A1 Report

Nhat Minh Hoang, Nadina Kapidzic

November 2024

1 Characterization of Function

a)

Given:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in R^2, \quad \mathbf{a} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}, \quad d = 2.5$$

we have

$$\begin{aligned} f(\mathbf{x}) &= (\mathbf{a}^T\mathbf{x} - d)^2 = (-x_1 + 3x_2 - 2.5)^2 \\ &= x_1^2 - 6x_1x_2 + 9x_2^2 + 5x_1 - 15x_2 + 6,25 \end{aligned}$$

Gradient of $f(\mathbf{x})$

$$\frac{\partial f}{\partial x_1} = 2x_1 - 6x_2 + 5$$
$$\frac{\partial f}{\partial x_2} = -6x_1 + 18x_2 - 15$$

Thus,

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2(x_1 - 3x_2 + 2.5) \\ -6(x_1 - 3x_2 + 2.5) \end{pmatrix}$$

Hessian of $f(\mathbf{x})$

$$\frac{\partial^2 f}{\partial x_1^2} = 2$$

$$\frac{\partial^2 f}{\partial x_2^2} = 18$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = -6$$

Using formula for Hessian, we get:

$$H_f(\mathbf{x}) = \begin{pmatrix} 2 & -6 \\ -6 & 18 \end{pmatrix}$$

Stationary Points

Stationary points occur when $\nabla f(\mathbf{x}) = 0$:

$$\begin{pmatrix} 2(x_1 - 3x_2 + 2.5) \\ -6(x_1 - 3x_2 + 2.5) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Both are same linear function with only different factors, which is:

$$x_1 = 3x_2 - 2.5$$

Thus, the stationary points are all the points that lie on the line $x_1 = 3x_2 - 2.5$.

Characterization of Stationary Points

To determine the nature of the stationary points, we analyze the eigenvalues of the Hessian $H_f(\mathbf{x})$. First we solve the characteristic equation $\det(H_f - \lambda I) = 0$:

$$\det \begin{pmatrix} 2-\lambda & -6 \\ -6 & 18-\lambda \end{pmatrix} = 0$$

$$\implies (2-\lambda)(18-\lambda) - (-6)(-6) = \lambda^2 - 20\lambda = (\lambda - 20)\lambda = 0$$
$$\implies \lambda_1 = 0, \lambda_2 = 20$$

The eigenvalues of $H_f(\mathbf{x})$ are positive semi definite, so it is a local minimum. So all points on the line $x_1 = 3x_2 - 2.5$ is local minimum. But note that:

$$f(\mathbf{x}) = (-x_1 + 3x_2 - 2.5)^2$$

Which means f(x) ≥ 0 . In fact, f(x) $= 0$ where :

$$(-x_1 + 3x_2 - 2.5)^2 = 0$$
$$\Leftrightarrow x_1 = 3x_2 - 2.5$$

Therefore, each point on the line $x_1 = 3x_2 - 2.5$ is also a global minimum.

b)

Given:

$$\begin{aligned} f(\mathbf{x}) &= (x_1 - 2)^2 + x_1 x_2^2 - 2 \\ &= x_1^2 - 4x_1 + x_1 x_2^2 + 2 \end{aligned}$$

Gradient of $f(\mathbf{x})$

$$\frac{\partial f}{\partial x_1} = 2x_1 - 4 + x_2^2$$
$$\frac{\partial f}{\partial x_2} = 2x_1 x_2$$

Thus,

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 - 4 + x_2^2 \\ 2x_1 x_2 \end{pmatrix}$$

Hessian of $f(\mathbf{x})$

$$\frac{\partial^2 f}{\partial x_1^2} = 2$$

$$\frac{\partial^2 f}{\partial x_2^2} = 2x_1$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = 2x_2$$

So the Hessian matrix is:

$$H_f(\mathbf{x}) = \begin{pmatrix} 2 & 2x_2 \\ 2x_2 & 2x_1 \end{pmatrix}$$

Stationary Points

Stationary points occur when $\nabla f(\mathbf{x}) = 0$:

$$2x_1 - 4 + x_2^2 = 0$$
$$2x_1 x_2 = 0$$

From the second equation, we have two cases: $x_1 = 0$ or $x_2 = 0$. Substitute these into the first equation, we will get 3 stationary points: $(0, 2)$, $(0, -2)$, and $(2, 0)$.

Characterization of Stationary Points

To determine the nature of each stationary point, we examine the Hessian $H_f(\mathbf{x})$ at each point.

1. **At $(0, 2)$:**

$$H_f(0, 2) = \begin{pmatrix} 2 & 4 \\ 4 & 0 \end{pmatrix}$$

The eigenvalues of $H_f(0, 2)$ are $\lambda_1 = 1 + \sqrt{17}$ and $\lambda_2 = 1 - \sqrt{17}$. Since we have both positive and negative eigenvalues, $(0, 2)$ is a saddle point.

2. **At $(0, -2)$:**

$$H_f(0, -2) = \begin{pmatrix} 2 & -4 \\ -4 & 0 \end{pmatrix}$$

The eigenvalues of $H_f(0, -2)$ are $\lambda_1 = 1 + \sqrt{17}$ and $\lambda_2 = 1 - \sqrt{17}$. So $(0, -2)$ is also a saddle point.

3. **At $(2, 0)$:**

$$H_f(2, 0) = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$$

The eigenvalues of $H_f(2, 0)$ are $\lambda_1 = 2$ and $\lambda_2 = 4$, both positive. Therefore, $(2, 0)$ is a strict local minimum.

In summary:

- The points $(0, 2)$ and $(0, -2)$ are saddle points.
- The point $(2, 0)$ is a strict local minimum.

Function f(x) goes to both positive and negative infinity, so there's no global min/max

Proof:

$$\begin{aligned} f(\mathbf{x}) &= x_1^2 - 4x_1 + x_1x_2^2 + 2 \\ &= x_1(x_1 + x_2^2 - 4) + 2 \end{aligned}$$

The term $(x_1 + x_2^2 - 4)$ has x_2^2 as dominant part, so it can goes to positive infinity. x_1 outside can be positive or negative, which makes f(x) goes to both positive and negative infinity.

c)

Given:

$$\begin{aligned} f(\mathbf{x}) &= x_1^2 + x_1\|\mathbf{x}\|^2 + \|\mathbf{x}\|^2 \\ &= x_1^3 + 2x_1^2 + x_1x_2^2 + x_2^2 \end{aligned}$$

Gradient of $f(\mathbf{x})$

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= 3x_1^2 + x_2^2 + 4x_1 \\ \frac{\partial f}{\partial x_2} &= 2x_1x_2 + 2x_2 \end{aligned}$$

Thus,

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 3x_1^2 + x_2^2 + 4x_1 \\ 2x_1x_2 + 2x_2 \end{pmatrix}$$

Hessian of $f(\mathbf{x})$

$$\frac{\partial^2 f}{\partial x_1^2} = 6x_1 + 4$$

$$\frac{\partial^2 f}{\partial x_2^2} = 2x_1 + 2$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = 2x_2$$

Therefore, the Hessian matrix is:

$$H_f(\mathbf{x}) = \begin{pmatrix} 6x_1 + 4 & 2x_2 \\ 2x_2 & 2x_1 + 2 \end{pmatrix}$$

Stationary Points

Stationary points occur when $\nabla f(\mathbf{x}) = 0$. So:

$$\begin{aligned} 3x_1^2 + x_2^2 + 4x_1 &= 0 \\ 2x_1x_2 + 2x_2 = 0 &\Leftrightarrow (x_1+1)x_2 = 0 \end{aligned}$$

From the second equation, we have two cases: $x_2 = 0$ or $x_1 = -1$.

Case 1: If $x_2 = 0$, substitute into the first equation:

$$3x_1^2 + 4x_1 = 0 \Rightarrow x_1(3x_1 + 4) = 0$$

Thus, $x_1 = 0$ or $x_1 = -\frac{4}{3}$. This gives two stationary points: $(0,0)$ and $\left(-\frac{4}{3}, 0\right)$.

Case 2: If $x_1 = -1$, substitute into the first equation:

$$3(1) + x_2^2 + 4(-1) = 0$$

$$\Rightarrow x_2^2 = 1 \Rightarrow x_2 = \pm 1$$

This gives two more stationary points: $(-1,1)$ and $(-1,-1)$
In summary, the stationary points are $(0,0)$, $\left(-\frac{4}{3}, 0\right)$, $(-1,1)$ and $(-1,-1)$

Characterization of Stationary Points

1. **At $(0,0)$:**

$$H_f(0,0) = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$$

The eigenvalues of $H_f(0,0)$ are $\lambda_1 = 4$ and $\lambda_2 = 2$, both positive. Therefore, $(0,0)$ is a strict local minimum.

2. **At $\left(-\frac{4}{3}, 0\right)$:**

$$H_f\left(-\frac{4}{3}, 0\right) = \begin{pmatrix} -4 & 0 \\ 0 & -\frac{2}{3} \end{pmatrix}$$

The eigenvalues of $H_f\left(-\frac{4}{3}, 0\right)$ are $\lambda_1 = -4$ and $\lambda_2 = -\frac{2}{3}$, both negative. Therefore, $\left(-\frac{4}{3}, 0\right)$ is a strict local maximum.

3. **At $(-1,1)$:**

$$H_f(-1,1) = \begin{pmatrix} -2 & 2 \\ 2 & 0 \end{pmatrix}$$

The eigenvalues of this Hessian are $\lambda_1 = \sqrt{5} - 1$ and $\lambda_2 = -\sqrt{5} - 1$. This is mixed (one positive and one negative), so $(-1,1)$ is a saddle point.

4. **At $(-1,-1)$:**

$$H_f(-1,-1) = \begin{pmatrix} -2 & -2 \\ -2 & 0 \end{pmatrix}$$

The eigenvalues of this Hessian are $\lambda_1 = \sqrt{5} - 1$ and $\lambda_2 = -\sqrt{5} - 1$. This is again a saddle point.

In summary:
- The point $(0,0)$ is a strict local minimum.
- The point $\left(-\frac{4}{3},0\right)$ is a strict local maximum.
- The points $(-1,1)$ and $(-1,-1)$ are saddle points.

The function $f(x) = x_1^3 + 2x_1^2 + x_1x_2^2 + x_2^2$ has x_1^3 as the dominant part, which can be positive or negative, hence f(x) goes to both positive and negative infinity. So no global minimum/maximum

d)

Given:

$$f(\mathbf{x}) = \alpha x_1^2 - 2x_1 + \beta x_2^2$$

Gradient of $f(\mathbf{x})$

$$\frac{\partial f}{\partial x_1} = 2\alpha x_1 - 2$$
$$\frac{\partial f}{\partial x_2} = 2\beta x_2$$

Thus,

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2\alpha x_1 - 2 \\ 2\beta x_2 \end{pmatrix}$$

Stationary Points

Stationary points occur when $\nabla f(\mathbf{x}) = 0$:

$$2\alpha x_1 - 2 = 0 \Rightarrow x_1 = \frac{1}{\alpha}$$
$$2\beta x_2 = 0$$

Stationary point only exists when $\alpha \neq 0$.
If $\beta \neq 0$, then $x_2 = 0$, so we have 1 stationary point at $\left(\frac{1}{\alpha},0\right)$.
Else if $\beta = 0$, there's infinite stationary points on the line $(x_1 = \frac{1}{\alpha})$

Hessian of $f(\mathbf{x})$

$$\frac{\partial^2 f}{\partial x_1^2} = 2\alpha, \quad \frac{\partial^2 f}{\partial x_2^2} = 2\beta, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = 0$$

The Hessian matrix is therefore:

$$H_f(\mathbf{x}) = \begin{pmatrix} 2\alpha & 0 \\ 0 & 2\beta \end{pmatrix}$$

The eigenvalues of this Hessian are $\lambda_1 = 2\alpha$ and $\lambda_2 = 2\beta$.

Characterization of Stationary Point

The function $f(\mathbf{x}) = \alpha x_1^2 - 2x_1 + \beta x_2^2$ has x_1^2, x_2^2 as dominant parts, so the signs of α and β will decide whether f(x) is bounded or not:

- $\alpha > 0$ and $\beta \geq 0$: f(x) $\geq 0 \Rightarrow$ There exists Global Minimum
- $\alpha < 0$ and $\beta \leq 0$: f(x) $\leq 0 \Rightarrow$ There exists Global Maximum

Since α and β are also the eigenvalues of Hessian, its sign will decide the characterization of the stationary point. As mentioned above, the stationary points only exist when $\alpha \neq 0$, so either $\alpha > 0$ or $\alpha < 0$. For β, there are 3 cases:

1. $\beta < 0$: the stationary point is $\left(\frac{1}{\alpha}, 0\right)$
   - $\alpha < 0$: strict global maximum
   - $\alpha > 0$: saddle point
2. $\beta = 0$: the stationary points are the on the line $(x_1 = \frac{1}{\alpha})$
   - $\alpha < 0$: global maximum
   - $\alpha > 0$: global minimum
2. $\beta > 0$: the stationary point is $\left(\frac{1}{\alpha}, 0\right)$
   - $\alpha < 0$: saddle point
   - $\alpha > 0$: strict global minimum

2 Matrix Calculus

a)

$$f(x) = \frac{1}{4}\|x-b\|^4; \quad x, b \in \mathbb{R}^n$$

$$\|x-b\|^4 = \left(\sum_{i=1}^{n}(x_i-b_i)^2\right)^2$$

$$f(x) = \frac{1}{4}\left(\sum_{i=1}^{n}(x_i-b_i)^2\right)^2$$

if $g(x) = \sum_{i=1}^{n}(x_i-b_i)^2$ then we can write $f(x) = \frac{1}{4}(g(x))^2$

1. Gradient

$\frac{\partial}{\partial x_k} g(x) = 2(x_k - b_k)$

in vector notation: $\nabla g(x) = 2(x-b)$

$$\nabla f(x) = \frac{\partial}{\partial x}\left[\frac{1}{4}(g(x))^2\right] = \frac{1}{4} \cdot 2g(x) \cdot \nabla g(x)$$

$$g(x) = \sum_{i=1}^{n}(x_i-b_i)^2 \text{ and } \nabla g(x) = 2(x-b)$$

$$\begin{aligned}\nabla f(x) &= \frac{1}{2}g(x)2(x-b) \\ &= \sum_{i=1}^{n}(x_i-b_i)^2(x-b)\end{aligned}$$

2. Hessian $H_f(x) \Rightarrow$ derive each component of $\nabla f(x)$ with respect to each x_j :

$$\nabla f(x) = \underbrace{\sum_{i=1}^{n}(x_i-b_i)^2}_{g(x)}(x-b)$$

$$H_f(x) = \nabla^2 f(x) = \underbrace{\nabla g(x) \cdot (x-b)^\top}_{\in \mathbb{R}^{n\times n}} + \underbrace{g(x) \cdot I}_{\substack{\text{diagonal matrix} \\ \text{with } g(x) \text{ on diagonal}}}$$

$$\Rightarrow \quad \nabla g(x) \cdot (x-b)^\top = 2(x-b) \cdot (x-b)^\top$$

$$\Rightarrow \quad g(x) \cdot I = \underbrace{\sum_{i=1}^{n}(x_i-b_i)^2}_{\text{this term on each diagonal entry}} \cdot I$$

$$H_f(x) = 2(x-b)(x-b)^\top + \left(\sum_{i=1}^{n}(x_i-b_i)^2\right) \cdot I$$

b)

1. Gradient

$$f(x) = \sum_{i=1}^{n} g((Ax)_i) \quad \text{with } g(z) = \frac{1}{2}z^2 + z \quad z \in \mathbb{R}, A \in \mathbb{R}^{n\times n}$$

$$Ax = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \\ \vdots & & \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots \cdot + a_{1n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^n a_{1i}x_i \\ \vdots \\ \sum_{i=1}^n a_{ni}x_i \end{pmatrix}; \quad (Ax)_i = \sum_{j=1}^{n} a_{ij}x_j$$

$$f(x) = \sum_{i=1}^{n} g((Ax)_i) = \sum_{i=1}^{n} g\left(\sum_{j=1}^{n} a_{ij}x_j\right)$$

$$\frac{\partial f(x)}{\partial x} = \frac{\partial}{\partial x} \sum_{i=1}^{n} g\left(\sum_{j=1}^{n} a_{ij}x_j\right) =$$

$$= \frac{\partial}{\partial x}\left[\sum_{i=1}^{n}\left(\frac{1}{2}\left(\sum_{j=1}^{n} a_{ij}x_j\right)^2 + \sum_{j=1}^{n} a_{ij}x_j\right)\right]$$

$$\rightarrow \frac{\partial}{\partial x_k}\left[\sum_{i=1}^{n}\frac{1}{2}\left(\sum_{j=1}^{n} a_{ij}x_j\right)^2\right] = 2\cdot\frac{1}{2}\left(\sum_{j=1}^{n} a_{ij}x_j\right)\cdot a_{ik}$$

$$\rightarrow \frac{\partial}{\partial x_k}\left[\sum_{j=1}^{n} a_{ij}x_j\right] = a_{ik}$$

$$\frac{\partial f}{\partial x_k} = \sum_{i=1}^{n}\left(a_{ik}\sum_{j=1}^{n} a_{ij}x_j + a_{ik}\right)$$

$$\nabla f(x) = \begin{pmatrix} \sum_{i=1}^n \left(a_{i1}\left(\sum_{j=1}^n a_{ij}x_j\right) + a_{i1}\right) \\ \vdots \\ \sum_{i=1}^n \left(a_{in}\left(\sum_{j=1}^n a_{ij}x_j\right) + a_{in}\right) \end{pmatrix}$$

$$\nabla f(x) = \left(A^\top(Ax)\right) + A^\top \cdot \overrightarrow{1}$$

2. Hessian

$$\nabla^2 f(x) = ?$$

$$\frac{\partial}{\partial x_k^2}\left[\sum_{i=1}^{n}\left(a_{ik}\left(\sum_{j=1}^{n} a_{ij}x_j\right)+a_{ik}\right)\right]$$

$$\hookrightarrow \frac{\partial}{\partial {x_k}^2} = \sum_{i=1}^{n} {a_{ik}}^2$$

$$\hookrightarrow \frac{\partial}{\partial x_k x_l} = \sum_{i=1}^{n} a_{ik}\cdot a_{il}$$

$$\nabla^2 f(x) = \begin{pmatrix} \sum_{i=1}^{n} {a_{i1}}^2 & \sum_{i=1}^{n} a_{i1}\cdot a_{i2} & \cdots & \sum_{i=1}^{n} a_{i1}\cdot a_{in} \\ \sum_{i=1}^{n} a_{i1}a_{i2} & \sum_{i=1}^{n} {a_{i2}}^2 & \cdots & \sum_{i=1}^{n} a_{i2}\cdot a_{in} \\ \vdots & & & \\ \sum_{i=1}^{n} a_{i1}a_{in} & \sum_{i=1}^{n} a_{in}a_{i2} & \cdots & \sum_{i=1}^{n} {a_{in}}^2 \end{pmatrix}$$

$$\to \sum_{i=1}^{n} a_{ik}a_{il} = \sum_{i=1}^{n} a_{ki}^{\top}a_{il} = \left(A^{\top}A\right)_{KL}$$

$$\nabla^2 f(x) = A^{\top}A$$

(c)

$$f(x) = (x \oslash b)^{\top} D (x \oslash b) \quad \text{for} \quad b \in \mathbb{R}^n, D \in \mathbb{R}^{n\times n}$$

$$f(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} (x\oslash b)_i D_{ij} (x\oslash b)_j$$

$$(x\oslash b)_i = \frac{x_i}{b_i}; \quad (x\oslash b)_j = \frac{x_j}{b_j}$$

$$f(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{x_i}{b_i} D_{ij} \frac{x_j}{b_j}$$

1. Gradient
$\nabla f(x) \leftarrow$ differentiate $f(x)$ with respect to x_k for each $k = 1, \ldots, n$

$$f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{x_i}{b_i} D_{ij} \frac{x_j}{b_j}$$

$$\text{if } x_i = x_k, \quad \frac{\partial}{\partial x_k} \left(\frac{x_k}{b_i} \right) = \frac{1}{b_i}$$

$$\text{if } x_j = x_k, \quad \frac{\partial}{\partial x_k} \frac{x_k}{b_j} = \frac{1}{b_j}$$

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{x_i}{b_i} D_{ij} \frac{x_j}{b_j} \right) = \sum_{j=1}^{n} \frac{1}{b_k} D_{kj} \frac{x_j}{b_j} + \sum_{i=1}^{n} \frac{x_i}{b_i} D_{ik} \frac{1}{b_k}$$

$$= \frac{1}{b_k} \sum_{j=1}^{n} D_{k_j} \cdot \frac{x_j}{b_j} + \frac{1}{b_k} \sum_{i=1}^{n} \frac{x_i}{b_i} D_{ik} = \frac{1}{b_k} \left(\sum_{j=1}^{n} D_{kj} \frac{x_j}{b_j} + \sum_{i=1}^{n} \frac{x_i}{b_i} D_{ik} \right)$$

$$\frac{\partial f(x)}{\partial x_k} = D \cdot (x \oslash b) \oslash b + D^\top (x \oslash b) \oslash b$$

2. Hessian

$$\frac{\partial f}{\partial x_k} = \frac{1}{b_k} \sum_{j=1}^{n} D_{kj} \frac{x_j}{b_j} + \frac{1}{b_k} \sum_{i=1}^{n} \frac{x_i}{b_i} D_{ik}$$

$$H_f(x) = \frac{\partial^2 f}{\partial x_k \partial x_l}$$

$$\to \frac{\partial}{\partial x_l} \left(\frac{1}{b_k} \frac{x_i}{b_i} D_{ik} \right)$$

$\hookrightarrow$ 1. if $x_i = x_l$ ie. $i = l \Rightarrow \frac{D_{lk}}{b_k b_l}$
$\hookrightarrow$ 2. If $x_i \neq x_l$ i.e. $i \neq l \Rightarrow$ derivative is equal to zero.

$$\to \frac{\partial}{\partial x_l} \left(\frac{1}{b_k} D_{k_j} \frac{x_j}{b_j} \right),$$

$\hookrightarrow$ 1. if $x_j = x_l$ ie. $j = l \Rightarrow \frac{D_{lk}}{b_k b_l}$
$\hookrightarrow$ 2. If $x_j \neq x_l$ i.e. $j \neq l \Rightarrow$ derivative is equal to zero.

So, each element H_{kl} of the Hessian matrix is:

$$H_{kl} = \frac{\partial^2 f(x)}{\partial x_k \partial x_l} = \frac{D_{kl}}{b_k b_l}$$

$$H_f(x) = \frac{D}{b \cdot b^\top} \quad ; \quad \to b \cdot b^\top \text{ reslts in an } n \times n$$

matrix where each dement is $b_i b_j$ in the (i, j)-position.

3 Numerical Gradient Verification

c)

In numerical approximation using central differences, choosing smaller ϵ (step size) will generally result in smaller error for the approximation. This is shown in 1c, 2a, 2b and 2c, where choosing $\epsilon = 0.01$ and smaller results in error of near 0.

But choosing a very small ϵ can actually increase error due to **round-off errors** in floating-point arithmetic. Especially functions with rough surface like 1a, 1b and 1d, a small round off error is enough to be a big difference between the approximation and analatically computed gradient

Therefore, the ideal ϵ is smaller ϵ, but not the smallest possible. Instead a **moderately small value** that balances round-off error and error from approximating the derivative.

Based on the plotted graph, the best ϵ is around 10^{-4}, which result in error of near/exact 0 in all functions.

4 The Diet Problem

(a) Decision Variables

Let:

- x_{milk}: amount of Milk (in grams per 100g serving)
- x_{tomatoes}: amount of Tomatoes (in grams per 100g serving)
- x_{bananas}: amount of Bananas (in grams per 100g serving)
- x_{apples}: amount of Apples (in grams per 100g serving)
- x_{noodles}: amount of Noodles (in grams per 100g serving)
- x_{lettuce}: amount of Lettuce (in grams per 100g serving)
- x_{bread}: amount of Bread (in grams per 100g serving)
- x_{eggs}: amount of Eggs (in grams per 100g serving)
- x_{meat}: amount of Meat (in grams per 100g serving)
- x_{fish}: amount of Fish (in grams per 100g serving)

Each decision variable x_i represents the amount of each food to be included in the diet, measured in servings of 100g.

(b) Objective Function

The objective is to minimize the total cost of the diet. Based on the cost data provided in the table, we formulate the objective function as follows:

$$\begin{aligned}\text{Minimize } Z = {} & 0.2x_{\text{milk}} + 0.6x_{\text{tomatoes}} + 0.25x_{\text{bananas}} + 0.3x_{\text{apples}} \\ & + 0.3x_{\text{noodles}} + 0.1x_{\text{lettuce}} + 0.4x_{\text{bread}} \\ & + 0.6x_{\text{eggs}} + 0.9x_{\text{meat}} + 1.2x_{\text{fish}}. \qquad (1)\end{aligned}$$

This equation represents the total cost, where each term is the cost per 100g serving of each food multiplied by its corresponding decision variable.

(c) Constraints

We can translate each nutritional requirement into a constraint.

1. Energy Intake (kJ)

The required energy intake is 8000 ± 150 kJ, meaning it should be between 7850 and 8150 kJ.

$$\begin{aligned}7850 \leq 266x_{\text{milk}} &+ 80x_{\text{tomatoes}} + 373x_{\text{bananas}} + 217x_{\text{apples}} \\ &+ 1521x_{\text{noodles}} + 49x_{\text{lettuce}} + 937x_{\text{bread}} \\ &+ 648x_{\text{eggs}} + 485x_{\text{meat}} + 932x_{\text{fish}} \leq 8150,\end{aligned} \tag{2}$$

2. Lipids (g)

The total lipid intake should be between 35g and 70g.

$$\begin{aligned}35 \leq 3.5x_{\text{milk}} &+ \\ &+ 2x_{\text{noodles}} + 0.2x_{\text{lettuce}} + 0.8x_{\text{bread}} \\ &+ 11x_{\text{eggs}} + 3.1x_{\text{meat}} + 16x_{\text{fish}} \leq 70,\end{aligned} \tag{3}$$

3. Carbohydrates (g)

The total carbohydrate intake should be between 180g and 300g.

$$\begin{aligned}180 \leq 4.6x_{\text{milk}} &+ 3.9x_{\text{tomatoes}} + 23x_{\text{bananas}} + 14x_{\text{apples}} \\ &+ 71x_{\text{noodles}} + 1.1x_{\text{lettuce}} + 46x_{\text{bread}} \\ &+ 0.7x_{\text{eggs}} \leq 300,\end{aligned} \tag{4}$$

4. Proteins (g)

The total protein intake should be between 40g and 160g.

$$\begin{aligned}40 \leq 3.4x_{\text{milk}} &+ 1.1x_{\text{bananas}} \\ &+ 13x_{\text{noodles}} + 1.3x_{\text{lettuce}} + 4.7x_{\text{bread}} \\ &+ 13x_{\text{eggs}} + 22x_{\text{meat}} + 20x_{\text{fish}} \leq 160,\end{aligned} \tag{5}$$

5. Fiber (g)

The minimum fiber intake required is 30g.

$$1.2x_{\text{tomatoes}} + 2.6x_{\text{bananas}} + 2.4x_{\text{apples}} + 3x_{\text{noodles}} + 1.3x_{\text{lettuce}} + 0.5.x_{\text{bread}} \geq 30$$

6. Vitamin C (mg)

The minimum vitamin C intake required is 80mg.

$$
\begin{aligned}
80 \leq\ & 13.7x_{\text{milk}} + 13.7x_{\text{tomatoes}} + 8.7x_{\text{bananas}} + 4.6x_{\text{apples}} \\
& + 9.2x_{\text{lettuce}},
\end{aligned} \tag{6}
$$

7. Sodium (NaCl) (mg)

The maximum sodium intake allowed is 3000mg.

$$130x_{\text{milk}} + 2200x_{\text{bread}} + 150x_{\text{meat}} + 100x_{\text{fish}} \leq 3000$$

8. Non-negativity Constraints

Each decision variable must be non-negative, meaning:

$$x_{\text{milk}}, x_{\text{tomatoes}}, x_{\text{bananas}}, x_{\text{apples}}, x_{\text{noodles}}, x_{\text{lettuce}}, x_{\text{bread}}, x_{\text{eggs}}, x_{\text{meat}}, x_{\text{fish}} \geq 0$$

(d) Refer to main.py

(e) Resulting Diet

Diet without fish requirement:
noodles : 397.0g
lettuce : 1391.6g
eggs : 220.7g

(f) Total cost of the diet

Total cost: €3.91 per day

(g) Diet Healthy?

The generated diet fulfills the constraints with the goal of minimizing the cost, it however completely neglects the sodium intake, and can not be considered a balanced diet with a variety of choices.

(h) Minimum Fish intake set to 30g

Diet with minimum 30g fish requirement:
noodles : 397.2g
lettuce : 1391.0g
eggs : 177.0g

fish : 30.0g

New total cost: €4.00 per day

The total cost of the diet went up, just to accommodate the new requirement and the egg amount lowered.