[SWCON253] Machine Learning – Lec.14b

Constrained Optimization

김휘용

hykim.v@khu.ac.kr

경희대학교
KYUNG HEE UNIVERSITY

Contents

References

- *Chapter 7, Mathematics for Machine Learning* by Deisenroth, Faisal, and Ong (https://mml-book.com)
- *Intro to Deep Learning & Generative Models* by Sebastian Raschka (http://pages.stat.wisc.edu/~sraschka/teaching/stat453-ss2020/)
- *패턴 인식* by 오일석, *기계 학습* by 오일석
- https://mml-book.github.io/

4. Lagrange Multipliers for *Multiple* Constraints

◆ **Multiple Constraints** (*including both equality & inequality constraints*)

$$\begin{aligned}&\text{minimize } f_0(\underline{x})\\ &\text{subject to } f_i(\underline{x}) \le 0 \quad (i = 1, \ldots, m)\\ &\qquad\qquad\quad h_j(\underline{x}) = 0 \quad (j = 1, \ldots, k)\end{aligned}$$

$$L(\underline{x}, \underline{\lambda}, \underline{\nu}) = f_0(\underline{x}) + \sum_{i=1}^{m} \lambda_i f_i(\underline{x}) + \sum_{j=1}^{k} \nu_j h_j(\underline{x}) = f_0(\underline{x}) + \underline{\lambda}^T \underline{f}(\underline{x}) + \underline{\nu}^T \underline{h}(\underline{x})$$

★ λ_i : Lagrange multiplier associated with $f_i(x_i) \le 0.$

★ ν_i : Lagrange multiplier associated with $h_i(x_i) = 0.$

● Note: Let $\boldsymbol{x}_0$ be a feasible point of the primal problem (that is a point that satisfies all the constraints), then we have:

$$\sum_{i=1}^{m} \underbrace{\lambda_i}_{\ge 0} \underbrace{f_i(x_0)}_{\le 0} + \sum_{i=1}^{k} \nu_j \underbrace{h_j(x_0)}_{=0} \le 0 \quad \Rightarrow \quad L(x_0, \lambda, \nu) = f_0(x_0) + \sum_{i=1}^{m} \lambda_i f_i(x_0) + \sum_{j=1}^{k} \nu_j h_j(x_0) \le f_0(x_0)$$

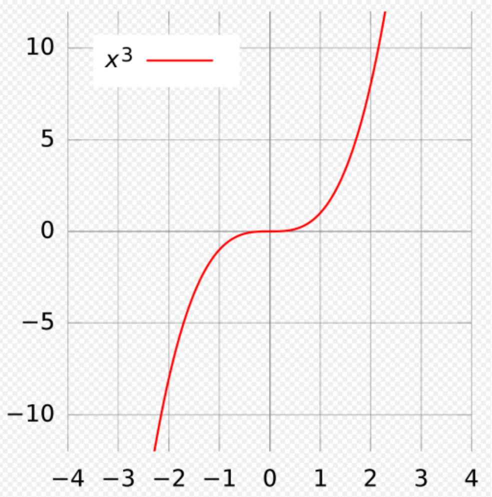
4. Lagrange Multipliers for *Multiple* Constraints (cont'd)

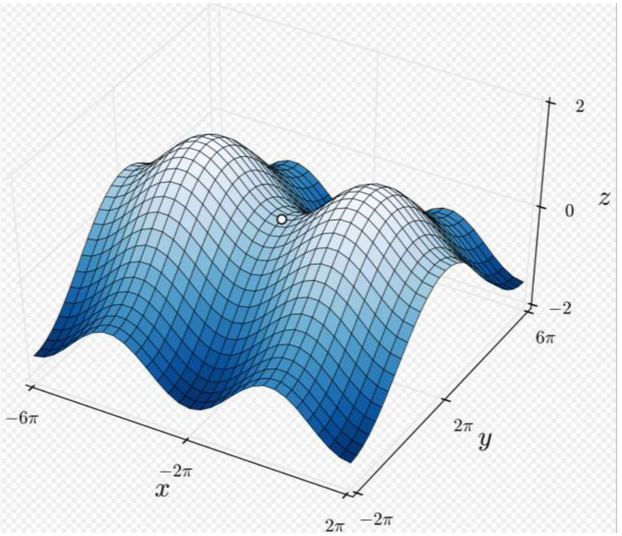
KKT (Karush-Kuhn-Tucker) Condition

- *First-order necessary conditions* (sometimes called first derivative tests) for a solution in nonlinear programming to be optimal, provided that some *regularity conditions* are satisfied.
- Generalizes the method of *Lagrange multipliers*, which (*originally*) allows only equality constraints.
- KKT Theorem is sometimes referred to as the ***saddle-point theorem***.

Saddle Point (or minimax point)

- A point on the surface of the graph of a function where the *derivatives* (slopes) in orthogonal directions *are all zero* (a critical point), but which is *not a local extremum* of the function.





4. Lagrange Multipliers for *Multiple* Constraints (cont'd)

Optimization with Multiple Constraints

- Optimization Problem

$$\begin{aligned} &\text{optimize } f(\mathbf{x}) \\ &\text{subject to} \\ &\qquad g_i(\mathbf{x}) \le 0, \\ &\qquad h_j(\mathbf{x}) = 0. \end{aligned}$$

$\mathbf{x} \in \mathbf{X}$ is the optimization variable chosen from a convex subset of $\mathbb{R}^n$

The numbers of inequalities and equalities are denoted by m and ℓ respectively.

- Lagrangian Function

$$\mathcal{L}(\mathbf{x}, \mu, \lambda) = f(\mathbf{x}) + \mu^\top \mathbf{g}(\mathbf{x}) + \lambda^\top \mathbf{h}(\mathbf{x}) = L(\mathbf{x}, \alpha) = f(\mathbf{x}) + \alpha^\top \begin{pmatrix} \mathbf{g}(\mathbf{x}) \\ \mathbf{h}(\mathbf{x}) \end{pmatrix}$$

$$\mathbf{g}(\mathbf{x}) = \begin{bmatrix} g_1(\mathbf{x}) \\ \vdots \\ g_i(\mathbf{x}) \\ \vdots \\ g_m(\mathbf{x}) \end{bmatrix}, \quad \mathbf{h}(\mathbf{x}) = \begin{bmatrix} h_1(\mathbf{x}) \\ \vdots \\ h_j(\mathbf{x}) \\ \vdots \\ h_\ell(\mathbf{x}) \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_i \\ \vdots \\ \mu_m \end{bmatrix}, \quad \lambda = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_j \\ \vdots \\ \lambda_\ell \end{bmatrix} \quad \text{and} \quad \alpha = \begin{bmatrix} \mu \\ \lambda \end{bmatrix}.$$

Quiz 1. Multiple Equality Constraints

minimize	x^2+y^2
subject to	$2x+y=1$
	$-x+2y=2$
	$4x+2y=3$

minimize	$f(x,y)=x^2+y^2$
subject to	$f_1(x,y)=2x+y-1=0$
	$f_2(x,y)=-x+2y-2=0$
	$f_3(x,y)=4x+2y-3=0$

풀이

$$L(\boldsymbol{x},\boldsymbol{\nu})=L(x,y,\boldsymbol{\nu})=f(x,y)+\sum_{i=1}^{3}\nu_i f_i(x,y)$$
$$=x^2+y^2+\nu_1(2x+y-1)+\nu_2(-x+2y-2)+\nu_3(4x+2y-3)$$
$$=x^2+(2\nu_1-\nu_2+4\nu_3)x+y^2+(\nu_1+2\nu_2+2\nu_3)y-(\nu_1+2\nu_2+3\nu_3)$$

...

해보세요!

참고

행렬 $A=\begin{bmatrix}2&1\\-1&2\\4&2\end{bmatrix}$와 벡터 $\boldsymbol{b}=\begin{bmatrix}1\\2\\3\end{bmatrix}$을 사용하면, 주어진 최적화 문제의 등식 제약조건을 $A\boldsymbol{x}-\boldsymbol{b}=0$로 나타낼 수 있습니다. 그러면 라그랑주 승수 $\boldsymbol{\nu}=\begin{bmatrix}\nu_1\\\nu_2\\\nu_3\end{bmatrix}$에 대하여, 라그랑주 함수 $L:\mathbb{R}^2\times\mathbb{R}^3\to\mathbb{R}$을 다음과 같이 쓸 수 있습니다.

$$L(\boldsymbol{x},\boldsymbol{\nu})=f(\boldsymbol{x})+\sum_{i=1}^{3}\nu_i f_i(\boldsymbol{x})=f(\boldsymbol{x})+\boldsymbol{\nu}^{\mathrm{T}}(A\boldsymbol{x}-\boldsymbol{b})$$

Quiz 1. Multiple Equality Constraints (Solution)

- Lagrangian

$$L(x, y, \boldsymbol{\nu}) = x^2 + y^2 + \nu_1(2x + y - 1) + \nu_2(-x + 2y - 2) + \nu_3(4x + 2y - 3)$$

- Gradient of L w.r.t. x and y

$$\frac{\partial}{\partial x} L(x, y, \boldsymbol{\nu}) = 2x + (2\nu_1 - \nu_2 + 4\nu_3) = 0 \qquad \qquad x = -\frac{1}{2}(2\nu_1 - \nu_2 + 4\nu_3)$$

$$\Rightarrow$$

$$\frac{\partial}{\partial y} L(x, y, \boldsymbol{\nu}) = 2y + (\nu_1 + 2\nu_2 + 2\nu_3) = 0 \qquad \qquad y = -\frac{1}{2}(\nu_1 + 2\nu_2 + 2\nu_3)$$

- Gradient of L w.r.t. $\boldsymbol{\nu}$

$$\frac{\partial L}{\partial \nu_1} = f_1(x, y) = 2x + y - 1 = 0 \qquad y = -2x + 1$$

$$\frac{\partial L}{\partial \nu_2} = f_2(x, y) = -x + 2y - 2 = 0 \quad \Rightarrow \quad y = \frac{1}{2}x + 1 \quad \Rightarrow \text{infeasible!}$$

$$\frac{\partial L}{\partial \nu_3} = f_3(x, y) = 4x + 2y - 3 = 0 \qquad y = -2x + 3$$

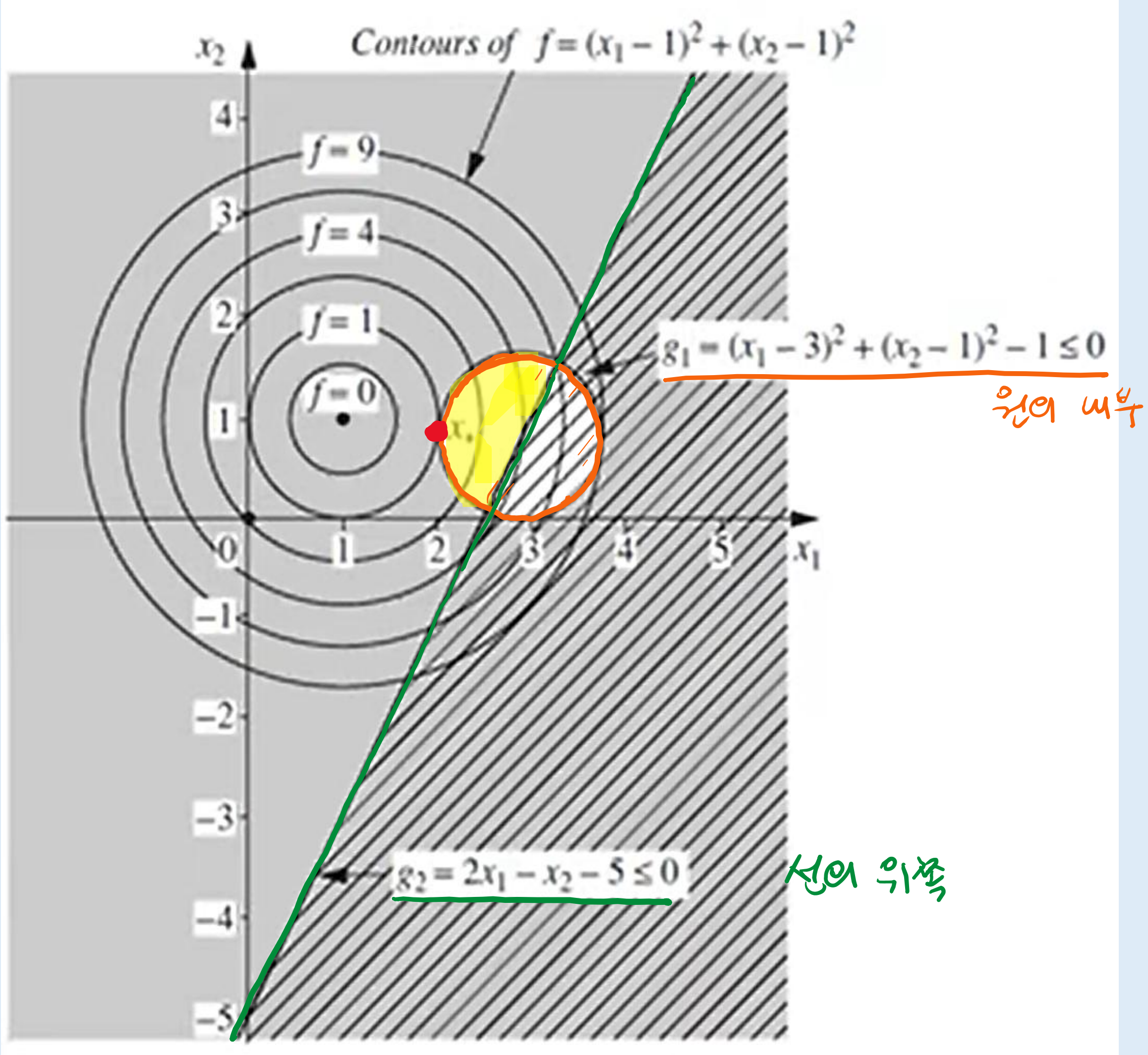
Quiz 2. Multiple Inequality Constraints

minimize $f = (x_1 - 1)^2 + (x_2 - 1)^2$
subject to $g_1 := (x_1 - 3)^2 + (x_2 - 1)^2 - 1 \le 0,$
$g_2 := 2x_1 - x_2 - 5 \le 0,$
$x_1 \ge 0, x_2 \ge 0.$

Where is feasible domain?

What is the minimum?

해보세요!



- feasible domain: 정의역(domain)의 원소들 중 모든 제약조건을 만족하는 원소들의 집합

Quiz 2. Multiple Inequality Constraints (Solution)

- Lagrangian

$$\mathcal{L}(x_1, x_2, \lambda_1, \lambda_2, \lambda_3, \lambda_4) = f(x_1, x_2) + \lambda_1 g_1(x_1, x_2) + \lambda_2 g_2(x_1, x_2) + \lambda_3 g_3(x_1, x_2) + \lambda_4 g_4(x_1, x_2)$$

$$\left(g_3(x_1, x_2) = -x_1 , \quad g_4(x_1, x_2) = -x_2 \right)$$

- Stationarity condition

$$\nabla_{\underline{x}} \mathcal{L} = \underline{0} : \quad \frac{\partial L}{\partial x_1} = (2x_1 - 6)\lambda_1 + 2\lambda_2 - \lambda_3 + (2x_1 - 2) = 0$$

$$\frac{\partial L}{\partial x_2} = (2x_2 - 2)\lambda_1 - \lambda_2 - \lambda_4 + (2x_2 - 2) = 0$$

- Complementary slackness

8 가지 cases : $\lambda_1 \begin{cases} = 0 \\ \neq 0 \end{cases}$, $\lambda_2 \begin{cases} = 0 \\ \neq 0 \end{cases}$, $\lambda_3 \begin{cases} = 0 \\ \neq 0 \end{cases}$, $\lambda_4 \begin{cases} = 0 \\ \neq 0 \end{cases}$

① $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0 \Rightarrow x_1 = 1, \ x_2 = 1 \Rightarrow g_1(1,1) = 3 \leq 0 \ (\times)$

② $\lambda_1 \neq 0, \ \lambda_2 = \lambda_3 = \lambda_4 = 0 \Rightarrow \begin{cases} g_1(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 1)^2 - 1 = 0 \\ (2x_1 - 6)\lambda_1 + (2x_1 - 2) = 0 \Rightarrow x_1 = \dfrac{3\lambda_1 + 1}{\lambda_1 + 1} \\ (2x_2 - 2)\lambda_1 + (2x_2 - 2) = 0 \Rightarrow x_2 = \dfrac{\lambda_1 + 1}{\lambda_1 + 1} = 1 \end{cases}$

$(\lambda_1 > 0)$

$\Rightarrow \lambda_1 = \begin{cases} -1 \ (\times) \\ 2 \ (\text{O}) \end{cases}$

$\Downarrow$

$x_1 = \dfrac{7}{3}, \ x_2 = 1$

③ ...

5. Lagrangian Duality

Duality in Optimization

- The idea of converting an optimization problem in one set of variables $\boldsymbol{x}$ (called the *primal variables*) into another set of variables $\boldsymbol{\lambda}$ (called the *dual variables*).

Primal Problem

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & g_i(\boldsymbol{x}) \leqslant 0 \quad \text{for all} \quad i = 1, \ldots, m \end{aligned}$$

Lagrangian Dual Problem

$$\begin{aligned} \max_{\boldsymbol{\lambda} \in \mathbb{R}^m} \quad & \mathfrak{D}(\boldsymbol{\lambda}) = \boxed{\min_{\boldsymbol{x} \in \mathbb{R}^d} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})} \\ \text{subject to} \quad & \boldsymbol{\lambda} \geqslant 0\,, \end{aligned}$$

$\mathcal{L}(\underline{x}, \underline{\lambda}) = f(\underline{x}) + \underline{\lambda}^T \underline{g}(\underline{x})$

$\nabla_{\underline{x}} \mathcal{L}(\underline{x}, \underline{\lambda}) = 0 \Rightarrow \underline{x}^* = \ldots$

$\therefore D(\underline{\lambda}) = \mathcal{L}(\underline{x}^*, \lambda)$

5. Lagrangian Duality – *Linear Programming*

Consider the special case when all the preceding functions are linear, i.e.,

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^d} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b}\,, \end{aligned} \tag{7.39}$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times d}$ and $\boldsymbol{b} \in \mathbb{R}^m$.

The Lagrangian is given by

$$\mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = (\boldsymbol{c} + \boldsymbol{A}^\top \boldsymbol{\lambda})^\top \boldsymbol{x} - \boldsymbol{\lambda}^\top \boldsymbol{b}\,.$$

Taking the derivative of $\mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{x}$ and setting it to zero $\left(\nabla_{\boldsymbol{x}} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{0} \right)$

$$\boldsymbol{c} + \boldsymbol{A}^\top \boldsymbol{\lambda} = \boldsymbol{0}.$$

Therefore, the dual Lagrangian is $\mathfrak{D}(\boldsymbol{\lambda}) = -\boldsymbol{\lambda}^\top \boldsymbol{b}$. $\Leftarrow$ $\mathfrak{D}(\boldsymbol{\lambda}) = \min_{\boldsymbol{x} \in \mathbb{R}^d} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$.

dual optimization problem

$$\begin{aligned} \max_{\boldsymbol{\lambda} \in \mathbb{R}^m} \quad & -\boldsymbol{b}^\top \boldsymbol{\lambda} \\ \text{subject to} \quad & \boldsymbol{c} + \boldsymbol{A}^\top \boldsymbol{\lambda} = \boldsymbol{0} \\ & \boldsymbol{\lambda} \geqslant \boldsymbol{0}. \end{aligned}$$

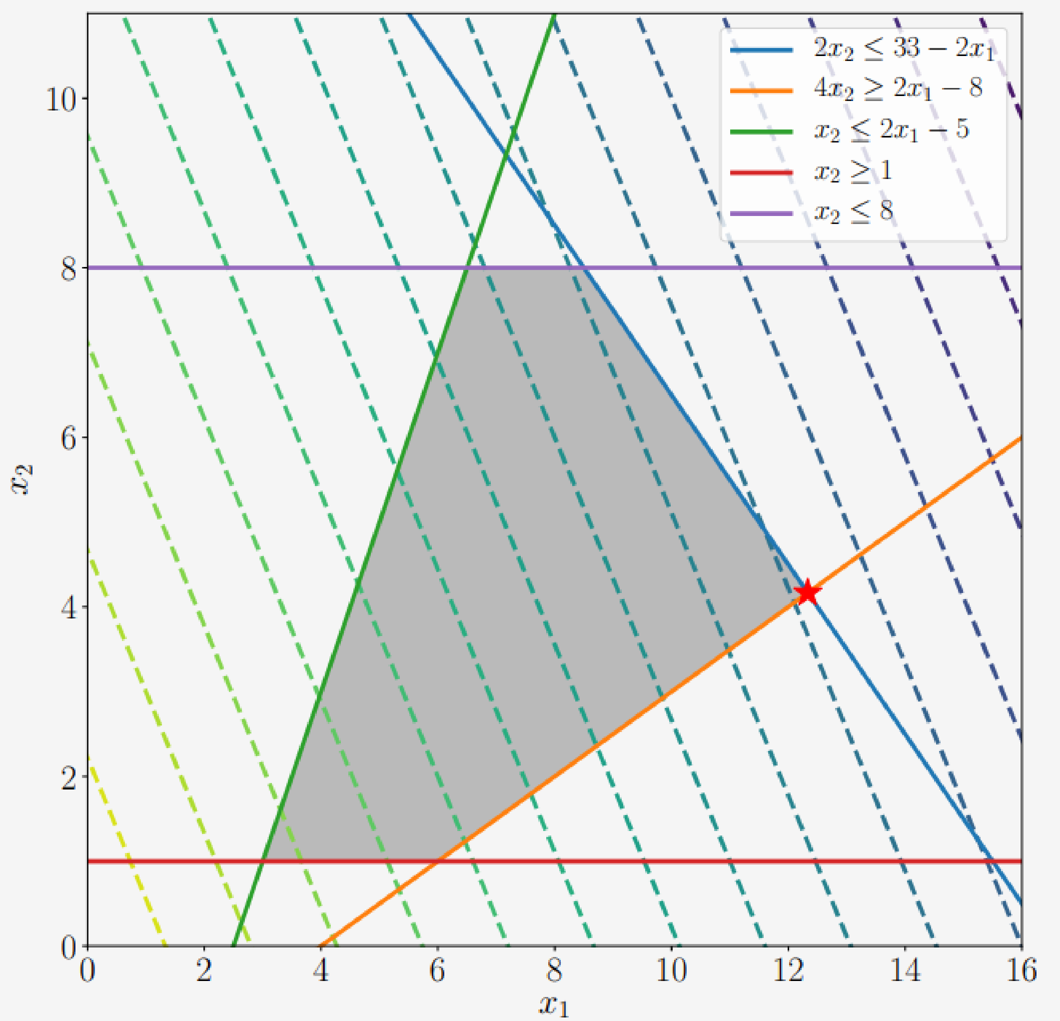
We have the choice of solving the primal or the dual program depending on whether *m* or *d* is larger.

5. Lagrangian Duality – *Linear Programming (cont'd)*

Example 7.5 (Linear Program)
Consider the linear program

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} \quad -\begin{bmatrix}5\\3\end{bmatrix}^\top \begin{bmatrix}x_1\\x_2\end{bmatrix}$$
$$\text{subject to} \quad \begin{bmatrix}2 & 2\\ 2 & -4\\ -2 & 1\\ 0 & -1\\ 0 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} \leqslant \begin{bmatrix}33\\8\\5\\-1\\8\end{bmatrix} \tag{7.44}$$

with two variables. This program is also shown in Figure 7.9. The objective function is linear, resulting in linear contour lines. The constraint set in standard form is translated into the legend. The optimal value must lie in the shaded (feasible) region, and is indicated by the star.



5. Lagrangian Duality – *Quadratic Programming*

Consider the case of a convex quadratic objctive function, where the constraints are affine, i.e.,

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \quad \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x} \tag{7.45}$$
$$\text{subject to} \quad \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b}\,,$$

where $\boldsymbol{A} \in \mathbb{R}^{m\times d}$, $\boldsymbol{b} \in \mathbb{R}^m$, and $\boldsymbol{c} \in \mathbb{R}^d$.

The square symmetric matrix $\boldsymbol{Q} \in \mathbb{R}^{d\times d}$ is positive definite, and therefore the objective function is convex.

The Lagrangian is given by $\mathfrak{L}(\boldsymbol{x},\boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + (\boldsymbol{c} + \boldsymbol{A}^\top\boldsymbol{\lambda})^\top \boldsymbol{x} - \boldsymbol{\lambda}^\top \boldsymbol{b}\,,$

Taking the derivative of $\mathfrak{L}(\boldsymbol{x},\boldsymbol{\lambda})$ with respect to $\boldsymbol{x}$ and setting it to zero gives $\boldsymbol{Q}\boldsymbol{x} + (\boldsymbol{c} + \boldsymbol{A}^\top\boldsymbol{\lambda}) = \boldsymbol{0}\,.$

Assuming that $\boldsymbol{Q}$ is invertible, we get $\boldsymbol{x} = -\boldsymbol{Q}^{-1}(\boldsymbol{c} + \boldsymbol{A}^\top\boldsymbol{\lambda})\,.$

Therefore, the dual optimization problem is given by

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^m} \quad -\frac{1}{2}(\boldsymbol{c} + \boldsymbol{A}^\top\boldsymbol{\lambda})^\top \boldsymbol{Q}^{-1}(\boldsymbol{c} + \boldsymbol{A}^\top\boldsymbol{\lambda}) - \boldsymbol{\lambda}^\top \boldsymbol{b}$$
$$\text{subject to} \quad \boldsymbol{\lambda} \geqslant \boldsymbol{0}\,.$$

We have the choice of solving the primal or the dual program depending on whether *m* or *d* is larger.

5. Lagrangian Duality – *Quadratic Programming (cont'd)*

- **Example: Converting into the standard QP form and then solving the dual problem**

  - Primal Problem

$$\boxed{\begin{aligned} \min_{\boldsymbol{x}} \quad & J(\boldsymbol{x}) = x_1^2 + 2x_2^2 \\ \text{s.t.} \quad & g(\boldsymbol{x}) = -2x_1 - x_2 + 1 \le 0 \end{aligned}}$$

$$\boxed{\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^d} \quad & \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b}\,, \end{aligned}}$$

$J(\underline{x})$

$A\underline{x} - \underline{b} \le 0 \Rightarrow g(\underline{x}) = A\underline{x} - \underline{b}$

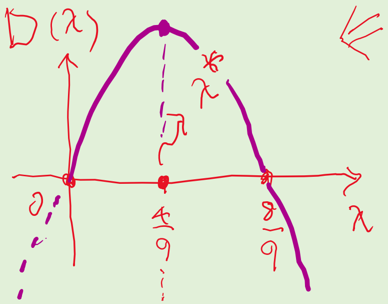
Let $Q = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$, $\underline{c} = \underline{0}$, $A = [-2 \ -1]$, $b = -1$.

Then we can represent the original problem in the standard QP form.

  - Lagrangian Dual Problem

$$\boxed{\begin{aligned} \max_{\boldsymbol{\lambda}\in\mathbb{R}^m} \quad & -\frac{1}{2}(\boldsymbol{c} + \boldsymbol{A}^\top\boldsymbol{\lambda})^\top \boldsymbol{Q}^{-1}(\boldsymbol{c} + \boldsymbol{A}^\top\boldsymbol{\lambda}) - \boldsymbol{\lambda}^\top \boldsymbol{b} \\ \text{subject to} \quad & \boldsymbol{\lambda} \geqslant \boldsymbol{0}\,. \end{aligned}}$$

$$\boxed{\begin{aligned} \max_{\lambda} \quad & -\frac{9}{8}\lambda^2 + \lambda \\ \text{s.t.} \quad & \lambda \ge 0 \end{aligned}}$$



$\Rightarrow$ Sol. $\lambda^* = \frac{4}{9} \Rightarrow \boldsymbol{x} = -\boldsymbol{Q}^{-1}(\boldsymbol{c} + \boldsymbol{A}^\top\boldsymbol{\lambda}) = \frac{1}{9}\begin{bmatrix} 8 \\ 4 \end{bmatrix}$.

출처: https://mml-book.github.io/

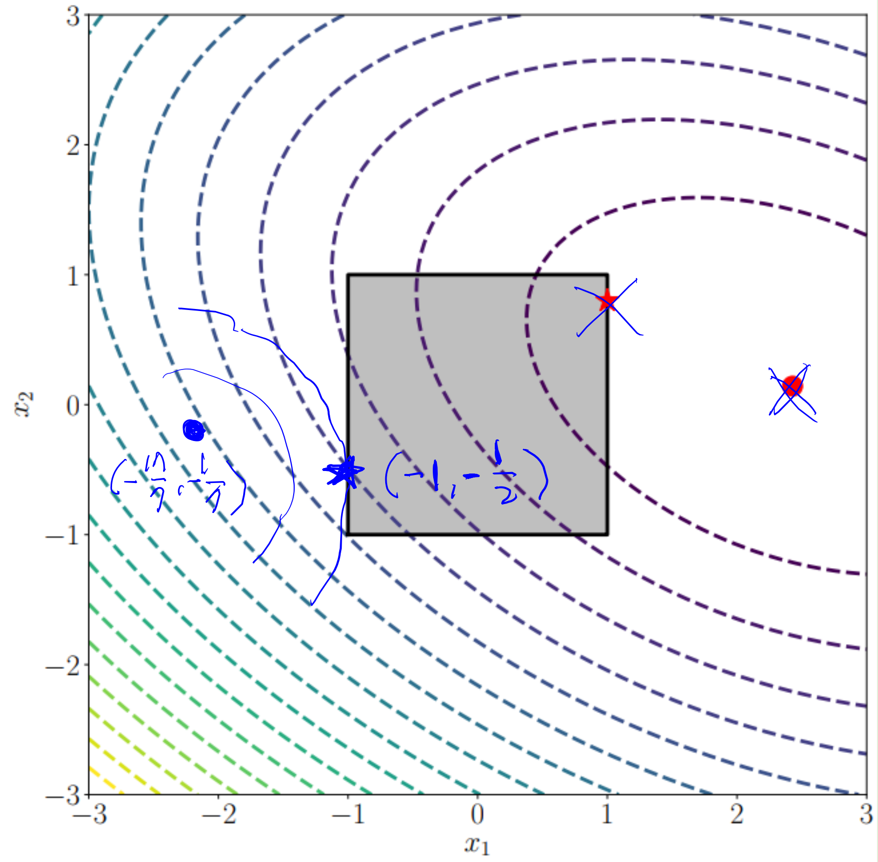
5. Lagrangian Duality – *Quadratic Programming (cont'd)*

Example 7.6 (Quadratic Program)
Consider the quadratic program

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} \quad \frac{1}{2}\begin{bmatrix}x_1\\x_2\end{bmatrix}^\top \begin{bmatrix}2 & 1\\1 & 4\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} + \begin{bmatrix}5\\3\end{bmatrix}^\top \begin{bmatrix}x_1\\x_2\end{bmatrix} \tag{7.46}$$

$$\text{subject to} \quad \begin{bmatrix}1 & 0\\-1 & 0\\0 & 1\\0 & -1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} \leqslant \begin{bmatrix}1\\1\\1\\1\end{bmatrix} \tag{7.47}$$

of two variables. The program is also illustrated in Figure 7.4. The objective function is quadratic with a positive semidefinite matrix $\boldsymbol{Q}$, resulting in elliptical contour lines. The optimal value must lie in the shaded (feasible) region, and is indicated by the star.



$$\underline{x} = \begin{bmatrix}x_1\\x_2\end{bmatrix} \qquad Q = \begin{bmatrix}2 & 1\\1 & 4\end{bmatrix} \qquad \underline{c} = \begin{bmatrix}5\\3\end{bmatrix}$$

$$A = \begin{bmatrix}1 & 0\\-1 & 0\\0 & 1\\0 & -1\end{bmatrix} \qquad \underline{b} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

The unconstrained problem (indicated by the contour lines) has a minimum on the right side (indicated by the circle).

The box constraints ($-1 \le x_1 \le 1$ and $-1 \le x_2 \le 1$) require that the optimal solution is within the box, resulting in an optimal value indicated by the star.

출처: https://mml-book.github.io/

5. Lagrangian Duality – *Quadratic Programming (cont'd)*

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \quad \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x}$$
$$\text{subject to} \quad \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b},$$

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad Q = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \qquad \underline{c} = \begin{bmatrix} 5 \\ 3 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix} \qquad \underline{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathcal{L}(\underline{x},\underline{\lambda}) = \frac{1}{2}\underline{x}^T Q \underline{x} + \underline{c}^T \underline{x} + \underline{\lambda}^T (A\underline{x} - \underline{b}), \qquad \underline{\lambda} = [\lambda_1\ \lambda_2\ \lambda_3\ \lambda_4]^T, \qquad \underline{g}(\underline{x}) = A\underline{x} - \underline{b}$$

- $\nabla_x \mathcal{L} = Q\underline{x} + \underline{c} + A^T\underline{\lambda} = 0 \rightarrow \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 5 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} = 0$

$$\Rightarrow \begin{cases} 2x_1 + x_2 + 5 + \lambda_1 - \lambda_2 = 0 & ① \\ x_1 + 4x_2 + 3 + \lambda_3 - \lambda_4 = 0 & ② \end{cases}$$

- $\underline{\lambda}^T (A\underline{x} - \underline{b}) = 0 \rightarrow [\lambda_1\ \lambda_2\ \lambda_3\ \lambda_4]\begin{bmatrix} x_1 - 1 \\ -x_1 - 1 \\ x_2 - 1 \\ -x_2 - 1 \end{bmatrix} = 0 \rightarrow \begin{cases} \lambda_1(x_1 - 1) = 0 & ③ \\ \lambda_2(x_1 + 1) = 0 & ④ \\ \lambda_3(x_2 - 1) = 0 & ⑤ \\ \lambda_4(x_2 + 1) = 0 & ⑥ \end{cases}$

③,⑤ : In case $x_1 = 1$ and $x_2 = 1$, $\lambda_2 = \lambda_4 = 0$ (∵ ④,⑤).

$\rightarrow$ ① : $\lambda_1 = -8 < 0$ (X)

③,⑥ : In case $x_1 = 1$ and $x_2 = -1$, $\lambda_2 = \lambda_3 = 0$ (∵ ④,⑥).

$\rightarrow$ ① : $\lambda_1 = -6 < 0$ (X)

...

③ : $x_1 = 1$, $x_2 \neq 1$, $x_2 \neq -1$ $\Rightarrow$ $\lambda_2 = \lambda_3 = \lambda_4 = 0$

$$\rightarrow \begin{cases} ① : 2 + x_2 + \lambda_1 + 5 = 0 \\ ② : 1 + 4x_2 + 3 = 0 \Rightarrow x_2 = -1 \end{cases}$$

Q & A