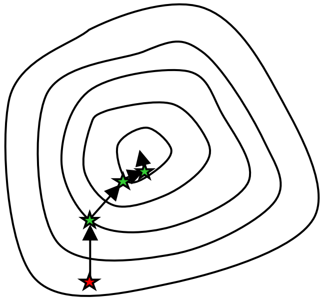
Optimierung

Vorlesung 4

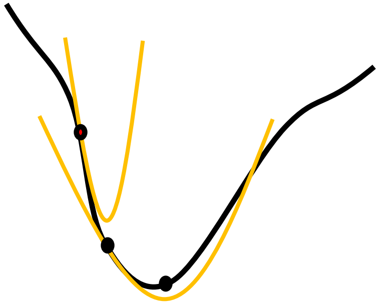
Optimierung mit Nebenbedingungen

In den bisherigen Vorlesungen



Gradientenabstieg

(allgemeines Verfahren für kontinuierliche Probleme)



Newton-Verfahren

(schnelle Konvergenz durch Ausnutzen der Krümmung)

$$d^k = -Q^k \nabla f(x^k)$$

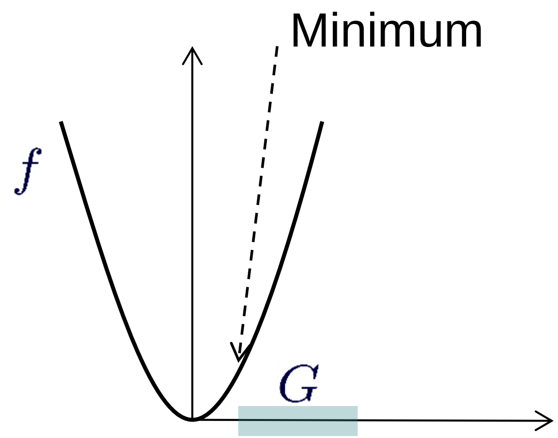
Quasi-Newton-Verfahren (BFGS)

(effiziente Approximation der Krümmung)

Nebenbedingungen

- Bisher war der Suchraum für ein Optimum der gesamte Raum $\mathbb{R}^n$
- In vielen Fällen ist der Suchraum $G \subset \mathbb{R}^n$ eingeschränkt durch Nebenbedingungen, z.B.

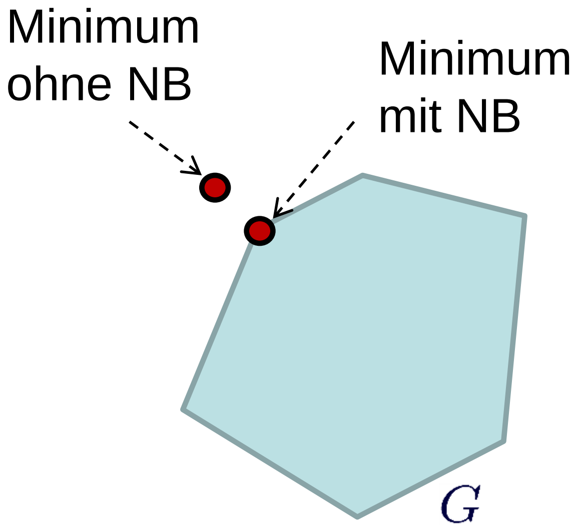
$$x \leq 5$$
$$x \geq 2$$



- Die Lösung des Problems ohne Nebenbedingung ist immer eine **untere Schranke** für das Problem mit Nebenbedingung:

$$f(x \in \mathbb{R}^n) \leq f(x \in G)$$

- Oft (aber nicht immer) liegt das Optimum auf dem Rand von G.

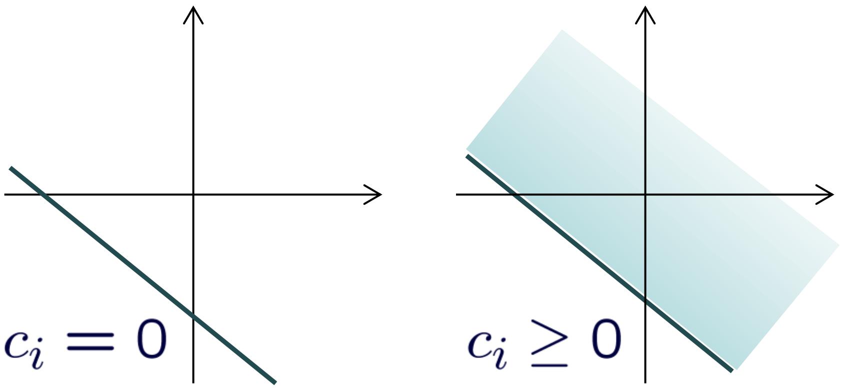


Optimierung mit Nebenbedingungen

- Jedes Optimierungsproblem lässt sich allgemein schreiben als:

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \begin{cases} c_i(x) = 0, \; i \in \mathcal{E} \\ c_i(x) \geq 0, \; i \in \mathcal{I} \end{cases}$$

- Zwei Arten von Bedingungen:
  - Gleichheitsbedingungen
  - Ungleichheitsbedingungen



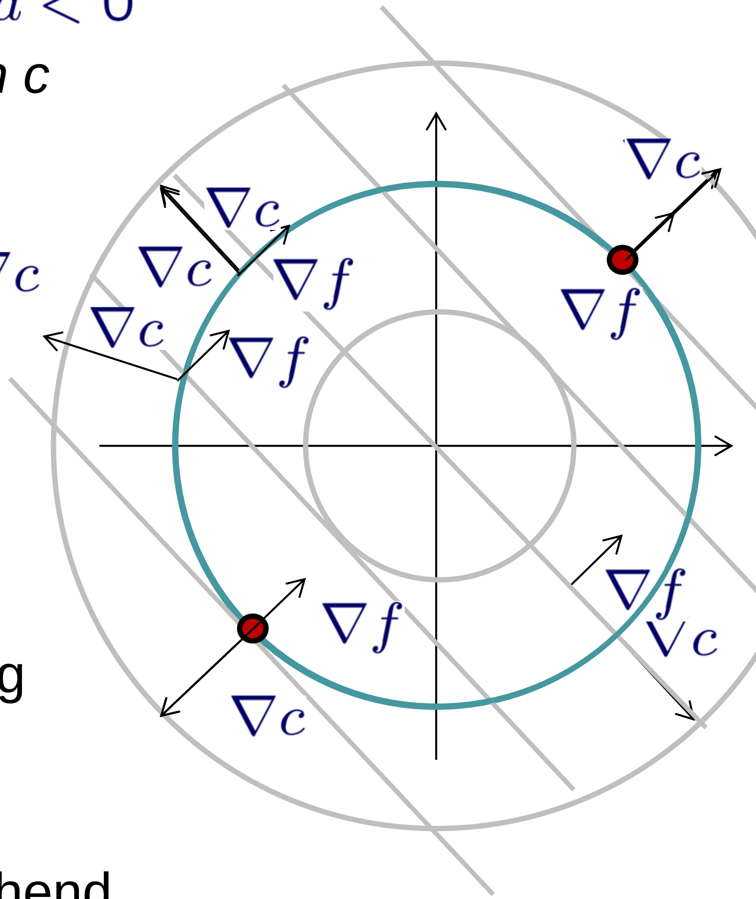
- Jede Bedingung lässt sich in diese Form bringen, z.B.

$$5x_1 + 3x_2 \leq 2 \quad \rightarrow \quad 2 - 5x_1 - 3x_2 \geq 0$$

Beispiel mit einer Gleichheitsbedingung

- Das Problem: $f(x) = x_1 + x_2 \qquad c(x) = x_1^2 + x_2^2 - 2 = 0$
- Gradienten von ∇f und ∇c
- Schritt d erlaubt falls $\nabla c^T d = 0$ und $\nabla f^T d < 0$

  also Abstieg in f und keine Änderung in c



- Im Minimum sind die Gradienten ∇f und ∇c offenbar parallel, d.h.

  $$\nabla f(x) = \lambda \nabla c(x)$$

  für einen bestimmten skalaren Wert

- In der Tat ist dies eine <u>notwendige</u> Bedingung für ein Minimum
- Offensichtlich ist die Bedingung nicht <u>hinreichend</u>, denn sie ist auch im Maximum erfüllt

Lagrange Funktion

- Wir können die notwendige Bedingung für ein Minimum mithilfe der **Lagrange Funktion** ausdrücken

$$\mathcal{L}(x,\lambda) = f(x) - \lambda c(x)$$

- Die Bedingung lautet dann

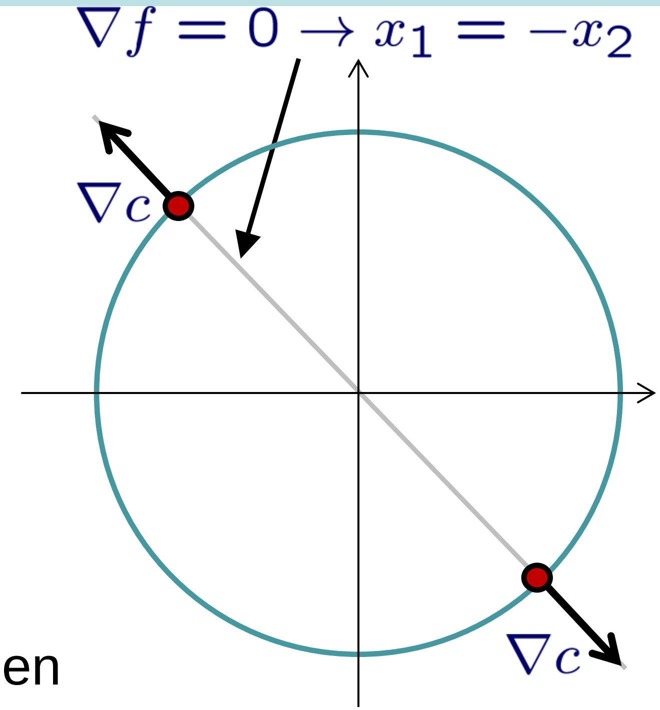
$$\nabla_x \mathcal{L}(x,\lambda) = \nabla f(x) - \lambda \nabla c(x) = 0$$

für ein geeignete Wahl von λ

- Man bezeichnet λ auch als **Lagrange Multiplikator**

- In manchen Fällen nimmt dieser auch den Wert 0 an.
Was bedeutet das?

Nebenbedingung ohne Relevanz



- Betrachten wir das Problem

$$f(x) = (x_1 + x_2)^2$$

$$c(x) = x_1^2 + x_2^2 - 2 = 0$$

- Wir haben zwei (gleich gute) Minima.

- In beiden Fällen ist $\nabla f = 0$, d.h. unabhängig von c kann f lokal nicht weiter minimiert werden

- Die notwendige Bedingung

$$\nabla_x \mathcal{L}(x, \lambda) = \nabla f(x) - \lambda \nabla c(x) = 0$$

wird für $\lambda = 0$ erfüllt.

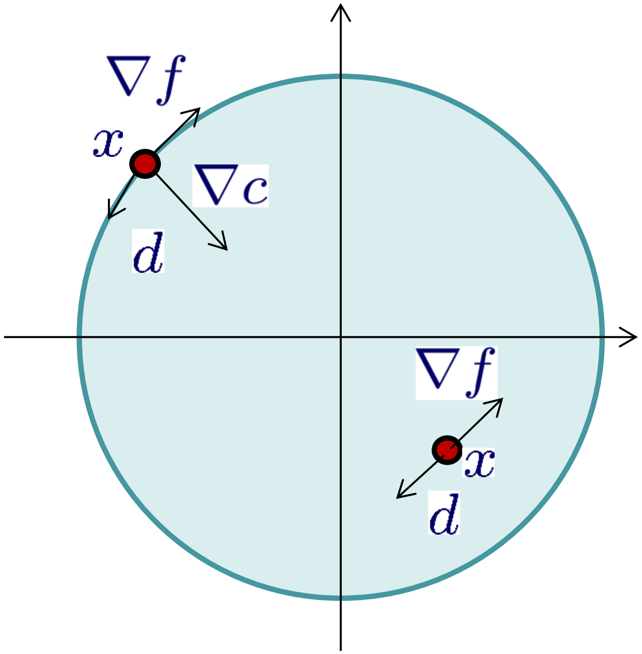
Eine Ungleichheitsbedingung

- Das Problem:

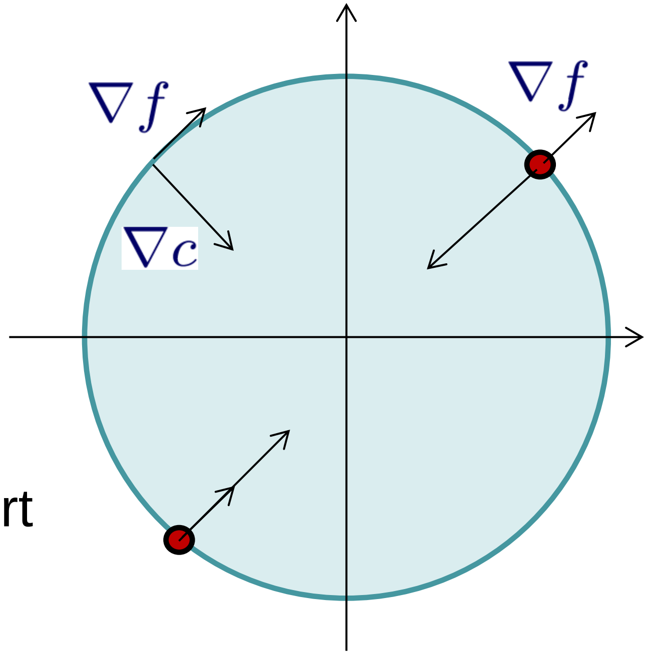
$$f(x) = x_1 + x_2$$

$$c(x) = 2 - x_1^2 - x_2^2 \geq 0$$

- Wir betrachten eine Punkt x und fragen uns wann ein kleiner Schritt d zulässig ist.



Ungleichheitsbedingung I: Minimum am Rand



- Das Problem:

$$f(x) = x_1 + x_2$$

$$c(x) = 2 - x_1^2 - x_2^2 \geq 0$$

- Minimum am Rand: wir wollen in Richtung $-\nabla f$ aber ein weiterer Abstieg wird durch c verhindert

$$c(x) = 0$$

- Die Gradienten ∇f und ∇c zeigen im Minimum notwendigerweise in die gleiche Richtung

$$\nabla f(x) = \lambda \nabla c(x), \quad \lambda \geq 0$$

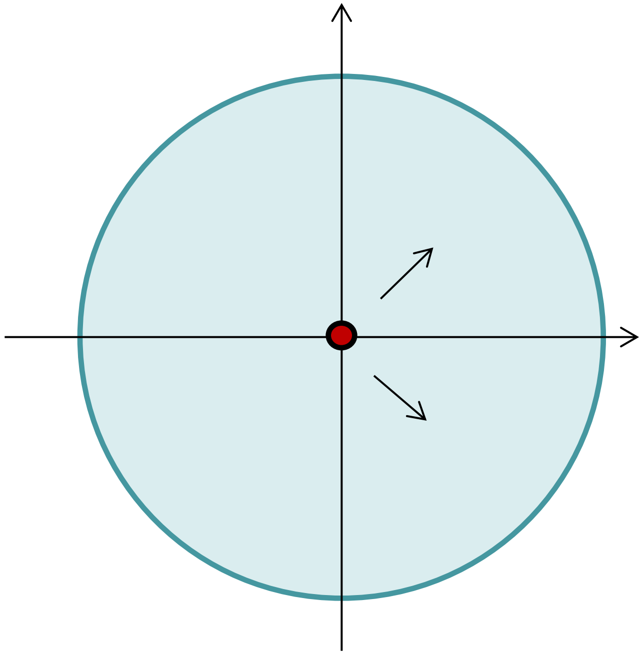
- Das Vorzeichen von λ spielt diesmal eine Rolle. Würden ∇f und ∇c in entgegengesetzte Richtung zeigen, könnte man f in Richtung von ∇c weiter minimieren.

Beispiel mit einer Ungleichheitsbedingung

- Neues Problem:

$$f(x) = x_1^2 + x_2^2$$

$$c(x) = 2 - x_1^2 - x_2^2 \geq 0$$



- Das Minimum liegt nun im Inneren von G, d.h. c hat keinen Einfluss und $\nabla f(x) = 0$

- Die Bedingung

$$\nabla f(x) = \lambda \nabla c(x), \quad \lambda \geq 0$$

ist für $\lambda = 0$ erfüllt.

- Die Nebenbedingung ist hier im Minimum **inaktiv**.

Komplementaritätsbedingung

- Wir können allgemein wieder die Lagrange Funktion

  $$\mathcal{L}(x,\lambda) = f(x) - \lambda c(x), \quad \lambda \geq 0$$

  verwenden und erhalten die **Optimalitätsbedingung**

  $$\nabla_x \mathcal{L}(x,\lambda) = \nabla f(x) - \lambda \nabla c(x) = 0, \quad \lambda \geq 0$$

- Zusätzlich muss die **Komplementaritätsbedingung** gelten:

  $$\lambda c(x) = 0$$

- Entweder liegt das Optimum auf dem Rand (dann ist $c(x) = 0$) oder c ist im Optimum inaktiv (dann ist $\lambda = 0$) oder beides.

- Sämtliche Fälle sind also sehr elegant über die Lagrange Funktion und die Komplementaritätsbedingung abgedeckt.

- **Strenge Komplementarität** liegt vor wenn c und λ nicht gleichzeitig null sind

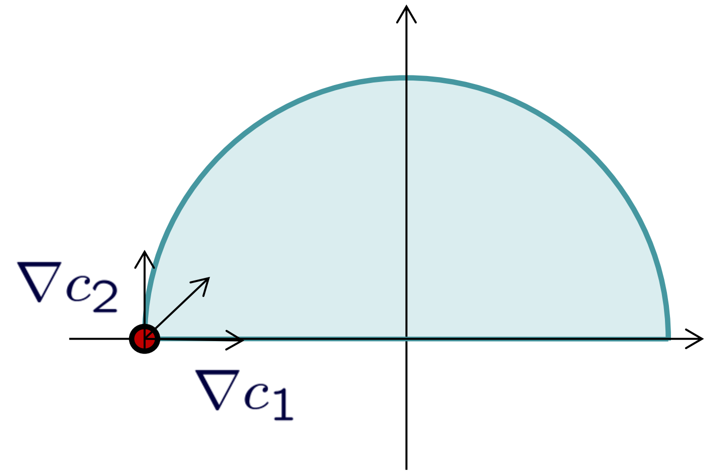
Mehrere Nebenbedingungen

- Problem mit zwei Bedingungen:

$$f(x) = x_1 + x_2$$

$$c_1(x) = 2 - x_1^2 - x_2^2 \geq 0$$

$$c_2(x) = x_2 \geq 0$$



- In diesem Fall sind beide Nebenbedingungen im Minimum aktiv

$$\nabla f(x) - \lambda_1 \nabla c_1(x) - \lambda_2 \nabla c_2(x) = 0, \quad \lambda_1 \geq 0, \lambda_2 \geq 0$$

d.h. $c_1(x) = c_2(x) = 0, \quad \lambda_1 > 0, \lambda_2 > 0$

- Die Menge der aktiven Nebenbedingungen wird als die **aktive Menge** (active set) bezeichnet

Redundante Nebenbedingungen

- Einige Nebenbedingungen können überflüssig sein, da sie bereits von anderen Nebenbedingungen abgedeckt sind, z.B.:

$$c_1(x) = x_2 \geq 0$$
$$c_2(x) = x_2 \geq 2$$

- c_2 deckt automatisch auch c_1 ab. c_1 ist daher nie aktiv.

- Da die Komplexität von der Anzahl der Nebenbedingungen abhängt, ist es meist ratsam redundante Bedingungen vorab zu eliminieren.

Karush-Kuhn-Tucker Bedingungen

- Zusammenfassend sind die notwendigen Bedingungen für ein lokales Optimum:

$$
\boxed{
\begin{aligned}
&\nabla_x \mathcal{L}(x, \lambda) = 0 \\
c_i(x) &= 0, \quad \forall i \in \mathcal{E} \\
c_i(x) &\geq 0, \quad \forall i \in \mathcal{I} \\
\lambda_i &\geq 0, \quad \forall i \in \mathcal{I} \\
\lambda_i c_i(x) &= 0, \quad \forall i \in \cancel{\mathcal{E}} \cup \mathcal{I}
\end{aligned}
}
$$

- Diese werden als Karush-Kuhn-Tucker (KKT) Bedingungen bezeichnet.

- Im folgenden werden wir alle Bedingungen $c_i(x)$ in einen Funktionsvektor $c(x) := (c_1(x), \ldots, c_m(x))^\top$ und entsprechend $\lambda \in \mathbb{R}^m$ zusammenfassen.

Dualität

- Durch die Nebenbedingungen kamen neben den primalen Variablen x die Lagrange Multiplikatoren λ als Variablen ins Spiel

$$\mathcal{L}(x,\lambda) = f(x) - \lambda^\top c(x), \quad \lambda \geq 0$$

- Sie geben an wie stark die Funktion gegen die Bedingungen drückt.

$$\nabla f(x) = \lambda^\top \nabla c(x), \quad \lambda \geq 0$$

- Alternativ können wir $\mathcal{L}(x,\lambda)$ als zu minimierende Funktion <u>ohne</u> Nebenbedingungen betrachten, indem wir λ festhalten

$$q(\lambda) = \inf_x \mathcal{L}(x,\lambda)$$

- In vielen Fällen erhält man $-\infty$ für einige Werte von λ. Daher schränken wir die Definitionsmenge von $q(\lambda)$ entsprechend ein:

$$\mathcal{D} = \{\lambda | q(\lambda) > -\infty\}$$

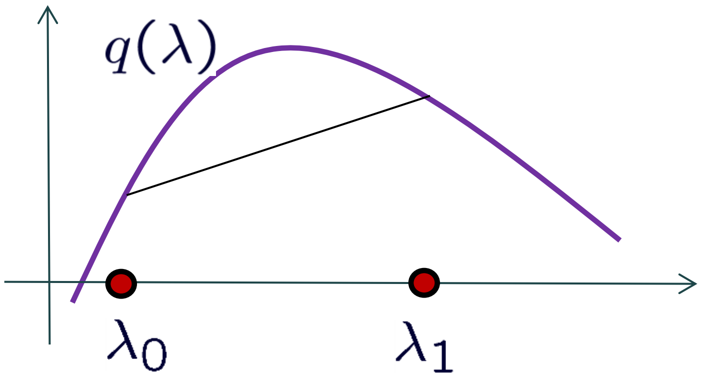
- **Anmerkung**: Wir betrachten nur Ungleichheitsbed.

Dualität (Konkav)

- Die Funktion

  $$q(\lambda) = \inf_x \mathcal{L}(x, \lambda)$$

  ist konkav.



- Beweis Schritt 1: Wir zeigen

$$\mathcal{L}(x, (1-\alpha)\lambda_0 + \alpha\lambda_1) = (1-\alpha)\mathcal{L}(x, \lambda_0) + \alpha\mathcal{L}(x, \lambda_1)$$

- Teilbeweis aus def. Lagrange:

$$\begin{aligned}\mathcal{L}(x, (1-\alpha)\lambda_0 + \alpha\lambda_1) &= f(x) - (1-\alpha)\lambda_0 c(x) - \alpha\lambda_1 c(x)\\ &= (1-\alpha)f(x) + \alpha f(x) - (1-\alpha)\lambda_0 c(x) - \alpha\lambda_1 c(x)\\ &= (1-\alpha)\mathcal{L}(x, \lambda) + \alpha\mathcal{L}(x, \lambda)\end{aligned}$$

Dualität (Konkav, cont.)

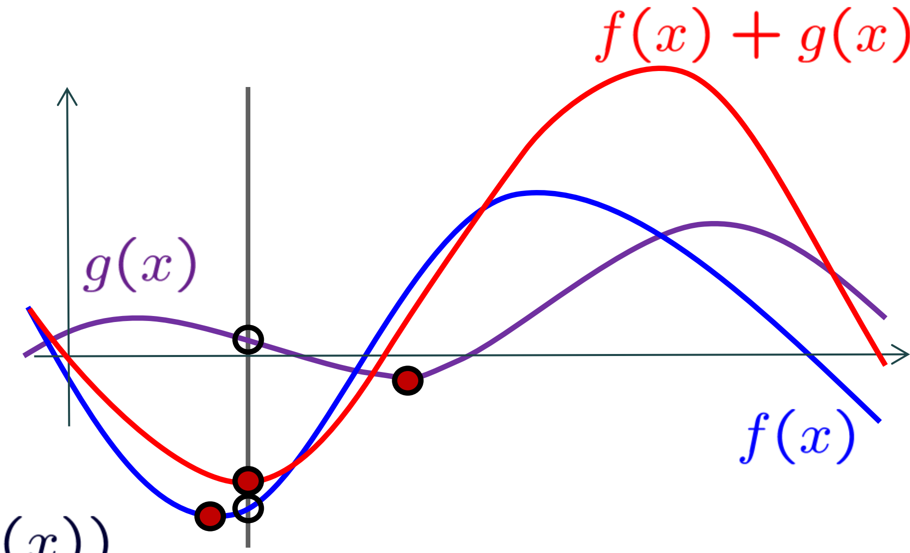
- Damit haben wir:

$$\mathcal{L}(x,(1-\alpha)\lambda_0+\alpha\lambda_1) = (1-\alpha)\mathcal{L}(x,\lambda_0)+\alpha\mathcal{L}(x,\lambda_1)$$

- Beweis (fort.)
- Das Infimum einer Summe ist größer gleich der Summe von Infima, daher:

$$q((1-\alpha)\lambda_0+\alpha\lambda_1) \geq (1-\alpha)q(\lambda_0)+\alpha q(\lambda_1)$$

- Damit ist Konkavität bewiesen (siehe Abb. oben)



$$\inf f(x) + \inf g(x) \geq \inf(f(x)+g(x))$$

Duales Problem (Lagrange Dualität)

- Aus dem primalen Optimierungsproblem

$$\min_x f(x), \quad c(x) \geq 0$$

können wir also ein **duales Optimierungsproblem** ableiten

$$\max_\lambda \inf_x \mathcal{L}(x, \lambda), \quad \lambda \geq 0$$

- In manchen Fällen lässt sich das duale Problem leichter lösen als das primale Problem.
- Sogenannte Primal-Dual-Verfahren optimieren gleichzeitig das primale und das duale Problem.
- Es gibt noch andere Formen der Dualität, z.B. die Fenchel Dualität.

Schwache und starke Dualität

- **Schwache Dualität:**
  Für alle gültigen Lösungen x und λ gilt $q(\lambda) \leq f(x)$

  Beweis: $q(\lambda) = \inf_x f(x) - \lambda^\top c(x) \leq f(x) - \underbrace{\lambda^\top}_{\geq 0} c(x) \leq f(x)$

  d.h. das duale Problem liefert immer eine untere Schranke für das primale Problem

- **Starke Dualität:**
  - $f(x)$ sei konvex und die gültige Menge sei eine konvexe Menge
  - $\hat{\lambda}$ bezeichne das Optimum von $q(\lambda)$ mit $\hat{x} = \mathrm{arginf}_x \mathcal{L}(x, \hat{\lambda})$
  - $\mathcal{L}(x, \hat{\lambda})$ sei streng konvex

  Dann gilt:

  $$q(\hat{\lambda}) = \mathcal{L}(\hat{x}, \hat{\lambda}) = f(\hat{x})$$

  d.h. das Optimum des dualen Problems minimiert auch das primale Problem.

Beispiel

- Konvexes Beispielproblem:

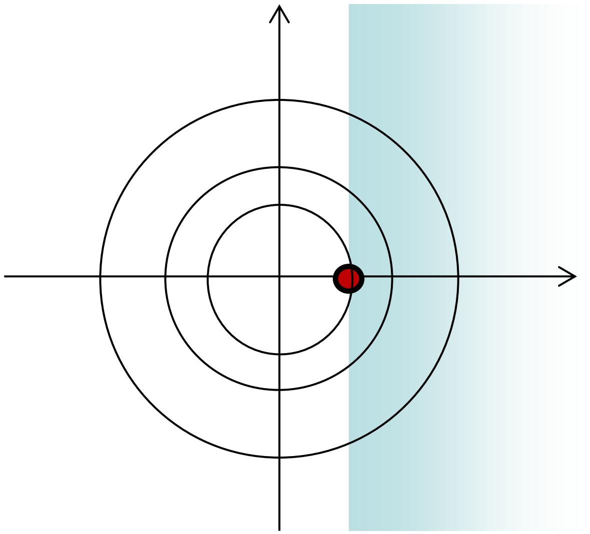
$$\min_{x_1,x_2} \frac{1}{2}(x_1^2 + x_2^2), \qquad x_1 - 1 \geq 0$$

- Lagrange-Funktion:

$$\mathcal{L}(x_1, x_2, \lambda) = \frac{1}{2}(x_1^2 + x_2^2) - \lambda(x_1 - 1)$$

- Bedingungen für ein Minimum:

$$x_1 - \lambda = 0, \qquad x_2 = 0,$$



- Damit lässt sich x_1, x_2 in der Lagrange-Funktion eliminieren (geht nicht immer analytisch). Es bleibt das duale Problem

$$\max q(\lambda) = \max -\frac{1}{2}\lambda^2 + \lambda, \qquad \lambda \geq 0$$

mit der Lösung $\lambda = 1$

- Lösung des primalen Problems (starke Dualität): $x_1 = \lambda = 1, \quad x_2 = 0$

Wichtige Probleme mit starker Dualität

- **Lineare Programme**

$$\min_x w^\top x, \quad Ax - b \geq 0$$

Duales Problem:

$$\max_\lambda b^\top \lambda, \quad A^\top \lambda = w, \ \lambda \geq 0$$

- **Quadratische Programme**

$$\min_x \frac{1}{2} x^\top Q x + w^\top x, \quad Ax - b \geq 0$$

mit Q symmetrisch und positiv definit

Duales Problem:

$$\max_\lambda -\frac{1}{2}(A^\top \lambda - w)^\top Q^{-1}(A^\top \lambda - w) + b^\top \lambda, \quad \lambda \geq 0$$

Zusammenfassung

- Optimalitätsbedingungen für Probleme mit Nebenbedingungen unterscheiden ob eine Bedingung aktiv oder inaktiv ist.
- Beide Fälle lassen sich mithilfe der Lagrange-Funktion ausdrücken.
- Über die Lagrange-Funktion lässt sich auch ein duales Problem herleiten.
- Die Lösung des dualen Problems bildet mindestens eine untere Schranke für das ursprüngliche Problem (schwache Dualität).
- Für konvexe Funktionen mit konvexen Nebenbedingungen führt die Lösung des dualen Problems zur exakten Lösung des ursprünglichen Problems (starke Dualität).

Übungsaufgabe

- Optimierungsproblem:

$$\min_{x} (x_1 - 3)^2 + (x_2 - 1)^2$$

$$1 - x_1 - x_2 \geq 0 \qquad 1 + x_1 - x_2 \geq 0$$
$$1 - x_1 + x_2 \geq 0 \qquad 1 + x_1 + x_2 \geq 0$$

- Machen Sie eine Skizze des erlaubten Suchraums.
- Berechnen Sie für verschiedene Punkte des Suchraums den Gradienten der Funktion und tragen Sie ihn in die Skizze ein.
- Wo befindet sich das Minimum?
- Wie sehen die Lagrange Multiplikatoren im Optimum aus? Welche Nebenbedingungen sind im Optimum aktiv, welche nicht?