Optimization Algorithms: HW2

Lo Chun, Chou
R13922136

May 10, 2025

1

We're given the following problem:

$$x_\star \in \arg\min_{x \in \Delta_d} f(x), \qquad f(x) = -\sum_{i=1}^{n} w_i \log\langle a_i, x\rangle$$

where:

1. $$x \in \Delta_d = \{x \in \mathbb{R}^d \mid x[i] \geq 0, \sum_{i=1}^{d} x[i] = 1\} \text{ (probabilit`y simplex)}$$

2. $$w_i > 0,\ \sum_{i=1}^{n} w_i = 1$$

3. $$a_i \in \mathbb{R}^d,\ a_i = \begin{bmatrix} a_i[1] \\ a_i[2] \\ \vdots \\ a_i[d] \end{bmatrix},$$
$$a_i[j] \geq 0\ \forall i = 1, \ldots, n,\ j = 1, \ldots, d$$
$$a_i \neq 0\ \forall i = 1, \ldots, n$$

We're asked to show that:

- f is 1-smooth relative to the log-barrier, which is defined as:

$$h(x) = -\sum_{i=1}^{d} \log x[i]$$

- specify value of L

And denote the Bregman divergence associated with h as D_h, i.e.,

$$D_h(y,x) = h(y) - [h(x) + \langle \nabla h(x), (y-x) \rangle]$$

Also, consider solving the optimization problem (1) y the following algorithm:

- Let $x_1 = (1/d, \ldots, 1/d) \in \Delta_d$
- For every $t \in \mathbb{N}$, compute:

$$x_{t+1} = \arg\min_{x \in \Delta_d} [\langle \nabla f(x_t), x - x_t \rangle + D_h(x, x_t)]$$

Solution. By the following proposition [1]:

PROPOSITION 1.1. *The following conditions are equivalent:*
(a-i) $f(\cdot)$ *is L-smooth relative to* $h(\cdot)$*;*
(a-ii) $Lh(\cdot) - f(\cdot)$ *is a convex function on* Q*;*
(a-iii) *under twice differentiability* $\nabla^2 f(x) \preceq L\nabla^2 h(x)$ *for any* $x \in \operatorname{int} Q$*;*
(a-iv) $\langle \nabla f(x) - \nabla f(y), x-y \rangle \leq L \langle \nabla h(x) - \nabla h(y), x-y \rangle$ *for all* $x, y \in \operatorname{int} Q$.

we knew that we could prove the required condition (which is (a-i)) by proving its equivalent condition (a-iii).

First calculate $\nabla f(x)$:

$$\begin{aligned}
\nabla f(x) &= \frac{d}{dx}\left(-\sum_{i=1}^{n} w_i \log\langle a_i, x\rangle\right) \\
&= -\sum_{i=1}^{n} w_i \cdot \frac{d}{dx}\left(\log\langle a_i, x\rangle\right) \\
&= -\sum_{i=1}^{n} w_i \cdot \left(\frac{a_i}{\langle a_i, x\rangle}\right) \\
&= -\sum_{i=1}^{n} w_i \frac{a_i}{\langle a_i, x\rangle}
\end{aligned}$$

Then the Hessian of f is:

[1] *Relative Smooth Convex Optimization by First-Order Methods, and Applications*, MIT Lecture Notes, available at: `https://dspace.mit.edu/bitstream/handle/1721.1/120867/16m1099546.pdf`, accessed: May. 9, 2025, p. 336.

$$
\begin{aligned}
\nabla^2 f(x) &= \frac{d}{dx}\left(-\sum_{i=1}^{n} w_i \frac{a_i}{\langle a_i, x\rangle}\right) \\
&= -\sum_{i=1}^{n} w_i \cdot \frac{d}{dx}\left(\frac{a_i}{\langle a_i, x\rangle}\right) \\
&= -\sum_{i=1}^{n} w_i \cdot \left(\frac{a_i}{\langle a_i, x\rangle^2} a_i^T\right)
\end{aligned}
$$

Then we shall do the same to $h(x)$

$$
\begin{aligned}
\nabla h(x) &= \frac{d}{dx}\left(-\sum_{i=1}^{d} \log x[i]\right) \\
&= \begin{bmatrix} -\frac{1}{x[1]} \\ -\frac{1}{x[2]} \\ \vdots \\ -\frac{1}{x[d]} \end{bmatrix}
\end{aligned}
$$

Then $\nabla^2 h(x)$ is:

$$
\begin{aligned}
\nabla^2 h(x) &= \nabla \begin{bmatrix} -\frac{1}{x[1]} \\ -\frac{1}{x[2]} \\ \vdots \\ -\frac{1}{x[d]} \end{bmatrix} \\
&= \begin{bmatrix}
\frac{d}{dx[1]}\left(-\frac{1}{x[1]}\right) & \frac{d}{dx[2]}\left(-\frac{1}{x[1]}\right) & \cdots & \frac{d}{dx[d]}\left(-\frac{1}{x[1]}\right) \\
\frac{d}{dx[1]}\left(-\frac{1}{x[2]}\right) & \frac{d}{dx[2]}\left(-\frac{1}{x[2]}\right) & \cdots & \frac{d}{dx[d]}\left(-\frac{1}{x[2]}\right) \\
\vdots & & \ddots & \vdots \\
\frac{d}{dx[1]}\left(-\frac{1}{x[d]}\right) & \frac{d}{dx[2]}\left(-\frac{1}{x[d]}\right) & \cdots & \frac{d}{dx[d]}\left(-\frac{1}{x[d]}\right)
\end{bmatrix} \\
&= \begin{bmatrix}
\frac{1}{x[1]^2} & 0 & \cdots & 0 \\
0 & \frac{1}{x[2]^2} & \cdots & 0 \\
\vdots & & \ddots & \vdots \\
0 & 0 & \cdots & \frac{1}{x[d]^2}
\end{bmatrix}
\end{aligned}
$$

We then first show that $\nabla^2 f(x)$ is symmetric, and $\nabla^2 h(x)$ is positive definite, in order to use the following generalized Rayleigh quotient to prove that $\nabla^2 f(x) \preceq L\nabla^2 h(x)$:

The generalized Rayleigh quotients

Def 0.2. For a fixed symmetric matrix $\mathbf{A} \in S^n(\mathbb{R})$ and a positive definite matrix $\mathbf{B} \in S^n_+(\mathbb{R})$ of the same size, a **generalized Rayleigh quotient** corresponding to them is a function $f : \mathbb{R}^n - \{\mathbf{0}\} \longmapsto \mathbb{R}$ defined by

$$f(\mathbf{x}) = \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}}.$$

By our previous calculation, we have:

$$\nabla^2 f(x) = -\sum_{i=1}^{n}\left(\frac{w_i}{\langle a_i, x\rangle^2}a_i a_i^T\right)$$

Since $a_i a_i^T$ is symmetric, and multiplying a scalar $\frac{w_i}{\langle a_i, x\rangle^2}$, summing up symmetric matrices also results in a symmetric matrix, we have $\nabla^2 f(x)$ is symmetric.

Then, since $\nabla^2 h(x)$ is a diagonal matrix, and we're given that $x[i] \geq 0$, with proposition $(a-iii)$ only requires dealing with int Δ_d, we can guarantee $x[i] > 0$ (so for each $\frac{1}{x[i]}$), and $\nabla^2 h(x)$ is positive definite.

Therefore, by letting $A = \nabla^2 f(x)$ and $B = \nabla^2 h(x)$, with $z \in \mathbb{R}^d - \{0\}$, we define the generalized Rayleigh quotient:

$$\begin{aligned}
R(z) &= \frac{z^T\nabla^2 f(x) z}{z^T\nabla^2 h(x) z} \\
&= \frac{z^T\left(-\sum_{i=1}^n\left(\frac{w_i}{\langle a_i,x\rangle^2}a_i a_i^T\right)\right)z}{-\sum_{i=1}^d\left(\frac{1}{x[i]}z[i]^2\right)} \\
&= \frac{\sum_{i=1}^n\left(\frac{w_i}{\langle a_i,x\rangle^2}z^T a_i a_i^T z\right)}{-\sum_{i=1}^d\left(\frac{1}{x[i]}z[i]^2\right)} \\
&= \frac{\sum_{i=1}^n\left(\frac{w_i}{\langle a_i,x\rangle^2}\|a_i^T z\|^2\right)}{-\sum_{i=1}^d\left(\frac{1}{x[i]}z[i]^2\right)} \\
&= \frac{\sum_{i=1}^n\left[\frac{w_i}{\langle a_i,x\rangle^2}\left(\sum_{j=1}^d a_i[j]z[j]\right)^2\right]}{-\sum_{i=1}^d\left(\frac{1}{x[i]}z[i]^2\right)}
\end{aligned}$$

Using Cauchy-Schwarz inequality in $\mathbb{R}^d$, which is [2]:

$$\left(\sum_{i=1}^{d} u_i v_i\right)^2 \leq \left(\sum_{i=1}^{d} u_i^2\right)\left(\sum_{i=1}^{d} v_i^2\right)$$

we have:

$$\left(\sum_{i=1}^{d} a_i[j]z[j]\right)^2 \leq \left(\sum_{i=1}^{d} a_i[j]^2\right)\left(\sum_{i=1}^{d} z[j]^2\right)$$

Thus we can bound the previous expression and get:

$$\frac{\sum_{i=1}^{n}\left[\frac{w_i}{\langle a_i,x\rangle^2}\left(\sum_{j=1}^{d} a_i[j]z[j]\right)^2\right]}{-\sum_{i=1}^{d}\left(\frac{1}{x[i]}z[i]^2\right)} \leq \frac{\sum_{i=1}^{n}\left[\frac{w_i}{\langle a_i,x\rangle^2}\left(\sum_{j=1}^{d} a_i[j]^2\right)\left(\sum_{j=1}^{d} z[j]^2\right)\right]}{-\sum_{i=1}^{d}\left(\frac{1}{x[i]}z[i]^2\right)}$$

■

[2] *Cauchy-Schwarz inequality*, available at: `https://en.wikipedia.org/wiki/Cauchy-Schwarz_inequality`, accessed: May. 10, 2025.