Optimalitätsbedingungen

Nadja Irmscher

28. Mai 2010

1 Nachweis von Suboptimalität und Abbruchkriterien

Über das gegebene Programm

$$\begin{aligned} \text{minimiere} \quad & f_0(x) \quad \text{über } x \in \mathcal{D} \\ \text{sodass} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{aligned}$$

ist nichts bezüglich seiner Konvexität bekannt. Das zugehörige duale Programm wurde ermittelt.
Falls ein zulässiger Punkt für das duale Programm existiert, zählt dieser als Nachweis für die Abschätzung:

$$p^* \geq g(\lambda, \nu)$$

Es ist also möglich, den primalen Optimalwert p^* nach unten abzuschätzen, ohne seinen genauen Wert zu kennen.
Ist x ein zulässiger Punkt für das primale Problem, so gilt:

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

x heißt ϵ-suboptimal mit $\epsilon = f_0(x) - g(\lambda, \nu)$.
(Analog wird (λ, ν) als ϵ-suboptimal für das duale Problem definiert.)
Die Dualitätslücke zweier zulässiger Punkte x und (λ, ν) lautet:

$$f_0(x) - g(\lambda, \nu)$$

Daraus folgt, dass

$$p^* \in [g(\lambda, \nu), f_0(x)]$$

und

$$d^* \in [g(\lambda, \nu), f_0(x)]$$

Falls die Dualitätslücke gerade Null ist für den primal zulässigen Punkt x und den dual zulässigen Punkt (λ, ν), so ist x eine Optimallösung des primalen und (λ, ν) eine Optimallösung des dualen Problems. Somit gilt (λ, ν) als Nachweis

für die Optimalität von x.
Diese Feststellung kann man zur Herleitung nicht-heuristischer Abbruchkriterien für Optimierungsalgorithmen verwenden:
Seien x^k eine Folge primal zulässiger und (λ^k, ν^k) eine Folge dual zulässiger, von einem Algorithmus erzeugte Punkte. $\epsilon_{abs} > 0$ sei die absolute Genauigkeit. Somit garantiert

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

dass, wenn der Algorithmus terminiert, dann x^k gerade ϵ_{abs}-suboptimal ist.
Eine ähnliche Bedingung kann verwendet werden, um eine gegebene relative Genauigkeit $\epsilon_{rel} \geq 0$ zu garantieren:
Wenn

$$g(\lambda^k, \nu^k) > 0, \quad \frac{f_0(x^k) - g(\lambda^k, \nu^k)}{g(\lambda^k, \nu^k)} \leq \epsilon_{rel}$$

oder

$$f_0(x^k) < 0, \quad \frac{f_0(x^k) - g(\lambda^k, \nu^k)}{-f_0(x^k)} \leq \epsilon_{rel}$$

gilt, so ist $p^* \neq 0$ und

$$\frac{f_0(x^k) - p^*}{|p^*|} \leq \epsilon_{rel}$$

Diese Abschätzung für ϵ_{rel} kann also zur problembezogenen Bestimmung eines ϵ verwendet werden, wenn man gegebene Punkte auf Suboptimalität testen möchte bzw. die Suboptimalität als Abbruchkriterium für einen Algorithmus verwenden will.

2 Komplementaritätsbedingung

Seien x^* die primal zulässige Lösung und (λ^*, ν^*) die dual zulässige Lösung und der duale Optimalwert ist gleich dem primalen, d.h. es herrscht starke Dualität. Dann gilt:

$$\begin{aligned}
f_0(x^*) &= g(\lambda^*, \nu^*) \\
&= \inf_x (f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x)) \\
&\leq f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) \\
&\leq f_0(x^*)
\end{aligned}$$

Die erste Zeile folgt, da die Dualitätslücke Null ist und die zweite ist die Definition der dualen Funktion. Die dritte Ungleichung gilt, da das Infimum über alle x gerade kleiner oder gleich dem Funktionswert an der Stelle $x = x^*$ ist. Die letzte Ungleichung folgt daraus, dass $\lambda_i^* \geq 0$, $f_i(x^*) \leq 0$, $i = 1, ..., m$ und $h_i(x^*) = 0$,$i = 1, ..., p$. Es ist offensichtlich, dass die Ungleichungen hier durch Gleichheit ersetzt werden können.
Hieraus ergeben sich verschiedene Schlüsse:
Aus der Gleichheit in der dritten Zeile folgt, dass x^* die Funktion $L(x, \lambda^*, \nu^*)$ über x minimiert. Dabei muss x^* nicht die einzige Optimallösung sein. Weiterhin gilt

$$\sum_{i=1}^{m} \lambda_i^* f_i(x^*) = 0$$

und da jeder Summand nichtpositiv ist, gilt

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

Dies ist eine Komplementaritätsbedingung. Sie gilt für alle primal optimalen x^* und dual optimalen (λ^*, ν^*), solange starke Dualität gilt. Anders ausgedrückt heißt das

$$\lambda_i^* > 0 \quad \Rightarrow \quad f_i(x^*) = 0$$

oder

$$f_i(x^*) < 0 \quad \Rightarrow \quad \lambda_i^* = 0$$

Solange also die i-te Nebenbedingung aktiv ist, ist der i-te optimale Lagrangemultiplikator gleich Null.

3 KKT-Optimalitätsbedingungen

Wir nehmen an, die Funktionen $f_0, ..., f_m, h_1, ..., h_p$ sind differenzierbar. Über die Konvexität ist nichts bekannt.

KKT-Bedingungen nichtkonvexer Funktionen

Wie bisher sind x^* und (λ^*, ν^*) optimale Punkte und die Dualitätslücke ist gleich Null. Da x^* die Funktion $L(x, \lambda^*, \nu^*)$ minimiert, muss der Gradient bei x^* gerade verschwinden (vgl. notwendige Optimalitätsbedingung 1. Ordnung), d.h.:

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

Dadurch ergeben sich die Karusch-Kuhn-Tucker-(KKT)-Bedingungen:

$$\begin{aligned} f_i(x^*) &\leq 0, \quad i = 1, ..., m \\ h_i(x^*) &= 0, \quad i = 1, ..., p \\ \lambda_i^* &\geq 0, \quad i = 1, ..., m \\ \lambda_i^* f_i(x^*) &= 0, \quad i = 1, ..., m \\ \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) &= 0 \end{aligned}$$

Für jedes Optimierungsproblem mit differenzierbaren Zielfunktionen und Nebenbedingungen, welches die starke Dualität erfüllt, müssen die dualen und primalen Optimalpunkte die KKT-Bedingungen erfüllen.

KKT-Bedingungen konvexer Aufgaben

Wenn das primale Problem konvex ist, sind die KKT-Bedingungen hinreichend für ein Optimum, d.h.: Seien f_i konvex, h_i affin und $\widetilde{x}$, $\widetilde{\lambda}, \widetilde{\nu}$ erfüllen die KKT-Bedingungen:

$$\begin{aligned} f_i(\widetilde{x}) &\leq 0, \quad i = 1, ..., m \\ h_i(\widetilde{x}) &= 0, \quad i = 1, ..., p \\ \widetilde{\lambda}_i &\geq 0, \quad i = 1, ..., m \\ \widetilde{\lambda}_i f_i(\widetilde{x}) &= 0, \quad i = 1, ..., m \\ \nabla f_0(\widetilde{x}) + \sum_{i=1}^{m} \widetilde{\lambda}_i \nabla f_i(\widetilde{x}) + \sum_{i=1}^{p} \widetilde{\nu}_i \nabla h_i(\widetilde{x}) &= 0 \end{aligned}$$

Man kann leicht zeigen, dass die Dualitätslücke hier gleich Null ist:
Die ersten beiden Bedingungen garantieren die primale Zulässigkeit von $\widetilde{x}$. Da $\widetilde{\lambda}_i \geq 0$ gilt, ist $L(x, \widetilde{\lambda}, \widetilde{\nu})$ konvex in x. Die letzte KKT-Bedingung besagt, dass der Gradient von $L(x, \widetilde{\lambda}, \widetilde{\nu})$ an der Stelle $x = \widetilde{x}$ gerade Null ist, d.h., $\widetilde{x}$ minimiert $L(x, \widetilde{\lambda}, \widetilde{\nu})$. Daraus folgt, dass:

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

Die letzte Zeile folgt aus $\widetilde{\lambda}_i f_i(\widetilde{x}) = 0$ und $h_i(\widetilde{x}) = 0$ gerade Null. Somit ist gezeigt, dass $\widetilde{x}$ und $(\widetilde{\lambda}, \widetilde{\nu})$ eine Dualitätslücke von Null haben und daher primal und dual optimal sind. Kurz gesagt sind also bei konvexen Optimalitätsproblemen mit differenzierbarer Zielfunktion und differenzierbaren Nebenbedingungen alle

Punkte, die die KKT-Bedingungen erfüllen, primal und dual optimal und die Dualitätslücke verschwindet. Wenn das konvexe Optimalitätsproblem zusätzlich die Slaterbedingung erfüllt, sind die KKT-Bedingungen notwendig und hinreichend für die Optimalität: Die Slaterbedingung besagt, dass die Dualitätslücke Null ist und das duale Optimum erreicht wird. Also ist x optimal genau dann, wenn (λ, ν) existieren, sodass die KKT-Bedingungen erfüllt sind.
Im Allgemeinen können viele Algorithmen für konvexe Optimierungsaufgaben als Methode zur Lösung der KKT-Bedingungen interpretiert werden.

Beispiel 3.1 (quadratisches Problem mit Gleichungsnebenbedingung) *Gegeben sei das Problem:*

$$\begin{aligned} \textit{minimiere} \quad & (1/2)x^T P x + q^T x + r \\ \textit{sodass} \quad & Ax = b \end{aligned}$$

wobei $P \in S^n_+$.
Die KKT-Bedingungen für dieses Problem lauten:

$$Ax^* = b, \quad Px^* + q + A^T \nu^* = 0$$

Anders geschrieben:

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

Diese n+m Gleichungen lassen sich nun leicht auflösen und man erhält x^* *als primales Optimum und* ν^* *als duales Optimum.*

Beispiel 3.2 (Das Water-Filling-Problem) *Wir betrachten das konvexe Optimierungsproblem:*

$$\begin{aligned} \textit{minimiere} \quad & -\sum_{i=1}^{n} \log(\alpha_i + x_i) \\ \textit{sodass} \quad & x \succeq 0 \\ & \mathbf{1}^T x = 1 \end{aligned}$$

mit $\alpha_i > 0$. *Das Problem entstammt der Informationstheorie beim Zuweisen von einer bestimmten Leistung zu einer Menge von n Kommunikationskanälen. Die Variable* x_i *repräsentiert die Sendeleistung des i-ten Kanals und* $\log(\alpha_i + x_i)$ *gibt die Kapazität der Kanäle an. Es geht also darum, eine Gesamtleistung auf verschiedene Kanäle zu verteilen, sodass die totale Kommunikationsrate maximiert wird.*
Mit $\lambda^* \in \mathbb{R}^n$ *als Lagarange-Multiplikator für die Ungleichungsnebenbedingungen* $x^* \succeq 0$ *und* $\nu^* \in \mathbb{R}$ *für die Gleichungsnebenbedingung* $1^T x = 1$ *erhalten wir die KKT-Bedingungen*

$$\begin{aligned} x^* \succeq 0, \quad \mathbf{1}^T x^* = 1, \quad \lambda^* \succeq 0, \quad \lambda_i^* x_i^* = 0, \quad i = 1, ..., n \\ -1/(\alpha_i + x_i^*) - \lambda_i^* + \nu^* = 0, \quad i = 1, ..., n \end{aligned}$$

Wir können diese Gleichungen lösen, um x^*, λ^* und ν^* zu finden. λ^* fungiert wie eine Schlupfvariable, weshalb sie auch weggelassen werden kann:

$$x^* \succeq 0, \quad \mathbf{1}^T x^* = 1, \quad x_i^*(\nu^* - 1/(\alpha_i + x_i^*)) = 0, \quad i = 1, ..., n$$
$$\nu^* \geq 1/(\alpha_i + x_i^*), \quad i = 1, ..., n$$

Falls $\nu^* < 1/\alpha_i$ ist, gilt die letzte Bedingung nur, wenn: $x_i^* > 0$. Dies wiederum impliziert in der dritten Bedingung, dass $\nu^* = 1/(\alpha_i + x_i^*)$. Wenn wir nach x_i^* auflösen, folgt, dass $x_i^* = 1/\nu^* - \alpha_i$, wenn $\nu^* < 1/\alpha_i$. Ist $\nu^* \geq 1/\alpha_i$, dann ist $x_i^* > 0$ nicht möglich, sonst gilt $\nu^* \geq 1/\alpha_i > 1/(\alpha_i + x_i)$. Das verletzt gerade die Komplementaritätsbedingung.
Das heißt, $x_i^* = 0$, wenn $\nu^* \geq 1/\alpha_i$. Das ergibt

$$x_i^* = \begin{cases} 1/\nu^* - \alpha_i & \textit{wenn } \nu^* < 1/\alpha_i \\ 0 & \textit{wenn } \nu^* \geq 1/\alpha_i, \end{cases}$$

oder einfacher: $x_i^* = \max(0, 1/nu^* - \alpha_i)$. Wenn man dieses x in die Gleichung $\mathbf{1}^T x^* = 1$ einsetzt, erhält man

$$\sum_{i=1}^{n} \max(0, 1/\nu^* - \alpha_i) = 1$$

Die linke Seite ist eine stückweise lineare Funktion in $1/\nu^*$ mit Sprungstellen in α_i, sodass die Gleichung eine eindeutige endliche Lösung hat.

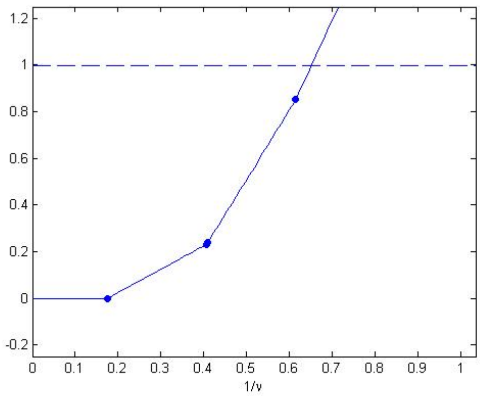

Abbildung 3.1: Zur Veranschaulichung der Sprungstellen von α

Diese Lösungsmethode heißt Water-Filling, weil man folgende Assoziationen finden kann:
α_i ist die jeweilige Bodenhöhe über dem Nullniveau an der zugehörigen Position i. Nun wird die Region mit Wasser geflutet bis wir einen Wasserstand von $1/\nu$ haben. Die totale Menge Wasser ist jetzt $\sum_{i=1}^{n} \max(0, 1/\nu^* - \alpha_i)$. Das Wasser wird weiter aufgefüllt, bis die totale Menge an Wasser gerade 1 ist. Die Tiefe des Wassers über der Erhöhung an der Position i ist gerade x_i^*.

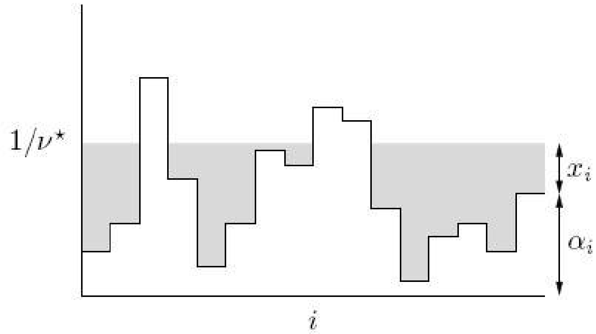

Abbildung 3.2: Water-Filling-Modell

4 Mechanische Interpretation der KKT-Optimalitätsbedingungen

Die KKT-Bedingungen haben eine anschauliche mechanische Interpretation, welche auch eine von Lagranges ursprünglichen Motivationen war.

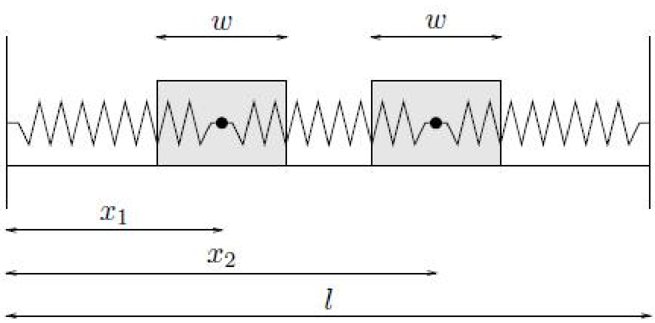

Abbildung 4.1: mechanische Interpretation der KKT-Punkte

In Bild ist folgendes Modell dargestellt:
Zwei Blöcke sind miteinander und rechts und links jeweils mit der Wand durch Federn verbunden. Die Federn sind mittig an den Blöcken befestigt, wobei ihre Position durch ein $x \in \mathbb{R}^2$ beschrieben wird:
x_1 ist der Abstand der linken Wand bis zur Befestigung der Feder am linken Blöck, x_2 ist der Abstand der linken Wand bis zur Befestigung der Feder am rechten Blöck.
Die linke Wand befindet sich an der Position Null, die rechte Wand an der Position l.
Die potenzielle Energie der Feder ist durch folgende Gleichung gegeben:

$$f_0(x_1, x_2) = 1/2k_1x_1^2 + 1/2k_2(x_2 - x_1)^2 + 1/2k_3(l - x_2)^2$$

k_i, $i \in \{1,2,3\}$, sind die Federkonstanten (Kraft pro Auslenkung der Feder).
x_1 ist die Auslenkung der linken Feder, $(x_2 - x_1)$ die Auslenkung der mittleren Feder und $(l - x_2)$ die Auslenkung der rechten Feder.
Bei x^* wird die potentielle Energie unter folgenden Nebenbedingungen minimiert:

$$w/2 - x_1 \leq 0, \quad w + x_1 - x_2 \leq 0, \quad w/2 - l + x_2 \leq 0$$

Diese Nebenbedingungen werden kinematische Nebenbedingungen genannt (Kinematik ist die Lehre der Bewegung von Punkten und Körpern im Raum). Sie beinhalten die Tatsache, dass $w > 0$ und sorgen dafür, dass die Körper sich nicht gegenseitig durchdringen bzw. nicht durch die Wand gehen.
Daraus ergibt sich folgende Optimierungsaufgabe:

$$\begin{aligned} \text{minimiere} \quad & 1/2k_1x_1^2 + 1/2k_2(x_2 - x_1)^2 + 1/2k_3(l - x_2)^2 \\ \text{sodass} \quad & w/2 - x_1 \leq 0 \\ & w + x_1 - x_2 \leq 0 \\ & w/2 - l + x_2 \leq 0 \end{aligned}$$

Mit λ_1, λ_2, λ_3 als Lagrangemultiplikatoren bestehen die KKT-Bedingungen dieses Problems aus den kinematischen Nebenbedingungen, den nichtnegativen Variablen $\lambda_i \geq 0$ und den Komplementaritätsbedingungen

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

Der Gradient wird gleich Null gesetzt:

$$\begin{pmatrix} k_1x_1 - k_2(x_2 - x_1) \\ k_2(x_2 - x_1) - k_3(l - x_2) \end{pmatrix} + \lambda_1 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \lambda_3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0$$

Die k_i, $i \in \{1,2,3\}$, haben die Einheit Newton/Meter und die x_i, $i \in \{1,2\}$ bzw. das l können in Metern angegeben werden. Hieraus ergibt sich, dass die Gleichung am Ende eine Kraft in Newton angibt und auch die λ_i, $i \in \{1,2,3\}$, Kräfte sein müssen. Sie lassen sich als Kontaktkräfte zwischen den Wänden und den Blöcken interpretieren.
Die obere Gleichung legt fest, dass die Summe der Kräfte an dem ersten Block gerade Null ist, dass heißt, sie heben sich gegenseitig auf bzw. sie sind ausgeglichen. $-k_1x_1$ ist die Kraft, mit der die linke Feder an den linken Block drückt. $k_2(x_2 - x_1)$ ist die Kraft, mit der die mittlere Feder gegen den ersten Block drückt. λ_1 ist die Kraft, mit der die linke Wand dagegen hält und $-\lambda_2$ ist die Kraft, die vom rechten Block ausgeübt wird. Die Kontaktkräfte müssen gerade von der Kontaktoberfläche wegzeigen, was dazu führt, dass $\lambda_1 \geq 0$ und $-\lambda_2 \leq 0$ sein müssen.
Analog dazu ist die zweite Gleichung die Kraftbalance für den zweiten Block und die letzte Komplementaritätsbedingung sorgt dafür, dass λ_3 gleich Null ist solange der zweite Block die Wand nicht berührt.
Im folgenden Bild wird das noch einmal anschaulich dargestellt:

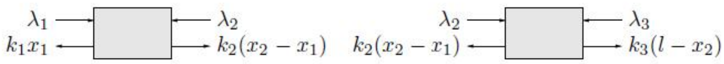

Abbildung 4.2: Krafteinwirkungen

In diesem Beispiel sind die Zielfunktion und die Nebenbedingungen konvex und die Slaterbedingung ist erfüllt, falls $2w \leq l$. Das bedeutet, dass der Raum zwischen den Wänden groß genug ist, damit die beiden Blöcke dazwischen passen. Daraus folgt, dass die Energieformulierung dasselbe Ergebnis liefert wie die Formulierung mit den Kraftverhältnissen, gegeben durch die KKT-Bedingungen.

5 Lösung des primalen Problems mithilfe des dualen

Wie bereits erwähnt ist jeder primale Optimalwert auch ein Minimum von $L(x, \lambda^*, \nu^*)$, wenn starke Dualität gilt und die optimale Lösung (λ^*, ν^*) existiert. Dieser Fakt hilft uns manchmal, eine primale Optimallösung mithilfe der dualen Optimallösung zu bestimmen:
Sei starke Dualität gegeben und das duale Optimum (λ^*, ν^*) bekannt. Nehmen wir an, die Lösung von

$$\text{minimiere} \quad f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \quad \text{über} \quad x \in \mathbb{R}^n$$

ist eindeutig. Dann ist diese Lösung das primale Optimum, wenn sie primal zulässig ist. Falls sie nicht primal zulässig ist, kann es keinen primal optimalen Punkt geben. Dieses Vorgehen ist interessant, wenn das duale Problem leichter zu lösen ist als das primale, beispielsweise, wenn es analytisch lösbar ist oder wenn es eine spezielle Struktur aufweist.

Beispiel 5.1 (Entropiemaximierung) *Wir betrachten das Entropiemaximierungsproblem*

$$\begin{aligned} \textit{minimiere} \quad & f_0(x) = \sum_{i=1}^{n} x_i \ln x_i \\ \textit{sodass} \quad & Ax \preceq b \\ \textit{und} \quad & \mathbf{1}^T x = 1 \end{aligned}$$

mit $x \in \mathbb{R}^n_{++}$
und das duale Problem

$$\begin{aligned} &maximiere \quad -b^T\lambda - \nu - e^{-\nu-1}\sum_{i=1}^{n} e^{-a_i^T\lambda} \\ &sodass \quad \lambda \succeq 0 \end{aligned}$$

wobei a_i *gerade die Spalten von* A *sind. Wir stellen fest, dass die schwache Slaterbedingung gilt, d.h.: Es existiert ein* $x \succ 0$ *mit* $Ax \preceq b$ *und* $1^T x = 1$. *Daraus folgt, dass die starke Dualität gilt und eine optimale Lösung* (λ^*, ν^*) *existiert. Angenommen, wir haben das duale Problem gelöst. Dann ist die Lagrangefunktion an der Stelle* (λ^*, ν^*) *gerade:*

$$L(x, \lambda^*, \nu^*) = \sum_{i=1}^{n} x_i \log x_i + \lambda^{*T}(Ax - b) + \nu^*(1^T x - 1)$$

$L(x, \lambda^*, \nu^*)$ *ist streng konvex auf* D *und nach unten beschränkt, sodass die eindeutige Lösung lautet:*

$$x_i^* = 1/exp(a_i^T\lambda^* + \nu^* + 1), \quad i = 1, ..., n$$

Wenn x^* *primal zulässig ist, muss es die Optimallösung vom primalen Problem sein.*
Falls es nicht primal zulässig ist, können wir darauf schließen, dass das Minimum nicht angenommen wird.

Beispiel 5.2 (Minimierung über die Summe separierbarer Funktionen)
Wir betrachten das Problem

$$\begin{aligned} &minimiere \quad f_0(x) = \sum_{i=1}^{n} f_i(x_i) \\ &sodass \quad a^T x = b \end{aligned}$$

mit $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ *und* $f_i : \mathbb{R} \to \mathbb{R}$ *differenzierbar und streng konvex.*
Die Zielfunktion wird separabel genannt, weil sie die Summe von Funktionen von individuellen Variablen $x_1, ..., x_n$ *ist. Wir nehmen an, dass ein Punkt* $x_0 \in dom f_0$ *mit* $a^T x_0 = b$ *existiert. Das bedeutet, das Problem hat einen eindeutigen Optimalpunkt* x^*.
Die Lagrangefunktion lautet

$$L(x, \nu) = \sum_{i=1}^{n} f_i(x_i) + \nu(a^T x - b) = -b\nu + \sum_{i=1}^{n}(f_i(x_i) + \nu a_i x_i)$$

Sie ist separabel, also lautet die duale Funktion:

$$\begin{aligned} g(\nu) &= -b\nu + \inf_x \Big(\sum_{i=1}^{n} (f_i(x_i) + \nu a_i x_i)\Big) \\ &= -b\nu + \Big(\sum_{i=1}^{n} \inf_{x_i} (f_i(x_i) + \nu a_i x_i)\Big) \\ &= -b\nu - \sum_{i=1}^{n} f_i^*(-\nu a_i) \end{aligned}$$

Das duale Problem sieht folgendermaßen aus:

$$maximiere \quad -b\nu - \sum_{i=1}^{n} f_i^*(-\nu a_i)$$

mit dem Skalar $\nu \in \mathbb{R}$.
Angenommen wir haben ein duales Optimum ν^* *gefunden.*
(Es gibt verschiedene Methoden, ein konvexes Problem mit einem Skalar zu lösen, eine davon ist die sogenannte Bisektion.)
Weil f_i *streng konvex ist, ist die Funktion* $L(x, \nu^*)$ *streng konvex in* x *und hat somit ein eindeutiges Minimum* $\widetilde{x}$. *Aber wir wissen ebenso, dass* x^* *gerade* $L(x, \nu^*)$ *minimiert, also muss* $\widetilde{x} = x^*$ *gelten. Wir können* x^* *aus der Bedingung* $\bigtriangledown_x L(x, \nu^*) = 0$ *berechnen, indem wir* $f_i'(x_i^*) = -\nu^* a_i$ *lösen.*