Continuous Optimisation, Chpt 4: Duality

Peter J.C. Dickinson

DMMP, University of Twente

p.j.c.dickinson@utwente.nl

http://dickinson.website/Teaching/2016CO.html

version: 05/12/16

Monday 17th October 2016

Table of Contents

Literature

For further reading:

- KRT: 3.1, 3.2, 3.3, 3.4
- FKS: 5.1, 5.2, 5.4

Lagrangian Function

Primal problem

$$\begin{aligned} \inf_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & g_j(\mathbf{x}) \leq 0 \quad \text{for all } j = 1, \ldots, m \\ & \mathbf{x} \in \mathcal{C}. \end{aligned} \tag{C}$$

$\mathcal{C} \subseteq \mathbb{R}^n$ is an open set, $\quad f, g_1, \ldots, g_m \in \mathrm{C}^1$,

$\mathcal{F} := \{\mathbf{x} \in \mathcal{C} : g_j(\mathbf{x}) \leq 0 \text{ for all } j = 1, \ldots, m\}$

Definition 4.1

Define the *Lagrangian function* $L : \mathcal{C} \times \mathbb{R}^m_+ \to \mathbb{R}$ *as*

$$L(\mathbf{x}, \mathbf{y}) := f(\mathbf{x}) + \sum_{i=1}^{m} y_i g_i(\mathbf{x}).$$

$L(\mathbf{x}, \mathbf{y})$ *is affine in* $\mathbf{y}$.

If (C) *is convex then* $L(\mathbf{x}, \mathbf{y})$ *is convex in* $\mathbf{x}$.

Derivative

Ex. 4.1 Find the gradient of the Lagrangian function.

Ex. 4.2 Assuming $f, g_1, \ldots, g_m \in C^2$, find the Hessian of the Lagrangian function.

Lemma 4.2

For $\mathrm{A} \in \mathcal{S}^n$, $\mathrm{B} \in \mathbb{R}^{n\times m}$ *we have*

$$\begin{pmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{B}^\mathsf{T} & \mathrm{O}_{m,m} \end{pmatrix} \succeq \mathrm{O} \qquad \Leftrightarrow \qquad \mathrm{A} \succeq \mathrm{O} \ \wedge \ \mathrm{B} = \mathrm{O}_{n,m}.$$

Ex. 4.3 For $f, g_1, \ldots, g_m \in C^2$, give a necessary and sufficient condition for L to be convex function.

Example

$$\min_x \quad x \quad \text{s. t.} \quad 2 \leq x, \quad x \in \mathbb{R}.$$

$$\begin{aligned} f(x) &= x, \qquad g(x) = 2 - x, \\ L(x,y) &= x + y(2-x), \\ \nabla L &= \begin{pmatrix} 1-y \\ 2-x \end{pmatrix}, \\ \nabla^2 L &= \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \not\succeq \mathrm{O}. \end{aligned}$$

https://ggbm.at/mM4gk3dM

Primal problem

Lemma 4.3

For $\mathbf{x} \in \mathcal{C}$ we have

$$\sup_{\mathbf{y} \in \mathbb{R}^m_+} \{L(\mathbf{x}, \mathbf{y})\} = \begin{cases} f(\mathbf{x}) & \text{if } \mathbf{x} \in \mathcal{F} \\ \infty & \text{if } \mathbf{x} \notin \mathcal{F}. \end{cases}$$

https://ggbm.at/tppBaWnS

Theorem 4.4

Problem (C) is equivalent to

$$\inf_{\mathbf{x} \in \mathcal{F}} \{f(\mathbf{x})\} = \inf_{\mathbf{x} \in \mathcal{C}} \sup_{\mathbf{y} \in \mathbb{R}^m_+} L(\mathbf{x}, \mathbf{y}).$$

inf-sup, sup-inf

Lemma 4.5

For $L : \mathcal{A} \times \mathcal{B} \to \mathbb{R}$ *we have*

$$\inf_{\mathbf{x}\in\mathcal{A}} \sup_{\mathbf{y}\in\mathcal{B}} L(\mathbf{x},\mathbf{y}) \geq \sup_{\mathbf{y}\in\mathcal{B}} \inf_{\mathbf{x}\in\mathcal{A}} L(\mathbf{x},\mathbf{y}).$$

Proof.

For all $\mathbf{x} \in \mathcal{A}$, $\mathbf{y} \in \mathcal{B}$ we have $\sup_{\mathbf{z}\in\mathcal{B}} L(\mathbf{x},\mathbf{z}) \geq L(\mathbf{x},\mathbf{y})$.
Therefore, for all $\mathbf{y} \in \mathcal{B}$ have $\inf_{\mathbf{x}\in\mathcal{A}} \sup_{\mathbf{z}\in\mathcal{B}} L(\mathbf{x},\mathbf{z}) \geq \inf_{\mathbf{x}\in\mathcal{A}} L(\mathbf{x},\mathbf{y})$.
Therefore $\inf_{\mathbf{x}\in\mathcal{A}} \sup_{\mathbf{z}\in\mathcal{B}} L(\mathbf{x},\mathbf{z}) \geq \sup_{\mathbf{y}\in\mathcal{B}} \inf_{\mathbf{x}\in\mathcal{A}} L(\mathbf{x},\mathbf{y})$. □

Corollary 4.6

We have

$$\inf_{\mathbf{x}\in\mathcal{F}} \{f(\mathbf{x})\} = \inf_{\mathbf{x}\in\mathcal{C}} \sup_{\mathbf{y}\in\mathbb{R}^m_+} L(\mathbf{x},\mathbf{y}) \geq \sup_{\mathbf{y}\in\mathbb{R}^m_+} \inf_{\mathbf{x}\in\mathcal{C}} L(\mathbf{x},\mathbf{y}).$$

Lagrangian Dual Problem

Definition 4.7

Define $\psi : \mathbb{R}^m_+ \to \mathbb{R} \cup \{-\infty\}$ *as* $\psi(\mathbf{y}) := \inf_{\mathbf{x}\in\mathcal{C}} L(\mathbf{x}, \mathbf{y})$.
The Lagrangian dual problem to (C) *is then given by*

$$\sup_{\mathbf{y}\in\mathbb{R}^m_+} \psi(\mathbf{y}) \tag{D}$$

Theorem 4.8

$\psi(\mathbf{y})$ *is concave function over its domain* $\mathbb{R}^m_+$ *(i.e.* $-\psi(\mathbf{y})$ *is convex function over* $\mathbb{R}^m_+$*) and thus* (D) *is convex maximisation problem.*

Theorem 4.9 (Weak duality)

We have $\inf_{\mathbf{x}\in\mathcal{F}}\{f(\mathbf{x})\} \geq \sup_{\mathbf{y}\in\mathbb{R}^m_+}\{\psi(\mathbf{y})\}$.

Corollary 4.10

For $\mathbf{x}_0 \in \mathcal{F}$ *and* $\mathbf{y}_0 \in \mathbb{R}^m_+$ *we have*

$$f(\mathbf{x}_0) \geq \inf_{\mathbf{x}\in\mathcal{F}}\{f(\mathbf{x})\} \geq \sup_{\mathbf{y}\in\mathbb{R}^m_+}\{\psi(\mathbf{y})\} \geq \psi(\mathbf{y}_0).$$

Example

Example

For $\mathbf{c}, \mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$, consider the primal problem

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^\mathsf{T}\mathbf{x} \\ \text{s. t.} \quad & \mathbf{a}_i^\mathsf{T}\mathbf{x} \geq b_i \text{ for all } i = 1, \ldots, m \\ & \mathbf{x} \in \mathbb{R}^n, \end{aligned}$$

The Lagrangian dual problem is equivalent to

$$\begin{aligned} \max_{\mathbf{y}} \quad & \mathbf{b}^\mathsf{T}\mathbf{y} \\ \text{s. t.} \quad & \mathbf{c} = \sum_{i=1}^m y_i \mathbf{a}_i \\ & \mathbf{y} \in \mathbb{R}^m_+. \end{aligned}$$

Examples

Ex. 4.4 The following two primal problems are equivalent.

$$\min_x \quad -x^2 \qquad\qquad \min_x \quad -x^2$$
$$\text{s.t.} \quad -1 \le x \le 1 \qquad\qquad \text{s.t.} \quad x^2 \le 1$$
$$x \in \mathbb{R} \qquad\qquad x \in \mathbb{R}$$

What are their optimal values and optimal solutions?

What are the functions ψ for these problems?

What are the optimal values and optimal solutions to their Lagrangian dual problems?

Strong duality

Theorem 4.11

If (C) *is convex and Slater's condition holds for* (C) *then*

$$\inf_{\mathbf{x}\in\mathcal{F}}\{f(\mathbf{x})\} = \sup_{\mathbf{y}\in\mathbb{R}^m_+}\{\psi(\mathbf{y})\},$$

i.e. there is Strong duality.

Furthermore, if in such a case we have $\inf_{\mathbf{x}\in\mathcal{F}}\{f(\mathbf{x})\} \neq \pm\infty$ *then the optimal solution to* $\sup_{\mathbf{y}\in\mathbb{R}^m_+}\{\psi(\mathbf{y})\}$ *is attained.*

We will now prove this.

Proof: Strong duality

$\inf_{\mathbf{x}\in\mathcal{F}}\{f(\mathbf{x})\} = \infty$ contradicts Slater's condition.

If $\inf_{\mathbf{x}\in\mathcal{F}}\{f(\mathbf{x})\} = -\infty$, then result follows from weak duality.

For case when $\gamma = \inf_{\mathbf{x}\in\mathcal{F}}\{f(\mathbf{x})\} \in \mathbb{R}$ we first require some preliminary results.

Proof: Strong duality, Singular and regular constraints

Definition 4.12

Set of singular and regular constraints are defined respectively as:

$$J_s := \{j \in \{1, \ldots, m\} : g_j(\mathbf{x}) = 0 \text{ for all } \mathbf{x} \in \mathcal{F}\},$$
$$J_r := \{1, \ldots, m\} \setminus J_s$$

Lemma 4.13

If Slater's condition holds then g_i is affine for all $i \in J_s$.

Lemma 4.14

If $g_1, \ldots, g_m, \mathcal{C}$ are convex and then $\exists \mathbf{z} \in \mathcal{F}$ s.t. $g_i(\mathbf{z}) < 0 \ \forall i \in J_r$.

If Slater's condition holds for (C) then such a point is referred to as an ideal Slater point.

Proof: Strong duality, Separation theorem

Lemma 4.15 (KRT, Theorem 2.23)

For $\mathcal{U} \subseteq \mathbb{R}^n$ *a convex set and* $\mathbf{w} \in \mathbb{R}^n \setminus \mathcal{U}$*, there exists* $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ *such that* $\mathbf{a}^\mathsf{T}\mathbf{w} \leq \mathbf{a}^\mathsf{T}\mathbf{u}$ *for all* $\mathbf{u} \in \mathcal{U}$ *and* $\mathbf{a}^\mathsf{T}\mathbf{w} < \mathbf{a}^\mathsf{T}\mathbf{v}$ *for all* $\mathbf{v} \in \operatorname{reint}(\mathcal{U})$*.*

Lemma 4.16

For $f, g_1, \ldots, g_m$ *convex,* $\mathcal{C} \subseteq \mathbb{R}^n$ *convex and Slater's condition holding, the following set is convex:*

$$\mathcal{U} = \left\{ \begin{pmatrix} u_0 \\ \mathbf{u} \end{pmatrix} \in \mathbb{R}^{m+1} : \exists \mathbf{x} \in \mathcal{C} \text{ s.t. } f(\mathbf{x}) < u_0, \begin{array}{l} g_i(\mathbf{x}) \leq u_i \ \forall i \in J_r, \\ g_j(\mathbf{x}) = u_j \ \forall j \in J_s \end{array} \right\}$$

Ex. 4.5 Prove that $\mathcal{U}$ given in Lemma 4.16 is a convex set.

Proof: Strong duality, Alternative system

We will prove Lemma 4.18, which together with the exercise proves the following theorem:

Theorem 4.17

For (C) convex, Slater's condition holding and $\gamma \in \mathbb{R}$, the following are alternative systems (i.e. exactly one holds):

1. $\exists \mathbf{x} \in \mathcal{C}$ *s.t.* $f(\mathbf{x}) < \gamma$, $g_i(\mathbf{x}) \leq 0$ *for all* i.
2. $\exists \mathbf{y} \in \mathbb{R}^m_+$ *s.t.* $f(\mathbf{x}) + \sum_{i=1}^m y_i g_i(\mathbf{x}) \geq \gamma$ *for all* $\mathbf{x} \in \mathcal{C}$.

Ex. 4.6 Show that both systems cannot be true.

Lemma 4.18

Consider (C) convex, Slater's condition holding and $\gamma \in \mathbb{R}$ such that statement (1) of Theorem 4.17 does not hold. Then statement (2) of Theorem 4.17 holds with $y_i > 0$ for all $i \in J_s$.

Proof: Strong duality, Lemma 4.18

Have $\begin{pmatrix}\gamma\\ \mathbf{0}\end{pmatrix} \notin \mathcal{U}$, as defined in Lemma 4.16.

From Lemma 4.15, $\exists \begin{pmatrix}y_0\\ \mathbf{y}\end{pmatrix} \in \mathbb{R}^{m+1}$ such that

$$y_0\gamma \leq y_0 u_0 + \sum_{i=1}^{m} y_i u_i \qquad \text{for all } \begin{pmatrix}u_0\\ \mathbf{u}\end{pmatrix} \in \mathcal{U},$$

$$y_0\gamma < y_0 v_0 + \sum_{i=1}^{m} y_i v_i \qquad \text{for all } \begin{pmatrix}v_0\\ \mathbf{v}\end{pmatrix} \in \text{reint}(\mathcal{U}).$$

We split the remainder of this proof into four parts:

I. $y_i \geq 0$ for all $i \in \{0\} \cup J_r$.

II. $y_0\gamma \leq y_0 f(\mathbf{x}) + \sum_{i=1}^m y_i g_i(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{C}$.

III. $y_0 > 0$, and thus w.l.o.g. $y_0 = 1$.

IV. W.l.o.g. $y_i > 0$ for all $i \in J_s$.

Ex. 4.7 Prove parts I and II.

Proof: Strong duality, Lemma 4.18, III

Part III:

- Assume for sake of contradiction that $y_0 = 0$.
- Considering the ideal Slater point $\mathbf{z}$, together with the result from II, implies that $y_i = 0$ for all $i \in J_r$.
- From II, now have $0 \leq \sum_{i \in J_s} y_i g(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{C}$.
- Consider arbitrary $\begin{pmatrix} v_0 \\ \mathbf{v} \end{pmatrix} \in \operatorname{reint}(\mathcal{U})$. Have $0 < \sum_{i \in J_s} y_i v_i$.

  $\exists \mathbf{x} \in \mathcal{C}$ such that $g_i(\mathbf{x}) = v_i$ for all $i \in J_s$.
- As $\mathcal{C}$ is open, $\exists \varepsilon > 0$ such that $\mathbf{w} = \mathbf{z} + \varepsilon(\mathbf{z} - \mathbf{x}) \in \mathcal{C}$.
- Letting $\theta = (1+\varepsilon)^{-1} \in (0,1)$ have $\mathbf{z} = (1-\theta)\mathbf{x} + \theta\mathbf{w}$.
- As g_i are affine for all $i \in J_s$, have the contradiction

$$0 = \sum_{i \in J_s} y_i \underbrace{g_i(\mathbf{z})}_{=0} = (1-\theta) \underbrace{\sum_{i \in J_s} y_i g(\mathbf{x})}_{>0} + \theta \underbrace{\sum_{i \in J_s} y_i g_i(\mathbf{w})}_{\geq 0} > 0.$$

Proof: Strong duality, Lemma 4.18, IV

Part IV:

- Want to show that w.l.o.g. we may assume $y_i > 0$ for all $i \in J_s$. Will prove this by induction on $|J_s|$.
- This is true for all infeasible systems with $|J_s| = 0$. Suppose true for all infeasible systems with $|J_s| \leq p$, then we will show it is also true for all infeasible systems with $|J_s| = p + 1$.
- Consider convex system with Slater's condition holding and $|J_s| = p + 1$. For all $t \in J_s$ have $\nexists \mathbf{x} \in \mathcal{C}$ s.t. $g_t(\mathbf{x}) < 0$, $g_i(\mathbf{x}) \leq 0$ for all $i \in J_r \cup J_s \setminus \{t\}$.
- $\exists \widehat{\mathbf{y}}^t \in \mathbb{R}^m_+$ with $\widehat{y}^t_t = 1$ s.t. for all $\mathbf{x} \in \mathcal{C}$ we have

$$0 \leq g_t(\mathbf{x}) + \sum_{i \in J_r \cup J_s \setminus \{t\}} \widehat{y}^t_i g_i(\mathbf{x}) = \sum_i \widehat{y}^t_i g_i(\mathbf{x}).$$

- For all $t \in J_s$, can add sufficiently high multiples of these $\widehat{\mathbf{y}}^t$'s onto the original $\mathbf{y}$.

Strong duality: Summary

For (C) convex and Slater's condition holding for (C) we considered the statements

1. $\exists \mathbf{x} \in \mathcal{C}$ s.t. $f(\mathbf{x}) < \gamma$, $g_i(\mathbf{x}) \leq 0$ for all i.
2. $\exists \mathbf{y} \in \mathbb{R}^m_+$ s.t. $f(\mathbf{x}) + \sum_{i=1}^m y_i g_i(\mathbf{x}) \geq \gamma$ for all $\mathbf{x} \in \mathcal{C}$.

Showed: (1) holds $\Rightarrow$ (2) cannot hold.
Also showed: (1) does not hold $\Rightarrow$ (2) must hold.
Therefore the statements are alternative systems (i.e. exactly one holds). Equivalently we have the following lemma:

Lemma 4.19

For (C) convex, Slater's condition holding and $\gamma \in \mathbb{R}$, the following are alternative systems (i.e. exactly one holds):

1. $\exists \mathbf{x} \in \mathcal{F}$ *s.t.* $f(\mathbf{x}) < \gamma$.
2. $\exists \mathbf{y} \in \mathbb{R}^m_+$ *s.t.* $\psi(\mathbf{y}) \geq \gamma$.

Proof: Strong Duality

Lemma 4.19

For (C) convex, Slater's condition holding and $\gamma \in \mathbb{R}$, the following are alternative systems (i.e. exactly one holds):

1. $\exists \mathbf{x} \in \mathcal{F}$ s.t. $f(\mathbf{x}) < \gamma$.
2. $\exists \mathbf{y} \in \mathbb{R}^m_+$ s.t. $\psi(\mathbf{y}) \geq \gamma$.

For $f, g_1, \ldots, g_m, \mathcal{C}$ convex and Slater's condition holding, consider $\gamma = \inf_{\mathbf{x} \in \mathcal{F}}\{f(\mathbf{x})\}$.

$\nexists \mathbf{x} \in \mathcal{F}$ s.t. $f(\mathbf{x}) < \gamma$, therefore $\exists \mathbf{y}^* \in \mathbb{R}^m_+$ s.t. $\psi(\mathbf{y}^*) \geq \gamma$.

$\gamma \leq \psi(\mathbf{y}^*) \leq \sup_{\mathbf{y} \in \mathbb{R}^m_+}\{\psi(\mathbf{y})\} \leq \inf_{\mathbf{x} \in \mathcal{F}}\{f(\mathbf{x})\} = \gamma$.

Therefore $\sup_{\mathbf{y} \in \mathbb{R}^m_+}\{\psi(\mathbf{y})\} = \inf_{\mathbf{x} \in \mathcal{F}}\{f(\mathbf{x})\}$ and $\mathbf{y}^*$ is an optimal solution to $\sup_{\mathbf{y} \in \mathbb{R}^m_+}\{\psi(\mathbf{y})\}$.

e.g.: Convex but no Slater condition

Example

Consider $\inf_{\mathbf{x}\in\mathbb{R}^2}\{(x_1-1)^2 : x_1^2\exp(x_2)\leq 0\}$.
Have $\mathcal{F}=\{\mathbf{0}\}\times\mathbb{R}$ and optimal value 1 (attained at all $\mathbf{x}\in\mathcal{F}$).

$f(\mathbf{x})=(x_1-1)^2$ with $\nabla^2 f(\mathbf{x})=\begin{pmatrix}2 & 0\\ 0 & 0\end{pmatrix}\succeq \mathrm{O}$, and

$g_1(\mathbf{x})=x_1^2\exp(x_2)$ with $\nabla^2 g_1(\mathbf{x})=2\exp(x_2)\begin{pmatrix}1\\ x_1\end{pmatrix}\begin{pmatrix}1\\ x_1\end{pmatrix}^\mathsf{T}\succeq \mathrm{O}$,

thus convex.

Slater's condition does not hold as $g_1(\mathbf{x})=0$ for all $\mathbf{x}\in\mathcal{F}$.

For $y\geq 0$ have $L(\mathbf{x};y)=(x_1-1)^2+yx_1^2\exp(x_2)\geq 0$ and
$0\leq\psi(y)=\inf_{\mathbf{x}\in\mathbb{R}^2}L(\mathbf{x},y)\leq\lim_{x_2\to-\infty}L(1,x_2;y)=0$.

Therefore $\sup_{y\geq 0}\psi(y)=0$.

Examples

Ex. 4.8 Analyse the following problems, determining whether they are convex, whether Slater's condition holds, what the optimal values to the primal and dual problems are, whether the optimal values are attained and whether there is strong duality:

1. $\min_{\mathbf{x}\in\mathbb{R}^2}\{x_1^2 + x_2^2 : x_1 + x_2 \geq 1\}$
2. $\min_{\mathbf{x}\in\mathbb{R}^2}\{x_2 : \exp(x_1) \leq x_2\}$
3. $\min_{x\in\mathbb{R}}\{x : x^2 \leq 0\}$
4. $\min_{x\in\mathbb{R}}\{(x-1)^2 : x^3 \geq 4x\}$

Table of Contents

Saddle points

Definition 4.20

A vector pair $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \in \mathcal{C} \times \mathbb{R}^m_+$ *is called a* *saddle point* *of the Lagrangian function L if*

$$L(\widehat{\mathbf{x}}, \mathbf{y}) \leq L(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \leq L(\mathbf{x}, \widehat{\mathbf{y}}) \qquad \text{for all } \mathbf{x} \in \mathcal{C},\ \mathbf{y} \in \mathbb{R}^m_+.$$

Equivalently:

$$\sup_{\mathbf{y} \in \mathbb{R}^m_+} L(\widehat{\mathbf{x}}, \mathbf{y}) = L(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) = \inf_{\mathbf{x} \in \mathcal{C}} L(\mathbf{x}, \widehat{\mathbf{y}}).$$

Theorem 4.21

A vector pair $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \in \mathcal{C} \times \mathbb{R}^m_+$ *is a saddle point of L iff* $\widehat{\mathbf{x}} \in \mathcal{F}$ *and* $f(\widehat{\mathbf{x}}) = \psi(\widehat{\mathbf{y}})$. *In this event we have*

$$\begin{aligned} L(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) = f(\widehat{\mathbf{x}}) &= \inf_{\mathbf{x} \in \mathcal{F}} f(\mathbf{x}) \\ &= \psi(\widehat{\mathbf{y}}) = \sup_{\mathbf{y} \in \mathbb{R}^m_+} \psi(\mathbf{y}). \end{aligned}$$

Saddle - KKT

Consider $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \in \mathcal{C} \times \mathbb{R}^m_+$ and recall $L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) + \sum_{i=1}^m y_i g_i(\mathbf{x})$, where $\mathcal{C}$ is an open set and $f, g_1, \ldots, g_m \in \mathrm{C}^1$.

Lemma 4.22

$L(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) = \sup_{\mathbf{y} \in \mathbb{R}^m_+} L(\widehat{\mathbf{x}}, \mathbf{y})$ *iff* $\widehat{\mathbf{x}} \in \mathcal{F}$ *and* $\widehat{y}_i g_i(\widehat{\mathbf{x}}) = 0 \ \forall i$.

Lemma 4.23

If $L(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) = \inf_{\mathbf{x} \in \mathcal{C}} L(\mathbf{x}, \widehat{\mathbf{y}})$ *then* $\mathbf{0} = \nabla_{\mathbf{x}} L(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$.
If (C) *convex and* $\mathbf{0} = \nabla_{\mathbf{x}} L(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ *then* $L(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) = \inf_{\mathbf{x} \in \mathcal{C}} L(\mathbf{x}, \widehat{\mathbf{y}})$.
Have $\nabla_{\mathbf{x}} L(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) = \nabla f(\widehat{\mathbf{x}}) + \sum_{i=1}^m \widehat{y}_i \nabla g_i(\widehat{\mathbf{x}})$.

Theorem 4.24

If $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ *is a saddle point then it fulfills the KKT conditions.*
If (C) *is convex and* $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ *fulfills the KKT conditions then* $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ *is a saddle point.*

Table of Contents

Wolfe-dual

How do we find $\psi(\mathbf{y}) = \inf_{\mathbf{x}\in\mathcal{C}} L(\mathbf{x},\mathbf{y})$?

For (C) convex, $\widehat{\mathbf{y}} \in \mathbb{R}^m_+$ and $\widehat{\mathbf{x}} \in \mathcal{C}$ we have

$$L(\widehat{\mathbf{x}},\widehat{\mathbf{y}}) = \psi(\widehat{\mathbf{y}}) \qquad \Leftrightarrow \qquad \mathbf{0} = \nabla_{\mathbf{x}} L(\widehat{\mathbf{x}},\widehat{\mathbf{y}}).$$

Therefore, when (C) is convex we have

$$\sup_{\mathbf{y}\in\mathbb{R}^m_+} \psi(\mathbf{y}) \geq \sup_{\mathbf{x},\mathbf{y}} \left\{ L(\mathbf{x},\mathbf{y}) : \mathbf{0} = \nabla_{\mathbf{x}} L(\mathbf{x},\mathbf{y}),\ \mathbf{y}\in\mathbb{R}^m_+,\ \mathbf{x}\in\mathcal{C} \right\} \tag{WD}$$

We refer to this optimisation problem, (WD), as the Wolfe-dual.

Weak/Strong duality

Theorem 4.25

If (C) *is convex,* $\widehat{\mathbf{x}} \in \mathcal{F}$ *and* $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ *is feasible for* (WD) *then* $L(\bar{\mathbf{x}}, \bar{\mathbf{y}}) = \psi(\bar{\mathbf{y}}) \leq f(\widehat{\mathbf{x}})$.

Theorem 4.26

Suppose (C) *is convex with Slater's condition holding and* $\widehat{\mathbf{x}} \in \mathcal{F}$ *is a minimiser of* (C)*. Then* $\exists \widehat{\mathbf{y}} \in \mathbb{R}^m_+$ *such that* $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ *is an optimal solution to* (WD)*, and* $L(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) = f(\widehat{\mathbf{x}})$.

Comparison of duals

Remark 4.27

The Lagrangian dual gives a lower bound for convex and nonconvex problems. The Wolfe-dual only guarantees a lower bound for convex differentiable functions.

Ex. 4.9 For the following primal problem, find the optimal values and optimal solutions to the primal problem, its Lagrangian dual and its Wolfe-dual.

$$\begin{aligned} \min_x \quad & \sin x \\ \text{s.t.} \quad & x \leq 0 \\ & x \in \mathbb{R} \end{aligned}$$

How difficult is the Wolfe-dual?

Ordinarily the Wolfe-dual is a difficult problem: The inequality constraints have moved into the objective and equality constraints, there are extra variables and there are gradients. However it can still be useful.

Ex. 4.10 For the parameters $\mathbf{c}, \mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$, give an explicit formulation for the Wolfe-Dual to

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^\mathsf{T}\mathbf{x} \\ \text{s. t.} \quad & \mathbf{a}_i^\mathsf{T}\mathbf{x} \geq b_i \text{ for } i = 1, \ldots, m, \\ & \mathbf{x} \in \mathbb{R}^m. \end{aligned}$$

Fixed $\widehat{\mathbf{x}}$

Lemma 4.28

The Wolfe-dual is lower bounded by the following linear problem for all fixed $\widehat{\mathbf{x}} \in \mathcal{C}$*:*

$$\begin{aligned} \max_{\mathbf{y}} \quad & f(\widehat{\mathbf{x}}) + \sum_{i=1}^{m} y_i g_i(\widehat{\mathbf{x}}) \\ \text{s. t.} \quad & \nabla f(\widehat{\mathbf{x}}) + \sum_{i=1}^{m} y_i \nabla g_i(\widehat{\mathbf{x}}) = \mathbf{0} \\ & \mathbf{y} \in \mathbb{R}^m_+ \end{aligned}$$

This in turn gives a lower bound on the original primal problem.

For a convex function $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1$, let ν be the optimal value to the following problem:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \quad f(\mathbf{x}) \qquad \text{s. t.} \quad -1 \le x_i \le 1 \text{ for all } i = 1, \ldots, n.$$

Ex. 4.11 Show that the Wolfe-dual to this problem is equivalent to

$$\sup_{\mathbf{x},\mathbf{u},\mathbf{v}} \left\{ f(\mathbf{x}) - \mathbf{x}^{\mathsf{T}}\nabla f(\mathbf{x}) - \mathbf{1}^{\mathsf{T}}\mathbf{u} - \mathbf{1}^{\mathsf{T}}\mathbf{v} : \begin{array}{l} \mathbf{v} - \mathbf{u} = \nabla f(\mathbf{x}), \\ \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{u}, \mathbf{v} \in \mathbb{R}^n_+ \end{array} \right\}.$$

This is equivalent to

$$\sup_{\mathbf{x}} \left\{ f(\mathbf{x}) - \mathbf{x}^{\mathsf{T}}\nabla f(\mathbf{x}) - \|\nabla f(\mathbf{x})\|_1 : \mathbf{x} \in \mathbb{R}^n \right\}.$$

Therefore for all $\mathbf{x} \in [-1, 1]^n$ we have the following, with equality at the optimal:

$$f(\mathbf{x}) - \mathbf{x}^{\mathsf{T}}\nabla f(\mathbf{x}) - \|\nabla f(\mathbf{x})\|_1 \le \nu \le f(\mathbf{x}).$$

Exercises

Ex. 4.12 Do KRT exercise 3.3.