

Numerical optimization with python - Dry HW 1

Roni Ben-Dom 207576463

Part 1- Modeling optimization problems:

- Denote unknown variables as x_1, x_2, x_3 where:

x_1 - # of kg of vanilla ice cream

x_2 - # of kg of chocolate ice cream

x_3 - # of kg of strawberry ice cream

The total cost to maximize: $7x_1 + 6x_2 + 5x_3$

Constraint of machine production: $x_1 + x_2 + x_3 \leq 100$

constraint on milk supplies: $0.5x_1 + 0.2x_2 + 0.4x_3 \leq 45$

constraint on cream supplies: $0.4x_1 + 0.7x_2 + 0.4x_3 \leq 60$

constraint on vanilla supplies: $0.1x_1 \leq 10$

constraint on chocolate supplies: $0.1x_2 \leq 10$

constraint on strawberries supplies: $0.2x_3 \leq 15$

non negativity constraint: $x_i \geq 0 \quad \forall i \in \{1, 2, 3\}$

Therefore, we arrived at the following formulation:

$\max(7x_1 + 6x_2 + 5x_3)$ subject to:

$$x_1 + x_2 + x_3 \leq 100$$

$$0.5x_1 + 0.2x_2 + 0.4x_3 \leq 45$$

$$0.4x_1 + 0.7x_2 + 0.4x_3 \leq 60$$

$$0.1x_1 \leq 10 \Rightarrow x_1 \leq 100$$

$$0.1x_2 \leq 10 \Rightarrow x_2 \leq 100$$

$$0.2x_3 \leq 15 \Rightarrow x_3 \leq 75$$

$$x_i \geq 0 \quad \forall i \in \{1, 2, 3\}$$

2. Denote unknown variables as p_i , $i \in \{1, 2, 3, 4\}$

The total value to maximize: $-\sum_{i=1}^4 p_i \log(p_i)$

non-negativity constraint: $\forall i \in \{1, 2, 3, 4\}: p_i \geq 0$

probability distribution constraint: $\sum_{i=1}^4 p_i = 1$

Expected value constraint: $\sum_{i=1}^4 p_i x_i = 0$

Mathematical formulation:

$\max \left(-\sum_{i=1}^4 p_i \log(p_i) \right)$ subject to:

- $p_i \geq 0 \quad \forall i \in \{1, 2, 3, 4\}$

- $\sum_{i=1}^4 p_i = 1$

- $-10.2 p_1 + 0.4 p_2 + 16.6 p_3 + 10.3 p_4 = 0$ (in general case: $\sum_{i=1}^4 p_i x_i = 0$)

unknown variable

3. Let $x = (x_1, \dots, x_n)$. We want to find x that minimizes $\|x - p\|_2$ for a given $p \in \mathbb{R}^n$, x is in the northern half closed unit ball.

Mathematical formulation:

$\min(\|x - p\|_2)$ subject to:

$\sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \leq 1$

$x_i \geq 0$ for all $i \in \{1, \dots, n\}$

a. fulfilled in the last step.

b. now we want $\|x - p\|_1$ to be minimal:

$\min(\|x - p\|_1)$ subject to:

$|x_1| + |x_2| + \dots + |x_n| \leq 1$

$x_i \geq 0$ for all $i \in \{1, \dots, n\}$

The definition of the absolute value is not differentiable at the origin. But, given the second constraint we can write all the $|x_i|$ as $x_i + \text{sign}(x_i) \cdot x_i$ and then the problem is smooth.

The problem is not smooth due to the absolute value.

3.C. now let $x = (x_1, x_2, \dots, x_n, x_{n+1}, \dots, x_m)$ be the unknown variable. $P = (P_1, \dots, P_n, P_{n+1}, \dots, P_m)$

We want to minimize $\|x - p\|_1$ and have a smooth problem with the same interpretation for x_1, \dots, x_n .

Mathematical Formulation:

$$|x_1| + |x_2| + \dots + |x_n| \leq 1$$

$$x_i \geq 0 \quad \forall i \in \{1, \dots, n\}$$

$$x_i = x_{j+n} \quad \forall i \in \{1, \dots, n\}$$

The third constraint creates the next situation:

$$|x_1 - p_1| + |x_2 - p_2| + \dots + |x_n - p_n| + |x_{n+1} - p_{n+1}| + \dots + |x_{2n} - p_{2n}| = \\ x_1 \quad p_1 \quad x_n \quad p_n$$

$$= (|x_1 - p_1|)^2 + (|x_2 - p_2|)^2 + \dots + (|x_n - p_n|)^2 = (x_1 - p_1)^2 + (x_2 - p_2)^2 + \dots + (x_n - p_n)^2$$

since all values bound to be ~~positive~~ non negative by the second constraint minimizing this problem is the same as minimizing $\|x-p\|_1$ in the previous section, because the square function is monotonically increasing, just like the absolute value.

Part 2- General preview material

$$1. a. \frac{\partial f(x)}{\partial x_i} = 1$$

$$b. \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} n \times 1 \text{ vector}$$

$$C \quad \frac{\partial f}{\partial u} = \langle \nabla f, u \rangle = \nabla f^\top u = \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}} + \dots + \frac{1}{\sqrt{n}} = \frac{n+1}{\sqrt{n}} = \sqrt{n}$$

it does not depend on x , f has no squared components

so any x_i does not appear in the gradient, hence the directional derivative does not depend on x .

d. $\nabla f(x) = \nabla a^T x = a^T$

2. a We want to find Q s.t $x^T Q x = x_1^2 + 2x_2^2 + x_1 x_2$.

$$Q = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \Rightarrow x^T Q x = (x_1 a_{11} + x_2 a_{21}, x_1 a_{12} + x_2 a_{22}) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= a_{11} x_1^2 + a_{21} x_2 x_1 + a_{12} x_1 x_2 + a_{22} x_2^2 \Rightarrow a_{11} = 1$$

$$Q = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix}$$

$$a_{22} = 2$$

$$a_{21} + a_{12} = 1 \Rightarrow a_{12} = a_{21} = 0.5$$

Q is symmetric

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

$$\frac{\partial f}{\partial x_2} = 4x_2 + x_1 \Rightarrow \nabla f = \begin{pmatrix} 2x_1 + x_2 \\ 4x_2 + x_1 \end{pmatrix}$$

c. assume we have $x^T Q x$ for some $Q \in \mathbb{R}^{n \times n}$ we can

write $x^T Q x$ as following: $x^T \left(\sum_{j=1}^n a_{ij} x_j \right)$

** can be written as:

$$\sum_{i=1}^n \left(a_{ii} x_i^2 + \sum_{j \neq i} x_i a_{ij} x_j \right) = \sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} x_j$$

Taking derivative with respect to some element K :

$$2a_{11} x_1 + \sum_{j \neq 1} a_{11j} x_j + \sum_{j \neq 1} x_j a_{11j} = \sum_{i=1}^n a_{ii} x_i + \sum_{i=1}^n x_i a_{ii}$$

$$\Rightarrow \nabla f(x) = \begin{pmatrix} \sum_{i=1}^n x_i a_{1i} + \sum_{i=1}^n a_{1i} x_i \\ \sum_{i=1}^n x_i a_{2i} + \sum_{i=1}^n a_{2i} x_i \\ \vdots \\ \sum_{i=1}^n x_i a_{ni} + \sum_{i=1}^n a_{ni} x_i \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n x_i a_{1i} \\ \sum_{i=1}^n x_i a_{2i} \\ \vdots \\ \sum_{i=1}^n x_i a_{ni} \end{pmatrix} + \begin{pmatrix} \sum_{i=1}^n a_{1i} x_i \\ \sum_{i=1}^n a_{2i} x_i \\ \vdots \\ \sum_{i=1}^n a_{ni} x_i \end{pmatrix}$$

$$= Q^T x + Q x = (Q^T + Q) x \stackrel{Q \text{ is symmetric}}{=} 2Q x$$

$$d. \frac{\partial^2 f}{\partial x_1^2} = \frac{\partial \nabla f}{\partial x_1} = 2 = \frac{\partial}{\partial x_1} \circ \frac{\partial f}{\partial x_1}$$

$$\frac{\partial^2 f}{\partial x_2^2} = \frac{\partial \nabla f}{\partial x_2} = 4 = \frac{\partial}{\partial x_2} \circ \frac{\partial f}{\partial x_2} \Rightarrow \nabla^2 f = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial}{\partial x_1} \circ \frac{\partial f}{\partial x_2} = \frac{\partial}{\partial x_1} (4x_2 + x_1) = 1 = \frac{\partial^2 f}{\partial x_2 \partial x_1}$$

$$e. \text{ based on c, } \nabla f = 2Q x \Rightarrow \nabla^2 f = 2Q$$

f. as J showed in c, the general form of ∇f is $(Q^T + Q)x$

The general form of $\nabla^2 f = Q^T + Q$

$$3.a. f(t) = x + tu$$

$$b. f(f(t)) = f(x + tu) = g(t)$$

$$c. g'(t) = \frac{d}{dt}(g(t)) = \frac{d}{dt}(f(x + tu)) = \nabla f^T(x + tu) \cdot u = \nabla f^T u$$

d. By chain rule after equation 1, $\nabla^2 f(x + tu) \cdot u u$

e. ~~From 2 we have: $\nabla f = \begin{pmatrix} 2x_1 + x_2 \\ 4x_2 + x_1 \end{pmatrix}$~~

f. From 2 we have: $\nabla f = \begin{pmatrix} 2x_1 + x_2 \\ 4x_2 + x_1 \end{pmatrix}$

The hessian is:

$$\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$$

Because we have: $\frac{\partial^2 f}{\partial t^2} = \begin{pmatrix} \cancel{2} & \cancel{1} \\ \cancel{1} & \cancel{4} \end{pmatrix} \begin{pmatrix} \cancel{2} & \cancel{1} \\ \cancel{1} & \cancel{4} \end{pmatrix} = \begin{pmatrix} \cancel{2} & \cancel{1} \\ \cancel{1} & \cancel{4} \end{pmatrix}$

d. $g''(t) = \frac{d^2}{dt^2}(f(x + tu)) = \frac{d}{dt}(g'(t)) = \frac{d}{dt}(\nabla f(x + tu) \cdot u) = u^T \nabla^2 f(x + tu) \cdot u = u^T H(f(x + tu)) \cdot u$

e. The same as we had in first directional derivative;
we want $t=0$:

$$\frac{\partial^2 f}{\partial u^2} = u^T H(f(x)) \cdot u = g''(0)$$

f. Continued:

$$\frac{\partial^2 f}{\partial u^2} = \left(\frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}}\right) \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \left(\frac{3}{\sqrt{2}} \quad \frac{5}{\sqrt{2}}\right) \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{3}{2} + \frac{5}{2} = 4$$

Index of comments

- 2.1 Only x_n should be non-negative. Requiring all components are non-negative gives the $1/2^n$ part of it, not a half (e.g. the upper right quarter in 2D). -1pts
 - 2.2 The above objective is not smooth as it has a square root, and so is the constraint. -1pts
 - 2.3 Note that we did not require to change the constraint to L1 unit ball, only the objective is required to switch to L1 -1pts
-
- 3.1 No, this is not the same objective function as L1 minimization and will not yield the same solutions. See posted solution for details. -3pts
 - 3.2 Just a , as a column vector (for gradient, not the differential linear operator which is indeed a row)