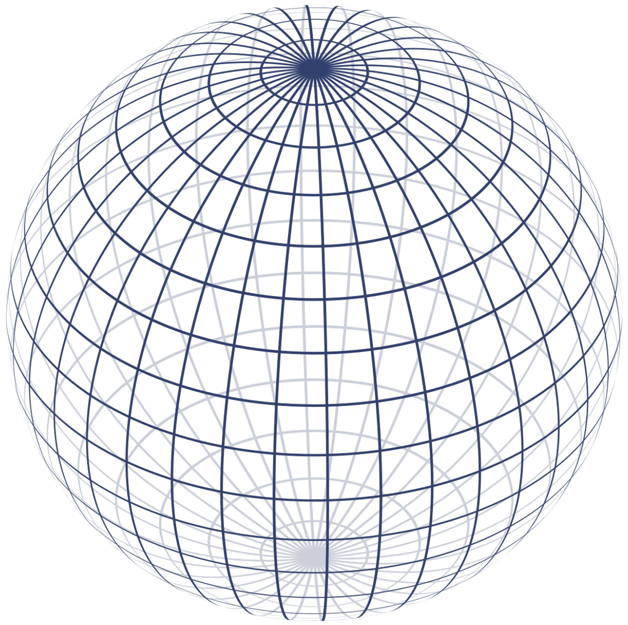


اگر درون گوی نبود: بهینه: گرادیان | | بردار

شاید نقطه بهینه در قیود صدق کند:

$$a = \frac{-\eta \hat{L}_{t-1}}{1 + \eta \|\hat{L}_{t-1}\|}$$

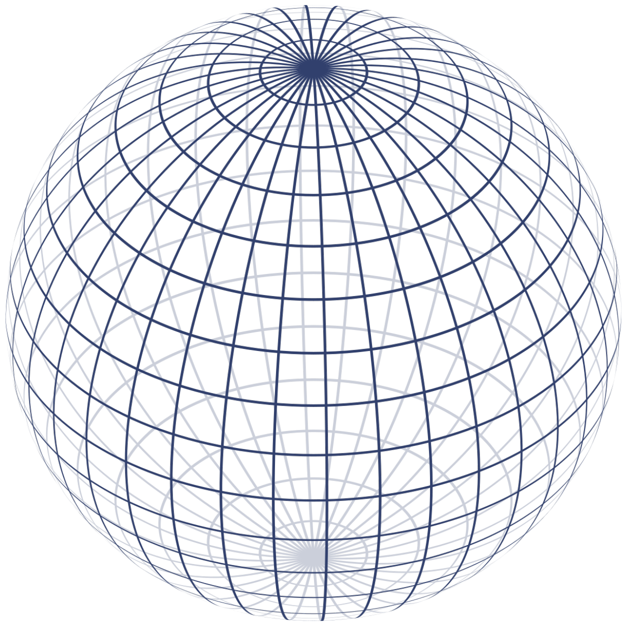
اگر درون گوی نبود: بهینه: گرادیان | | بردار



$$\eta \sum \hat{Y}_s + \frac{a}{1 - \|a\|} = \beta a$$

شاید نقطه بهینه در قیود صدق کند:

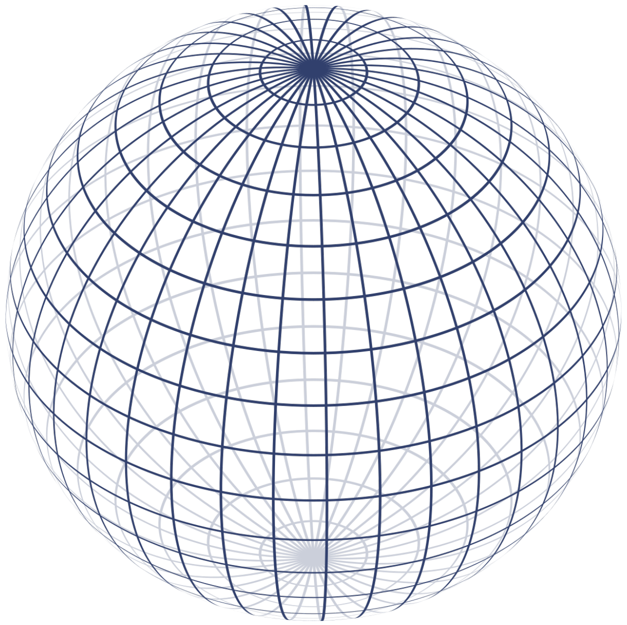
$$a = \frac{-\eta \hat{L}_{t-1}}{1 + \eta \|\hat{L}_{t-1}\|}$$



اگر درون گوی نبود: بهینه: گرادیان | | بردار

$$\eta \sum \hat{Y}_s + \frac{a}{1 - \|a\|} = \beta a \qquad \eta \sum \hat{Y}_s = a(\beta - \frac{1}{1 - \|a\|})$$

شاید نقطه بهینه در قیود صدق کند: $a = \frac{-\eta \hat{L}_{t-1}}{1 + \eta \|\hat{L}_{t-1}\|}$



اگر درون گوی نبود: بهینه: گرادیان | | بردار

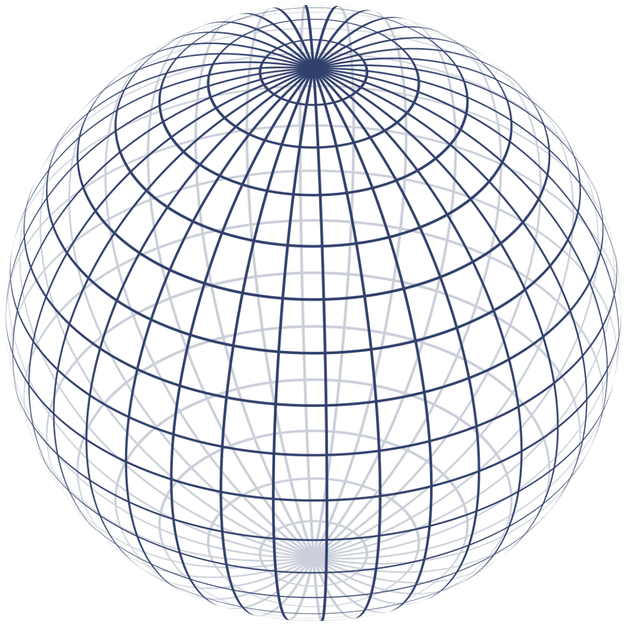
$$\eta \sum \hat{Y}_s + \frac{a}{1 - \|a\|} = \beta a$$

$$\eta \sum \hat{Y}_s = a(\beta - \frac{1}{1 - \|a\|})$$

موازی و روی سطح

شاید نقطه بهینه در قیود صدق کند:

$$a = \frac{-\eta \hat{L}_{t-1}}{1 + \eta \|\hat{L}_{t-1}\|}$$



اگر درون گوی نبود: بهینه: گرادیان | | بردار

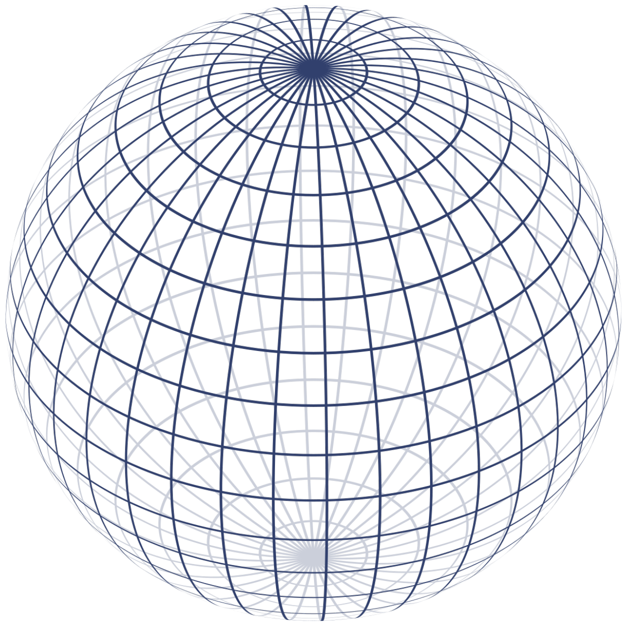
$$\eta \sum \hat{Y}_s + \frac{a}{1 - \|a\|} = \beta a$$

$$\eta \sum \hat{Y}_s = a(\beta - \frac{1}{1 - \|a\|})$$

موازی و روی سطح

تصویر روی کره

شاید نقطه بهینه در قیود صدق کند: $$a = \frac{-\eta \hat{L}_{t-1}}{1 + \eta \|\hat{L}_{t-1}\|}$$



اگر درون گوی نبود: بهینه: گرادیان | | بردار

$$\eta \sum \hat{Y}_s + \frac{a}{1 - \|a\|} = \beta a \qquad \eta \sum \hat{Y}_s = a(\beta - \frac{1}{1 - \|a\|})$$

موازی و روی سطح

تصویر روی کره

$$\bar{A}_t = \Pi \left(\frac{-\eta \hat{L}_{t-1}}{1 + \eta \|\hat{L}_{t-1}\|} \right) \qquad \hat{L}_{t-1} = \sum_{s=1}^{t-1} \hat{Y}_s,$$

مثال: بندیت خطی روی گوی واحد

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|$$

$$\tilde{\mathcal{A}} = \{x \in \mathbb{R}^d : \|x\|_2 \leq r\}$$

$$E_t \in \{0, 1\}$$

$$\mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$$

$$U_t \sim \{\pm e_1, \ldots, \pm e_d\}$$

$$A_t = E_t U_t + \frac{(1 - E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

Input Legendre potential F, action set $\mathcal{A}$ and learning

Choose $\bar{A}_1 = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} F(a)$

for $t = 1, \ldots, n$ **do**

Choose measure P_t on $\mathcal{A}$ with mean $\bar{A}_t$

Sample action A_t from P_t and observe $\langle A_t, y_t \rangle$

Compute estimate $\hat{Y}_t$ of the loss vector y_t

Update:

$\bar{A}_{t+1} = \text{argmin}$ … $, \bar{A}_t)$ (Mirror descent)

$\bar{A}_{t+1} = \text{arg}$ … (a) (follow-the-regularised-leader)

end for

$$\hat{Y}_t = \frac{d\, E_t\, A_t \langle A_t, y_t \rangle}{1 - \|\bar{A}_t\|}$$

$$\bar{A}_t = \Pi\left(\frac{-\eta \hat{L}_{t-1}}{1 + \eta \|\hat{L}_{t-1}\|}\right)$$

$$\hat{L}_{t-1} = \sum_{s=1}^{t-1} \hat{Y}_s,$$

$$) = -\log(1 - \|a\|) - \|a\|$$

تحلیل پشیمانی الگوریتم

$$R_n = \mathbb{E}\left[\sum_{t=1}^{n} \langle A_t - ra^*, y_t \rangle\right] + \sum_{t=1}^{n} \langle ra^* - a^*, y_t \rangle$$

تحلیل پشیمانی الگوریتم

$$R_n = \mathbb{E}\left[\sum_{t=1}^{n}\langle A_t - ra^*, y_t\rangle\right] + \sum_{t=1}^{n}\langle ra^* - a^*, y_t\rangle \ \leq R_n(ra^*) + (1-r)n$$

$$R_n(a) \leq \mathbb{E}\left[\frac{F(a) - F(\bar{A}_1)}{\eta} + \sum_{t=1}^{n}\langle \bar{A}_t - \bar{A}_{t+1}, \hat{Y}_t\rangle - \frac{1}{\eta}\sum_{t=1}^{n} D(\bar{A}_{t+1}, \bar{A}_t)\right]$$

$$R_n(a) \leq \mathbb{E}\left[\frac{F(a) - F(\bar{A}_1)}{\eta} + \sum_{t=1}^{n}\langle \bar{A}_t - \bar{A}_{t+1}, \hat{Y}_t\rangle - \frac{1}{\eta}\sum_{t=1}^{n} D(\bar{A}_{t+1}, \bar{A}_t)\right]$$

THEOREM 26.13. *Let $\eta > 0$ and f be Legendre and twice differentiable in $A = \operatorname{int}(\operatorname{dom}(f))$, $x, y \in A$, and let $z \in [x, y]$ be the point such that $D_f(x, y) = \frac{1}{2}\|x - y\|^2_{\nabla^2 f(z)}$. Then, for all $u \in \mathbb{R}^d$,*

$$\langle x - y, u\rangle - \frac{D_f(x, y)}{\eta} \leq \frac{\eta}{2}\|u\|^2_{(\nabla^2 f(z))^{-1}}.$$

$$R_n(a) \leq \mathbb{E}\left[\frac{F(a) - F(\bar{A}_1)}{\eta} + \sum_{t=1}^{n}\langle \bar{A}_t - \bar{A}_{t+1}, \hat{Y}_t\rangle - \frac{1}{\eta}\sum_{t=1}^{n} D(\bar{A}_{t+1}, \bar{A}_t)\right]$$

THEOREM 26.13. *Let $\eta > 0$ and f be Legendre and twice differentiable in $A = \operatorname{int}(\operatorname{dom}(f))$, $x, y \in A$, and let $z \in [x, y]$ be the point such that $D_f(x, y) = \frac{1}{2}\|x - y\|^2_{\nabla^2 f(z)}$. Then, for all $u \in \mathbb{R}^d$,*

$$\langle x - y, u\rangle - \frac{D_f(x, y)}{\eta} \leq \frac{\eta}{2}\|u\|^2_{(\nabla^2 f(z))^{-1}}.$$

$$R_n(ra^*) \leq \frac{\operatorname{diam}_F(\tilde{\mathcal{A}})}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{n}\|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right]$$

$$R_n(a) \leq \mathbb{E}\left[\frac{F(a) - F(\bar{A}_1)}{\eta} + \sum_{t=1}^{n} \langle \bar{A}_t - \bar{A}_{t+1}, \hat{Y}_t \rangle - \frac{1}{\eta}\sum_{t=1}^{n} D(\bar{A}_{t+1}, \bar{A}_t)\right]$$

THEOREM 26.13. *Let $\eta > 0$ and f be Legendre and twice differentiable in $A = \operatorname{int}(\operatorname{dom}(f))$, $x, y \in A$, and let $z \in [x, y]$ be the point such that $D_f(x, y) = \frac{1}{2}\|x - y\|^2_{\nabla^2 f(z)}$. Then, for all $u \in \mathbb{R}^d$,*

$$\langle x - y, u \rangle - \frac{D_f(x, y)}{\eta} \leq \frac{\eta}{2}\|u\|^2_{(\nabla^2 f(z))^{-1}}.$$

$$R_n(ra^*) \leq \frac{\operatorname{diam}_F(\tilde{\mathcal{A}})}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{n} \|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right]$$

$$\frac{1}{\eta}\log\left(\frac{1}{1-r}\right)$$

$$\mathbb{E}\left[\|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right] \leq$$

$$\mathbb{E}\left[\|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right] \leq \; E[\lambda_{\max} \|\hat{Y}_t\|^2]$$

$$\mathbb{E}\left[\|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right] \leq \; E[\lambda_{\max}\|\hat{Y}_t\|^2]$$

$$\nabla F(a) = \frac{a}{1 - \|a\|}$$

$$\mathbb{E}\left[\|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right] \leq \; E[\lambda_{\max}\|\hat{Y}_t\|^2]$$

$$\nabla F(a) = \frac{a}{1 - \|a\|}$$

$$\nabla^2 F(a) = \frac{I}{1 - \|a\|} + \frac{aa^t}{(1 - \|a\|)^2}$$

$$\mathbb{E}\left[\|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right] \leq \ E[\lambda_{\max}\|\hat{Y}_t\|^2]$$

$$\nabla F(a) = \frac{a}{1 - \|a\|}$$

$$\nabla^2 F(a) = \frac{I}{1 - \|a\|} + \frac{aa^t}{(1 - \|a\|)^2} \succeq \frac{I}{1 - \|a\|}$$

$$\mathbb{E}\left[\|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right] \leq E[\lambda_{\max}\|\hat{Y}_t\|^2]$$

$$\nabla F(a) = \frac{a}{1 - \|a\|}$$

$$\nabla^2 F(a) = \frac{I}{1 - \|a\|} + \frac{aa^t}{(1 - \|a\|)^2} \succeq \frac{I}{1 - \|a\|}$$

$$\nabla^2 F(a)^{-1} \preceq \frac{I}{1 - \|a\|}$$

$$\mathbb{E}\left[\|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right] \leq \; E[\lambda_{\max}\|\hat{Y}_t\|^2]$$

$$\nabla F(a) = \frac{a}{1 - \|a\|}$$

$$\nabla^2 F(a) = \frac{I}{1 - \|a\|} + \frac{aa^t}{(1 - \|a\|)^2} \succeq \frac{I}{1 - \|a\|}$$

$$\nabla^2 F(a)^{-1} \preceq \frac{I}{1 - \|a\|}$$

$$\mathbb{E}\left[\|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right] \leq E[\lambda_{\max}\|\hat{Y}_t\|^2] \leq \mathbb{E}\left[(1 - \|Z_t\|)\|\hat{Y}_t\|^2\right]$$

$$\nabla F(a) = \frac{a}{1 - \|a\|}$$

$$\nabla^2 F(a) = \frac{I}{1 - \|a\|} + \frac{aa^t}{(1 - \|a\|)^2} \succeq \frac{I}{1 - \|a\|}$$

$$\nabla^2 F(a)^{-1} \preceq \frac{I}{1 - \|a\|}$$

$$\mathbb{E}\left[\|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right] \leq \; E[\lambda_{\max}\|\hat{Y}_t\|^2] \; \leq \mathbb{E}\left[(1 - \|Z_t\|)\|\hat{Y}_t\|^2\right]$$

$$\hat{Y}_t = \frac{d\, E_t\, A_t \langle A_t, y_t \rangle}{1 - \|\bar{A}_t\|}$$

$$A_t = E_t U_t + \frac{(1 - E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

$$\nabla F(a) = \frac{a}{1 - \|a\|}$$

$$\nabla^2 F(a) = \frac{I}{1 - \|a\|} + \frac{aa^t}{(1 - \|a\|)^2} \succeq \frac{I}{1 - \|a\|}$$

$$\nabla^2 F(a)^{-1} \preceq \frac{I}{1 - \|a\|}$$

$$\mathbb{E}\left[\|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right] \leq E[\lambda_{\max}\|\hat{Y}_t\|^2] \leq \mathbb{E}\left[(1 - \|Z_t\|)\|\hat{Y}_t\|^2\right]$$

$$= d^2\mathbb{E}\left[\frac{(1 - \|Z_t\|)E_t\langle U_t, y_t\rangle^2}{(1 - \|\bar{A}_t\|)^2}\right]$$

$$\hat{Y}_t = \frac{d\, E_t\, A_t\langle A_t, y_t\rangle}{1 - \|\bar{A}_t\|}$$

$$A_t = E_t U_t + \frac{(1 - E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

$$= d^2 \mathbb{E}\left[\frac{(1-\|Z_t\|)E_t\langle U_t, y_t\rangle^2}{(1-\|\bar{A}_t\|)^2}\right] \quad \leq 2d.$$

$$= d^2 \mathbb{E}\left[\frac{(1 - \|Z_t\|) E_t \langle U_t, y_t \rangle^2}{(1 - \|\bar{A}_t\|)^2}\right] \leq 2d.$$

$$= d^2\mathbb{E}\left[\frac{(1-\|Z_t\|)E_t\langle U_t, y_t\rangle^2}{(1-\|\bar{A}_t\|)^2}\right] \leq 2d.$$

$$\frac{1-\|Z_t\|}{1-\|\bar{A}_t\|}$$

$$= d^2\mathbb{E}\left[\frac{(1-\|Z_t\|)E_t\langle U_t, y_t\rangle^2}{(1-\|\bar{A}_t\|)^2}\right] \leq 2d.$$

$$\frac{1-\|Z_t\|}{1-\|\bar{A}_t\|} \leq \sup_{\alpha\in[0,1]} \frac{1-\|\alpha\bar{A}_t+(1-\alpha)\bar{A}_{t+1}\|}{1-\|\bar{A}_t\|}$$

$$= d^2\mathbb{E}\left[\frac{(1-\|Z_t\|)E_t\langle U_t, y_t\rangle^2}{(1-\|\bar{A}_t\|)^2}\right] \leq 2d.$$

$$\frac{1-\|Z_t\|}{1-\|\bar{A}_t\|} \leq \sup_{\alpha\in[0,1]} \frac{1-\|\alpha\bar{A}_t+(1-\alpha)\bar{A}_{t+1}\|}{1-\|\bar{A}_t\|} = \max\left\{1, \frac{1-\|\bar{A}_{t+1}\|}{1-\|\bar{A}_t\|}\right\}$$

$$= d^2\mathbb{E}\left[\frac{(1-\|Z_t\|)E_t\langle U_t, y_t\rangle^2}{(1-\|\bar{A}_t\|)^2}\right] \leq 2d.$$

$$\frac{1-\|Z_t\|}{1-\|\bar{A}_t\|} \leq \sup_{\alpha\in[0,1]} \frac{1-\|\alpha\bar{A}_t+(1-\alpha)\bar{A}_{t+1}\|}{1-\|\bar{A}_t\|} = \max\left\{1, \frac{1-\|\bar{A}_{t+1}\|}{1-\|\bar{A}_t\|}\right\}$$

اگر ۱ بیشینه نباشد، At+1 نزدیک‌تر است

برای At، تصویر رخ نداده

$$\bar{A}_t = \Pi\left(\frac{-\eta\hat{L}_{t-1}}{1+\eta\|\hat{L}_{t-1}\|}\right)$$

$$= d^2\mathbb{E}\left[\frac{(1-\|Z_t\|)E_t\langle U_t, y_t\rangle^2}{(1-\|\bar{A}_t\|)^2}\right] \leq 2d.$$

$$\frac{1-\|Z_t\|}{1-\|\bar{A}_t\|} \leq \sup_{\alpha\in[0,1]} \frac{1-\|\alpha\bar{A}_t + (1-\alpha)\bar{A}_{t+1}\|}{1-\|\bar{A}_t\|} = \max\left\{1, \frac{1-\|\bar{A}_{t+1}\|}{1-\|\bar{A}_t\|}\right\}$$

$$\leq \max\left\{1, \frac{1+\eta\|\hat{L}_{t-1}\|}{1+\eta\|\hat{L}_t\|}\right\}$$

اگر ۱ بیشینه نباشد، At+1 نزدیک‌تر است

برای At، تصویر رخ نداده

$$\bar{A}_t = \Pi\left(\frac{-\eta\hat{L}_{t-1}}{1+\eta\|\hat{L}_{t-1}\|}\right)$$

$$= d^2\mathbb{E}\left[\frac{(1-\|Z_t\|)E_t\langle U_t, y_t\rangle^2}{(1-\|\bar{A}_t\|)^2}\right] \leq 2d.$$

$$\frac{1-\|Z_t\|}{1-\|\bar{A}_t\|} \leq \sup_{\alpha\in[0,1]} \frac{1-\|\alpha\bar{A}_t+(1-\alpha)\bar{A}_{t+1}\|}{1-\|\bar{A}_t\|} = \max\left\{1, \frac{1-\|\bar{A}_{t+1}\|}{1-\|\bar{A}_t\|}\right\}$$

$$\eta\|\hat{Y}_t\| \leq \eta d/(1-r) = 1/2.$$

$$\leq \max\left\{1, \frac{1+\eta\|\hat{L}_{t-1}\|}{1+\eta\|\hat{L}_t\|}\right\}$$

اگر ۱ بیشینه نباشد، At+1 نزدیک‌تر است

برای At، تصویر رخ نداده

$$\bar{A}_t = \Pi\left(\frac{-\eta\hat{L}_{t-1}}{1+\eta\|\hat{L}_{t-1}\|}\right)$$

$$= d^2\mathbb{E}\left[\frac{(1-\|Z_t\|)E_t\langle U_t, y_t\rangle^2}{(1-\|\bar{A}_t\|)^2}\right] \leq 2d.$$

$$\frac{1-\|Z_t\|}{1-\|\bar{A}_t\|} \leq \sup_{\alpha\in[0,1]} \frac{1-\|\alpha\bar{A}_t+(1-\alpha)\bar{A}_{t+1}\|}{1-\|\bar{A}_t\|} = \max\left\{1, \frac{1-\|\bar{A}_{t+1}\|}{1-\|\bar{A}_t\|}\right\}$$

$$\hat{Y}_t = \frac{d\,E_t\,A_t\langle A_t, y_t\rangle}{1-\|\bar{A}_t\|}$$

$$\eta\|\hat{Y}_t\| \leq \eta d/(1-r) = 1/2.$$

$$\leq \max\left\{1, \frac{1+\eta\|\hat{L}_{t-1}\|}{1+\eta\|\hat{L}_t\|}\right\}$$

اگر ۱ بیشینه نباشد، At+1 نزدیک‌تر است

برای At، تصویر رخ نداده

$$\bar{A}_t = \Pi\left(\frac{-\eta\hat{L}_{t-1}}{1+\eta\|\hat{L}_{t-1}\|}\right)$$

$$= d^2\mathbb{E}\left[\frac{(1-\|Z_t\|)E_t\langle U_t, y_t\rangle^2}{(1-\|\bar{A}_t\|)^2}\right] \le 2d.$$

$$\frac{1-\|Z_t\|}{1-\|\bar{A}_t\|} \le \sup_{\alpha\in[0,1]} \frac{1-\|\alpha\bar{A}_t + (1-\alpha)\bar{A}_{t+1}\|}{1-\|\bar{A}_t\|} = \max\left\{1, \frac{1-\|\bar{A}_{t+1}\|}{1-\|\bar{A}_t\|}\right\}$$

$$\hat{Y}_t = \frac{d\,E_t\,A_t\langle A_t, y_t\rangle}{1-\|\bar{A}_t\|}$$

$$r = 1 - 2\eta d$$

$$\eta\|\hat{Y}_t\| \le \eta d/(1-r) = 1/2.$$

$$\le \max\left\{1, \frac{1+\eta\|\hat{L}_{t-1}\|}{1+\eta\|\hat{L}_t\|}\right\}$$

اگر ۱ بیشینه نباشد، At+1 نزدیک‌تر است

برای At، تصویر رخ نداده

$$\bar{A}_t = \Pi\left(\frac{-\eta\hat{L}_{t-1}}{1+\eta\|\hat{L}_{t-1}\|}\right)$$

$$= d^2\mathbb{E}\left[\frac{(1-\|Z_t\|)E_t\langle U_t, y_t\rangle^2}{(1-\|\bar{A}_t\|)^2}\right] \leq 2d.$$

$$\frac{1-\|Z_t\|}{1-\|\bar{A}_t\|} \leq \sup_{\alpha\in[0,1]} \frac{1-\|\alpha\bar{A}_t + (1-\alpha)\bar{A}_{t+1}\|}{1-\|\bar{A}_t\|} = \max\left\{1, \frac{1-\|\bar{A}_{t+1}\|}{1-\|\bar{A}_t\|}\right\}$$

$$\hat{Y}_t = \frac{d\,E_t\,A_t\langle A_t, y_t\rangle}{1-\|\bar{A}_t\|}$$

$$r = 1 - 2\eta d$$

$$\eta\|\hat{Y}_t\| \leq \eta d/(1-r) = 1/2.$$

$$\leq \max\left\{1, \frac{1+\eta\|\hat{L}_{t-1}\|}{1+\eta\|\hat{L}_t\|}\right\} \leq \max\left\{1, \frac{1+\eta\|\hat{L}_{t-1}\|}{1/2+\eta\|\hat{L}_{t-1}\|}\right\}$$

اگر ۱ بیشینه نباشد، At+1 نزدیک‌تر است

برای At، تصویر رخ نداده

$$\bar{A}_t = \Pi\left(\frac{-\eta\hat{L}_{t-1}}{1+\eta\|\hat{L}_{t-1}\|}\right)$$

$$= d^2\mathbb{E}\left[\frac{(1-\|Z_t\|)E_t\langle U_t, y_t\rangle^2}{(1-\|\bar{A}_t\|)^2}\right] \le 2d.$$

$$\frac{1-\|Z_t\|}{1-\|\bar{A}_t\|} \le \sup_{\alpha\in[0,1]} \frac{1-\|\alpha\bar{A}_t + (1-\alpha)\bar{A}_{t+1}\|}{1-\|\bar{A}_t\|} = \max\left\{1, \frac{1-\|\bar{A}_{t+1}\|}{1-\|\bar{A}_t\|}\right\}$$

$$\hat{Y}_t = \frac{d\,E_t\,A_t\langle A_t, y_t\rangle}{1-\|\bar{A}_t\|}$$

$$r = 1 - 2\eta d$$

$$\eta\|\hat{Y}_t\| \le \eta d/(1-r) = 1/2.$$

$$\le \max\left\{1, \frac{1+\eta\|\hat{L}_{t-1}\|}{1+\eta\|\hat{L}_t\|}\right\} \le \max\left\{1, \frac{1+\eta\|\hat{L}_{t-1}\|}{1/2+\eta\|\hat{L}_{t-1}\|}\right\} \le 2.$$

اگر ۱ بیشینه نباشد، At+1 نزدیک‌تر است

برای At، تصویر رخ نداده

$$\bar{A}_t = \Pi\left(\frac{-\eta\hat{L}_{t-1}}{1+\eta\|\hat{L}_{t-1}\|}\right)$$

$$R_n \leq R_n(ra^*) + (1-r)n$$

$$R_n \leq R_n(ra^*) + (1 - r)n$$

$$R_n \leq R_n(ra^*) + (1-r)n$$

$$R_n(ra^*) \leq \frac{\operatorname{diam}_F(\tilde{\mathcal{A}})}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{n} \|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right]$$

$$R_n \le R_n(ra^*) + (1-r)n$$

$$R_n(ra^*) \le \frac{\mathrm{diam}_F(\tilde{\mathcal{A}})}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{n} \|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right]$$

$$\frac{1}{\eta}\log\left(\frac{1}{1-r}\right)$$

$$R_n \leq R_n(ra^*) + (1-r)n$$

$$R_n(ra^*) \leq \frac{\mathrm{diam}_F(\tilde{\mathcal{A}})}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{n} \|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right]$$

$$\frac{1}{\eta}\log\left(\frac{1}{1-r}\right)$$

$$\mathbb{E}\left[\|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right] \leq 2d.$$

$$R_n \leq R_n(ra^*) + (1-r)n$$

$$R_n(ra^*) \leq \frac{\operatorname{diam}_F(\tilde{\mathcal{A}})}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{n} \|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right]$$

$$\frac{1}{\eta}\log\left(\frac{1}{1-r}\right)$$

$$\mathbb{E}\left[\|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right] \leq 2d.$$

$$\leq (1-r)n + \frac{1}{\eta}\log\left(\frac{1}{1-r}\right) + \eta n d$$

$$R_n \leq R_n(ra^*) + (1-r)n$$

$$R_n(ra^*) \leq \frac{\operatorname{diam}_F(\tilde{\mathcal{A}})}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{n} \|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right]$$

$$\frac{1}{\eta}\log\left(\frac{1}{1-r}\right)$$

$$\mathbb{E}\left[\|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right] \leq 2d.$$

$$\leq (1-r)n + \frac{1}{\eta}\log\left(\frac{1}{1-r}\right) + \eta n d$$

$$r = 1 - 2\eta d$$

$$R_n \leq R_n(ra^*) + (1-r)n$$

$$R_n(ra^*) \leq \frac{\mathrm{diam}_F(\tilde{\mathcal{A}})}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{n} \|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right]$$

$$\frac{1}{\eta}\log\left(\frac{1}{1-r}\right)$$

$$\mathbb{E}\left[\|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right] \leq 2d.$$

$$\leq (1-r)n + \frac{1}{\eta}\log\left(\frac{1}{1-r}\right) + \eta n d$$

$$r = 1 - 2\eta d$$

$$= \frac{1}{\eta}\log\left(\frac{1}{2\eta d}\right) + 3\eta n d$$

$$R_n \leq R_n(ra^*) + (1-r)n$$

$$R_n(ra^*) \leq \frac{\operatorname{diam}_F(\tilde{\mathcal{A}})}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{n} \|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right]$$

$$\frac{1}{\eta}\log\left(\frac{1}{1-r}\right)$$

$$\mathbb{E}\left[\|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right] \leq 2d.$$

$$\leq (1-r)n + \frac{1}{\eta}\log\left(\frac{1}{1-r}\right) + \eta n d$$

$$r = 1 - 2\eta d$$

$$= \frac{1}{\eta}\log\left(\frac{1}{2\eta d}\right) + 3\eta n d$$

$$\eta = \sqrt{\log(n)/(3dn)}.$$

$$R_n \le R_n(ra^*) + (1-r)n$$

$$R_n(ra^*) \le \frac{\operatorname{diam}_F(\tilde{\mathcal{A}})}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^n \|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right]$$

$$\frac{1}{\eta}\log\left(\frac{1}{1-r}\right)$$

$$\mathbb{E}\left[\|\hat{Y}_t\|^2_{(\nabla^2 F(Z_t))^{-1}}\right] \le 2d.$$

$$\le (1-r)n + \frac{1}{\eta}\log\left(\frac{1}{1-r}\right) + \eta n d$$

$$r = 1 - 2\eta d$$

$$= \frac{1}{\eta}\log\left(\frac{1}{2\eta d}\right) + 3\eta n d$$

$$\le 2\sqrt{3nd\log(n)},$$

$$\eta = \sqrt{\log(n)/(3dn)}.$$

مثال: بندیت خطی روی گوی واحد

$R_n \leq$

$2\sqrt{3nd\log(n)}.$

$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|$

$\tilde{\mathcal{A}} = \{x \in \mathbb{R}^d : \|x\|_2 \leq r\}$

$E_t \in \{0, 1\}$

$\mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$

$U_t \sim \{\pm e_1, \ldots, \pm e_d\}$

Input Legendre potential F, action set $\mathcal{A}$ and learning

Choose $\bar{A}_1 = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} F(a)$

for $t = 1, \ldots, n$ **do**

Choose measure P_t on $\mathcal{A}$ with mean $\bar{A}_t$

Sample action A_t from P_t and observe $\langle A_t, y_t \rangle$

Compute estimate $\hat{Y}_t$ of the loss vector y_t

Update:

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \langle a, \hat{Y}_t \rangle + D_F(a, \bar{A}_t)$$

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \sum_{s=1}^{t} \langle a, \hat{Y}_s \rangle + F(a) \quad \text{(follow-the-regularised-leader)}$$

end for

$F(a) = -\log(1 - \|a\|) - \|a\|$

مثال: بندیت خطی روی گوی واحد

$$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|$$

$$R_n \leq 2\sqrt{3nd\log(n)}.$$

$$\tilde{\mathcal{A}} = \{x \in \mathbb{R}^d : \|x\|_2 \leq r\}$$

$$E_t \in \{0,1\}$$

$$\mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$$

$$U_t \sim \{\pm e_1, \ldots, \pm e_d\}$$

$$A_t = E_t U_t + \frac{(1-E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

Input Legendre potential F, action set $\mathcal{A}$ and learning

Choose $\bar{A}_1 = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} F(a)$

for $t = 1, \ldots, n$ **do**

Choose measure P_t on $\mathcal{A}$ with mean $\bar{A}_t$

Sample action A_t from P_t and observe $\langle A_t, y_t \rangle$

Compute estimate $\hat{Y}_t$ of the loss vector y_t

Update:

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \langle a, \hat{Y}_t \rangle + D_F(a, \bar{A}_t)$$

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \sum_{s=1}^{t} \langle a, \hat{Y}_s \rangle + F(a) \quad \text{(follow-the-regularised-leader)}$$

end for

$$F(a) = -\log(1 - \|a\|) - \|a\|$$

مثال: بندیت خطی روی گوی واحد

$R_n \leq 2\sqrt{3nd\log(n)}.$

$\mathscr{L} = \mathcal{A} = B_2^d = \{a \in \mathbb{R}^d : \|a\|$

$\tilde{\mathcal{A}} = \{x \in \mathbb{R}^d : \|x\|_2 \leq r\}$

$E_t \in \{0,1\}$

$\mathbb{E}_t[E_t] = 1 - \|\bar{A}_t\|$

$U_t \sim \{\pm e_1, \ldots, \pm e_d\}$

$$A_t = E_t U_t + \frac{(1 - E_t)\bar{A}_t}{\|\bar{A}_t\|}$$

$$\hat{Y}_t = \frac{d\, E_t\, A_t \langle A_t, y_t\rangle}{1 - \|\bar{A}_t\|}$$

Input Legendre potential F, action set $\mathcal{A}$ and learning

Choose $\bar{A}_1 = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} F(a)$

for $t = 1, \ldots, n$ **do**

Choose measure P_t on $\mathcal{A}$ with mean $\bar{A}_t$

Sample action A_t from P_t and observe $\langle A_t, y_t\rangle$

Compute estimate $\hat{Y}_t$ of the loss vector y_t

Update:

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \langle a, \hat{Y}_t\rangle + D_F(a, \bar{A}_t)$$

$$\bar{A}_{t+1} = \operatorname{argmin}_{a \in \mathcal{A} \cap \operatorname{dom}(F)} \eta \sum_{s=1}^{t} \langle a, \hat{Y}_s\rangle + F(a) \quad \text{(follow-the-regularised-leader)}$$

end for

$F(a) = -\log(1 - \|a\|) - \|a\|$

پایان