Dualität bei konvexer Optimierung

Seminar zur Numerik I im SS 2016

Laslo Hunhold

10. Mai 2016

Ausarbeitung zum Seminarvortrag vom 2. Mai 2016

Mathematisches Institut
Mathematisch-Naturwissenschaftliche Fakultät
Universität zu Köln

Betreuung: Prof. Dr.-Ing. Gregor Gassner

Inhaltsverzeichnis

1 Einleitung

Nicht immer ist es möglich, Optimierungsprobleme direkt zu betrachten. Selbst wenn man eine vermutete optimale Lösung findet, ist man vor allem daran interessiert, ob es sich wirklich um eine optimale Lösung handelt.
Unter gewissen Voraussetzungen bietet sich für diese Fälle die Betrachtung dualer Probleme an. Wir werden sehen, daß jedes Optimierungsproblem ein duales Optimierungsproblem hat, und unter welchen Bedingungen wir die Optimalität einer Lösung, falls diese existiert, zeigen können.

In dieser Ausarbeitung blicken wir auf zwei Dualitäten: Die Lagrange- und die Wolfe-Dualität, die unter gewissen Umständen äquivalent sind. Jede dieser Dualitäten hat ein Konzept der schwachen und starken Dualität, deren Erfüllung allgemeine Aussagen für ein gegebenes Optimierungsproblem ermöglichen, und einen Satz an Voraussetzungen, die erfüllt werden müssen.

Für den Einstieg in diesen Bereich der Optimierungstheorie nutzen wir die Ergebnisse der vorherigen Vorträge, vor allen Dingen die grundlegenden Aussagen über die konvexe Analysis und Optimalitäts- und Slater-Regularitätsbedingungen. Die Karush-Kuhn-Tucker-Theorie spielt eine zentrale Rolle in der starken Dualität von dualen Lagrange- und Wolfe-Problemen. Die Betrachtung von Slater-regulären Problemen im Rahmen der Karush-Kuhn-Tucker-Theorie erlaubt in erster Linie die starken Dualitätsaussagen.

Zur Verdeutlichung der theoretischen Ergebnisse werden Beispiele für die relativ leichte Lösung von komplexen Problemen über beide Dualitätskonzepte gebracht. Häufige Mißverständnisse zur Stärke der schwachen und starken Dualitätsaussagen werden beispielhaft behandelt.

Die wichtigsten Ergebnisse der vorherigen Vorträge, die für diese Ausarbeitung relevant sind, werden zur Vereinfachung des Einstiegs im Folgenden noch einmal wiederholt.

Definition 1.1 (Nichtlineares Optimierungsproblem ohne Gleichheitsnebenbedingungen)**.** *Sei $\mathcal{C} \subseteq \mathbb{R}^n$ und $f : \mathcal{C} \to \mathbb{R}$. Dann ist das nichtlineare Optimierungsproblem ohne Gleichheitsnebenbedingungen für die Zielfunktion f definiert als*

$$\begin{aligned} \inf_x \quad & f(x) \\ \textit{gemäß} \quad & g_j(x) \leq 0, \quad g_j : \mathcal{C} \to \mathbb{R}, \quad j \in J = \{1, \dots, m\} \\ & x \in \mathcal{C}. \end{aligned} \tag{NLOP}$$

Definition 1.2 (Zulässiger Bereich)**.** *Der zulässige Bereich für ein* (NLOP) *X ist definiert als*

$$\mathcal{F}_X = \left\{ x \in \mathcal{C} \,\middle|\, \forall j \in J : g_j(x) \leq 0 \right\}.$$

Definition 1.3 (Konvexes Optimierungsproblem)**.** *Ein* (NLOP) *ist genau dann* konvex*, wenn*

$$\begin{aligned} \mathcal{C} \quad & \textit{konvex}, \\ f, g_1, \dots, g_m \quad & \textit{differenzierbar \& konvex} \end{aligned} \tag{KOP}$$

gilt.

Definition 1.4 (Relativ innerer Punkt). *Sei $\mathcal{C}$ konvex. $x \in \mathcal{C}$ ist genau dann ein relativ innerer Punkt von $\mathcal{C}$, wenn*

$$\forall \bar{x} \in \mathcal{C} : \exists \tilde{x} \in \mathcal{C} : \exists \lambda \in (0,1) : x = \lambda \bar{x} + (1-\lambda)\tilde{x}.$$

Die Menge der relativ inneren Punkte, das relative Innere von $\mathcal{C}$, wird mit $\mathrm{ri}(\mathcal{C})$ bezeichnet.

Definition 1.5 (Slater-Regularität). *Ein* (KOP) *ist genau dann Slater-Regulär bzw. erfüllt die Slater-Regularitätsbedingung, wenn für dieses ein Slaterpunkt existiert.*
Ein $s \in \mathrm{ri}(\mathcal{C})$ ist genau dann ein Slaterpunkt für das (KOP)*, wenn*

$$\begin{aligned} &\forall j : g_j \text{ nichtlinear} &&: g_j(s) < 0 \\ &\forall j : g_j \text{ linear} &&: g_j(s) \leq 0 \end{aligned}$$

gilt.

Bemerkung 1.6 (Gleichheitsnebenbedingungen). *Für das* (NLOP) *werden keine Gleichheitsnebenbedingungen betrachtet, weil ein solches Problem dann nicht Slater-regulär wäre. Dies folgt daraus, daß sich jede Gleichheitsnebenbedingung*

$$h_i(x) = 0$$

umschreiben läßt in zwei im Allgemeinen nichtlineare Ungleichheitsnebenbedingungen

$$\begin{aligned} h_i(x) &\leq 0 \\ -h_i(x) &\leq 0, \end{aligned}$$

für die kein $\bar{x} \in \mathcal{C}$ existiert, sodaß

$$\begin{aligned} h_i(\bar{x}) &< 0 \\ -h_i(\bar{x}) &< 0 \end{aligned}$$

gilt. Die Betrachtung Slater-regulärer Probleme wird aber genau zentraler Gegenstand bei der Einbeziehung der Karush-Kuhn-Tucker-Theorie für die starke Lagrange- und Wolfe-Dualität sein.

Definition 1.7 (Lagrange-Funktion). *Sei $y = (y_1, \ldots, y_m)^T \in \mathbb{R}^m_+$. Die Lagrange-Funktion $L : \mathcal{C} \times \mathbb{R}^m_+ \to \mathbb{R}$ für ein* (NLOP) *ist definiert als*

$$L(x,y) := f(x) + \sum_{j=1}^{m} y_j g_j(x).$$

Satz 1.8 (Konvexität der Lagrange-Funktion). *Die Lagrange-Funktion für ein* (KOP) *ist in x konvex.*

Beweis. Laut Voraussetzung sind $f, g_1, \ldots, g_m$ konvex und $\forall j \in \{1, \ldots, m\} : y_j \geq 0$. Damit ist nach [KRT, Lemma 1.40] die Lagrange-Funktion konvex. □

Definition 1.9 (Sattelpunkt). *Ein Sattelpunkt ist ein Tupel $(\bar{x}, \bar{y}) \in \mathcal{C} \times \mathbb{R}^m_+$ mit der Eigenschaft*

$$\forall (x,y) \in \mathcal{C} \times \mathbb{R}^m_+ : L(\bar{x}, y) \leq L(\bar{x}, \bar{y}) \leq L(x, \bar{y}).$$

In einem Sattelpunkt wird folglich die Lagrange-Funktion auf ihrem Definitionsbereich in $\bar{x}$ minimiert und in $\bar{y}$ maximiert.

Theorem 1.10 (Karush-Kuhn-Tucker). *Sei ein* (KOP) *Slater-regulär. Dann sind äquivalent*

- $\bar{x} \in \mathcal{F}_{(\mathrm{KOP})}$ *optimale Lösung*
- $\exists \bar{y} \in \mathbb{R}^m_+ : (\bar{x}, \bar{y})$ *Sattelpunkt*

Beweis. Siehe [KRT, Theorem 2.30]. □

Korollar 1.11. *Sei ein* (KOP) *Slater-regulär. Dann sind äquivalent*

- $\bar{x} \in \mathcal{F}_{(\mathrm{KOP})}$ *optimale Lösung*
- $\exists \bar{y} \in \mathbb{R}^m_+ : \begin{array}{l} f(\bar{x}) = \inf_{x \in \mathcal{C}} \{L(x, \bar{y})\} \\ \sum_{j=1}^m \bar{y}_j g_j(\bar{x}) = \max_{y \in \mathbb{R}^m_+} \left\{ \sum_{j=1}^m y_j g_j(\bar{x}) \right\} \end{array}$

Beweis. Siehe [KRT, Korollar 2.31]. □

Korollar 1.12. *Sei ein* (KOP) *mit* $\mathcal{C} = \mathbb{R}^n$ *Slater-regulär,* $\{f, g_1, \ldots, g_m\} \subset \mathcal{C}^1(\mathbb{R}^n)$*. Dann sind äquivalent*

- $\bar{x} \in \mathcal{F}_{(\mathrm{KOP})}$ *optimale Lösung*
- $\exists \bar{y} \in \mathbb{R}^m_+ : \begin{array}{l} \nabla_x L(\bar{x}, \bar{y}) = 0 \\ \sum_{j=1}^m \bar{y}_j g_j(\bar{x}) = 0 \end{array}$

Beweis. Siehe [KRT, Korollar 2.33]. □

2 Duales Lagrange-Problem

Betrachte nun im Folgenden ein (KOP).

Definition 2.1 (Duale Lagrange-Funktion). *Die duale Lagrange-Funktion* $\psi : \mathbb{R}^m_+ \to \mathbb{R}$ *ist definiert als*

$$\psi(y) = \inf_{x \in \mathcal{C}} \{L(x, y)\}.$$

Definition 2.2 (Duales Lagrange-Problem). *Das zum* (KOP) *duale Lagrange-Problem ist definiert als*

$$\begin{array}{rl} \sup\limits_{y} & \psi(y) \\ \textit{gemäß} & y \in \mathbb{R}^m_+. \end{array} \tag{DLP}$$

Satz 2.3 (Konvexität des dualen Lagrange-Problems). *Das* (DLP) *ist ein* (KOP)*, selbst wenn* $f, g_1, \ldots, g_m$ *nichtkonvex sind.*

Beweis. Das Problem der Maximierung von $\psi(y)$ ist äquivalent zur Minimierung von $-\psi(y)$. Es genügt folglich zu zeigen, daß $-\psi(y)$ konvex, also $\psi(y)$ konkav ist.

Sei $\bar{y}, \hat{y} \in \mathbb{R}^m_+$, $\lambda \in [0,1]$. Dann gilt

$$\begin{aligned}
\psi(\lambda\bar{y} + (1-\lambda)\hat{y}) &= \inf_{x\in\mathcal{C}} \left\{ f(x) + \sum_{j=1}^{m} [\lambda\bar{y}_j + (1-\lambda)\hat{y}_j] g_j(x) \right\} \\
&= \inf_{x\in\mathcal{C}} \left\{ \lambda \left[f(x) + \sum_{j=1}^{m} \bar{y}_j g_j(x) \right] + (1-\lambda) \left[f(x) + \sum_{j=1}^{m} \hat{y}_j g_j(x) \right] \right\} \\
&\geq \inf_{x\in\mathcal{C}} \left\{ \lambda \left[f(x) + \sum_{j=1}^{m} \bar{y}_j g_j(x) \right] \right\} + \inf_{x\in\mathcal{C}} \left\{ (1-\lambda) \left[f(x) + \sum_{j=1}^{m} \hat{y}_j g_j(x) \right] \right\} \\
&= \lambda\psi(\bar{y}) + (1-\lambda)\psi(\hat{y}).
\end{aligned}$$

Somit ist $\psi(y)$ konkav. □

Definition 2.4 (Dualitätslücke)**.** *Die Dualitätslücke ist für* $(\bar{x}, \bar{y}) \in \mathcal{F}_{(\mathrm{KOP})} \times \mathbb{R}^m_+$ *definiert als*

$$f(\bar{x}) - \psi(\bar{y}).$$

Satz 2.5 (Positivität der Dualitätslücke)**.** *Sei* $(\bar{x}, \bar{y}) \in \mathcal{F}_{(\mathrm{KOP})} \times \mathbb{R}^m_+$. *Für die Dualitätslücke gilt*

$$f(\bar{x}) - \psi(\bar{y}) \geq 0.$$

Beweis. Sei $(\bar{x}, \bar{y}) \in \mathcal{F}_{(\mathrm{KOP})} \times \mathbb{R}^m_+$. Betrachte die Ungleichungskette

$$\psi(\bar{y}) = \inf_{x\in\mathcal{C}} (L(x, \bar{y})) \leq L(\bar{x}, \bar{y}) = f(\bar{x}) + \sum_{j=1}^{m} \bar{y}_j g_j(\bar{x}) \leq f(\bar{x}), \tag{2.1}$$

woraus direkt folgt, daß $f(\bar{x}) - \psi(\bar{y}) \geq 0$. □

Was dies auch zeigt ist, daß der optimale Wert der Zielfunktion des dualen Lagrange-Problems eine untere Schranke des optimalen Werts der Zielfunktion des (KOP) ist.
Wir suchen also nach der größten unteren Schranke, idealerweise mit $f(\bar{x}) - \psi(\bar{y}) = 0$, also dem Verschwinden der Dualitätslücke.

2.1 Theoreme über Lagrange-Dualität

Im Folgenden werden wir Bedingungen für das Verschwinden der Dualitätslücke aufstellen und die Karush-Kuhn-Tucker-Theorie auf Slater-reguläre duale Lagrange-Probleme anwenden.

Theorem 2.6 (Schwache Lagrange-Dualität)**.** *Sei* $(\bar{x}, \bar{y}) \in \mathcal{F}_{(\mathrm{KOP})} \times \mathbb{R}^m_+$. *Dann gilt*

1. $\psi(\bar{y}) \leq f(\bar{x})$
2. $\psi(\bar{y}) = f(\bar{x}) \Leftrightarrow \inf_{x\in\mathcal{C}}(L(x, \bar{y})) = f(\bar{x})$

Beweis. Betrachte die Ungleichungskette (2.1).

1. $\psi(\bar{y}) \leq f(\bar{x})$ ist direkt ersichtlich.
2. Betrachte zuerst die Hinrichtung. Falls $\psi(\bar{y}) = f(\bar{x})$, werden in (2.1) alle „$\leq$" zu „$=$", da $\inf_{x\in\mathcal{C}}(L(x, \bar{y})) \not> L(\bar{x}, \bar{y})$ und $f(\bar{x}) + \sum_{j=1}^{m} \bar{y}_j g_j(\bar{x}) \not> f(\bar{x})$.
   Folglich gilt $\inf_{x\in\mathcal{C}}(L(x, \bar{y})) = f(\bar{x})$.

   Die Rückrichtung gilt, da $\psi(\bar{y}) \stackrel{\text{def}}{=} \inf_{x\in\mathcal{C}}(L(x, \bar{y})) = f(\bar{x})$.

$\square$

Korollar 2.7. *Sei* $(\bar{x}, \bar{y}) \in \mathcal{F}_{(\text{KOP})} \times \mathbb{R}^m_+$*. Dann sind äquivalent*

- $\psi(\bar{y}) = f(\bar{x})$
- $\bar{x}$ *optimale Lösung für* (KOP) $\wedge\, \bar{y}$ *optimale Lösung für* (DLP)

Beweis. Wir wissen, daß $\psi(\bar{y}) \leq f(\bar{x})$. Folglich ist bei Gleichheit die duale Lagrange-Funktion maximiert und damit eine optimale Lösung des (DLP) gegeben.
Weiterhin gilt auch $f(\bar{x}) \geq \psi(\bar{y})$, womit folglich bei Gleichheit $f(\bar{x})$ minimal ist und damit eine optimale Lösung des (KOP) vorliegt. $\square$

Korollar 2.8. *Sei* $\mathcal{C} = \mathbb{R}^n$*,* $\{f, g_1, \ldots, g_m\} \subset \mathcal{C}^1(\mathbb{R}^n)$*,* $(\bar{x}, \bar{y}) \in \mathcal{F}_{(\text{KOP})} \times \mathbb{R}^m_+$*,* $\psi(\bar{y}) = f(\bar{x})$*. Dann ist* $(\bar{x}, \bar{y})$ *ein KKT-Punkt für das* (KOP)*.*

Beweis. Die Bedingungen $\mathcal{C} = \mathbb{R}^n$ und $\{f, g_1, \ldots, g_m\} \subset \mathcal{C}^1(\mathbb{R}^n)$ sind notwendig für die Definition eines KKT-Punktes. Betrachte nun die weiteren Bedingungen an einen KKT-Punkt:

- $\forall j \in J : g_j(\bar{x}) \leq 0$ ist erfüllt, da $\bar{x} \in \mathcal{F}_{(\text{KOP})}$.
- $\nabla_x L(\bar{x}, \bar{y}) = 0$ ist erfüllt, da $L(\cdot, \bar{y})$ konvex ist und von $\bar{x}$ minimiert wird. Der Satz von Fermat für konvexe Funktionen definiert auf $\mathbb{R}^n$ liefert $\nabla_x L(\bar{x}, \bar{y}) = 0$.
- $\sum_{j=1}^m \bar{y}_j g_j(\bar{x}) = 0$ ist erfüllt, da $\bar{x}$ $L(\cdot, \bar{y})$ minimiert und laut Voraussetzung $\psi(\bar{y}) \overset{\text{def}}{=} \inf_{x \in \mathcal{C}} \left\{ f(x) + \sum_{j=1}^m \bar{y}_j g_j(x) \right\} = f(\bar{x}) + \sum_{j=1}^m \bar{y}_j g_j(\bar{x}) = f(\bar{x})$ gilt.
- $\bar{y} \in \mathbb{R}^m_+$ ist erfüllt.

$\square$

Theorem 2.9 (Starke Lagrange-Dualität)**.** *Sei das* (KOP) *Slater-regulär und* $\bar{x} \in \mathcal{F}_{(\text{KOP})}$*. Dann sind äquivalent*

- $\bar{x}$ *optimale Lösung für* (KOP)
- $\exists \bar{y} \in \mathbb{R}^m_+ : \bar{y}$ *optimale Lösung für* (DLP) $\wedge\, \psi(\bar{y}) = f(\bar{x})$

Beweis. Wir beginnen mit der Hinrichtung. Betrachte Korollar 1.11, dessen Voraussetzungen erfüllt sind, womit es nur noch zu zeigen gilt, daß die Aussagen $f(\bar{x}) = \inf_{x \in \mathcal{C}} \{L(x, \bar{y})\}$ und $\sum_{j=1}^m \bar{y}_j g_j(\bar{x}) = \max_{y \in \mathbb{R}^m_+} \left\{ \sum_{j=1}^m y_j g_j(\bar{x}) \right\}$ zu $\psi(\bar{y}) = f(\bar{x})$ führen, da $\bar{y}$ als optimale Lösung für das (DLP) durch $\psi(\bar{y}) = f(\bar{x})$ nach Theorem 2.7 impliziert wird.
Betrachte die Ungleichungskette (2.1) und beobachte wieder, daß alle „$\leq$“ zu „$=$“ werden, da $\inf_{x \in \mathcal{C}}(L(x, \bar{y})) \not> L(\bar{x}, \bar{y})$ und $f(\bar{x}) + \sum_{j=1}^m \bar{y}_j g_j(\bar{x}) \not> f(\bar{x})$. Folglich gilt $\psi(\bar{y}) \overset{\text{def}}{=} \inf_{x \in \mathcal{C}}(L(x, \bar{y})) = f(\bar{x})$.
Die Rückrichtung folgt aus Korollar 2.7. $\square$

Bemerkung 2.10. *Es gilt jedoch zwei Aspekte zu beachten, um möglichen Mißverständnissen vorzubeugen:*

- *Falls das* (KOP) *nicht Slater-regulär ist, heißt dies nicht, daß die Dualitätslücke nicht verschwindet; siehe auch Beispiel 2.11.*
- *Das* (KOP) *hat nicht immer eine optimale Lösung* $\bar{x}$*.*

Beispiel 2.11 (Slater-Regularität und Dualitätslücke)**.** *Betrachte das folgende* (KOP)

$$\begin{aligned} \inf_x \quad & f(x) = x \\ \textit{gemäß} \quad & g(x) = x^2 \leq 0 \\ & x \in \mathbb{R}. \end{aligned}$$

Das Problem ist nicht Slater-regulär, da für die nichtlineare Nebenbedingung $g(x) \leq 0$ gilt

$$\nexists \tilde{x} \in \mathrm{ri}(\mathbb{R}) = \mathbb{R} : g(\tilde{x}) = \tilde{x}^2 < 0.$$

Betrachtet man die duale Lagrange-Funktion $\psi : \mathbb{R}_+ \to \mathbb{R}$ des (KOP)*, so sieht man*

$$\psi(y) = \inf_{x \in \mathbb{R}} (x + yx^2) = \begin{cases} \psi(-\frac{1}{2y}) = -\frac{1}{4y} & y > 0 \\ \psi(-\infty) = -\infty & y = 0. \end{cases}$$

Das duale Lagrange-Problem lautet folglich

$$\begin{aligned} \sup_y \quad & \psi(y) \\ \textit{gemäß} \quad & y \in \mathbb{R}_+ \end{aligned}$$

Die Dualitätslücke verschwindet, da

$$\psi(\bar{y}) = \sup_{y \in \mathbb{R}_+} \psi(y) = \psi(\infty) = 0 = f(0) = f\left(-\frac{1}{2\bar{y}}\right) = f(\bar{x}).$$

Folglich ist das gegebene Problem ein Beispiel für ein nicht Slater-reguläres (KOP)*, bei dem trotzdem die Dualitätslücke verschwindet (siehe Bemerkung 2.10).*

3 Duales Wolfe-Problem

Betrachte nun im Folgenden ein (KOP) mit $\mathcal{C} = \mathbb{R}^n$ und $\{f', g_1', \ldots, g_m'\} \subset \mathcal{C}^0(\mathcal{C})$.

Definition 3.1 (Duales Wolfe-Problem)**.** *Das zum* (KOP) *duale Wolfe-Problem ist definiert als*

$$\begin{aligned} \sup_{x,y} \quad & L(x,y) \\ \textit{gemäß} \quad & \nabla_x L(x,y) = 0_n \\ & (x,y) \in \mathcal{F}_{KOP} \times \mathbb{R}_+^m. \end{aligned} \tag{DWP}$$

Satz 3.2 (Äquivalenz von (DWP) und (DLP))**.** *Falls ein Sattelpunkt existiert, sind das* (DLP) *und das* (DWP) *äquivalent.*

Beweis. Wir wissen, daß $\forall y \in \mathbb{R}_+^m : L(\cdot, y) \in \mathcal{C}^1(\mathbb{R}^n)$ konvex. Es folgt daher mit dem konvexen Satz von Fermat für Funktionen definiert auf $\mathbb{R}^n$

$$\forall (\bar{x}, y) \in \mathbb{R}^n \times \mathbb{R}_+^m : \nabla_x L(\bar{x}, y) = 0 \Leftrightarrow L(\bar{x}, y) \text{ minimal}$$

Angenommen der laut Voraussetzung existente Sattelpunkt hat die Form $(\bar{x}, y)$, dann hat die duale Lagrange-Funktion die Gestalt

$$\psi(y) = \inf_{x \in \mathbb{R}^n} (L(x,y)) = L(\bar{x}, y).$$

Das (DWP) läßt sich folglich umschreiben in

$$\begin{aligned} \sup_y \quad & \psi(y) \\ \text{gemäß} \quad & y \in \mathbb{R}_+^m, \end{aligned}$$

wobei es sich aber genau um das (DLP) handelt. □

3.1 Theoreme über Wolfe-Dualität

Auch für das duale Wolfe-Problem lassen sich Kriterien für optimale Lösungen und Abschätzungen ableiten.

Theorem 3.3 (schwache Wolfe-Dualität)**.** *Sei* $\hat{x} \in \mathcal{F}_{(\text{KOP})}$, $(\bar{x}, \bar{y}) \in \mathcal{F}_{(\text{DWP})}$. *Dann gilt*

$$L(\bar{x}, \bar{y}) \leq f(\hat{x})$$

Beweis. Wir wissen, daß $\forall y \in \mathbb{R}^m_+ : L(\cdot, y) \in \mathcal{C}^1(\mathbb{R}^n)$ konvex ist und daß $\forall (\bar{x}, \bar{y}) \in \mathcal{F}_{DWP}$: $\nabla_x L(\bar{x}, \bar{y}) = 0$. Es folgt mit dem Satz von Fermat für konvexe Funktionen definiert auf $\mathbb{R}^n$, daß

$$\inf_{x \in \mathbb{R}^n} L(x, \bar{y}) = L(\bar{x}, \bar{y}) \leftrightarrow \forall x \in \mathbb{R}^n : L(\bar{x}, \bar{y}) \leq L(x, \bar{y})$$

Insbesondere gilt dies $\forall \hat{x} \in \mathcal{F}_{(\text{KOP})} \subseteq \mathbb{R}^n$ und es folgt

$$L(\bar{x}, \bar{y}) \leq L(\hat{x}, \bar{y}) = f(\hat{x}) + \sum_{j=1}^{m} \bar{y}_j g_j(\hat{x}) \leq f(\hat{x}).$$

□

Aus diesem Theorem wird ersichtlich, daß die Beschränkungen des (DWP) bezogen auf die optimale Lösung genau die KKT-Bedingungen sind.

Theorem 3.4 (starke Wolfe-Dualität)**.** *Sei das* (KOP) *Slater-regulär und* $\bar{x} \in \mathcal{F}_{(\text{KOP})}$. *Dann sind äquivalent*

- $\bar{x}$ *optimale Lösung für (KOP)*
- $\exists \bar{y} \in \mathbb{R}^m_+ : (\bar{x}, \bar{y})$ *optimale Lösung für* (DWP)

Beweis. Betrachte Korollar 1.12, dessen Voraussetzungen erfüllt sind, womit es nur noch zu zeigen gilt, daß die Aussagen $\nabla_x L(\bar{x}, \bar{y}) = 0$ und $\sum_{j=1}^m \bar{y}_j g_j(\bar{x}) = 0$ zur Existenz eines $\bar{y} \in \mathbb{R}^m_+$ äquivalent sind, sodaß $(\bar{x}, \bar{y})$ eine optimale Lösung für das (DWP) ist.
Die Hinrichtung folgt daraus, daß für $\sum_{j=1}^m \bar{y}_j g_j(\bar{x}) = 0$ die Aussage $L(\bar{x}, \bar{y}) = f(\bar{x})$ gilt. Weiterhin ist $\nabla_x L(\bar{x}, \bar{y}) = \nabla_x f(\bar{x}) = 0$, womit die Lagrange-Funktion nach dem Satz von Fermat für konvexe Funktionen definiert auf $\mathbb{R}^n$ im Punkt $(\bar{x}, \bar{y})$ ihr Maximum annimmt und durch Gleichheit mit $f(\bar{x})$ die Optimalität dieser Lösung gezeigt ist.
Liegt in der Rückrichtung eine optimale Lösung $(\bar{x}, \bar{y})$ für (DWP) vor, gilt $L(\bar{x}, \bar{y}) = f(\bar{x})$, da $\bar{x}$ eine optimale Lösung für (KOP) ist. Daraus folgt direkt $\sum_{j=1}^m \bar{y}_j g_j(\bar{x}) = 0$. Weiterhin nimmt nach dem Satz von Fermat für konvexe Funktionen definiert auf $\mathbb{R}^n$ eine konvexe Funktion wie $L(\cdot, \bar{y})$ ihr Maximum nur dort an, wo der Gradient $\nabla_x L(\cdot, \bar{y})$ verschwindet, also gilt $\nabla_x L(\bar{x}, \bar{y}) = 0$. □

4 Beispiele

Beispiel 4.1 (Wolfe-Lösungsverfahren)**.** *Betrachte das folgende* (KOP)

$$\begin{aligned} \inf_{x} \quad & f(x) = x_1 + \exp(x_2) \\ \textit{gemäß} \quad & 3x_1 - 2\exp(x_2) \geq 10 \\ & x_2 \geq 0 \\ & x \in \mathbb{R}^2. \end{aligned} \tag{KOP2}$$

Bilde zuerst die normalisierten Nebenbedingungen

$$\begin{aligned} g_1(x) &= 10 - 3x_1 + 2\exp(x_2) && \leq 0 \\ g_2(x) &= -x_2 && \leq 0 \end{aligned}$$

und prüfe auf Slater-Regularität. Man findet $\tilde{x} = (5, 0)$*, für das* $g_1(\tilde{x}) < 0 \wedge g_2(\tilde{x}) \leq 0$ *erfüllt ist, womit Slater-Regularität folgt.*
Es bietet sich daher an, das Problem Wolfe-dual zu betrachten, weil nach Theorem 3.4 die optimale Lösung des (DWP) *hinreichend ist für die optimale Lösung des* (KOP).
Das duale Wolfe-Problem für das gegebene Problem hat die Gestalt

$$\begin{aligned} \sup_{x,y} \quad & L(x,y) = x_1 + \exp(x_2) + y_1(10 - 3x_1 + 2\exp(x_2)) - y_2 x_2 \\ \textit{gemäß} \quad & \frac{\partial L(x,y)}{\partial x_1} = 1 - 3y_1 = 0 \\ & \frac{\partial L(x,y)}{\partial x_2} = \exp(x_2) + 2\exp(x_2)y_1 - y_2 = 0 \\ & (x,y) \in \mathcal{F}_{(\text{KOP2})} \times \mathbb{R}^2_+. \end{aligned}$$

Die erst Nebenbedingung läßt sich direkt umformen in $y_1 = \frac{1}{3}$ *und in die zweite Nebenbedingung einsetzen, die sich zu* $\frac{5}{3}\exp(x_2) - y_2 = 0$ *vereinfacht. Einfaches Umformen ergibt* $y_2 = \frac{5}{3}\exp(x_2)$. *Beide Ausdrücke für* y_1 *und* y_2 *setzt man in die Zielfunktion ein und erhält*

$$\begin{aligned} L(x,y) &= x_1 + \exp(x_2) + \frac{1}{3}(10 - 3x_1 + 2\exp(x_2)) - \frac{5}{3}\exp(x_2)x_2 \\ &= x_1 + \exp(x_2) + \frac{10}{3} - x_1 + \frac{2}{3}\exp(x_2) - \frac{5}{3}\exp(x_2)x_2 \\ &= \frac{5}{3}\exp(x_2) - \frac{5}{3}\exp(x_2)x_2 + \frac{10}{3} \\ &= \frac{5}{3}\exp(x_2)(1 - x_2) + \frac{10}{3} := \tilde{f}(x_2). \end{aligned}$$

Betrachte die partielle Ableitung

$$\frac{\partial \tilde{f}}{\partial x_2} = \frac{5}{3}\exp(x_2)(1 - x_2 - 1) = -\frac{5}{3}\exp(x_2)x_2,$$

die für $x_2 = 0$ *verschwindet und an dieser Stelle nach dem Vorzeichenwechselkriterium ein Maximum indiziert. Daraus folgt für das* (DWP)

$$\sup_{x,y} L(x,y) = \sup_{x_2} \tilde{f}(x_2) = \tilde{f}(0) = 5,$$

und mit Theorem 3.4, daß dies auch der optimale Wert des (KOP) *ist.*
Für die Bestimmung der optimalen Lösung gilt es noch x_1 *zu bestimmen. Dazu betrachtet man wieder das gegebene* (KOP) *und setzt* $x_2 = 0$ *ein.* $f(x_2)$ *wird minimal, wenn* x_1 *minimal wird, d.h., man nimmt jenes kleinste* x_1*, welches noch die Einschränkungen erfüllt. In diesem Fall ist die Einschränkung* $3x_1 - 2\exp(0) \geq 10 \leftrightarrow x_1 \geq 4$*, woraus* $x_1 = 4$ *folgt. Die optimale Lösung des* (KOP) *ist folglich* $(4, 0)$.
Die Auswertung der Lagrange-Multiplikatoren $y = (y_1, y_2) = (\frac{1}{3}, \frac{5}{3})$ *ist nicht Gegenstand dieser Ausarbeitung.*

Beispiel 4.2 (Quadratisches konvexes Optimierungsproblem)**.** *Sei $Q \in \mathbb{R}^{n\times n}$ symmetrisch positiv semidefinit, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$. Betrachte das folgende* (KOP)

$$\begin{aligned} \inf_x \quad & f(x) = \frac{1}{2}x^T Qx + c^T x \\ \textit{gemäß} \quad & Ax \geq b \Leftrightarrow Ax - b \in \mathbb{R}^m_+ \\ & x \in \mathbb{R}^n_+. \end{aligned} \tag{QKOP}$$

Es handelt sich um ein konvexes Optimierungsproblem, weil Q positiv semidefinit ist und damit f konvex ist. $-Ax + b$ ist als lineare Funktion konvex in jeder Komponente.
Zuerst bildet man die normalisierten Nebenbedingungen

$$g_j(x) = \begin{cases} (-a^j)^T x + b_j & j = 1, \dots, m \\ -x_{j-m} & j = m+1, \dots, m+n \end{cases} \leq 0$$

und notiert, daß die Slater-Regularität trivialerweise gegeben ist.
Betrachte nun das duale Wolfe-Problem

$$\begin{aligned} \inf_{x,y,z} \quad & L(x,y,z) = \frac{1}{2}x^T Qx + c^T x + y^T(-Ax + b) + z^T(-x) \\ \textit{gemäß} \quad & \nabla_x L(x,y,z) = \frac{1}{2}(Q^T + Q)x + c - A^T y - z \overset{Q^T=Q}{=} Qx + c - A^T y - z = 0_n \\ & (x,y,z) \in \mathcal{F}_{(\text{QKOP})} \times \mathbb{R}^m_+ \times \mathbb{R}^n_+ \end{aligned}$$

und substituiere $c = -Qx + A^T y + z$ in $L(x,y,z)$. Man erhält

$$\begin{aligned} \inf_{x,y,z} \quad & -\frac{1}{2}x^T Qx + b^T y \\ \textit{gemäß} \quad & Qx + c - A^T y - z = 0_n \\ & (x,y,z) \in \mathcal{F}_{(\text{QKOP})} \times \mathbb{R}^m_+ \times \mathbb{R}^n_+. \end{aligned}$$

Da Q symmetrisch positiv definit ist, können wir unter anderem die Cholesky-Zerlegung bilden. Wir erhalten $Q = LGL^T$, wobei L eine untere Dreiecksmatrix mit 1-Diagonaleinträgen und G eine Diagonalmatrix mit positiven Einträgen ist.
Über die Matrixurzeldefinition $G := G^{\frac{1}{2}}G^{\frac{1}{2}}$ bilden wir $D := G^{\frac{1}{2}}L$ und erhalten $Q = D^T D$. Weiterhin definieren wir $d = Dx$. Damit läßt sich nun x in dem Problem eliminieren:

$$\begin{aligned} \inf_{y,d} \quad & -\frac{1}{2}d^T d + b^T y \\ \textit{gemäß} \quad & D^T z + c - A^T y - z = 0_n \\ & (y,z,d) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \times \mathbb{R}^n_+ \end{aligned}$$

Die Optimalitätsbedingungen lassen sich aus der ursprünglichen Problemstellung ablesen und sind

- $y^T(Ax - b) = 0$
- $z^T x = 0$
- $d = Dx$

5 Konklusion

Die duale Betrachtung konvexer Optimierungsprobleme erlaubt unter bestimmten Gütekriterien, das ursprüngliche Problem zu lösen oder zumindest eine untere Schranke für den optimalen Wert zu finden.

Wie wir gesehen haben, sind das duale Lagrange- und Wolfe-Problem sehr eng miteinander verknüpft. Zweiteres stellt jedoch die strengeren Voraussetzungen, daß der Lösungsraum der ganze $\mathbb{R}^n$ ist und die Funktionen $f, g_1, \ldots, g_m$ stetige Ableitungen haben.
Sind diese Voraussetzungen aber erfüllt, so erlauben diese die Anwendung der Karush-Kuhn-Tucker-Theorie, was die theoretische Betrachtung des dualen Wolfe-Problems vereinfacht.
Die Aussagen zu dem dualen Lagrange-Problem sind allgemein für beliebige konvexe Optimierungsprobleme und daher für eine größere Bandbreite an Problemen zulässig.
Die Tatsache, daß das duale Lagrange-Problem auch für nichtkonvexe Problem konvex ist, erlaubt die Anwendung der schwachen und starken Lagrange-Dualität auf ein breites Feld von Optimierungsproblemen.

Wir haben in den Beispielen gesehen, wie die dualen Lagrange- und Wolfe-Probleme die Lösung von komplizierten Optimierungsproblemen stark vereinfachen. Das Beispiel 4.2 ist von besonderer numerischer Bedeutung, da die Untersuchung von quadratischen Optimierungsproblemen ein weites Feld verschiedener Disziplinen eröffnet, unter anderem die Betrachtung von quadratischen Optimierungsproblemen über Bilinearformen auf Sobolev- oder Hilberträumen.

Numerisch betrachtet sind die Gradienten der Wolfe-Nebenbedingungen algorithmisch zwar schlecht zu bestimmen, wenn jedoch gezeigt werden kann, daß ein Algorithmus sowohl für ein Problem als auch für dessen duales Problem dieselbe Lösung liefert, folgt daraus direkt die Optimalität der Lösung. Folglich sind duale Probleme ein mächtiges Werkzeug in der theoretischen Betrachtung der Güteklasse von Optimierungsalgorithmen.

In den folgenden Vorträgen werden diverse Optimierungsalgorithmen vorgestellt, deren Optimalkonvergenz bei Existenz einer optimalen Lösung des ursprünglichen Problems mithilfe von dualen Problemen gezeigt werden kann.

6 Notationsverzeichnis

In der folgenden Übersicht werden die gebräuchlichsten Zeichen und Variablen, die in dieser Ausarbeitung verwendet werden, kurz erläutert.

$\mathcal{C}$	Lösungsraum, siehe Definition 1.1
$\mathcal{F}_X$	Zulässiger Bereich des Optimierungsproblems X, siehe Definition 1.2
$\mathrm{ri}(A)$	Relatives Inneres einer Menge A, siehe Definition 1.4
$\mathbb{R}^m_+$	Halbraum der positiven Vektoren in $\mathbb{R}^n$
$\mathcal{C}^0(A)$	Raum der stetigen Funktionen auf A
$\mathcal{C}^1(A)$	Raum der stetig-differenzierbaren Funktionen auf A
∇_x	Gradient nach $x \in \mathbb{R}^n$ definiert als $\left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\right)$

Literatur

[KRT] E. de Klerk, C. Roos, T. Terlaky, *Nonlinear Optimization*, 2006, Lecture Notes, University of Waterloo