

Gliederung:

I. Die Konjugierte Funktion

- I.1. Definition und Beispiele
- I.2. Grundlegende Eigenschaften

II. Die Lagrange-duale Funktion

- II.1. Die Lagrange Funktion
- II.2. Die Lagrange-duale Funktion
- II.3. Untere Grenzen an optimalen Werten
- II.4. Lineare Approximation
- II.5. Beispiele
- II.6. Die Lagrange-duale Funktion und die konjugierte Funktion

I. Die konjugierte Funktion

I.1. Definition und Beispiele

$f: \mathbb{R}^n \rightarrow \mathbb{R}$. Die Funktion $f^*: \mathbb{R}^n \rightarrow \mathbb{R}$, definiert als $f^*(y) = \sup_{x \in \text{dom } f} y^T x - f(x)$ (I.1)

wird die Konjugierte der Funktion f genannt.

- $\text{dom } f^* = \{y \in \mathbb{R}^n : \sup_{x \in \text{dom } f} y^T x - f(x) < \infty\}$

- f^* ist konvex

Bsp.: $f: \mathbb{R} \rightarrow \mathbb{R}$, $y \in \mathbb{R}$

Die konjugierte Funktion $f^*(y)$ ist der größte gerichtete Abstand der Funktionswerte der linearen Funktion yx und $f(x)$.

Wenn f differenzierbar ist, ergibt sich dies an einem der Punkte x , wo $f'(x) = y$.

\Rightarrow Wenn f differenzierbar: $\frac{d}{dx}(y^T x - f(x)) = 0$

Beispiel I.1.

- Affine Funktion $f(x) = ax + b \Rightarrow \text{dom } f^* = \{a\}$ und $f^*(a) = -b$
- Negativer Logarithmus $f(x) = -\log(x)$, mit $\text{dom } f = \mathbb{R}_{++} \Rightarrow \text{dom } f^* = \{y : y < 0\} = -\mathbb{R}_{++}$ und $f^*(y) = -\log(-y) - 1$ für $y < 0$
- Exponentialfunktion $f(x) = e^x \Rightarrow \text{dom } f^* = \mathbb{R}_+$ und $f^*(y) = y \log(y) - y$ für $y \geq 0$

- Negative Entropie $f(x) = x \log x$, mit $\text{dom } f = \mathbb{R}_+$ \Rightarrow
 $\text{dom } f^* = \mathbb{R}$ und $f^*(y) = e^{y-1} \quad \forall y$
- Inverse $f(x) = \frac{1}{x}$ auf \mathbb{R}_{++} \Rightarrow $\text{dom } f^* = -\mathbb{R}_+$ und $f^*(y) = -2(-y)^{-1/2}$ für $y \leq 0$

Beispiel I.2.

Strikt konvexe quadratische Funktion $f(x) = \frac{1}{2} x^T Q x$, mit $Q \in S_{++}^n$ \Rightarrow
 $f^*(y) = \frac{1}{2} y^T Q^{-1} y$.

Beispiel I.3.

Logarithmus-Determinante. Wir betrachten $f(X) = \log \det(X^{-1})$ auf S_{++}^n . Die konjugierte Funktion ist definiert als $f^*(Y) := \sup_{X>0} (\text{tr}(YX) + \log \det X)$.
 $\Rightarrow f^*(Y) = \log \det[(-Y)^{-1}] - n$, mit $\text{dom } f^* = -S_{++}^n$ für $Y < 0$.

Beispiel I.4.

Indikator Funktion. I_S sei die Indikator Funktion von einer (nicht unbedingt konvexen) Menge $S \subseteq \mathbb{R}^n$, d.h. $I_S(x) = 0$ auf $\text{dom } I_S = S$. Ihre Konjugierte ist $I_S^*(y) = \sup_{x \in S} (y^T x)$, was die Stützfunktion der Menge S ist.

Beispiel I.5.

Norm. Sei $\|\cdot\|$ eine Norm in \mathbb{R}^n , mit dualer Norm $\|\cdot\|_*$. Die Konjugierte von $f(x) = \|x\|$ ist
 $f^*(y) = \begin{cases} 0, & \|y\|_* \leq 1 \\ \infty, & \text{sonst} \end{cases}$, d.h. die Konjugierte einer Norm ist die Indikatorfunktion von der Einheitskugel bzgl. der dualen Norm.

Beispiel I.6.

Quadratische Norm. $f(x) = \frac{1}{2} \|x\|^2$, wobei $\|\cdot\|$ ein Norm mit der dualen Norm $\|\cdot\|_*$.
 $\Rightarrow f^*(y) = \frac{1}{2} \|y\|_*^2$

I.2. Grundlegende Eigenschaften

- Fenchel's Ungleichheit: $f(x) + f^*(y) \geq x^T y \quad \forall x, y$
- Z.B. mit $f(x) = \frac{1}{2} x^T Q x$, $Q \in S_{++}^n$, erhalten wir $x^T y \leq \frac{1}{2} x^T Q x + \frac{1}{2} y^T Q^{-1} y$.

- Konjugierte der Konjugierten

Wenn f konvex und $\text{epi } f$ eine abgeschlossene Menge, dann $f^{***} = f$.
Z.B. wenn $\text{dom } f = \mathbb{R}^n$, dann haben wir $f^{**} = f$.

- Differenzierbare Funktionen

Wir nehmen an, dass f konvex und differenzierbar ist mit $\text{dom } f = \mathbb{R}^n$.
Sei $x \in \mathbb{R}^n$ beliebig und wir definieren $y = \nabla f(x)$. Dann haben wir
 $f^*(y) = x^T \nabla f(x) - f(x)$.

- Skalierung und Vergleich mit affiner Transformation

Für $a > 0$ und $b \in \mathbb{R}$ ist die Konjugierte von $g(x) = a f(x) + b$: $g^*(y) = a f^*\left(\frac{y}{a}\right) - b$.

Wir nehmen an, dass $A \in \mathbb{R}^{n \times n}$ nonsingular und $b \in \mathbb{R}^n$.

Dann ist die Konjugierte von $g(x) = f(Ax + b)$:

$$g^*(y) = f^*(A^{-T}y) - b^T A^{-T}y \text{ mit } \text{dom } g^* = A^T \text{dom } f^*$$

- Summen von unabhängigen Funktionen

Wenn $f(u, v) = f_1(u) + f_2(v)$, wo f_1 und f_2 konvexe Funktionen mit Konjugierten f_1^* und f_2^* sind, dann $f^*(w, z) = f_1^*(w) + f_2^*(z)$.

II. Die Lagrange-duale Funktion

II.1. Die Lagrange Funktion

Wir betrachten ein Optimalitätsproblem in Standardform:

$$\begin{aligned} \min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0 \quad i=1, \dots, m \\ & h_i(x) = 0 \quad i=1, \dots, p \end{aligned} \quad (\text{II.1})$$

- $x \in \mathbb{R}^n$

- $D = \text{dom } f_1 \cap \dots \cap \text{dom } f_m \cap \text{dom } h_1 \cap \dots \cap \text{dom } h_p \neq \emptyset$

- optimaler Wert von (5.1): p^*

Die grundlegende Idee in der Lagrange Dualität ist es, die Einschränkungen in (II.1) zu berücksichtigen, durch Vergrößern der Zielfunktion mit einer gewichteten Summe der Einschränkungsfunktionen.

Definition:

Die Funktion $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$, $L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \mu_i h_i(x)$ mit

$\text{dom } L = D \times \mathbb{R}^m \times \mathbb{R}^p$ heißt Lagrangefunktion zum Problem (II.1).

- λ_i Lagrange Multiplikator bzgl. der i -ten Ungleichheitsbeschränkung $f_i(x) \leq 0$

- μ_i Lagrange Multiplikator bzgl. der i -ten Gleichheitsbeschränkung $h_i(x) = 0$

II.2. Die Lagrange-duale Funktion

Definition:

Die Lagrange-duale Funktion (oder nur duale Funktion) $g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ ist der Minimalwert der Lagrangefunktion über x : Für $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$:

$$g(\lambda, \mu) = \inf_{x \in D} L(x, \lambda, \mu) = \inf_{x \in D} [f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \mu_i h_i(x)]$$

Wenn die Lagrangefunktion nach unten unbeschränkt ist in x , dann nimmt die duale Funktion den Wert $-\infty$ an.

Da die duale Funktion das punktweise Infimum von einer Familie von affinen Funktionen von (λ, μ) ist, ist sie konkav, sogar wenn das Problem (II.1) nicht konvex ist.

II.3. Untere Grenzen an optimalen Werten

Durch die duale Funktion kann man untere Grenzen an den optimalen Werten p^* des Problems (II.1) bestimmen:

Für jedes $\lambda \geq 0$ und jedes μ gilt $g(\lambda, \mu) \leq p^*$ (II.2)

Für $g(\lambda, \mu) = -\infty$ liefert uns diese Aussage allerdings keine neuen Erkenntnisse.

Die duale Funktion gibt nur dann eine nicht triviale untere Grenze zu p^* an,

wenn $\lambda \geq 0$ und $(\lambda, \mu) \in \text{dom } g$, d.h. $g(\lambda, \mu) > -\infty$.

Ein Paar (λ, μ) mit $\lambda \geq 0$ und $(\lambda, \mu) \in \text{dom } g$ heißt dual zulässig.

Bsp.: Für ein einfaches Problem mit $x \in \mathbb{R}$ und einer Ungleichungsbeschränkung.

II.4. Lineare Approximation

Hier: Lineare Approximation von (II.1) mithilfe der Indikatorfunktionen der Mengen $\{0\}$ und $-\mathbb{R}_+$.

Zuerst schreiben wir das Originalproblem (II.1) als ein unbeschränktes Problem,

$$\min f_0(x) + \sum_{i=1}^m I_-(f_i(x)) + \sum_{i=1}^p I_0(h_i(x)) \quad (\text{II.3})$$

wobei $I : \mathbb{R} \rightarrow \mathbb{R}$ die Indikatorfunktion für die nicht positiven reellen Zahlen $-\mathbb{R}_+$ ist,

$$I_-(u) = \begin{cases} 0 & , \text{für } u \leq 0 \\ \infty & , \text{für } u > 0 \end{cases}$$

und I_0 die Indikatorfunktion von $\{0\}$,

$$I_0(u) =$$

Nun können wir in (II.3) die Funktion $I_-(u)$ durch die lineare Funktion $\lambda_i u$ ($\lambda_i \geq 0$) und die Funktion $I_0(u)$ durch $\mu_i u$ ersetzen.

Aus der Zielfunktion wird dann die Lagrangefunktion $L(x, \lambda, \mu)$, und der duale Funktionswert $g(\lambda, \mu)$ ist dann der optimale Wert des Problems

$$\min L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \mu_i h_i(x) \quad (\text{II.4}).$$

Offensichtlich ist die Approximation der Indikatorfunktionen $I_-(u)$ und $I_0(u)$ mit linearen Funktionen $\lambda_i u$ und $\mu_i u$ eher schlecht.

Da aber $\lambda_i u \leq I_-(u)$ und $\mu_i u \leq I_0(u)$ für alle u , sehen wir sofort, dass die duale Funktion eine untere Grenze am optimalen Wert des Originalproblems liefert (\rightarrow II.3.).

II.5. Beispiele

- Lösen von linearen Gleichungen mithilfe der Methode der kleinsten Quadrate

$$\min x^T x \quad (\text{II.5}), \text{ wobei } A \in \mathbb{R}^{p \times n} .$$

$$\text{s.t. } Ax=b$$

• Lagrangefunktion: $L(x, \mu) = x^T x + \mu^T (Ax - b)$, mit Domain $\mathbb{R}^n \times \mathbb{R}^p$.

• duale Funktion: $g(\mu) = \inf_x L(x, \mu)$.

• $L(x, \mu)$ konvexe quadratische Funktion von $x \Rightarrow$

$$\exists x : \nabla_x L(x, \mu) = 2x + A^T \mu = 0 \Rightarrow x = -\frac{1}{2} A^T \mu \Rightarrow$$

$$g(\mu) = L\left(-\frac{1}{2} A^T \mu, \mu\right) = -\frac{1}{4} \mu^T A A^T \mu - b^T \mu, \text{ eine konkave quadratische Funktion}$$

mit Domain \mathbb{R}^p

• (II.2) $\Rightarrow \forall \mu \in \mathbb{R}^p : -\frac{1}{4} \mu^T A A^T \mu - b^T \mu \leq \inf \{x^T x : Ax=b\}$

- LP in Standardform

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax=b \quad (\text{II.6}), \text{ mit Ungleichheitsbeschränkungsfunktionen } f_i(x) = -x_i, \quad i=1, \dots, n \\ & x \geq 0 \end{aligned}$$

Um die Lagrangefunktion zu bilden, führen wir Multiplikatoren λ_i für die n Ungleichungsbeschränkungen und Multiplikatoren μ_i für die Gleichheitsbeschränkungen ein, und erhalten

$$L(x, \lambda, \mu) = c^T x - \sum_{i=1}^n \lambda_i x_i + \mu^T (Ax - b) = -b^T \mu + (c + A^T \mu - \lambda)^T x$$

Die duale Funktion ist $g(\lambda, \mu) = \inf_x L(x, \lambda, \mu) = -b^T \mu + \inf_x (c + A^T \mu - \lambda)^T x$,

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

Wir erhalten nur dann eine nicht triviale untere Grenze nach (II.2), wenn λ und μ $\lambda \geq 0$ und $A^T \mu - \lambda + c = 0$ erfüllen. Dann ist $-b^T \mu$ eine untere Grenze am Optimalwert des LP (II.6).

- Zwei-Wege-Teilungs Problem

Wir betrachten das (nicht konvexe) Problem $\min x^T W x$ (II.7), mit $W \in S^n$.
 $\text{s.t. } x_i^2 = 1, \quad i=1, \dots, n$

Aus der Beschränkung ergibt sich, dass $x_i = \pm 1$, also könnte man äquivalent auch den Vektor mit Komponenten ± 1 suchen, der $x^T W x$ minimiert.

Man kann das Problem (II.7) als ein Zwei-Wege-Teilungs Problem auf einer Menge von n Elementen, hier: $\{1, \dots, n\}$, interpretieren:

Sei $x = (x_1, \dots, x_n)$ zulässig, dann: $\{1, \dots, n\} = \{i : x_i = -1\} \cup \{i : x_i = 1\}$.

Der Matrixkoeffizient W_{ij} beschreibt die „Kosten“, die Elemente i und j im selben Abschnitt zu haben, und $-W_{ij}$ die „Kosten“, i und j in verschiedenen Teilen zu haben.

Die Zielfunktion in (II.7) sind die totalen Kosten, über alle Paare von Elementen, und das gesamte Problem (II.7) ist die Teilung mit den geringsten totalen Kosten zu finden.

Die Lagrangefunktion zu diesem Problem ist dann:

$$\begin{aligned} L(x, \mu) &= x^T W x + \sum_{i=1}^n \mu_i (x_i^2 - 1) \\ &= x^T (W + \text{diag}(\mu)) x - \mathbf{1}^T \mu \end{aligned}$$

Wir erhalten die Lagrange-duale Funktion durch Minimieren über x:

$$\begin{aligned} g(\mu) &= \inf_x x^T (W + \text{diag}(\mu)) x - \mathbf{1}^T \mu \\ &= \begin{cases} -\mathbf{1}^T \mu & , W + \text{diag}(\mu) \geq 0 \\ -\infty & , \text{sonst} \end{cases} \end{aligned}$$

(Da das Infimum eines quadratischen Ausdrucks entweder Null (wenn der Ausdruck positiv semidefinit ist) oder $-\infty$ (wenn er nicht positiv semidefinit ist) ist.)

Diese duale Funktion liefert wieder untere Grenzen am optimalen Wert des schwierigen Problems (II.7).

II.6. Die Lagrange-duale Funktion und die konjugierte Funktion

Zwischen der konjugierten Funktion und der Lagrange-dualen Funktion besteht eine enge Verbindung.

Wir betrachten das Problem
$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x=0 \end{array} .$$

Dieses Problem hat die Lagrange Funktion $L(x, \mu) = f(x) + \mu^T x$ und die duale Funktion
$$g(\mu) = \inf_x (f(x) + \mu^T x) = - \sup_x ((-\mu)^T x - f(x)) = -f^*(-\mu) .$$

Allgemeiner betrachten wir ein Optimalitätsproblem mit linearen Ungleichheits- und Gleichheitsbeschränkungen:
$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & Ax \leq b \quad (\text{II.8}) \\ & Cx = d \end{array} .$$

Dann können wir die duale Funktion für das Problem (II.8) schreiben als:

$$\begin{aligned} g(\lambda, \mu) &= \inf_x (f_0(x) + \lambda^T (Ax - b) + \mu^T (Cx - d)) \\ &= -b^T \lambda - d^T \mu + \inf_x (f_0(x) + (A^T \lambda + C^T \mu)^T x) \quad (\text{II.9}) \\ &= -b^T \lambda - d^T \mu - f_0^*(-A^T \lambda - C^T \mu) \end{aligned}$$

Den eigentlichen Definitionsbereich von g können wir aus dem eigentlichen Definitionsbereich von f_0^* folgern: $dom\ g = \{(\lambda, \mu) : -A^T \lambda - C^T \mu \in dom\ f_0^*\} .$

- Normminimierung

Wir betrachten das Problem
$$\begin{array}{ll} \min & \|x\| \\ \text{s.t.} & Ax = b \end{array} \quad (\text{II.10}) , \text{ wobei } \|\cdot\| \text{ eine beliebige Norm ist.}$$

Die Konjugierte von $f_0 = \|\cdot\|$ ist gegeben durch

$$f_0^*(y) = \begin{cases} 0 & , \|y\|_* \leq 1 \\ \infty & , \text{sonst} \end{cases} , \text{ die Indikatorfunktion der Einheitskugel bzgl. der dualen Norm.}$$

(II.9) \Rightarrow die duale Funktion für das Problem (II.10) ist:

$$g(\mu) = -b^T \mu - f_0^*(-A^T \mu) = \begin{cases} -b^T \mu & , \|A^T \mu\|_* \leq 1 \\ -\infty & , \text{sonst} \end{cases}$$

- Entropiemaximierung

Wir betrachten das Entropiemaximierungsproblem
$$\begin{array}{ll} \min & f_0(x) = \sum_{i=1}^n x_i \log x_i \\ \text{s.t.} & Ax \leq b \\ & \mathbf{1}^T x = 1 \end{array} \quad (\text{II.11}) ,$$

wobei $dom\ f_0 = \mathbb{R}_{++}^n$. Da f_0 eine Summe von negativen Entropiefunktionen von

differenzierbaren Variablen ist \Rightarrow ihre Konjugierte ist $f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}$, mit $dom\ f_0^* = \mathbb{R}^n$.

(II.9) \Rightarrow dass die duale Funktion von (II.11) gegeben ist durch:

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

wobei a_i die i -te Spalte von A ist.

- Minimales Volumen eines Ellipsoids

Wir betrachten das Problem mit Variable $X \in S^n$, $\min f_0(X) = \log \det X^{-1}$ (II.12),
 $s.t. a_i^T X a_i \leq 1, i=1, \dots, m$

wobei $dom f_0 = S_{++}^n$. Dieses Problem hat eine einfache geometrische Interpretation.

Zu jedem $X \in S_{++}^n$ kann man den Ellipsoid $E_X = \{z : z^T X z \leq 1\}$, mit Ursprung als Mittelpunkt, konstruieren. Das Volumen dieses Ellipsoids ist proportional zu $(\det X^{-1})^{1/2}$.

Somit ist die Zielfunktion von (II.12), bis auf eine Konstante, der Logarithmus des Volumens von E_X . $a_i \in E_X$ sind dann die Beschränkungen des Problems (II.12).

Statt dem Problem (II.12) kann man also alternativ das minimale Volumen des Ellipsoids, mit Ursprung als Mittelpunkt, der die Punkte a_1, \dots, a_m enthält, bestimmen.

Die Ungleichheitsbeschränkungen des Problems sind affin; sie können durch $tr((a_i a_i^T) X) \leq 1$ ausgedrückt werden.

Im Beispiel I.3 haben wir herausgefunden, dass die Konjugierte von f_0 :

$$f_0^* = \log \det (Y^{-1})^{-1} - n, \text{ mit } dom f_0^* = -S_{++}^n \text{ ist.}$$

Aus (II.9) folgt, dass die duale Funktion für (II.12) gegeben ist durch:

$$g(\lambda) = \begin{cases} \log \det \left(\sum_{i=1}^m \lambda_i a_i a_i^T \right) - \mathbf{1}^T \lambda + n & , \sum_{i=1}^m \lambda_i a_i a_i^T > 0 \\ -\infty & , \text{sonst} \end{cases} \quad (\text{II.13}).$$

Folglich ist, für jedes $\lambda \geq 0$ mit $\sum_{i=1}^m \lambda_i a_i a_i^T > 0$, die Zahl $\log \det \left(\sum_{i=1}^m \lambda_i a_i a_i^T \right) - \mathbf{1}^T \lambda + n$ eine untere Grenze am optimalen Wert des Problems (II.12).