Das Lagrange-duale Problem

Tobias Kulke

29. April 2010

1 Einführung

Für jedes Paar (λ, ν) mit $\lambda \succeq 0$ liefert die Langrange-duale Funktion

$$g(\lambda, \nu) = \inf_{x \in D} L(x, \lambda, \nu) = \inf_{x \in D} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right), \tag{1.1}$$

mit $\lambda \in \mathbb{R}^m$ und $\nu \in \mathbb{R}^p$, eine untere Schranke des Optimalwertes p^* des Optimierungsproblems

$$\begin{aligned} \text{Minimiere} \quad & f_0(x) \\ \text{sodass} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ \text{und} \quad & h_i(x) = 0, \quad i = 1, \ldots, p. \end{aligned} \tag{1.2}$$

Wir haben also eine untere Schranke, die von den Parametern λ und ν abhängig ist. Nun wollen wir die Beste dieser unteren Schranken finden mittels des folgenden Optimierungsproblems.

Definition 1.1 *Das Problem*

$$\begin{aligned} \textit{Maximiere} \quad & g(\lambda, \nu) \\ \textit{sodass} \quad & \lambda \succeq 0 \end{aligned} \tag{1.3}$$

heißt ***Lagrange-duales Problem*** *bezüglich des Problems (1.2).*
Ein Paar (λ, ν) mit $\lambda \succeq 0$ und $g(\lambda, \nu) > -\infty$ heißt **dual zulässig**.
Wir nennen (λ^*, ν^*) ***dual optimal*** *oder* ***optimale Langrange-Multiplikatoren****, falls sie optimal sind für das Problem (1.3).*

Bemerkung 1.2 Das Problem (1.3) ist immer konvex, da die Beschränkungen konvex und die zu maximierende Funktion konkav ist.

Bemerkung 1.3 Es kommt häufig vor, dass der Defintionsbereich der Lagrange-dualen Funktion,

$$\textbf{\textit{dom}}\ g = \{(\lambda, \nu) \,|\, g(\lambda, \nu) > -\infty\},$$

eine geringere Dimension als $m+p$ hat, d.h. **dom** g liegt in einem affinen Unterraum. Meistens kann man die affine Hülle von **dom** g bestimmen und als Menge linearer Gleichungsbeschränkungen schreiben. Wir können dann ein äquivalentes Problem aufstellen, indem diese Gleichungsbeschränkungen explizit vorkommen.

Beispiel 1.4 (Langrange-duales LP eines LP in Normalform)

Wie wir wissen, ist die Langrange-duale Funktion eines LP in Normalform,

$$\begin{aligned} \text{Minimiere} \quad & c^T x \\ \text{sodass} \quad & Ax = b \\ \text{und} \quad & x \succeq 0, \end{aligned}$$

gegeben durch

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

Das Lagrange-duale Problem wäre nun

$$\begin{aligned} \text{Maximiere} \quad & g(\lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & \text{sonst} \end{cases} \\ \text{sodass} \quad & \lambda \succeq 0. \end{aligned}$$

Da (λ, ν) nur dann zulässig ist, wenn g endlich ist, also $A^T \nu - \lambda + c = 0$, können wir ein äquivalentes Problem aufstellen, indem wir die Gleichungsbeschränkungen explizit machen, d.h.

$$\begin{aligned} \text{Maximiere} \quad & -b^T \nu \\ \text{sodass} \quad & A^T \nu - \lambda + c = 0 \\ \text{und} \quad & \lambda \succeq 0, \end{aligned}$$

bzw.

$$\begin{aligned} \text{Maximiere} \quad & -b^T \nu \\ \text{sodass} \quad & A^T \nu + c \succeq 0. \end{aligned}$$

Man erhält ein LP mit Ungleichungsbeschränkungen.

2 Schwache Dualität

Satz 2.1 *Der Optimalwert d^* des Lagrange-dualen Problems (1.3) ist die größte untere Schranke für den Optimalwert p^* des primalen Problems (1.2), die man durch die Lagrange-duale Funktion erhalten kann. Insbesondere gilt*

$$d^* \leq p^*. \tag{2.1}$$

Diese Eigenschaft wird ***schwache Dualität*** *genannt.*

Bemerkung 2.2 Es können folgende Situationen auftreten

	$d^* \in \mathbb{R}$	dual unbeschränkt $d^* = +\infty$	dual unzulässig $d^* = -\infty$
$p^* \in \mathbb{R}$	(i)		(ii)
primal unbeschränkt $p^* = -\infty$			(iv)
primal unzulässig $p^* = +\infty$	(iii)	(v)	(vi)

Definition 2.3 *Die Differenz* $p^* - d^*$ *wird als* ***Optimalitätslücke*** *bezeichnet.*

Bemerkung 2.4 Man kann die Abschätzung (2.1) nutzen um eine untere Schranke für den Optimalwert schwieriger Probleme zu finden, da das duale Problem immer konvex ist (vgl. Bemerkung 1.2) und daher effektiv gelöst werden kann.

3 Starke Dualität und Slater's constraint qualification

Definition 3.1 *Falls die schwache Dualität mit Gleichheit erfüllt ist, also falls gilt*

$$d^* = p^*, \tag{3.1}$$

spricht man von ***starker Dualität****.*

Bemerkung 3.2 Im Allgemeinen gilt starke Dualität nicht. Falls jedoch das primale Problem konvex ist

$$\begin{aligned} \textit{Maximiere} \quad & f_0(x) \quad \textit{über } x \in \mathbb{R}^n \\ \textit{sodass} \quad & f_i(x) \leq 0, \quad i = 1, \dots, m \\ \textit{und} \quad & Ax = b \end{aligned} \tag{3.2}$$

mit $f_{0,\dots,m}$ konvex, so gilt meistens die starke Dualität.
Es gibt Bedingungen an (3.2), die starke Dualität garantieren. Man nennt diese Bedingungen **constraint qualifications**. Oft ist es einfacher diese nachzuweisen, als direkt die starke Dualität zu zeigen.

Definition 3.3 *Eine einfache constraint qualification für die konvexe Optimierungsaufgabe (3.2) ist die **Slater-Bedingung**: Es existiert ein $x \in$ **relint** D, so dass gilt*

$$f_i(x) < 0, \quad i = 1, \ldots, m, \qquad Ax = b. \tag{3.3}$$

Also muss x strikt zulässig bezüglich der Ungleichungen und zulässig bezüglich der Gleichungen sein.

Definition 3.4 *Falls die ersten k Ungleichungsbeschränkungen $f_i, \ldots, f_k$ affin-linear sind, kann man die Slater-Bedingung wie folgt verfeinern:*
*Es existiert ein $x \in$ **relint** D, so dass gilt*

$$f_i(x) \leq 0, \quad i = 1, \ldots, k, \qquad f_i(x) < 0, \quad i = k+1, \ldots, m, \qquad Ax = b. \tag{3.4}$$

Für die affin-inearen Ungleichungsbeschränkungen muss als keine strikte Ungleichheit gelten.

Bemerkung 3.5 Man beachte, dass die verfeinerte Slater-Bedingung gerade der Zulässigkeit entspricht für $k = m$ und **dom** f_0 offen.
Somit gilt für jedes LP, dass starke Dualität herrscht, falls das primale oder das duale LP zulässig sind. Es gibt also nur in dem Fall, dass beide unzulässig sind, keine starke Dualität. Dieser pathologische Fall kann jedoch tatsächlich auftreten.

Satz 3.6 (Satz von Slater) *Falls bei einem konvexen Optimierungsproblem die (verfeinerte) Slater-Bedingung erfüllt ist, so gilt starke Dualität.*

4 Beispiele

Beispiel 4.1 (Lösung eines LGS durch die Methode der kleinsten Quadrate)

Wir erinnern uns an das Problem

$$\begin{aligned} \text{Minimiere} \quad & x^T x \quad \text{über } x \in \mathbb{R}^n \\ \text{sodass} \quad & Ax = b \end{aligned} \tag{4.1}$$

mit $A \in \mathbb{R}^{pxn}$ und $b \in \mathbb{R}^p$. Das duale Problem ist

$$\text{Maximiere} \quad -\frac{1}{4}\nu^T AA^T \nu - b^T \nu, \tag{4.2}$$

welches ein konkaves quadratisches Maximierungsproblem ohne Nebenbedingungen ist. Die Slater-Bedingung 3.3 bedeutet hier gerade, dass das primale Problem zulässig ist. Es gilt dann also $\mathrm{d}^* = p^*$.

Beispiel 4.2 (Quadratische Programme mit quadratischen Nebenbedingungen(QCQP))

Wir betrachten das QCQP

$$\begin{aligned} \text{Minimiere} \quad & \frac{1}{2}x^T P_0 x + q_0^T x + r_0 \\ \text{sodass} \quad & \frac{1}{2}x^T P_i x + q_i^T x + r_i \leq 0 \quad i = 1, \ldots, m \end{aligned} \tag{4.3}$$

mit $\mathbb{P}_0 \in S^n_{++}$ und $\mathbb{P}_i \in S^n_+, \forall i = 1, \ldots, m$. Die Lagrangefunktion ist

$$L(\lambda, \nu) = \frac{1}{2}x^T P(\lambda)x + q(\lambda)^T x + r(\lambda),$$

wobei

$$P(\lambda) = P_0 + \sum_{i=1}^m \lambda_i P_i, \qquad q(\lambda) = q_0 + \sum_{i=1}^m \lambda_i q_i, \qquad r(\lambda) = r_0 + \sum_{i=1}^m \lambda_i r_i$$

Da es zwar möglich, aber sehr schwierig ist für allgemeine λ die Lagrange-duale Funktion aufzustellen, werden wir im Folgenden $\lambda \succeq 0$ annehmen, wodurch $P(\lambda)$ weiterhin positiv definit und symmetrisch ist. Somit erhalten wir das duale Optimierungsproblem

$$\begin{aligned} \text{Maximiere} \quad & g(\lambda) = -\frac{1}{2}q(\lambda)^T P(\lambda)^{-1} q(\lambda) + r(\lambda) \quad \text{über } \lambda \in \mathbb{R}^m \\ \text{sodass} \quad & \lambda \succeq 0 \end{aligned}$$

Die Slater-Bedingung 3.3 besagt jetzt, dass starke Dualität herrscht, wenn es ein x gibt, welches strikt zulässig bezüglich der Ungleichungsbeschränkungen ist.

Beispiel 4.3 (Entropiemaximierung)

Wir greifen erneut das Entropiemaximierungsproblem

$$\begin{aligned} \text{Minimiere} \quad & \sum_{i=1}^n x_i \log xi \quad \text{über } x \in \mathbb{R}^n_+ \\ \text{sodass} \quad & Ax \preceq b \\ \text{und} \quad & 1^T x = 1 \end{aligned} \tag{4.4}$$

auf. Das Lagrange-duale Problem ist

$$\begin{aligned} \text{Maximiere} \quad & -b^T\lambda - \nu - e^{-\nu-1} \sum_{i=1}^n e^{-a_i^T x} \\ \text{sodass} \quad & \lambda \succeq 0. \end{aligned} \tag{4.5}$$

Die schwächere Slater-Bedingung 3.4 sagt jetzt aus, dass die Dualitätslücke null ist, wenn es ein $x \succ 0$ gibt mit $Ax \preceq b$ und $1^T x = 1$ gibt.
Wir können das duale Problem vereinfachen, indem wir über die duale Variable ν analytisch maximieren. Somit folgt

$$\nu = \log \sum_{i=1}^{n} e^{-a_i^T x} - 1.$$

Wenn wir nun den optimalen Wert von ν in das duale Problem einsetzen, erhalten wir

$$\begin{aligned} \text{Maximiere} \quad & -b^T\lambda - \log \sum_{i=1}^{n} e^{-a_i^T x} \\ \text{sodass} \quad & \lambda \succeq 0. \end{aligned}$$

Dies ist ein sogenanntes geometrisches Programm mit nichtnegativen Beschränkungen.

Beispiel 4.4 (Minimales Volumen eines Ellipsoids)
Wir betrachten

$$\begin{aligned} \text{Minimiere} \quad & \log\det X^{-1} \quad \text{über } X \in \mathbb{S}^n_{++} \\ \text{sodass} \quad & a_i^T X a_i \leq 1, \quad i = 1, \ldots, m. \end{aligned} \tag{4.6}$$

Das Lagrange-duale Problem ist dann

$$\begin{aligned} \text{Maximiere} \quad & \log\det(\sum_{i=1}^{m} \lambda_i a_i a_i^T) - 1^T\lambda + n \\ \text{sodass} \quad & \lambda \succeq 0, \end{aligned} \tag{4.7}$$

wobei wir $\log\det X = -\infty$ setzen, falls $X \not\succ 0$
Die verfeinerte Slater-Bedingung 3.4 ist nun gerade, dass es ein $X \in \mathbb{S}^n_{++}$ gibt mit $a_i^T X a_i \leq 1 \quad \forall i = 1, \ldots, m$. Da dies jedoch immer erfüllt ist, herrscht auch immer starke Dualität.

Beispiel 4.5 (Ein nichtkonvexes quadratisches Programm mit starker Dualität)
In seltenen Fällen gilt sogar bei nichtkonvexen Optimierungsaufgaben starke Dualität. Als wichtiges Beispiel betrachten wir jetzt das Problem eine nichtkonvexe quadratische Funktion über die Einheitskugel zu minimieren:

$$\begin{aligned} \text{Minimiere} \quad & x^T A x + 2b^T x \\ \text{sodass} \quad & x^T x \leq 1, \end{aligned} \tag{4.8}$$

mit $A \in S^n, A \not\succeq 0$ und $b \in \mathbb{R}^n$. Das Problem wir manchmal *Trust-Region-Problem* genannt. Es wird dabei eine Approximation 2.Ordnung einer Funktion über der Einheitskugel minimiert.

Die Lagrange-Funktion ist

$$L(x,\lambda) = x^T Ax + 2b^T x + \lambda(x^T x - 1) = x^T(A - \lambda I)x + 2b^T x - \lambda,$$

wodurch sich die Lagrange-duale Funktion ergibt als

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

wobei $(A+\lambda I)^+$ die Pseudoinverse von $(A+\lambda I)$ ist. Somit ist das Lagrange-duale Problem

$$\begin{aligned} \text{Maximiere} \quad & -b^T(A+\lambda I)^+ b - \lambda \quad \text{über} \lambda \in \mathbb{R} \\ \text{sodass} \quad & A + \lambda I \succeq 0 \\ \text{und} \quad & b \in (A+\lambda I). \end{aligned} \tag{4.9}$$

Dieses Programm ist konvex, auch wenn das in dieser Darstellung nicht offensichtlich ist. Obwohl das primale Problem nicht konvex war, so haben wir hier trotzdem immer Dualitätslücke null.
Tatsächlich gilt hier sogar ein allgemeineres Resultat:

Bei jedem Optimierungsproblem mit quadratischer Zielfunktion und einer quadratischen Ungleichungsbeschränkung gilt starke Dualität, falls die Slater-Bedingung erfüllt ist.

Beispiel 4.6 (Gemischte Strategien für Matrixspiele)

Im folgenden Beispiel werden wir mittels der starken Dualität ein wichtiges Resultat für Matrixspiele erhalten.

Wir betrachten ein Spiel mit zwei Spielern P_1 und P_2 und der Auszahlungsmatrix $A = (A_{kl}) \in \mathbb{R}^{mxn}$ wobei:

- P_1 eine Zeile von A auswürfelt, d.h. eine Zahl $k \in 1, \ldots, m$
- P_2 eine Spalte von A auswürfelt, d.h. eine Zahl $l \in 1, \ldots, n$
- P_1 zahlt dann A_{kl} an P_2

Selbstverständlich möchte P_1 diese Auszahlung minimieren und P_2 möchte sie maximieren.
Die Spieler haben dabei *gemischte Strategien*, d.h. sie würfeln unabhängig voneinander und zufällig, können jedoch vorher die Wahrscheinlichkeitsverteilung

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

mit der gewürfelt wird festlegen. Die erwartete Auszahlung von P_1 an P_2 beträgt dann

$$\sum_{k=1}^{n}\sum_{l=1}^{m} u_k v_l A_{kl} = u^T A v.$$

Annahme:
P_2 kennt die Strategie u von P_1. Dies ist offensichtlich ein Vorteil für P_2. Dieser wird nun v so wählen, dass $u^T A v$ maximiert wird, was zu folgender Auszahlung führt

$$\sup\{u^T A v | v \succeq 0, 1^T v = 1\} = \max_{i=1,\ldots,m} (u^T A v)_i$$

$\Rightarrow$P$_1$ sollte u so wählen, dass diese Auszahlung minimiert wird. Dies entspricht der Lösung des Problems

$$\begin{aligned} \text{Minimiere} \quad & \max_{i=1,\ldots,m} (u^T A v)_i \\ \text{sodass} \quad & u \succeq 0 \\ \text{und} \quad & 1^T u = 1. \end{aligned} \tag{4.10}$$

Dies ist ein stückweise lineare, konvexe Optimierungsaufgabe. Der optimale Wert u^* dieses Problems ist somit die geringste erwartete Auszahlung, die P_1 erreichen kann, falls P_2 seine Strategie kennt.

Analog betrachten wir folgende **Annahme**:
P_1 kennt die Strategie v von P_2. P_1 wird nun u so wählen, dass $u^T A v$ minimiert wird, was zu folgender Auszahlung führt

$$\inf\{u^T A v | u \succeq 0, \mathbf{1}^T u = 1\} = \min_{i=1,\ldots,m} (Av)_i$$

$\Rightarrow$P$_2$ sollte v so wählen, dass diese Auszahlung maximiert wird. Dies entspricht der Lösung des Problems

$$\begin{aligned} \text{Maximiere} \quad & \min_{i=1,\ldots,m} (Av)_i \\ \text{sodass} \quad & u \succeq 0 \\ \text{und} \quad & \mathbf{1}^T u = 1. \end{aligned} \tag{4.11}$$

Dies ist eine konvexe Optimierungsaufgabe mit stückweise linearer, konkaver Zielfunktion. Der optimale Wert v^* dieses Problems ist somit die größte erwartete Auszahlung, die P_2 erreichen kann, falls P_1 seine Strategie kennt.

Nun ist intuitiv klar, dass es nicht schaden kann die Startegie des Gegners zu kennen. In der Tat ist leicht zu zeigen, dass $u^* \leq v^*$. Die Differenz $u^* - v^*$ können wir als den Vorteil auffassen, den ein Spieler hat, falls er die Strategie seines Gegners kennt.
Mit der starken Dualität kann man ein erstaunliche Aussage erreichen: $u^* = v^*$. Dass heißt man hat keinen Vorteil in einem Matrixspiel, wenn man die Strategie des Gegners kennt.
Dies werden wir zeigen, indem wir nachweisen, dass die Programme (4.10)

und (4.11) Lagrange-duale Probleme sind. Für diese gilt dann aufgrund ihrer Konvexität und der verfeinerten Slater-Bedingung 3.4 starke Dualität und somit die Gleichheit von u^* und v^*.

Wir schreiben dazu (4.10) als LP

$$\begin{aligned} \text{Minimiere} \quad & t \quad \text{über} t \in \mathbb{R} \\ \text{sodass} \quad & u \succeq 0, \quad \mathbf{1}^T u = 1 \\ \text{und} \quad & A^T u \preceq t\mathbf{1}. \end{aligned}$$

Mit den Multiplikatoren λ für $A^T u \preceq t\mathbf{1}$, μ für $u \succeq 0$ und ν für $\mathbf{1}^T u = 1$ ist die Lagrange-Funktion

$$\begin{aligned} L(t, u, \lambda, \mu, \nu) &= t \, + \, \lambda^T (A^T u - t\mathbf{1}) \, - \, \mu^T u \, + \, \nu(1 - \mathbf{1}^T u \\ &= u \, + \, (1 - \mathbf{1}^T \lambda)t \, + \, (A\lambda - \nu\mathbf{1} - \mu)^T u. \end{aligned}$$

Somit ist die duale Funktion

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

Das duale Problem ist also

$$\begin{aligned} \text{Maximiere} \quad & \nu \\ \text{sodass} \quad & \lambda \succeq 0, \quad \mathbf{1}^T \lambda = 1, \quad \mu \succeq 0 \\ \text{und} \quad & A\lambda - \nu\mathbf{1} = \mu. \end{aligned}$$

Wenn wir μ eliminieren erhalten wir das folgende Lagrange-duale Problem von (4.10):

$$\begin{aligned} \text{Maximiere} \quad & \nu \\ \text{sodass} \quad & \lambda \succeq 0, \quad \mathbf{1}^T \lambda = 1 \\ \text{und} \quad & A\lambda \succeq \nu\mathbf{1}. \end{aligned}$$

Dies ist wie man leicht erkennt äquivalent zu (4.11).