5. Duality

최영준, 박정현

The Lagrange dual function

Lagrangian이란 objective function과 constraint function을 하나의 다항식으로 합치는 것!

Lagrange multiplier vector를 이용한 weighted sum 형태로

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \le 0, \quad i = 1, \ldots, m \\
& h_i(x) = 0, \quad i = 1, \ldots, p
\end{array}
$$

반드시 nonpositive

만약 positive라면 람다에 음수붙이기

variable $x \in \mathbf{R}^n$, domain $\mathcal{D}$, optimal value $p^\star$

Lagrangian: $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$, with $\mathbf{dom}\, L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$,

$$
L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)
$$

The Lagrange dual function

$$
\begin{aligned}
g(\lambda, \nu) &= \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \\
&= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)
\end{aligned}
$$

affine function

g is concave, can be $-\infty$ for some λ, ν

↳ affine function의
pointwise infimum 이기 때문

The Lagrange dual function

Dual function은 lower bound optimal value를 보장함

if $\lambda \succeq 0$, then $g(\lambda, \nu) \leq p^\star$

This important property is easily verified. Suppose $\tilde{x}$ is a feasible point for the problem (5.1), *i.e.*, $f_i(\tilde{x}) \leq 0$ and $h_i(\tilde{x}) = 0$, and $\lambda \succeq 0$. Then we have

$$\sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \leq 0,$$

since each term in the first sum is nonpositive, and each term in the second sum is zero, and therefore

$$L(\tilde{x}, \lambda, \nu) = f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \leq f_0(\tilde{x}).$$

Hence

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x}).$$

Since $g(\lambda, \nu) \leq f_0(\tilde{x})$ holds for every feasible point $\tilde{x}$, the inequality (5.2) follows. The lower bound (5.2) is illustrated in figure 5.1, for a simple problem with $x \in \mathbf{R}$ and one inequality constraint.

The inequality (5.2) holds, but is vacuous, when $g(\lambda, \nu) = -\infty$. The dual function gives a nontrivial lower bound on $p^\star$ only when $\lambda \succeq 0$ and $(\lambda, \nu) \in \mathbf{dom}\, g$, *i.e.*, $g(\lambda, \nu) > -\infty$. We refer to a pair (λ, ν) with $\lambda \succeq 0$ and $(\lambda, \nu) \in \mathbf{dom}\, g$ as *dual feasible*, for reasons that will become clear later.

$\sum \lambda_i f_i(\tilde{x}) \leq 0$

$\sum \nu_i h_i(\tilde{x}) = 0$

$$f_0(\tilde{x}) + \underbrace{\sum \lambda_i f_i(\tilde{x})}_{\leq 0} + \underbrace{\sum \nu_i h_i(\tilde{x})}_{=0} \leq f_0(\tilde{x})$$

↳ 모든 feasible point에 대해 성립

The Lagrange dual function

Examples

- Least-squares solution of linear equations

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b, \end{array}$$

inequality constraint 없으니까 λ 생략

$$L(x, \nu) = x^T x + \nu^T (Ax - b),$$

$\hookrightarrow$ quadratic form

$$g(\nu) = inf_x L(x, \nu)$$

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0$$

$$x = -(1/2) A^T \nu$$

The Lagrange dual function

Examples

- Least-squares solution of linear equations

$$
\begin{aligned}
L(x,\nu) &= x^T x + \nu^T (Ax - b) \\
&= \frac{1}{4}\nu^T A A^T \nu + \nu^T \left\{ A\left(-\frac{1}{2}A^T \nu\right) - b \right\} \\
&= \frac{1}{4}\nu^T A A^T \nu - \frac{1}{2}\nu^T A A^T \nu - b^T \nu \\
&= -\frac{1}{4}\nu^T A A^T \nu - b^T \nu
\end{aligned}
$$

$$g(\nu) = L(-(1/2)A^T\nu, \nu) = \boxed{-(1/4)\nu^T A A^T \nu - b^T \nu,}$$

$$-(1/4)\nu^T A A^T \nu - b^T \nu \leq \inf\{x^T x \mid Ax = b\}.$$

for any ν

The Lagrange dual function

Examples

- Standard form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \quad x \succeq 0 \end{array}$$

$$\begin{aligned} L(x, \lambda, \nu) &= c^T x + \nu^T (Ax - b) - \lambda^T x \\ &= -b^T \nu + (c + A^T \nu - \lambda)^T x \end{aligned}$$

L is affine in x

$c + A^T\nu - \lambda \neq 0$ 일때 unbounded below

The Lagrange dual function

Examples

- Standard form LP

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -b^T\nu & A^T\nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

g is linear on affine domain $\{(\lambda, \nu) \mid A^T\nu - \lambda + c = 0\}$

$p^\star \geq -b^T\nu$ if $A^T\nu + c \succeq 0$

The Lagrange dual function

non-convex 문제

Examples

- Two-way partitioning

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

$$L(x, \nu) = x^T W x + \sum_i \nu_i (x_i^2 - 1)$$

dual function

$$g(\nu) = \inf_x \left(x^T W x + \sum_i \nu_i (x_i^2 - 1)\right) = \inf_x x^T (W + \operatorname{diag}(\nu)) x - \mathbf{1}^T \nu$$

$$= \begin{cases} -\mathbf{1}^T \nu & W + \operatorname{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise} \end{cases}$$

$$\begin{bmatrix} \nu_1 & & \\ & \nu_2 & \\ & & \ddots & \\ & & & \nu_p \end{bmatrix}$$

quadratic form

The Lagrange dual function

Examples

- Two-way partitioning

lower bound property: $p^\star \geq -\mathbf{1}^T\nu$ if $W + \mathbf{diag}(\nu) \succeq 0$

example: $\nu = -\lambda_{\min}(W)\mathbf{1}$ gives bound $p^\star \geq n\lambda_{\min}(W)$

The Lagrange dual function

Lagrange dual and conjugate function

핵심은 알려져 있는 conjugate로 dual function을 단순화한다는 것

recall: $f^*(y) = \sup_{x \in \mathrm{dom} f} (y^T x - f(x))$

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & Ax \preceq b, \quad Cx = d
\end{array}
$$

$$L(x, \lambda, \nu) = f_0(x) + \lambda(Ax - b) + \nu(Cx - d)$$

$$
\begin{aligned}
g(\lambda, \nu) &= \inf_{x \in \mathrm{dom}\, f_0} \left(f_0(x) + (A^T\lambda + C^T\nu)^T x - b^T\lambda - d^T\nu \right) \\
&= -f_0^*(-A^T\lambda - C^T\nu) - b^T\lambda - d^T\nu
\end{aligned}
$$

The Lagrange dual function

Example

- Entropy maximization

$$
\begin{aligned}
g(\lambda, \nu) &= \inf_x \left(f_0(x) + \lambda^T (Ax - b) + \nu^T (Cx - d) \right) \\
&= -b^T \lambda - d^T \nu + \inf_x \left(f_0(x) + (A^T \lambda + C^T \nu)^T x \right) \\
&= -b^T \lambda - d^T \nu - f_0^*(-A^T \lambda - C^T \nu).
\end{aligned}
$$

Inf => -Sup

The Lagrange dual function

$\frac{d}{dx}(yx - x\log x) = 0$

$x = e^{y-1}$ ⟶ $f^*(y)$ 에 x값 대입하면 e^{y-1}

Example

- Entropy maximization(negative entropy)

$$\begin{array}{ll} \text{minimize} & f_0(x) = \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Ax \preceq b \\ & \mathbf{1}^T x = 1 \end{array}$$

$L(x, \lambda, \nu) = f_0(x) + \lambda(Ax - b) + \nu(\mathbf{1}^T x - 1)$

$$g(\lambda, \nu) = -b^T\lambda - \nu - \sum_{i=1}^n e^{-a_i^T\lambda - \nu - 1} = -b^T\lambda - \nu - e^{-\nu-1}\sum_{i=1}^n e^{-a_i^T\lambda}$$

The Lagrange dual problem

Dual problem :

Objective function - dual function

Constraint function – dual variable

$\Rightarrow$ Convex optimization

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- finds best lower bound on $p^\star$, obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted $d^\star$ ~ (λ^*, ν^*) optimal for the problem
- λ, ν are dual feasible if $\lambda \succeq 0$, $(\lambda, \nu) \in \mathbf{dom}\, g$

The Lagrange dual problem

Example

- Lagrange dual of standard form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array}$$

$\rightarrow L(x,\lambda,\nu) = c^T x - \lambda x + \nu(Ax-b)$

$= -b^T\nu + (A^T\nu - \lambda + c)x$

$$g(\lambda,\nu) = \begin{cases} -b^T\nu & A^T\nu - \lambda + c = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

$$\begin{array}{ll} \text{maximize} & g(\lambda,\nu) = \begin{cases} -b^T\nu & A^T\nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases} \\ \text{subject to} & \lambda \succeq 0. \end{array}$$

equivalent

The Lagrange dual problem

Example

- Lagrange dual of standard form LP

implicit 한 constraint를
explicit 하도록

$$
\begin{array}{ll}
\text{maximize} & -b^T\nu \\
\text{subject to} & A^T\nu - \lambda + c = 0 \; * \\
& \lambda \succeq 0.
\end{array}
$$

$\rightsquigarrow$

$$
\begin{array}{ll}
\text{maximize} & -b^T\nu \\
\text{subject to} & A^T\nu + c \succeq 0
\end{array}
$$

The Lagrange dual problem

($p^* - d^*$: duality gap)

weak duality: $d^\star \le p^\star$

Can be used to find nontrivial lower for difficult problems (ex. SDP)

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T\nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0 \end{array}$$

strong duality: $d^\star = p^\star$

Does not hold in general

(usually) holds for convex problems

Conditions that guarantee strong duality in convex problems are called **constraint qualifications**

The Lagrange dual problem

Slater's conditions(one of the constraint qualification)

E $x \in \textbf{relint}\ \mathcal{D}, \quad f_i(x) < 0, \quad i = 1, \dots, m, \quad Ax = b$ => strong duality 보장

strict inequalities

when some of the inequality constraint functions f_i are affine

E $x \in \textbf{relint}\ \mathcal{D}, \quad f_i(x) \le 0, \quad i = 1, \dots, k, \quad f_i(x) < 0, \quad i = k+1, \dots, m, \quad Ax = b$

affine한 inequalities

=> strong duality 보장

The Lagrange dual problem

Example

- Mixed strategies for matrix games

zero sum matrix games에서 strong duality를 이용해 간단한 결과를 도출할 수 있음

Player 1 : make a choice $k \in \{1, ..., n\}$

Player 2 : make a choice $l \in \{1, ..., m\}$

Player 1은 P_{kl}을 Player 2에게 지불해야 함

Player 1은 minimize, Player 2는 maximize가 목표

The Lagrange dual problem

Example

- Mixed strategies for matrix games

player들은 randomized or mixed strategies를 사용, 자신의 선택을 random하고 상대방의 선택과 독립적으로 구성할 수 있음 아래 확률분포를 따르는

$$\mathbf{prob}(k = i) = u_i, \quad i = 1, \ldots, n, \quad \mathbf{prob}(l = i) = v_i, \quad i = 1, \ldots, m$$

P1 : $\underset{u}{\text{minimize}} \; u^T P v$

그렇다면 지불해야 하는 값의 기댓값을 구한다면 다음과 같다.

$$\sum_{k=1}^{n} \sum_{l=1}^{m} u_k v_l P_{kl} = \boxed{u^T P v.}$$

P2 : $\underset{v}{\text{maximize}} \; u^T P v$

The Lagrange dual problem

Example

- Mixed strategies for matrix games

이때의 optimal : p_1^*

만약 player 1의 전략인 u를 이미 player 2가 아는 상황이라면 u^TPv를 maximize하는 v를 고를 것이고 그 식은 다음과 같다.

이 상황에서 P2 전략

$$\sup\{u^TPv \mid v \succeq 0,\ \mathbf{1}^Tv = 1\} = \max_{i=1,\dots,m}(P^Tu)_i.$$

이 상황에서 P1 전략

$$\begin{array}{ll} \text{minimize} & \max_{i=1,\dots,m}(P^Tu)_i \\ \text{subject to} & u \succeq 0, \quad \mathbf{1}^Tu = 1, \end{array}$$

worst case 피하기

The Lagrange dual problem

Example

- Mixed strategies for matrix games

이때의 optimal : p_2^*

$p_1^* - p_2^* \geq 0$ (직관적으로)

player2입장에서의 베스트는

p_1

$$\inf\{u^T P v \mid u \succeq 0,\ \mathbf{1}^T u = 1\} = \min_{i=1,\ldots,n} (Pv)_i.$$

p_2

$$\begin{array}{ll} \text{maximize} & \min_{i=1,\ldots,n}(Pv)_i \\ \text{subject to} & v \succeq 0, \quad \mathbf{1}^T v = 1, \end{array}$$

The Lagrange dual problem

Example

- Mixed strategies for matrix games

player 1의 선택에 대한 식을 LP형태로 구체화하면 다음과 같다.

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & u \succeq 0, \quad \mathbf{1}^T u = 1 \\ & P^T u \preceq t\mathbf{1}, \end{array}$$

$L(x, \lambda, \mu, \nu)$

$$= t + \lambda^T(P^T u - t\mathbf{1}) - \mu^T u + \nu(1 - \mathbf{1}^T u) = \nu + (1 - \mathbf{1}^T \lambda)t + (P\lambda - \nu\mathbf{1} - \mu)^T u,$$

The Lagrange dual problem

Example

- Mixed strategies for matrix games

unbounded below가 되지 않기위한...

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \mathbf{1}^T\lambda = 1, \quad P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{otherwise.} \end{cases}$$

$$\begin{array}{ll} \text{maximize} & \nu \\ \text{subject to} & \lambda \succeq 0, \quad \mathbf{1}^T\lambda = 1, \quad \mu \succeq 0 \\ & P\lambda - \nu\mathbf{1} = \mu. \end{array} \rightsquigarrow \left\{ \begin{array}{ll} \text{maximize} & \nu \\ \text{subject to} & \lambda \succeq 0, \quad \mathbf{1}^T\lambda = 1 \\ & P\lambda \succeq \nu\mathbf{1}, \end{array} \right.$$

Player 2 전략과 equivalent하다.

$\therefore p_1^* = p_2^*$

The Lagrange dual problem

5.3 *Problems with one inequality constraint.* Express the dual problem of

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & f(x) \leq 0, \end{array}$$

with $c \neq 0$, in terms of the conjugate f^*. Explain why the problem you give is convex. We do not assume f is convex.

3

Geometric Interpretation

Duality 및 Strong duality의 특징과 조건에 대한 기하적 해석

5.3 Geometric Interpretation

Simple example

$minimize\ f_0(x)$

$subject\ to\ f_1(x) \le 0$

$G = \{(u,t) \mid f_1(x) = u,\ f_0(x) = t, x \in D\}$

하나의 Inequality constraint만을 가진 Primal Problem을 가정

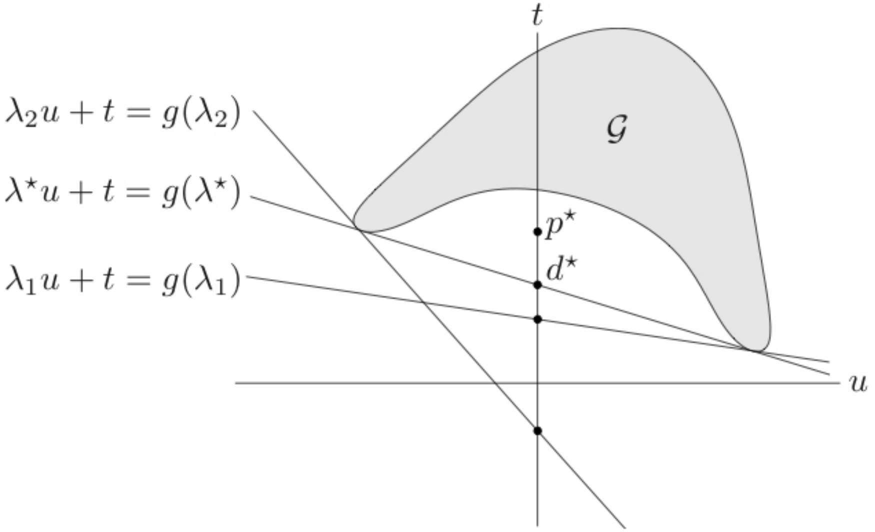
동일한 문제에 대한 dual problem을 정의

$$\inf_{x \in D}\{\lambda f_1(x) + f_0(x)\} = \inf_{(u,t)\in G}\{\lambda u + t\} = \inf_{(u,t)\in G}\{(\lambda,1)^T(u,t)\}$$

모든 가능한 람다 값에 대해 dual을 도출하는 과정을 적용한다면,

1. 모든 g값은 p*보다 작거나 같은 부분에 위치
2. g가 계속 증가하다가 어느 순간 다시 감소

이 변화 지점, 즉 g의 최대값이 dual problem의 optimal point d*



Duality Gap: p* - d*, 이것이 0인 경우가 Strong Duality

5.3 Geometric Interpretation

일반적인 Primal Problem

$$G = \{(f_1(x), ..., f_m(x), h_1(x), ..., h_p(x), f_0(x)) \in R^m \times R^p \times R \mid x \in D\}$$

Inequality constraint, equality constraint, objective function 전부의 함숫값의 집합

$$p^* = \inf\{t \mid (u, v, t) \in G,\ u \preceq 0,\ v = 0\}$$

Feasible한 영역 내에서의 t의 최저점

$$(\lambda, \nu, 1)^T (u, v, t) = \sum_{i=1}^{m} \lambda_i u_i + \sum_{i=1}^{p} \nu_i v_i + t$$ 벡터 형태의 라그랑지안

$$g(\lambda, \nu) = \inf\{(\lambda, \nu, 1)^T (u, v, t) \mid (u, v, t) \in G\}$$ Dual Function

$$(\lambda, \nu, 1)^T (u, v, t) \geq g(\lambda, \nu)$$ G의 supporting hyperplane

$$\begin{aligned} p^\star &= \inf\{t \mid (u, v, t) \in \mathcal{G},\ u \preceq 0,\ v = 0\} \\ &\geq \inf\{(\lambda, \nu, 1)^T (u, v, t) \mid (u, v, t) \in \mathcal{G},\ u \preceq 0,\ v = 0\} \\ &\geq \inf\{(\lambda, \nu, 1)^T (u, v, t) \mid (u, v, t) \in \mathcal{G}\} \\ &= g(\lambda, \nu), \end{aligned}$$

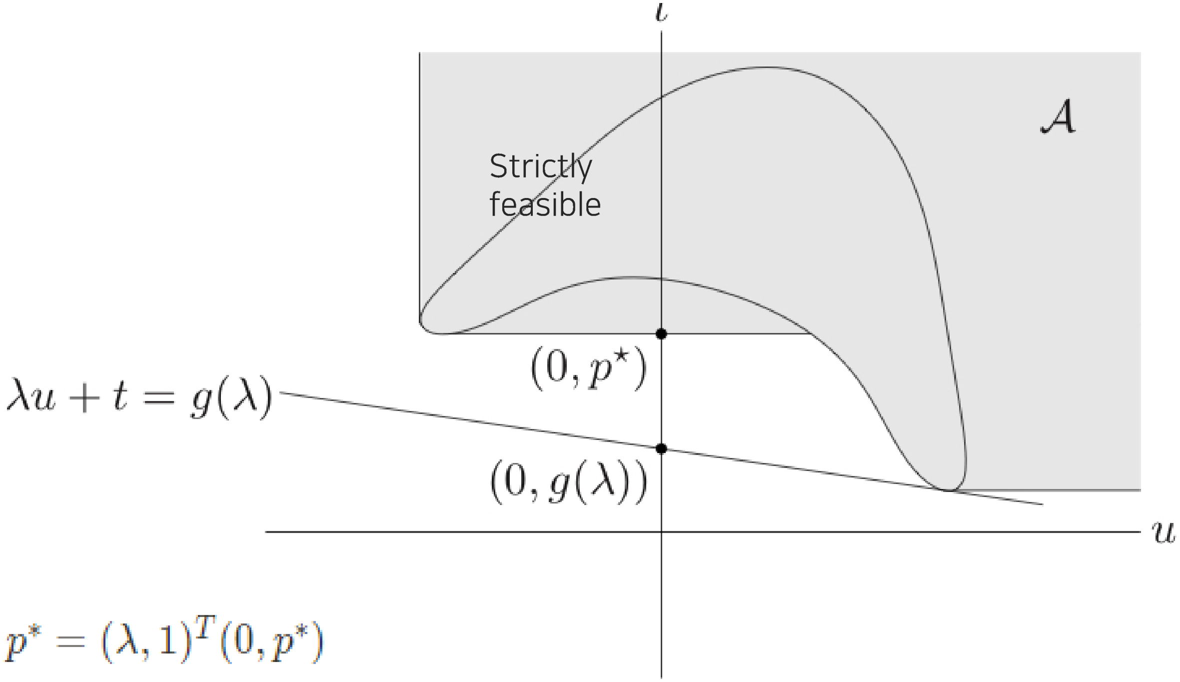
제약 조건이 있을 때보다 없을 때 하한을
더 작거나 같게 만들 수 있다는 아이디어

5.3 Geometric Interpretation

Convexity와 Strong Duality

Slater's theorem: Primal problem이 convex이고, strictly feasible한 x가 하나 이상 존재하면 strong duality가 만족된다.

$$A = \{(u,t) \mid \exists x \in D,\ f_0(x) \le t,\ f_1(x) \le u\}$$ Set G에 대한 일종의 epigraph



Primal problem이 convex해서 epigraph가 convex 했다면 (0, p*)에서의 non-vertical supporting hyperplane이 존재했을 것

Strong duality

Slater's theorem에 대한 증명은 5.3.2절 참고

$$p^* = (\lambda, 1)^T (0, p^*)$$

$$p^* = (\lambda, 1)^T (0, p^*) \ge g(\lambda, \nu)$$ 등호가 성립하는 경우(Strong duality) set A의 non-vertical supporting hyperplane

5.3 Geometric Interpretation

Multi-criterion interpretation

제약 조건이 없는 Multi-criterion problem의 scalarization = equality constraint가 없는 primal problem의 라그랑지안
(4.7) Objective function이 여러 개인 경우 scalarization 시 weight를 줌으로써 trade-off 다룰 수 있음

$$\text{minimize (w.r.t. } \mathbf{R}_+^{m+1}) \quad F(x) = (f_1(x), \ldots, f_m(x), f_0(x))$$

$$\tilde{\lambda}^T F(x) = f_0(x) + \sum_{i=1} \lambda_i f_i(x),$$ 위의 Multi-criterion problem에 대한 scalarization (람다0이 1)

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \ldots, m, \end{array}$$

이와 같은 primal problem의 라그랑지안과 정확하게 일치

만약 Primal problem이 convex 하다면 바로 앞 슬라이드의 예시와 동일한 상황(p*=d*)이 되고,
따라서 모든 Pareto optimal point가 어떤 람다 값에 대해 Scalarization 값을 최소화할 수 있을 것

4

Saddle Point Interpretation

Saddle Point의 개념을 차용한 Lagrange Duality의 이해

5.4 Saddle Point Interpretation

Max-min inequality and duality

$$\sup_{z \in Z} \inf_{w \in W} f(w,z) \le \inf_{w \in W} \sup_{z \in Z} f(w,z)$$ General max-min inequality. 이것이 duality와 어떻게 관련 있는가?

Equality constraint가 없는 simple problem에서 다음과 같은 등식에 대해 생각해 볼 수 있음

$$\begin{aligned} \sup_{\lambda \succeq 0} L(x,\lambda) &= \sup_{\lambda \succeq 0} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right) \\ &= \begin{cases} f_0(x) & f_i(x) \le 0, \quad i = 1, \ldots, m \\ \infty & \text{otherwise.} \end{cases} \end{aligned}$$

Lambda 값이 non-negative일 때 각 f 값들이 0보다 작거나 같다면 (inequality constraint을 만족한다면)

supremum은 f_0(x)가 됨, 이외의 경우 무한대. 즉 f_0(x)는 다음 등식의 infimum!

$$p^\star = \inf_{x} \sup_{\lambda \succeq 0} L(x,\lambda) \quad d^\star = \sup_{\lambda \succeq 0} \inf_{x} L(x,\lambda) \quad \sup_{\lambda \succeq 0} \inf_{x} L(x,\lambda) \le \inf_{x} \sup_{\lambda \succeq 0} L(x,\lambda)$$

위 사실과 dual function의 정의를 이용하면 p*와 d*를 각각 다음과 같이 나타낼 수 있고 따라서 위 부등식이 성립

Strong duality가 만족되어 등호가 성립하는 경우 '*saddle point property를 만족한다*'고 함

5.4 Saddle Point Interpretation

Saddle Point interpretation

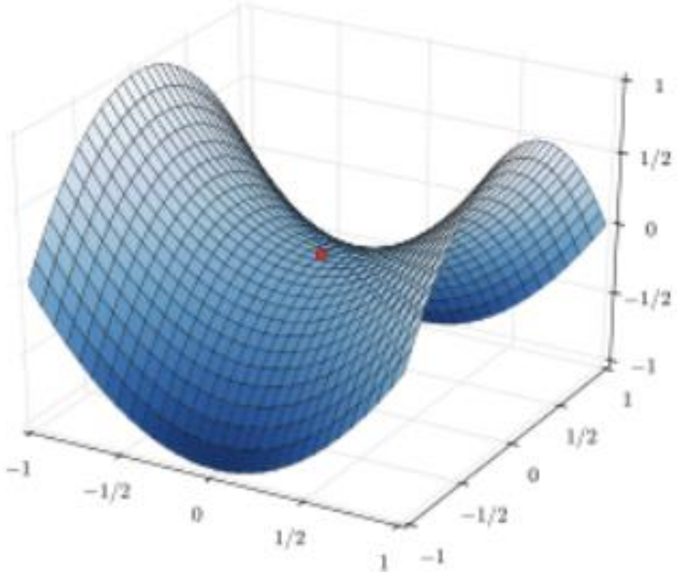
Saddle point란?

We refer to a pair $\tilde{w} \in W$, $\tilde{z} \in Z$ as a *saddle-point* for f (and W and Z) if

$$f(\tilde{w}, z) \le f(\tilde{w}, \tilde{z}) \le f(w, \tilde{z})$$

for all $w \in W$ and $z \in Z$. In other words, $\tilde{w}$ minimizes $f(w, \tilde{z})$ (over $w \in W$) and $\tilde{z}$ maximizes $f(\tilde{w}, z)$ (over $z \in Z$):

$$f(\tilde{w}, \tilde{z}) = \inf_{w \in W} f(w, \tilde{z}), \qquad f(\tilde{w}, \tilde{z}) = \sup_{z \in Z} f(\tilde{w}, z).$$



어느 방향에서 바라보면 극댓값을, 어느 방향에서 보면 극솟값을 갖는 경우

Lagrange Duality) Primal problem은 x*에서 최솟값 p*를 갖고, dual problem은 lambda*에서 최댓값 d*를 가짐

Strong duality가 충족될 경우 (p*=d*), 이 점은 어떤 관점에서는 최솟값, 어떤 관점에서는 최댓값이므로 일종의 saddle point

5.4.3, 5.4.4의 예시 참고 (제로섬 게임, 오퍼레이션 관리)

5

Optimality Conditions

Strong Duality와 Karush-Kuhn-Tucker (KKT) Conditions

5.5 Optimality Conditions

Stopping Criteria and subobtimality

알고리즘을 '적당히 돌아가게' 하기 위해 stopping criteria를 사용할 때, 그 기준을 정하는데 duality gap의 컨셉을 이용

대소 관계에 의해 다음의 조건을 만족

$$f_0(x) - p^\star \leq f_0(x) - g(\lambda, \nu).$$

$$p^\star \in [g(\lambda, \nu), f_0(x)], \quad d^\star \in [g(\lambda, \nu), f_0(x)],$$

주어진 feasible point가 얼마나 suboptimal한 지 보여주는 지표로 작용

$$f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)}) \leq \epsilon_{\mathrm{abs}} \qquad \frac{f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)})}{g(\lambda^{(k)}, \nu^{(k)})} \leq \epsilon_{\mathrm{rel}}$$

목표로 하는 gap을 stopping criteria인 epsilon으로 입력하여, 둘의 차이가 허용 오차인 epsilon보다 크지 않아지는 순간 알고리즘이 멈추도록 설계

5.5 Optimality Conditions

Complementary slackness

Primal problem과 Dual problem의 feasible한 값이 서로의 optimal solution이 되기 위한 조건

$$\begin{aligned} f_0(x^\star) &= g(\lambda^\star, \nu^\star) \\ &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\ &\le f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \quad \text{X*가 라그랑지안을 최소화} \\ &\le f_0(x^\star). \quad \text{Complementary slackness} \end{aligned}$$

첫 줄과 마지막 줄 사이에 등호가 성립하므로 모든 줄의 값이 같게 됨, 따라서 $\sum_{i=1}^m \lambda_i^\star f_i(x^\star) = 0.$

$$\lambda_i^\star f_i(x^\star) = 0, \quad i = 1, \ldots, m.$$

$$\lambda_i^\star > 0 \implies f_i(x^\star) = 0,$$

$$f_i(x^\star) < 0 \implies \lambda_i^\star = 0.$$

5.5 Optimality Conditions

KKT Optimality conditions

1. Nonconvex & Convex problem (general)

$$\nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(x^\star) = 0.$$

Strong duality에서의 primal and dual optimal points를 가정, 라그랑지안을 minimize 하므로 gradient가 0

따라서 다음의 모든 조건을 만족 (necessity)

$$\begin{array}{rcll} f_i(x^\star) & \leq & 0, & i = 1, \dots, m \\ h_i(x^\star) & = & 0, & i = 1, \dots, p \\ \lambda_i^\star & \geq & 0, & i = 1, \dots, m \\ \lambda_i^\star f_i(x^\star) & = & 0, & i = 1, \dots, m \\ \nabla f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^p \nu_i^\star \nabla h_i(x^\star) & = & 0, & \end{array}$$

feasibility

Complementary slackness

Stationarity

which are called the *Karush-Kuhn-Tucker* (KKT) conditions.

2. Convex problem

다음 조건을 모두 만족할 경우

$$\begin{array}{rcll} f_i(\tilde{x}) & \leq & 0, & i = 1, \dots, m \\ h_i(\tilde{x}) & = & 0, & i = 1, \dots, p \\ \tilde{\lambda}_i & \geq & 0, & i = 1, \dots, m \\ \tilde{\lambda}_i f_i(\tilde{x}) & = & 0, & i = 1, \dots, m \\ \nabla f_0(\tilde{x}) + \sum_{i=1}^m \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^p \tilde{\nu}_i \nabla h_i(\tilde{x}) & = & 0, & \end{array}$$

$$\begin{array}{rcl} g(\tilde{\lambda}, \tilde{\nu}) & = & L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \\ & = & f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{x}) \\ & = & f_0(\tilde{x}), \end{array}$$

Strong duality 성립, KKT condition을 만족하는 point가 primal and dual optimal (sufficiency)

5.5 Optimality Conditions

Example

Equality constrained convex quadratic minimization

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Ax = b, \end{array} \qquad (5.50)$$

where $P \in \mathbf{S}^n_+$. The KKT conditions for this problem are

$$Ax^\star = b, \qquad Px^\star + q + A^T\nu^\star = 0,$$

which we can write as

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ \nu^\star \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}.$$

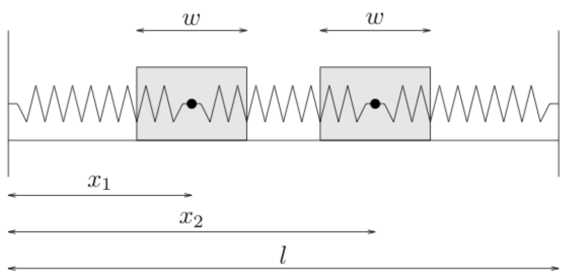


Figure 5.8 Two blocks connected by springs to each other, and the left and right walls. The blocks have width $w > 0$, and cannot penetrate each other or the walls.

Interpretation in mechanics

The potential energy in the springs, as a function of the block positions, is given by

$$f_0(x_1, x_2) = \frac{1}{2}k_1 x_1^2 + \frac{1}{2}k_2(x_2 - x_1)^2 + \frac{1}{2}k_3(l - x_2)^2,$$

where $k_i > 0$ are the stiffness constants of the three springs. The equilibrium position $x^\star$ is the position that minimizes the potential energy subject to the inequalities

$$w/2 - x_1 \le 0, \qquad w + x_1 - x_2 \le 0, \qquad w/2 - l + x_2 \le 0. \qquad (5.51)$$

$$\begin{array}{ll} \text{minimize} & (1/2)\left(k_1 x_1^2 + k_2(x_2 - x_1)^2 + k_3(l - x_2)^2\right) \\ \text{subject to} & w/2 - x_1 \le 0 \\ & w + x_1 - x_2 \le 0 \\ & w/2 - l + x_2 \le 0, \end{array} \qquad (5.52)$$

With λ_1, λ_2, λ_3 as Lagrange multipliers, the KKT conditions for this problem consist of the kinematic constraints (5.51), the nonnegativity constraints $\lambda_i \ge 0$, the complementary slackness conditions

$$\lambda_1(w/2 - x_1) = 0, \qquad \lambda_2(w - x_2 + x_1) = 0, \qquad \lambda_3(w/2 - l + x_2) = 0, \qquad (5.53)$$

and the zero gradient condition

$$\begin{bmatrix} k_1 x_1 - k_2(x_2 - x_1) \\ k_2(x_2 - x_1) - k_3(l - x_2) \end{bmatrix} + \lambda_1 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \lambda_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0. \qquad (5.54)$$

5.5 Optimality Conditions

Solving the primal problem via the dual

(5.5) Strong duality가 성립하고 dual optimal의 해가 존재하는 경우 primal optimal이 라그랑지안을 minimize

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^{m} \lambda_i^\star f_i(x) + \sum_{i=1}^{p} \nu_i^\star h_i(x),$$ Convex한 경우 유일해 x*를 갖는다

1. x*가 primal problem의 feasible set에 속한다면 이 점은 반드시 primal optimal에 해당
2. 그렇지 않다면 최적화 문제는 해를 갖지 않는다

<기억할 내용>

Strong duality 만족하면 dual problem의 해를 구하여 primal problem 다룰 수 있음

-> Strong duality(duality gap=0)는 다음의 경우 충족됨

a) Slater’s condition 만족 (strictly feasible 점 존재) + optimization problem이 convex

b) KKT condition 만족하는 점 존재 + optimization problem이 convex

: Primal feasible, Dual feasible, Complementary slackness, Stationarity

특히 이 경우 KKT condition의 해가 primal 및 dual optimal point가 됨

= Convexity가 이래서 좋구나, 앞으로는 KKT condition 활용한 최적화를 공부하겠구나!

Homework

5.26 Consider the QCQP

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & (x_1 - 1)^2 + (x_2 - 1)^2 \leq 1 \\ & (x_1 - 1)^2 + (x_2 + 1)^2 \leq 1 \end{array}$$

with variable $x \in \mathbf{R}^2$.

(a) Sketch the feasible set and level sets of the objective. Find the optimal point $x^\star$ and optimal value $p^\star$.

(b) Give the KKT conditions. Do there exist Lagrange multipliers $\lambda_1^\star$ and $\lambda_2^\star$ that prove that $x^\star$ is optimal?

(c) Derive and solve the Lagrange dual problem. Does strong duality hold?

감사합니다!