University of Cyprus
MAI613 - Research Methodologies and Professional Practices in AI
Lecturer: Stelios Timotheou

Exercises on Optimization

Exercises

1. Solve Exercises 7.3, 7.4, 7.6, 7.7, 7.8, from the book:
M.P. Deisenroth, A. A. Faisal, and C. S. Ong, "Mathematics for Machine Learning," Cambridge University Press, 2020.
https://mml-book.github.io/book/mml-book.pdf

2. Consider the optimization problem:

$$\begin{aligned}
&\text{minimize } \; f(\mathbf{x}) = 2x_1^2 + 4x_2^2 + 2x_1x_2 - 2x_2 \\
&\text{subject to: } x_1 + x_2 \geq 4 \\
&\qquad\qquad\quad\;\; x_1 - x_2 \leq 2
\end{aligned}$$

   i. Prove that the above problem is convex.
   ii. Using the Karesh-Kuhn-Tucker (KKT) conditions compute the optimal solution $\boldsymbol{x}^*$ and the optimal objective value $f(\boldsymbol{x}^*)$.

Note: These exercises are intended for self-assessment purposes.

7

Continuous Optimization

Exercises

7.3 Consider whether the following statements are true or false:

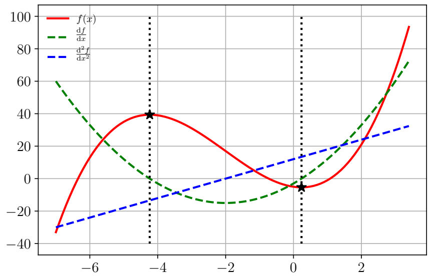

Figure 7.1 A plot of the function $f(x)$ along with its gradient and Hessian.

a. The intersection of any two convex sets is convex.
b. The union of any two convex sets is convex.
c. The difference of a convex set A from another convex set B is convex.

a. true
b. false
c. false

7.4 Consider whether the following statements are true or false:

a. The sum of any two convex functions is convex.
b. The difference of any two convex functions is convex.
c. The product of any two convex functions is convex.
d. The maximum of any two convex functions is convex.

a. true
b. false
c. false
d. true

7.6 Consider the linear program illustrated in Figure 7.9,

$$\min_{\boldsymbol{x} \in \mathbb{R}^2} - \begin{bmatrix} 5 \\ 3 \end{bmatrix}^\top \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\text{subject to} \quad \begin{bmatrix} 2 & 2 \\ 2 & -4 \\ -2 & 1 \\ 0 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leqslant \begin{bmatrix} 33 \\ 8 \\ 5 \\ -1 \\ 8 \end{bmatrix}$$

Derive the dual linear program using Lagrange duality.

Write down the Lagrangian:

$$\mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = -\begin{bmatrix} 5 \\ 3 \end{bmatrix}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top \left(\begin{bmatrix} 2 & 2 \\ 2 & -4 \\ -2 & 1 \\ 0 & -1 \\ 0 & 1 \end{bmatrix} \boldsymbol{x} - \begin{bmatrix} 33 \\ 8 \\ 5 \\ -1 \\ 8 \end{bmatrix} \right)$$

Rearrange and factorize $\boldsymbol{x}$:

$$\mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \left(-\begin{bmatrix} 5 \\ 3 \end{bmatrix}^\top + \boldsymbol{\lambda}^\top \begin{bmatrix} 2 & 2 \\ 2 & -4 \\ -2 & 1 \\ 0 & -1 \\ 0 & 1 \end{bmatrix} \right) \boldsymbol{x} - \boldsymbol{\lambda}^\top \begin{bmatrix} 33 \\ 8 \\ 5 \\ -1 \\ 8 \end{bmatrix}$$

Differentiate with respect to $\boldsymbol{x}$ and set to zero:

$$\nabla_{\boldsymbol{x}} \mathfrak{L} = -\begin{bmatrix} 5 \\ 3 \end{bmatrix}^\top + \boldsymbol{\lambda}^\top \begin{bmatrix} 2 & 2 \\ 2 & -4 \\ -2 & 1 \\ 0 & -1 \\ 0 & 1 \end{bmatrix} = 0$$

Then substitute back into the Lagrangian to obtain the dual Lagrangian:

$$\mathfrak{D}(\boldsymbol{\lambda}) = -\boldsymbol{\lambda}^\top \begin{bmatrix} 33 \\ 8 \\ 5 \\ -1 \\ 8 \end{bmatrix}$$

The dual optimization problem is therefore

$$\begin{aligned} \max_{\boldsymbol{\lambda} \in \mathbb{R}^m} \quad & -\boldsymbol{\lambda}^\top \begin{bmatrix} 33 \\ 8 \\ 5 \\ -1 \\ 8 \end{bmatrix} \\ \text{subject to} \quad & -\begin{bmatrix} 5 \\ 3 \end{bmatrix}^\top + \boldsymbol{\lambda}^\top \begin{bmatrix} 2 & 2 \\ 2 & -4 \\ -2 & 1 \\ 0 & -1 \\ 0 & 1 \end{bmatrix} = 0 \\ & \boldsymbol{\lambda} \geqslant \boldsymbol{0}. \end{aligned}$$

7.7 Consider the quadratic program illustrated in Figure 7.4,

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} \frac{1}{2}\begin{bmatrix}x_1\\x_2\end{bmatrix}^\top \begin{bmatrix}2 & 1\\1 & 4\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} + \begin{bmatrix}5\\3\end{bmatrix}^\top \begin{bmatrix}x_1\\x_2\end{bmatrix}$$

$$\text{subject to} \begin{bmatrix}1 & 0\\-1 & 0\\0 & 1\\0 & -1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} \leqslant \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

Derive the dual quadratic program using Lagrange duality.

Let $\boldsymbol{Q} = \begin{bmatrix}2 & 1\\1 & 4\end{bmatrix}, \boldsymbol{c} = \begin{bmatrix}5\\3\end{bmatrix}, \boldsymbol{A} = \begin{bmatrix}1 & 0\\-1 & 0\\0 & 1\\0 & -1\end{bmatrix}$, and $\boldsymbol{b} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$. Then by (7.45) and (7.52) the dual optimization problem is

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^4} -\frac{1}{2}\left(\begin{bmatrix}5\\3\end{bmatrix} + \begin{bmatrix}1 & 0\\-1 & 0\\0 & 1\\0 & -1\end{bmatrix}^\top \boldsymbol{\lambda}\right)^\top \begin{bmatrix}2 & 1\\1 & 4\end{bmatrix}^{-1}\left(\begin{bmatrix}5\\3\end{bmatrix} + \begin{bmatrix}1 & 0\\-1 & 0\\0 & 1\\0 & -1\end{bmatrix}^\top \boldsymbol{\lambda}\right) - \boldsymbol{\lambda}^\top \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

subject to $\boldsymbol{\lambda} \geqslant \boldsymbol{0}$.

which expands to

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^4} -\frac{1}{14}\left(88 + \begin{bmatrix}33\\-35\\1\\-3\end{bmatrix}^\top \boldsymbol{\lambda} + \boldsymbol{\lambda}^\top \begin{bmatrix}4 & -4 & -1 & 1\\-4 & 4 & 1 & -1\\-1 & 1 & 2 & -2\\1 & -1 & -2 & 2\end{bmatrix}\boldsymbol{\lambda}\right)$$

subject to $\boldsymbol{\lambda} \geqslant \boldsymbol{0}$.

Alternative derivation

Write down the Lagrangian:

$$\mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^\top \begin{bmatrix}2 & 1\\1 & 4\end{bmatrix}\boldsymbol{x} + \begin{bmatrix}5\\3\end{bmatrix}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top\left(\begin{bmatrix}1 & 0\\-1 & 0\\0 & 1\\0 & -1\end{bmatrix}\boldsymbol{x} - \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\right)$$

Differentiate with respect to $\boldsymbol{x}$ and set to zero:

$$\nabla_{\boldsymbol{x}}\mathfrak{L} = \boldsymbol{x}^\top \begin{bmatrix}2 & 1\\1 & 4\end{bmatrix} + \begin{bmatrix}5\\3\end{bmatrix}^\top + \boldsymbol{\lambda}^\top \begin{bmatrix}1 & 0\\-1 & 0\\0 & 1\\0 & -1\end{bmatrix} = 0$$

Solve for $\boldsymbol{x}$:

$$\boldsymbol{x}^\top = -\left(\begin{bmatrix}5\\3\end{bmatrix}^\top + \boldsymbol{\lambda}^\top \begin{bmatrix}1 & 0\\-1 & 0\\0 & 1\\0 & -1\end{bmatrix}\right)\begin{bmatrix}2 & 1\\1 & 4\end{bmatrix}^{-1}$$

Then substitute back to get the dual Lagrangian:

$$\mathfrak{D}(\boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^\top \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 5 \\ 3 \end{bmatrix}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top \left(\begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix} \boldsymbol{x} - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right)$$

7.8 Consider the following convex optimization problem

$$\begin{aligned} \min_{\boldsymbol{w} \in \mathbb{R}^D} \quad & \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{w} \\ \text{subject to} \quad & \boldsymbol{w}^\top \boldsymbol{x} \geqslant 1\,. \end{aligned}$$

Derive the Lagrangian dual by introducing the Lagrange multiplier λ.
First we express the convex optimization problem in standard form,

$$\begin{aligned} &\min_{\boldsymbol{w} \in \mathbb{R}^D} \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{w} \\ &\text{subject to } 1 - \boldsymbol{w}^\top \boldsymbol{x} \leqslant 0\,. \end{aligned}$$

By introducing a Lagrange multiplier $\lambda \geqslant 0$, we obtain the following Lagrangian

$$\mathfrak{L}(\boldsymbol{w}) = \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{w} + \lambda(1 - \boldsymbol{w}^\top \boldsymbol{x})$$

Taking the gradient of the Lagrangian with respect to $\boldsymbol{w}$ gives

$$\frac{\mathrm{d}\mathfrak{L}(\boldsymbol{w})}{\mathrm{d}\boldsymbol{w}} = \boldsymbol{w}^\top - \lambda \boldsymbol{x}^\top\,.$$

Setting the gradient to zero and solving for $\boldsymbol{w}$ gives

$$\boldsymbol{w} = \lambda \boldsymbol{x}\,.$$

Substituting back into $\mathfrak{L}(\boldsymbol{w})$ gives the dual Lagrangian

$$\begin{aligned} \mathfrak{D}(\lambda) &= \frac{\lambda^2}{2}\boldsymbol{x}^\top \boldsymbol{x} + \lambda - \lambda^2 \boldsymbol{x}^\top \boldsymbol{x} \\ &= -\frac{\lambda^2}{2}\boldsymbol{x}^\top \boldsymbol{x} + \lambda\,. \end{aligned}$$

Therefore the dual optimization problem is given by

$$\begin{aligned} &\max_{\lambda \in \mathbb{R}} -\frac{\lambda^2}{2}\boldsymbol{x}^\top \boldsymbol{x} + \lambda \\ &\text{subject to } \lambda \geqslant 0\,. \end{aligned}$$

2. Consider the optimization problem:

$$\text{minimize} \quad f(\mathbf{x}) = 2x_1^2 + 4x_2^2 + 2x_1x_2 - 2x_2$$

$$\text{subject to: } x_1 + x_2 \geq 4$$

$$x_1 - x_2 \leq 2$$

i. Prove that the above problem is convex.

ii. Using the Karesh-Kuhn-Tucker (KKT) conditions compute the optimal solution $\boldsymbol{x}^*$ and the optimal objective value $f(\boldsymbol{x}^*)$.

Θέμα 1

$$\text{minimize} \quad \overbrace{2x_1^2 + 4x_2^2 + 2x_1x_2 - 2x_2}^{f(x)}$$

$$\text{subject to: } \left.\begin{array}{l} x_1 + x_2 \geq 4 \quad (1) \\ x_1 - x_2 \leq 2 \quad (2) \end{array}\right] \text{primal feasibility}$$

i) $\nabla f(x) = \begin{bmatrix} 4x_1 + 2x_2 \\ 8x_2 + 2x_1 - 2 \end{bmatrix}$, $\nabla^2 f(x) = \begin{bmatrix} 4 & 2 \\ 2 & 8 \end{bmatrix}$

$\Delta_1 = 4 \geq 0, \ \Delta_2 = 32 - 4 = 28 \geq 0.$

$\Rightarrow f(x)$ is convex.

The constraints are affine $\Rightarrow$ the problem is convex.

ii) From the KKT conditions we have:

$$L(x,\mu) = f(x) + \mu_1(4 - x_1 - x_2) + \mu_2(x_1 - x_2 - 2)$$

$$\left.\begin{array}{l} \dfrac{\partial L}{\partial x_1} = 4x_1 + 2x_2 - \mu_1 + \mu_2 = 0 \quad (3) \\[2ex] \dfrac{\partial L}{\partial x_2} = 8x_2 + 2x_1 - 2 - \mu_1 - \mu_2 = 0 \quad (4) \end{array}\right] \text{Opt. Cond.}$$

$$\mu_1 \geq 0 \ (5), \quad \mu_2 \geq 0 \ (6) \quad \text{] dual feasibility.}$$

$$\left.\begin{array}{l} \mu_1(4 - x_1 - x_2) = 0 \quad (7) \\ \mu_2(x_1 - x_2 - 2) = 0 \quad (8) \end{array}\right] \text{Complementary slackness}$$

~~Case 1~~

Assume $\mu_1 > 0$ and $\mu_2 > 0$. $\xRightarrow{(7),(8)}$ $\begin{array}{l} x_1 + x_2 = 4 \\ x_1 - x_2 = 2 \\ \hline 2x_1 = 6 \Rightarrow x_1 = 3 \Rightarrow x_2 = 1 \end{array}$

$$\begin{array}{l} (3) \ 12 + 2 - \mu_1 + \mu_2 = 0. \\ \underline{8 + 6 - 2 - \mu_1 - \mu_2 = 0} \ + \end{array} \Rightarrow -2\mu_1 + 26 = 0 \Rightarrow \mu_1 = 13 \Rightarrow \mu_2 = -1 \overset{<0}{\Rightarrow} \text{Άτοπ.}$$

Case 2

$\mu_1 > 0, \; \mu_2 = 0$
$\Downarrow (7)$
$x_1 + x_2 = 4 \quad (9)$

$(3) \Rightarrow 4x_1 + 2x_2 = \mu_1$

$(4) \Rightarrow 8x_2 + 2x_1 - 2 - 4x_1 - 2x_2 = 0$

$-2x_1 + 6x_2 = 2$

$(9) \Rightarrow \; 2x_1 + 2x_2 = 8 \quad +$

$8x_2 = 10 \Rightarrow \boxed{x_2 = 1.25 \Rightarrow x_1 = 2.75}$

$\mu_1 = 4 \times 2.75 + 2 \times 1.25 = 13.5 \geq 0$

Primal feasibility?

$x_1 + x_2 \geq 4$ ✓

$x_1 - x_2 \leq 2$ ✓ $\quad f(x^*) = 25.75$ ✓

Case 3

$\mu_1 = 0, \; \mu_2 > 0$
$\Downarrow$
$x_1 - x_2 = 2$

$4x_1 + 2x_2 + \mu_2 = 0$

$8x_2 + 2x_1 - 2 - \mu_2 = 0 \quad +$

$6x_1 + 10x_2 = 2$

$10x_1 - 10x_2 = 20 \; +$

$16x_1 = 22 \Rightarrow x_1 = \frac{22}{16} \Rightarrow x_2 = \frac{22}{16} - 2 = -\frac{10}{16} \Rightarrow \mu_2 = +\frac{40}{16} - \frac{44}{16} = -\frac{1}{4} < 0$

✗

Άτοπο.

Case 4: $\mu_1 = 0, \mu_2 = 0$

$4x_1 + 2x_2 = 0$

$2x_1 + 8x_2 = 2$

$\Rightarrow -4x_1 - 16x_2 = -4 \; +$

$-14x_2 = -4 \Rightarrow x_2 = \frac{4}{14}$

$x_1 = -\frac{x_2}{2} = -\frac{2}{14}$

From (1) $x_1 + x_2 \ngeq 4 \Rightarrow$ Infeasible

✗