

Sattelpunkte und Optimalitätsbedingungen bei restringierten Optimierungsproblemen

Sandro Grunert

WS 08/09

Seminar Optimierung
Technische Universität Chemnitz

Inhaltsverzeichnis

0	Grundlegende Situation und Notation	3
1	Interpretationen der Lagrange-Dualität	4
1.1	Max.-Min.-Charakterisierung von schwacher und starker Dualität	4
1.2	Interpretation mit Hilfe von Sattelpunkten	5
1.3	Interpretation in der Spieltheorie	5
1.4	Ökonomische Interpretation	6
2	Optimalitätsbedingungen	7
2.1	Suboptimalität und Abbruchbedingung	7
2.2	Komplementärer Schlupf	8
2.3	KKT-Optimalitätsbedingungen	9
2.3.1	KKT-Bedingungen für konvexe Probleme	10
2.3.2	Beispiel: Water-Filling-Methode	10
2.3.3	Lösung des primalen Problems über das Duale	12

0 Grundlegende Situation und Notation

Primales Problem in Standardform:

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1, \dots, m, \\ & h_i(x) = 0, \quad i = 1, \dots, p, \\ & x \in \mathbb{R}^n. \end{aligned} \tag{0.1}$$

Falls die zulässige Menge

$$\mathcal{D} = \bigcap_{i=1}^m \text{dom } f_i \cap \bigcap_{i=1}^p \text{dom } h_i$$

von (0.1) nichtleer ist, existiert der optimale Wert p^* von (0.1). Über die Lagrange-Funktion

$$\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}, \quad \mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$$

mit $\text{dom } \mathcal{L} = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$, $\lambda_i \geq 0$ als Lagrange-Multiplikator für die i -te Ungleichungsrestriktion, ν_i als Lagrange-Multiplikator für die i -te Gleichungsrestriktion, definiert man die Lagrange-duale Funktion

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

Existiert keine untere Schranke für \mathcal{L} , dann ist $g(\lambda, \nu) = -\infty$. Die Lagrange-duale Funktion ist als punktweises Infimum affiner Funktionen von λ und ν konkav. Es gilt stets $g(\lambda, \nu) \leq p^*$. Nun kann man das duale Problem wie folgt formulieren:

$$\begin{aligned} \max \quad & g(\lambda, \nu) \\ \text{s.t.} \quad & \lambda \succeq 0. \end{aligned} \tag{0.2}$$

Den optimalen Wert von (0.2) bezeichnet man mit d^* . Es gilt stets

$$d^* \leq p^*,$$

was als schwache Dualität mit Dualitätslücke $p^* - d^* \geq 0$ bezeichnet wird. Die Frage ist, wann gilt $d^* = p^*$ (starke Dualität)?

1 Interpretationen der Lagrange-Dualität

1.1 Max.-Min.-Charakterisierung von schwacher und starker Dualität

Zunächst sollen das primale und das duale Optimierungsproblem in einer symmetrischen Form beschrieben werden. Zur Vereinfachung wird angenommen, es seien keine Gleichungsrestriktionen vorhanden. Die Lagrange-duale Funktion lautet nun also $g(\lambda)$. Man bezeichnet mit

$$p(x) = \sup_{\lambda \succeq 0} \mathcal{L}(x, \lambda) = \sup_{\lambda \succeq 0} f_0(x) + \sum_{i=1}^m \lambda_i f_i(x)$$

die primale Funktion, welche als punktweises Supremum affiner Funktionen von λ konvex ist. Es gilt

$$p(x) = \begin{cases} f_0(x), & \text{falls } f_i(x) \leq 0, \quad i = 1, \dots, m, \\ +\infty, & \text{sonst} \end{cases}.$$

Nun kann man den optimalen Wert des primalen Problems (0.1) als

$$p^* = \inf_{x \in \mathcal{D}} \sup_{\lambda \succeq 0} \mathcal{L}(x, \lambda) = \inf_{x \in \mathcal{D}} p(x)$$

schreiben. Zusätzlich folgt aus der Definition der Lagrange-dualen Funktion

$$d^* = \sup_{\lambda \succeq 0} \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda) = \inf_{x \in \mathcal{D}} g(\lambda).$$

Also kann schwache Dualität geschrieben werden als

$$\sup_{\lambda \succeq 0} \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda) \leq \inf_{x \in \mathcal{D}} \sup_{\lambda \succeq 0} \mathcal{L}(x, \lambda) \quad (1.1)$$

und starke Dualität als

$$\sup_{\lambda \succeq 0} \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda) = \inf_{x \in \mathcal{D}} \sup_{\lambda \succeq 0} \mathcal{L}(x, \lambda).$$

Starke Dualität bedeutet demnach, die Reihenfolge von Minimierung über $x \in \mathcal{D}$ und Maximierung über $\lambda \succeq 0$ kann vertauscht werden, ohne das optimale Resultat zu verändern. Tatsächlich hängt die Ungleichung (1.1) aber nicht von Eigenschaften der Lagrange-Funktion ab, sondern man kann diese viel allgemeiner formulieren:

Beobachtung 1 (*Max.-Min.-Ungleichung*)

Seien $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, $W \subseteq \mathbb{R}^n$ und $Z \subseteq \mathbb{R}^m$ gegeben. Dann gilt

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) \leq \inf_{w \in W} \sup_{z \in Z} f(w, z). \quad (1.2)$$

Diese allgemeine Ungleichung nennt man *Max.-Min.-Ungleichung*.

Falls in (1.2) Gleichheit gilt, das heißt

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) = \inf_{w \in W} \sup_{z \in Z} f(w, z), \quad (1.3)$$

sagt man: f erfüllt auf $W \times Z$ die strenge Max.-Min.-Eigenschaft (bzw. Max.-Min.-Gleichung, Sattelpunkteigenschaft). (1.3) ist nur in speziellen Fällen erfüllt, zum Beispiel wenn $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ die Lagrange-Funktion eines Optimierungsproblems ist, für das starke Dualität herrscht, $W = \mathbb{R}^n$ und $Z = \mathbb{R}_+^m$ ist.

1.2 Interpretation mit Hilfe von Sattelpunkten

Definition 1 (*Sattelpunkt*)

Ein Paar $(\tilde{w}, \tilde{z}) \in W \times Z \subseteq \text{dom } f$ heißt Sattelpunkt der Funktion $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, falls

$$f(\tilde{w}, z) \leq f(\tilde{w}, \tilde{z}) \leq f(w, \tilde{z}) \quad \forall w \in W, z \in Z. \quad (1.4)$$

Diese Ungleichung nennt man Sattelpunkteigenschaft bzw. Sattelpunktungleichung.

Anders formuliert: Ist (\tilde{w}, \tilde{z}) ein Sattelpunkt von f , gilt

$$f(\tilde{w}, \tilde{z}) = \inf_{w \in W} f(w, \tilde{z}), \quad f(\tilde{w}, \tilde{z}) = \sup_{z \in Z} f(\tilde{w}, z).$$

Beobachtung 2 (*Eigenschaften von Sattelpunkten*)

Ist (\tilde{w}, \tilde{z}) ein Sattelpunkt von f , so gilt Folgendes:

- (a) Die strenge Max.-Min.-Eigenschaft (1.3) ist erfüllt.
- (b) Der Wert $f(\tilde{w}, \tilde{z})$ ist derselbe für alle Sattelpunkte (\tilde{w}, \tilde{z}) . Sind $f(\tilde{w}_1, \tilde{z}_1)$ und $f(\tilde{w}_2, \tilde{z}_2)$ Sattelpunkte von f , dann auch $f(\tilde{w}_1, \tilde{z}_2)$ und $f(\tilde{w}_2, \tilde{z}_1)$.

Aus obigen Beobachtungen folgt für die Lagrange-Dualität: Sind x^* und λ^* primal und dual zulässige Optimallösungen für ein Optimierungsproblem mit starker Dualität, so bilden diese einen Sattelpunkt der Lagrange-Funktion. Die Umkehrung gilt auch: Ist (x, λ) ein Sattelpunkt der Lagrange-Funktion, dann ist x primal optimal, λ dual optimal und die Dualitätslücke für dieses Problem ist gleich Null.

1.3 Interpretation in der Spieltheorie

Gegeben sei ein stetiges 2-Personen-Nullsummenspiel, das heißt 2 Spieler, deren überabzählbare Strategiemengen W und Z , eine Auszahlungsfunktion $f : W \times Z \rightarrow \mathbb{R}$ und der Gewinn eines Spielers (positive Auszahlung) impliziert einen Verlust beim anderen Spieler (negative Auszahlung).

Spieler 1 wähle eine Strategie $w \in W$ und Spieler 2 wähle $z \in Z$. Dann zahlt Spieler 1 $f(w, z)$ an Spieler 2. Hierbei will Spieler 1 f über w minimieren, Spieler 2 dagegen f über z maximieren.

Zunächst wird angenommen, Spieler 1 entscheidet sich zuerst, anschließend entscheidet Spieler 2 im Wissen von der Entscheidung von Spieler 1. Spieler 2 will die Auszahlung maximieren, also wählt er $\tilde{z} \in Z$ so, dass die Auszahlung gerade $f(w, \tilde{z}) = \sup_{z \in Z} f(w, z)$ beträgt (Annahme: Das Supremum werde angenommen, ansonsten sei die Auszahlung ein Wert nahe des Supremums). $\sup_{z \in Z} f(w, z)$ hängt aber von $w \in W$ ab, der Entscheidung von Spieler 1. Dieser weiß (oder nimmt an), dass Spieler 2 so handeln wird, also wählt er $w \in W$ so, dass sich sein Schaden in Grenzen hält bzw. so klein wie möglich ist, das heißt

$$\operatorname{argmin}_{w \in W} \sup_{z \in Z} f(w, z),$$

was die Auszahlung

$$\inf_{w \in W} \sup_{z \in Z} f(w, z)$$

zur Folge hat.

Nimmt man nun an, die Spielreihenfolge sei umgekehrt, das heißt, zuerst wählt Spieler 2 und anschließend Spieler 1, dann erhält man die Auszahlung

$$\sup_{z \in Z} \inf_{w \in W} f(w, z).$$

Aus der Max.-Min.-Ungleichung (1.2) folgt somit, dass es für einen Spieler von Vorteil ist, als Zweiter zu wählen bzw. vor der eigenen Wahl die des Gegenspielers zu kennen. Wenn die Sattelpunkteigenschaft (1.3) gilt, gibt es als zweiter Auswählender keinen Vorteil.

Definition 2 (Optimale Lösung und optimale Strategien)

Ist (\tilde{w}, \tilde{z}) ein Sattelpunkt für die Auszahlungsfunktion f auf $W \times Z$, dann heißt dieser optimale Lösung für das Spiel; \tilde{w} heißt optimale Strategie für Spieler 1 und \tilde{z} heißt optimale Strategie für Spieler 2 (In diesem Fall gibt es keinen Anzugsnachteil).

Kehrt man wieder zur Lagrange-Dualität zurück, so ist die Auszahlungsfunktion gerade die Lagrange-Funktion, $W = \mathbb{R}^n$, $Z = \mathbb{R}_+^m$, Spieler 1 wählt x und Spieler 2 wählt $\lambda \succeq 0$. Die Auszahlung beträgt p^* , falls Spieler 1 beginnt und d^* falls Spieler 2 beginnt.

1.4 Ökonomische Interpretation

Lagrange-Dualität lässt sich in interessanter Art und Weise ökonomisch interpretieren. x bezeichne eine Unternehmensstrategie (Handlungsweise des Unternehmens in einer bestimmten Situation) und $f_0(x)$ seien die dabei entstehenden Aufwendungen und Kosten (Fixkosten) in dem Sinne, dass man gerade mit $-f_0(x)$ den Gewinn bei Handlungsstrategie x des Unternehmens erhält (zum Beispiel in €). Eine jede Ungleichungsrestriktion $f_i(x) \leq 0$ repräsentiere gewisse Grenzen, zum Beispiel Kapazitätsbegrenzungen in der Produktion (begrenzte Lagerfläche, begrenzte Kapazität an vorhandener Arbeitskraft). Man kann nun die Strategie des Unternehmens finden, die den Gewinn maximiert und dabei alle Grenzen einhält, in dem man das folgende Optimierungsproblem löst:

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

Als optimalen Wert erhält man den optimalen Gewinn $-p^*$.

Nun stelle man sich ein zweites Szenario vor: Das Unternehmen darf die Grenzen verletzen, in dem es die dabei entstehenden zusätzlichen Kosten bezahlen muss. Annahme: Diese zusätzlichen Kosten steigen linear mit steigender Übertretung der Grenzen, gemessen von f_i , das heißt, das Unternehmen zahlt für die i -te Grenzüberletzung $\lambda_i f_i(x)$ (variable Kosten, da von λ_i abhängig). Bei inaktiven Nebenbedingungen, das heißt $f_i(x) < 0$, bekommt das Unternehmen gerade eine Auszahlung in Höhe von $\lambda_i f_i(x)$.

Interpretation von λ_i : Der Koeffizient λ_i kann als Preis für die Verletzung der Grenze $f_i(x) \leq 0$ interpretiert werden (Einheit von λ_i : zum Beispiel € pro Einheit der Verletzung, gemessen von f_i). Für den gleichen Preis kann das Unternehmen sämtliche ungenutzten Kapazitäten verkaufen (im Falle $f_i(x) < 0$).

Annahme: $\lambda_i \geq 0 \forall i$, das heißt, das Unternehmen muss für Grenzüber tretungen zahlen

und erhält Einkommen für inaktive Nebenbedingungen. Die Gesamtkosten des Unternehmens für die Strategie x mit Marktpreisen λ_i stellt folgende Kostenfunktion dar:

$$\mathcal{L}(x, \lambda) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x).$$

Offenbar wird das Unternehmen so handeln, dass es seine Gesamtkosten $\mathcal{L}(x, \lambda)$ minimiert. Als Resultat für die optimalen Kosten des Unternehmens in Abhängigkeit von λ erhält man $g(\lambda)$, die Lagrange-duale Funktion. Der optimale Wert d^* des dualen Problems (0.2) entspricht hierbei den optimalen Kosten des Unternehmens unter dem günstigsten Preisvektor λ .

Schwache Dualität: Die optimalen Kosten d^* im zweiten Szenario (Verletzungen der Nebenbedingungen können „gemietet“ und „vermietet“ werden) sind kleiner oder gleich den optimalen Kosten p^* im ursprünglichen Szenario, selbst bei ungünstigsten Preisen am Markt!

Starke Dualität: Das duale Optimum λ^* werde angenommen. Dann kann man λ^* als einen Preisvektor interpretieren, für den es für das Unternehmen keinen Vorteil gibt, Verletzungen seiner Kapazitäten zu begehen. Aus diesem Grund bezeichnet man das duale Optimum λ^* auch als eine Kombination von Schattenpreisen für das ursprüngliche Problem.

2 Optimalitätsbedingungen

Es wird das Problem (0.1) betrachtet, welches nicht notwendigerweise konvex ist, außer falls ausdrücklich darauf hingewiesen wird.

2.1 Suboptimalität und Abbruchbedingung

Aus Überlegungen über Suboptimalität soll zunächst eine Abbruchbedingung für Optimierungsalgorithmen hergeleitet werden. Dazu sei an die Definition der ε -Suboptimalität erinnert:

Definition 3 (ε -Suboptimalität)

Ein zulässiger Punkt x mit

$$f_0(x) \leq \varepsilon$$

($\varepsilon > 0$) heißt ε -suboptimal.

Existiert ein dual zulässiger Punkt (λ, ν) , so hat man eine untere Schranke für den Optimalwert des primalen Problems:

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

Ist p^* unbekannt, liefert (λ, ν) also ein Kriterium für die Begrenzung des primalen Optimalwertes. Bei starker Dualität ist diese Begrenzung beliebig „gut“, das heißt, der Abstand zu p^* kann infinitesimal gewählt werden.

Hat man nun irgendeinen primal zulässigen Punkt x und „strebt nach p^* “, so erlauben entsprechende dual zulässige Punkte (λ, ν) nach oben abzugrenzen, wie suboptimal der gegebene primal zulässige Punkt ist:

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

das heißt, x ist ε -suboptimal mit $\varepsilon = f_0(x) - g(\lambda, \nu)$. Außerdem ist (λ, ν) ε -suboptimal für das duale Problem:

$$f_0(x) - g(\lambda, \nu) \geq d^* - g(\lambda, \nu).$$

$f_0(x) - g(\lambda, \nu)$ kann man als Dualitätslücke auffassen. Ein primal und dual zulässiges Paar $x, (\lambda, \nu)$ gibt also ein Intervall für die optimalen Werte des primalen und dualen Problems an:

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

Bei Dualitätslücke gleich Null ist x primal optimal und (λ, ν) dual optimal.

Diese Beobachtungen kann man nun nutzen, um Abbruchkriterien in Optimierungsalgorithmen zu erhalten. Ein Algorithmus erzeuge primal und dual zulässige Folgen $\{x^{(k)}\}$ und $\{\lambda^{(k)}, \nu^{(k)}\}$, $k = 1, 2, \dots$ und $\varepsilon_{abs} > 0$ sei eine gegebene absolute Genauigkeit. Dann garantiert die Abbruchbedingung

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

ε_{abs} -Suboptimalität von $x^{(k)}$ bei Algorithmus-Ende. Für beliebig kleine $\varepsilon_{abs} > 0$ würde starke Dualität benötigt werden.

Analog dazu lässt sich eine Abbruchbedingung für eine relative Genauigkeit $\varepsilon_{rel} > 0$ formulieren: Falls entweder

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

oder

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

gilt, dann ist $p^* \neq 0$ und der relative Fehler ist garantiert kleiner oder gleich ε_{rel} :

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

2.2 Komplementärer Schlupf

Annahme: Die primalen und dualen Optimalwerte werden angenommen und stimmen überein, also starke Dualität liegt vor. Seien dazu x^* primal optimal und (λ^*, ν^*) dual optimal. Dann gilt Folgendes:

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

Also minimiert $x^* \mathcal{L}(x, \lambda^*, \nu^*)$ über x . Außerdem folgt

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

und da jeder Summand kleiner gleich Null ist, folgert man sofort

$$\lambda_i^* f_i(x^*) = 0, \quad i = 1, \dots, m. \quad (2.1)$$

Diese nennt man Komplementarität und der Satz vom komplementären Schlupf hat jene zum Inhalt. Anders formuliert:

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

Gilt die Rückrichtung auch, das heißt $\lambda_i^* > 0 \Leftrightarrow f_i(x^*) = 0$, so spricht man von strikter Komplementarität.

2.3 KKT-Optimalitätsbedingungen

Annahme: Die Funktionen $f_0, f_1, \dots, f_m, h_1, \dots, h_p$ sind differenzierbar, (0.1) sei nicht notwendigerweise konvex!

Seien wie zuletzt x^* und (λ^*, ν^*) primal und dual optimal mit Dualitätslücke gleich Null. Da x^* $\mathcal{L}(x, \lambda^*, \nu^*)$ über x minimiert, muss der Gradient der Lagrange-Funktion bei x^* verschwinden, das heißt

$$\nabla_x \mathcal{L}(x^*, \lambda^*, \nu^*) = \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p \nu_i^* \nabla h_i(x^*) = 0.$$

Nun kann man Folgendes formulieren:

Definition 4 (KKT-Bedingungen, KKT-Punkt)

Die Bedingungen

$$\begin{aligned} \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p \nu_i^* \nabla h_i(x^*) &= 0, \\ f_i(x^*) &\leq 0, \quad i = 1, \dots, m, \\ h_i(x^*) &= 0, \quad i = 1, \dots, p, \\ \lambda_i &\geq 0, \quad i = 1, \dots, m, \\ \lambda_i f_i(x^*) &= 0, \quad i = 1, \dots, m. \end{aligned} \quad (2.2)$$

heißen Karush-Kuhn-Tucker-Bedingungen. Jeder Punkt (x^*, λ^*, ν^*) , der die KKT-Bedingungen (2.2) erfüllt, heißt KKT-Punkt von (0.1).

Für ein Optimierungsproblem mit differenzierbarer Zielfunktion und differenzierbaren Nebenbedingungen, für das starke Dualität gilt, muss jedes Paar primal und dual optimaler Punkte die KKT-Bedingungen erfüllen. Die KKT-Bedingungen sind also notwendige Optimalitätsbedingungen.

2.3.1 KKT-Bedingungen für konvexe Probleme

Ist das primale Problem (0.1) konvex, so sind die KKT-Bedingungen hinreichend für primale und duale Optimalität. Konkret:

Satz 1 (*KKT-Bedingungen für konvexe Probleme*)

Gegeben sei Problem (0.1). Sind die f_i konvex und die h_i affin und genügen die Punkte $(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$ den KKT-Bedingungen (2.2), dann ist \tilde{x} primal optimal, $(\tilde{\lambda}, \tilde{\nu})$ dual optimal und die Dualitätslücke verschwindet.

Man kann zeigen: Falls ein konvexes Optimierungsproblem mit differenzierbarer Zielfunktion und differenzierbaren Nebenbedingungen die Regularitätsbedingung von Slater erfüllt, dann sind die KKT-Bedingungen notwendig und hinreichend für Optimalität, das heißt: Die Regularitätsbedingung von Slater impliziert: Dualitätslücke gleich Null und duales Optimum wird angenommen, also ist x primal optimal $\Leftrightarrow \exists(\lambda, \nu)$, so dass $x, (\lambda, \nu)$ die KKT-Bedingungen erfüllen.

Beispiel 1 (*Quadratische Minimierung*)

Betrachtet wird

$$\begin{aligned} \min \quad & \frac{1}{2}x^T P x + q^T x + r \\ \text{s.t.} \quad & Ax = b, \\ & P \in S_+^n, A \in \mathbb{R}^{m \times n}. \end{aligned}$$

KKT-Bedingungen:

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

was geschrieben werden kann als

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

Dies ist ein symmetrisches Problem mit $m + n$ Gleichungen und Variablen, dessen Lösung die primal und dual optimalen Variablen bereitstellt.

2.3.2 Beispiel: Water-Filling-Methode

Betrachtet wird das konvexe Optimierungsproblem

$$\begin{aligned} \min \quad & -\sum_{i=1}^n \log(\alpha_i + x_i) \\ \text{s.t.} \quad & x \succeq 0, \\ & \mathbf{1}^T x = 1, \\ & \alpha_i > 0, \quad i = 1, \dots, n. \end{aligned}$$

Dieses Problem tritt in der Informationstheorie auf, genauer bei der Verteilung elektrischer Leistung auf n Kommunikationskanäle. x_i repräsentiert hierbei die Sendeleistung des i -ten Kanals und $\log(\alpha_i + x_i)$ beschreibt die Kapazität bzw. Kommunikationsrate auf dem i -ten Kanal. Das Problem lässt sich nun wie folgt beschreiben: Eine gewisse elektrische Gesamtleistung von 100 % soll auf n Kanäle so verteilt werden, dass die gesamte Kommunikationsrate maximiert wird.

Seien $\lambda^* \in \mathbb{R}^n$ für $x^* \succeq 0$ und $\nu^* \in \mathbb{R}$ für $\mathbf{1}^T x = 1$ die Lagrange-Multiplikatoren, dann erhält man folgende KKT-Bedingungen:

$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \lambda^*, \nu^*) = 0 &\Leftrightarrow -\frac{1}{\alpha_i + x_i^*} - \lambda_i^* + \nu^*, \quad i = 1, \dots, n, \\ x^* &\succeq 0, \\ \mathbf{1}^T x^* &= 1, \\ \lambda^* &\succeq 0, \\ \lambda_i^* x_i^* &= 0, \quad i = 1, \dots, n. \end{aligned}$$

Man kann diese Gleichungen nun wie folgt lösen, um x^* , λ^* , ν^* zu erhalten: Zunächst wird λ^* als Schlupfvariable aufgefasst, um es zu eliminieren:

$$\begin{aligned} x^* &\succeq 0, \\ \mathbf{1}^T x^* &= 1, \\ x_i^* \left(\nu^* - \frac{1}{\alpha_i + x_i^*} \right) &= 0, \quad i = 1, \dots, n \\ \nu^* &\geq \frac{1}{\alpha_i + x_i^*}, \quad i = 1, \dots, n. \end{aligned}$$

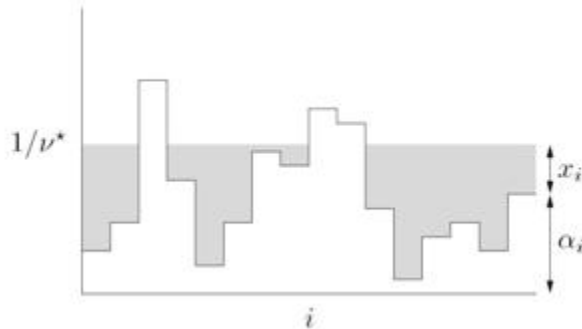
Hieraus folgt

$$x_i^* = \begin{cases} \frac{1}{\nu^*} - \alpha_i, & \text{falls } \nu^* < \frac{1}{\alpha_i}, \quad i = 1, \dots, n, \\ 0, & \text{falls } \nu^* \geq \frac{1}{\alpha_i}, \quad i = 1, \dots, n, \end{cases}$$

oder einfacher $x_i^* = \max\left\{0, \frac{1}{\nu^*} - \alpha_i\right\}$. Dies setzt man jetzt in die Bedingung $\mathbf{1}^T x^* = 1$ ein und erhält

$$\sum_{i=1}^n \max\left\{0, \frac{1}{\nu^*} - \alpha_i\right\} = 1.$$

Da auf der linken Seite der Gleichung eine in $\frac{1}{\nu^*}$ monoton wachsende und stückweise lineare Funktion mit Knickstellen α_i steht, hat diese Gleichung eine eindeutige Lösung. Diese Lösungsmethode bezeichnet man auch als Water-Filling-Methode.



2.3.3 Lösung des primalen Problems über das Duale

Falls starke Dualität gilt und eine dual optimale Lösung (λ^*, ν^*) existiert, dann minimiert jeder primal optimale Punkt $\mathcal{L}(x, \lambda^*, \nu^*)$ über x . Diese Tatsache erlaubt es manchmal, eine primal optimale Lösung über die dual optimale Lösung zu errechnen.

Annahmen: Starke Dualität gilt, duales Optimum (λ^*, ν^*) existiert und die Lösung von

$$\min f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x) \quad (2.3)$$

sei eindeutig (Für ein konvexes Problem ist dies der Fall, zum Beispiel falls $\mathcal{L}(x, \lambda^*, \nu^*)$ eine strikt konvexe Funktion in x ist).

Falls nun die Lösung von (2.3) primal zulässig ist, muss sie auch primal optimal sein. Falls diese nicht primal zulässig ist, so kann man davon ausgehen, dass das primale Optimum nicht angenommen wird. Diese Situation ist genau dann interessant, wenn das duale Problem leichter zu lösen ist als das primale Problem.

Beispiel 2 (*Entropie-Maximierung*)

Betrachtet wird

$$\begin{aligned} \min \quad & f_0(x) = \sum_{i=1}^n x_i \log x_i \\ \text{s.t.} \quad & Ax \preceq b, \\ & \mathbf{1}^T x = 1 \end{aligned} \quad (2.4)$$

mit Domain $\mathcal{D} = \mathbb{R}_{++}^n$ und das zugehörige duale Problem

$$\begin{aligned} \max \quad & -b^T \lambda - \nu - e^{-\nu-1} \sum_{i=1}^n e^{-a_i^T \lambda} \\ \text{s.t.} \quad & \lambda \succeq 0, \\ & \mathbf{1}^T x = 1 \end{aligned}$$

(a_i ist die i -te Spalte von A). Annahme: Die schwache Slater-Bedingung ist erfüllt, das heißt, $\exists x \succ 0 : Ax \preceq b$ und $\mathbf{1}^T x = 1$. \Rightarrow Starke Dualität liegt vor und die dual optimale Lösung (λ^*, ν^*) existiert. Angenommen, man habe das duale Problem gelöst: Die Lagrange-Funktion an der Stelle (λ^*, ν^*) lautet

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

Diese Funktion ist strikt konvex auf \mathcal{D} und nach unten beschränkt, also hat \mathcal{L} eine eindeutige minimierende Lösung x^* , gegeben durch

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

Ist x^* primal zulässig, dann ist es die optimale Lösung von (2.4), ansonsten kann man schlussfolgern, dass das primale Optimum nicht angenommen wird.

Beispiel 3 (Minimierung einer separablen Funktion)

Betrachtet wird

$$\begin{aligned} \min \quad & f_0(x) = \sum_{i=1}^n f_i(x_i) \\ \text{s.t.} \quad & a^T x = b, \\ & a \in \mathbb{R}^n, b \in \mathbb{R} \end{aligned} \tag{2.5}$$

mit $f_i : \mathbb{R} \rightarrow \mathbb{R}$ differenzierbar und strikt konvex. Annahme: $\exists x_0 \in \text{dom } f_0 : a^T x_0 = b$.
 \Rightarrow Das Problem hat einen eindeutigen optimalen Punkt x^* .

Lagrange-Funktion:

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

$\Rightarrow \mathcal{L}(x, \nu)$ ist separabel.

Duale Funktion:

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

mit der zu f_i konjugierten Funktion f_i^* .

Duales Problem:

$$\begin{aligned} \max \quad & -b\nu - \sum_{i=1}^n f_i^*(-\nu a_i) \\ \text{s.t.} \quad & \nu \in \mathbb{R}. \end{aligned}$$

Angenommen, man hat nun eine dual optimale Lösung ν^* gefunden: Da die f_i strikt konvex sind, ist auch $\mathcal{L}(x, \nu^*)$ strikt konvex in x und hat somit eine eindeutige minimierende Lösung \tilde{x} . Zusätzlich ist bekannt, dass $x^* \mathcal{L}(x, \nu^*)$ über x minimiert. $\Rightarrow \tilde{x} = x^*$. Man erhält x^* durch $\nabla_x \mathcal{L}(x, \nu^*) = 0$, das heißt

$$f'_i(x_i^*) = -\nu^* a_i, \quad i = 1, \dots, n$$

ist zu lösen.