Optimization Algorithms: HW1

Lo Chun, Chou
R13922136

April 17, 2025

1

2

(1)

Given a twice differentiable function $\varphi : \mathbb{R}^d \to [-\infty, \infty]$, assume that it is logarithmically homogeneous, then by the definition, the following holds:

$$\varphi(\gamma x) = \varphi(x) - \log \gamma, \quad \forall x \in \mathbb{R}^d, \gamma > 0 \tag{1}$$

Claim: $\langle \nabla \varphi(x), x \rangle = -1$

To derive the first equation, we first define the following:

$$F(\gamma) = \varphi(\gamma x)$$

Then the original equation (1) would become:

$$F(\gamma) = \varphi(x) - \log \gamma$$

Taking the derivative w.r.t. γ on both sides, we get:

$$\frac{dF}{d\gamma} = \frac{d}{d\gamma}\varphi(\gamma x) = \nabla \varphi(\gamma x) \cdot x = \langle \nabla \varphi(\gamma x), x \rangle \tag{2}$$

$$\frac{dF}{d\gamma} = \frac{d}{d\gamma}(\varphi(x) - \log \gamma) = -\frac{1}{\gamma} \tag{3}$$

Thus by (2) and (3), we have:

$$\langle \nabla \varphi(\gamma x), x \rangle = -\frac{1}{\gamma}$$

Then by plugging in $\gamma = 1$, we have:

$$\langle \nabla \varphi(x), x \rangle = -1 \qquad \square$$

Claim: $\nabla \varphi(x) = -\nabla^2 \varphi(x) x$

From the previous part, we have:

$$\nabla \varphi(x)^T x = -1$$

Compute the gradient of both sides, for the left hand side, we have:

$$\begin{aligned} \nabla(\nabla \varphi(x)^T x) &= \nabla(\nabla \varphi(x))^T x + \nabla \varphi(x)^T \nabla x \\ &= \nabla^2 \varphi(x) x + \nabla \varphi(x)^T \nabla x \end{aligned}$$

For the right hand side, we have:

$$\nabla(-1) = 0$$

Thus we have:

$$\begin{aligned} &\nabla^2 \varphi(x) x + \nabla \varphi(x)^T \nabla x = 0 \\ \Rightarrow &\nabla \varphi(x)^T \nabla x = -\nabla^2 \varphi(x) x \\ \Rightarrow &\nabla \varphi(x)^T I_d = -\nabla^2 \varphi(x) x \\ \Rightarrow &\nabla \varphi(x) = -\nabla^2 \varphi(x) x \qquad \square \end{aligned}$$

Claim: $\langle x, \nabla^2 \varphi(x) x \rangle = 1$

From the previous part, we have:

$$\nabla \varphi(x) = -\nabla^2 \varphi(x) x$$

Multiply both sides by x^T, we have:

$$x^T \nabla \varphi(x) = -x^T \nabla^2 \varphi(x) x$$

Which is equivalent to the following by using $\langle \nabla \varphi(x), x \rangle = -1$:

$$\langle x, \nabla^2 \varphi(x) x \rangle = -\langle x, \nabla \varphi(x) \rangle = (-1) \times (-1) = 1 \qquad \square$$

(2)

Suppose that $\varphi : \mathbb{R}^d \to [-\infty, \infty]$ is a twice differentiable function, and is strictly convex and logarithmically homogeneous, then the following holds by the definition:

$$\begin{aligned}
&\nabla^2 \varphi(x) > 0 \quad \forall x \in \mathbb{R}^d \\
&\varphi(\gamma x) = \varphi(x) - \log \gamma, \quad \forall x \in \mathbb{R}^d, \gamma > 0.
\end{aligned}$$

Also, we have the following properties from the previous subsection:

$$\langle \nabla \varphi(x), x \rangle = -1 \tag{1}$$
$$\nabla \varphi(x) = -\nabla^2 \varphi(x) x \tag{2}$$
$$\langle x, \nabla^2 \varphi(x) x \rangle = 1 \tag{3}$$

Claim: $\nabla^2 \varphi(x) \geq \nabla \varphi(x) \left(\nabla \varphi(x)\right)^T, \quad \forall x \in \text{dom } \varphi$

The claim is equivalent to proving that:

$$\nabla^2 \varphi(x) - \nabla \varphi(x) \left(\nabla \varphi(x)\right)^T \succeq 0$$

where $\succeq 0$ denotes positive semidefinite.
Let z be any vector in $\mathbb{R}^d$, then we have:

$$\begin{aligned}
z^T \left(\nabla^2 \varphi(x) - \nabla \varphi(x) \left(\nabla \varphi(x)\right)^T\right) z &= z^T \nabla^2 \varphi(x) z - z^T \nabla \varphi(x) \left(\nabla \varphi(x)\right)^T z \\
&= z^T \nabla^2 \varphi(x) z - \left(\nabla \varphi(x)^T z\right)^2 \qquad (*)
\end{aligned}$$

Case 1: $x = z$

If $x = z$, then $(*)$ becomes the following using (1) and (2):

$$\begin{aligned}
z^T \left(\nabla^2 \varphi(x) - \nabla \varphi(x) \left(\nabla \varphi(x)\right)^T\right) z &= z^T \nabla^2 \varphi(x) x - \left(\langle \nabla \varphi(x), x \rangle\right)^2 \\
&= x^T (-\nabla \varphi(x)) - (-1)^2 \\
&= x^T (-\nabla \varphi(x)) - 1 \\
&= -\langle \nabla \varphi(x), x \rangle - 1 \\
&= -(-1) - 1 \\
&= 0 \geq 0
\end{aligned}$$

Case 2: $x \neq z$

Using (2) to replace $\nabla\varphi(x)$ with $-\nabla^2\varphi(x)x$ in $(*)$, and using the fact that $(\nabla^2\varphi(x))^T = \nabla^2\varphi(x)$ (the Hessian is symmetric):

$$\begin{aligned}
z^T\left(\nabla^2\varphi(x) - \nabla\varphi(x)\left(\nabla\varphi(x)\right)^T\right)z &= z^T\nabla^2\varphi(x)z - \left(\nabla\varphi(x)^T z\right)^2 \\
&= z^T\nabla^2\varphi(x)z - \left((-\nabla^2\varphi(x)x)^T z\right)^2 \\
&= z^T\nabla^2\varphi(x)z - \left(-\underbrace{x^T}_{A^T}\underbrace{(\nabla^2\varphi(x))^T z}_{B}\right)^2 \\
&= z^T\nabla^2\varphi(x)z - \underbrace{[(\nabla^2\varphi(x))^T z]^T x x^T [(\nabla^2\varphi(x))^T z]}_{B^T A A^T B} \\
&= z^T\nabla^2\varphi(x)z - [x^T(\nabla^2\varphi(x))^T z]^T[x^T(\nabla^2\varphi(x))^T z] \\
&= z^T\nabla^2\varphi(x)z - ||x^T(\nabla^2\varphi(x))^T z||^2 \\
&= z^T\nabla^2\varphi(x)z - ||x^T(\nabla^2\varphi(x))z||^2
\end{aligned}$$

Let $H = \nabla^2\varphi(x)$, then the above expression is equivalent to:

$$z^T Hz - (x^T Hz)^2$$

Let's first check that we can define the function

$$h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, \quad h(u,v) = u^T Hv$$

as the inner product on $\mathbb{R}^d$. [1]

Using the fact that we assumed that $\nabla^2\varphi(x) > 0$, so H is positive definite, thus by theorem [2], there exists a one and only one positive definite matrix $H^{1/2}$ (also symmetric) such that $H = H^{1/2}H^{1/2}$.

- **Symmetry:** For any $u, v \in \mathbb{R}^d$, we have:

$$\begin{aligned}
h(u,v) &= u^T Hv \\
&= u^T H^{1/2}H^{1/2}v \\
&= (H^{1/2}u)^T(H^{1/2}v) \\
&= (H^{1/2}v)^T(H^{1/2}u) \\
&= v^T H^{1/2}H^{1/2}u \\
&= v^T Hu \\
&= h(v,u)
\end{aligned}$$

[1] H. Amann and J. Escher, *Analysis I*, 1st ed., Birkhäuser Basel, 2005, p. 153.

[2] "Square root of a matrix", Wikipedia, `https://en.wikipedia.org/wiki/Square_root_of_a_matrix`

- **Linearity:** For any $\lambda, \mu \in \mathbb{R}$ and $t, u, v \in \mathbb{R}^d$, we have:

$$
\begin{aligned}
h(t, \lambda u + \mu v) &= t^T H(\lambda u + \mu v) \\
&= t^T H(\lambda u + \mu v) \\
&= t^T H(\lambda u) + t^T H(\mu v) \\
&= \lambda t^T H u + \mu t^T H v \\
&= \lambda h(t, u) + \mu h(t, v)
\end{aligned}
$$

- **Positive definiteness:** For any $u \in \mathbb{R}^d$, we have:

$$h(u, u) = u^T H u > 0 \quad \text{since } H \text{ is positive definite}$$

[3]

Therefore, we have:

$$z^T H z - (x^T H z)^2 = \langle z, z\rangle_H - \langle x, z\rangle_H^2$$

Using the Cauchy-Schwarz inequality [4]:

Cauchy-Schwarz inequality

Let $(E, (\cdot \mid \cdot))$ be an inner product space. Then

$$|(x \mid y)|^2 \leq (x \mid x)(y \mid y), \qquad x, y \in E$$

We can derive the later equation using the fact that:

$$\langle x, x\rangle_H = x^T H x = x^T \nabla^2 \varphi(x) x = 1$$

(this is because property (3))
So we have:

$$
\begin{aligned}
\langle x, z\rangle_H^2 &\leq \langle x, x\rangle_H \langle z, z\rangle_H \\
&= 1 \times \langle z, z\rangle_H \\
&= \langle z, z\rangle_H
\end{aligned}
$$

Thus we have:

$$z^T H z - (x^T H z)^2 \geq 0 \qquad \square$$

[3] I later found that we have "A function $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is an inner product on $\mathbb{R}^n$ if and only if there exists a symmetric positive-definite matrix $\mathbf{M}$ such that $\langle x, y\rangle = x^\top \mathbf{M} y$ for all $x, y \in \mathbb{R}^n$." on "Inner product space", Wikipedia, `https://en.wikipedia.org/wiki/Inner_product_space`

[4] H. Amann and J. Escher, *Analysis I*, 1st ed., Birkhäuser Basel, 2005, p. 154.