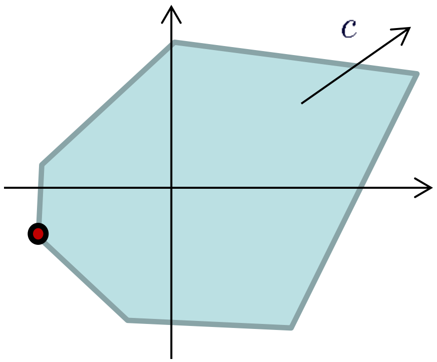
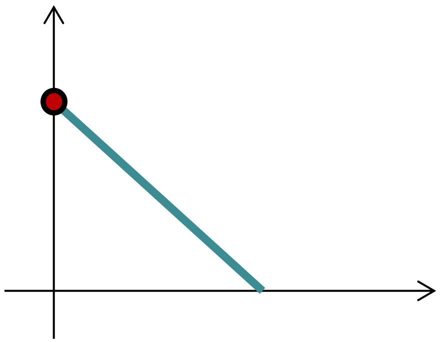
Optimierung

Vorlesung 6
Quadratische Programmierung

In der letzten Vorlesung



Lineare Programmierung



Simplex-Algorithmus

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \end{pmatrix}$$

Active-Set-Verfahren

Quadratische Programme

- Lineares Programm:

$$\min_x c^\top x, \qquad Ax - b \geq 0$$

- Quadratisches Programm:

$$\min_x \frac{1}{2} x^\top Q x + c^\top x$$

$$a_i^\top x = b_i, \qquad i \in \mathcal{E}$$

$$a_i^\top x \geq b_i, \qquad i \in \mathcal{I}$$

→ quadratische Zielfunktion $f(x)$ mit linearen Nebenbedingungen

- Die $n \times n$ Matrix Q ist symmetrisch und offensichtlich die Hesse-Matrix von $f(x)$

Konvexe vs. nicht-konvexe quadratische Programme

$$\min_x \frac{1}{2} x^\top Q x + c^\top x$$

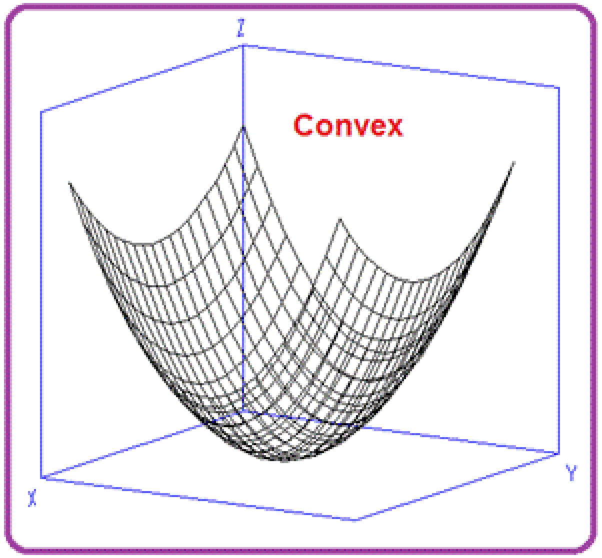
- Wenn Q <u>positiv semi-definit</u> ist, ist das quadratische Programm konvex.

  Jedes x, das die KKT Bedingungen erfüllt, ist ein globales Minimum.

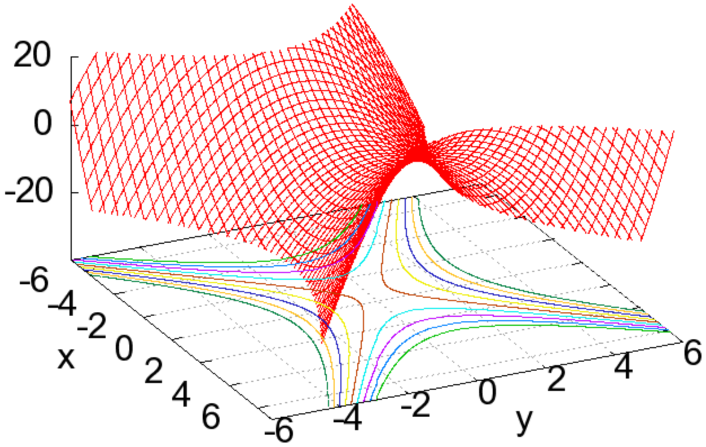
- Ist Q <u>positiv definit</u>, ist das globale Minimum sogar eindeutig.

- Wenn Q <u>indefinit</u> ist, ist das quadratische Programm nicht konvex.

  Es gibt stationäre Punkte, welche die KKT Bedingungen erfüllen, aber kein Minimum sind.



$f(x,y) = x^2 - 2y^2 + xy + x + y + 1$



Quelle: Google Images

Anwendungsbeispiel: Support Vector Machine

- Gegeben: Datenpunkte $a_i \in \mathbb{R}^n$ mit Label $t_i \in \{-1, 1\}$

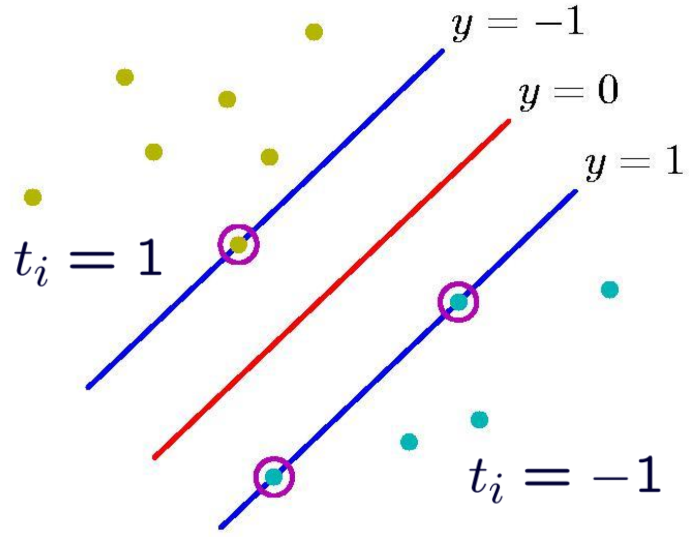
- Ziel: Finden der einfachsten Hyperebene

  $$y = a^\top x + b$$

  so dass alle Datenpunkte auf der richtigen Seite der Hyperebene liegen

  $$\mathrm{sign}(y(a_i)) = t_i$$

  und einen Mindestabstand von 1 zur Hyperebene haben.



Quelle: Christopher Bishop

- Quadratisches Programm:

$\min\limits_x x^\top x$	(Einfachheit der Hyperebene)
$t_i(a_i^\top x + b) \geq 1, \quad \forall i$	(Vorzeichen und Mindestabstand zur Hyperebene)

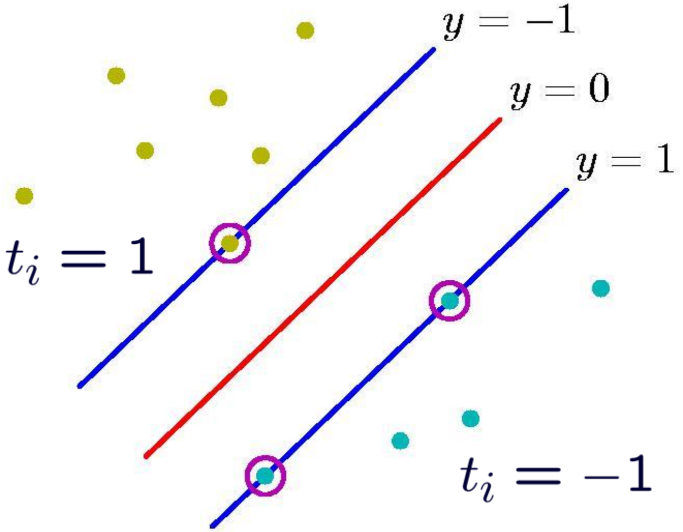
Anwendungsbeispiel: Support Vector Machine

- Das Optimierungsproblem ist hier offensichtlich konvex

$$\min_x x^\top x \qquad\qquad t_i(a_i^\top x + b) \geq 1, \quad \forall i$$

da die Hesse Matrix Q die Einheitsmatrix ist.

→ eindeutiges globales Minimum



Quelle: Christopher Bishop

- Die Nebenbedingungen sind meist redundant
→ Reduktion der Nebenbedingungen

- In der Praxis:
  – Millionen Datenpunkte, Reduktion relativ teuer
  – Duales Problem hat spezielle Form mit sehr einfachen Nebenbedingungen

  Daher einfaches iteratives Verfahren (Projected Coordinate Descent) effizienter als das folgende (allgemeinere) Verfahren zur quadratischen Programmierung
  → Projektionsmethoden in der nächsten Vorlesung

Quadratische Programme mit Gleichheitsbedingungen

- In dieser Vorlesung nur streng konvexe quadratische Programme

- Außerdem betrachten wir zunächst nur Probleme mit Gleichheitsbedingungen:

$$\min_x \frac{1}{2} x^\top Q x + c^\top x, \qquad Ax = b$$

Annahme: A hat vollen Rang (Reduktion der Nebenbedingungen)

- Lösung hilft uns beim Lösen des (schwierigeren) Problems mit Ungleichheitsbedingungen.

- Optimalitätsbedingungen (KKT Bedingungen):

$$Qx - A^\top \lambda = -c$$

$$Ax = b$$

Lösen des KKT Systems

- KKT System in Matrixform:

$$\begin{pmatrix} Q & -A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix}$$

- Da Q positiv definit ist, müssen wir nur dieses Gleichungssystem lösen, z.B. effizient mit der **Schur-Komplement-Methode**:
  - Multipliziere erste Gleichung mit AQ^{-1}

$$A\cancel{Q^{-1}Q}x - AQ^{-1}A^\top\lambda = -AQ^{-1}c$$

  - Ziehe zweite Gleichung von erster Gleichung ab

$$-AQ^{-1}A^\top\lambda = -AQ^{-1}c - b$$

  - Löse Gleichungssystem in λ

$$AQ^{-1}A^\top\lambda = AQ^{-1}c + b$$

  - Löse Gleichungssystem in x

$$Qx = A^\top\lambda - c$$

Anmerkungen

- Es gibt verschiedene Ansätze ein quadratisches Programm mit Gleichheitsbedingungen zu lösen, die je nach Problemstellung effizienter oder allgemeiner einsetzbar sind als die Schur-Komplement-Methode.

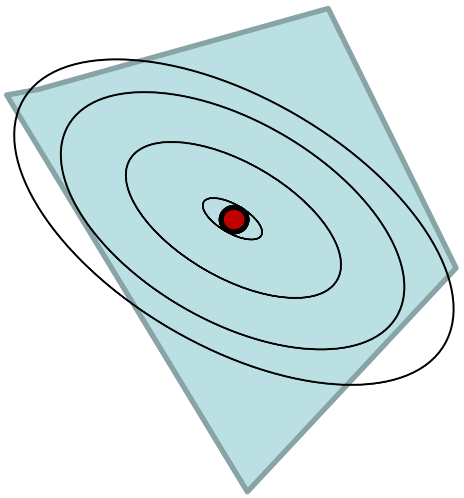
  Beispiele:
  - **Symmetrisch Indefinite Faktorisierung**: eine spezielle Faktorisierungsmethode, die auch mit indefiniten Matrizen umgehen kann
  - Eliminierung von Variablen anhand der Gleichheitsbedingungen und Optimierung (ohne Nebenbedingungen) im verbleibenden Nullraum (**Null-Space-Methode**)

- Das Grundproblem ist nicht nur für die nun folgenden quadratischen Programme mit Ungleichheitsbedingungen wichtig sondern auch für allgemeine nichtlineare Programme.

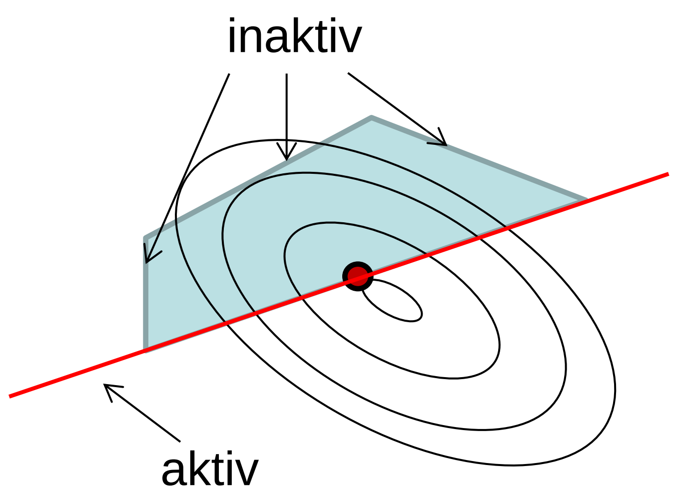
  → Sequential Quadratic Programming in der nächsten Vorlesung

Quadratische Programme mit Ungleichheitsbedingungen

$$\min_x \frac{1}{2} x^\top Q x + c^\top x, \qquad Ax \geq b$$



- Wie bei der linearen Programmierung können wir ein Active-Set-Verfahren verwenden.
- Unterschied: Die optimale Lösung muss nicht auf dem Rand der gültigen Menge liegen
- Beobachtung: Wüssten wir welche Nebenbedingungen im Minimum aktiv sind, hätten wir wieder Gleichheitsbedingungen.
- Die Schwierigkeit liegt also darin, die Menge der aktiven Nebenbedingungen zu finden.



Active Set

- DEFINITION. Geg. Eine Menge von Nebenbedingungen (NB)

$$g_1(x) \geq 0, \ldots, g_m(x) \geq 0$$

und ein Punkt x^* in der zulässigen Region. Eine NB $g_i(x) \geq 0$ wird **aktiv** genannt falls $g_i(x^*) = 0$ ist, ansonsten ist sie **inaktiv.**

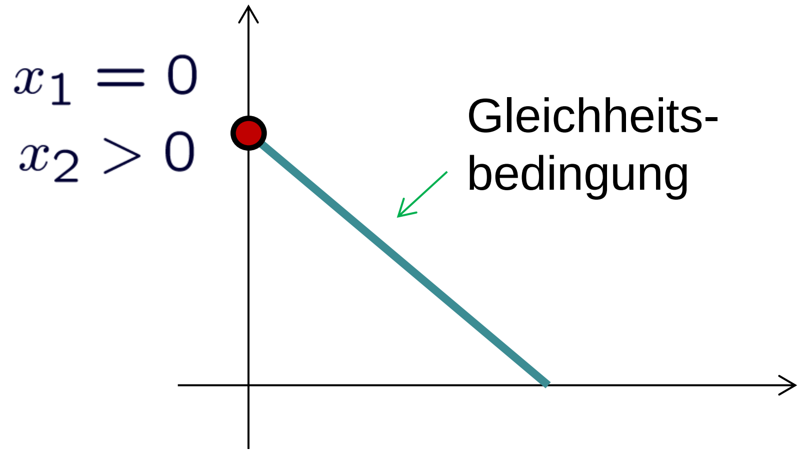
- BEMERKUNG: Gleichheitsbedingungen sind immer aktiv.

- Bsp Simplex-Verfahren letzte Stunde: $\min_x c^\top x, \quad Ax = b, \ Ex \geq 0$ wobei E die Einheitsmatrix ist. Die aktive Menge $\mathcal{N}$ bestand aus den Indizes $i \in \mathcal{N}$ für die $x_i = 0$ gilt. Damit sind das die i-ten Zeilen von $Ex \geq 0$ (entspricht i-ter Nebenbedingung).

Unterschied zwischen linearem und quadratischem Programm

- Mithilfe der Standardform erhielten wir beim linearen Programm Gleichheitsbedingungen

$$\min_x c^\top x, \quad Ax = b, \boxed{Ex \geq 0}$$



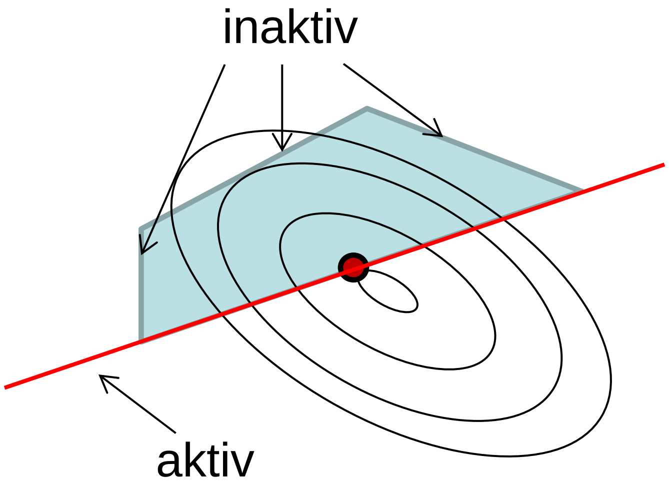
und nur einfache Ungleichheit in x

Die aktive Menge ließ sich durch Austausch der aktiven x_i (= i-te Zeile von $Ex \geq 0$) steuern.

- Das quadratische Programm

$$\min_x \frac{1}{2} x^\top Q x + c^\top x, \quad \boxed{Ax \geq b}$$

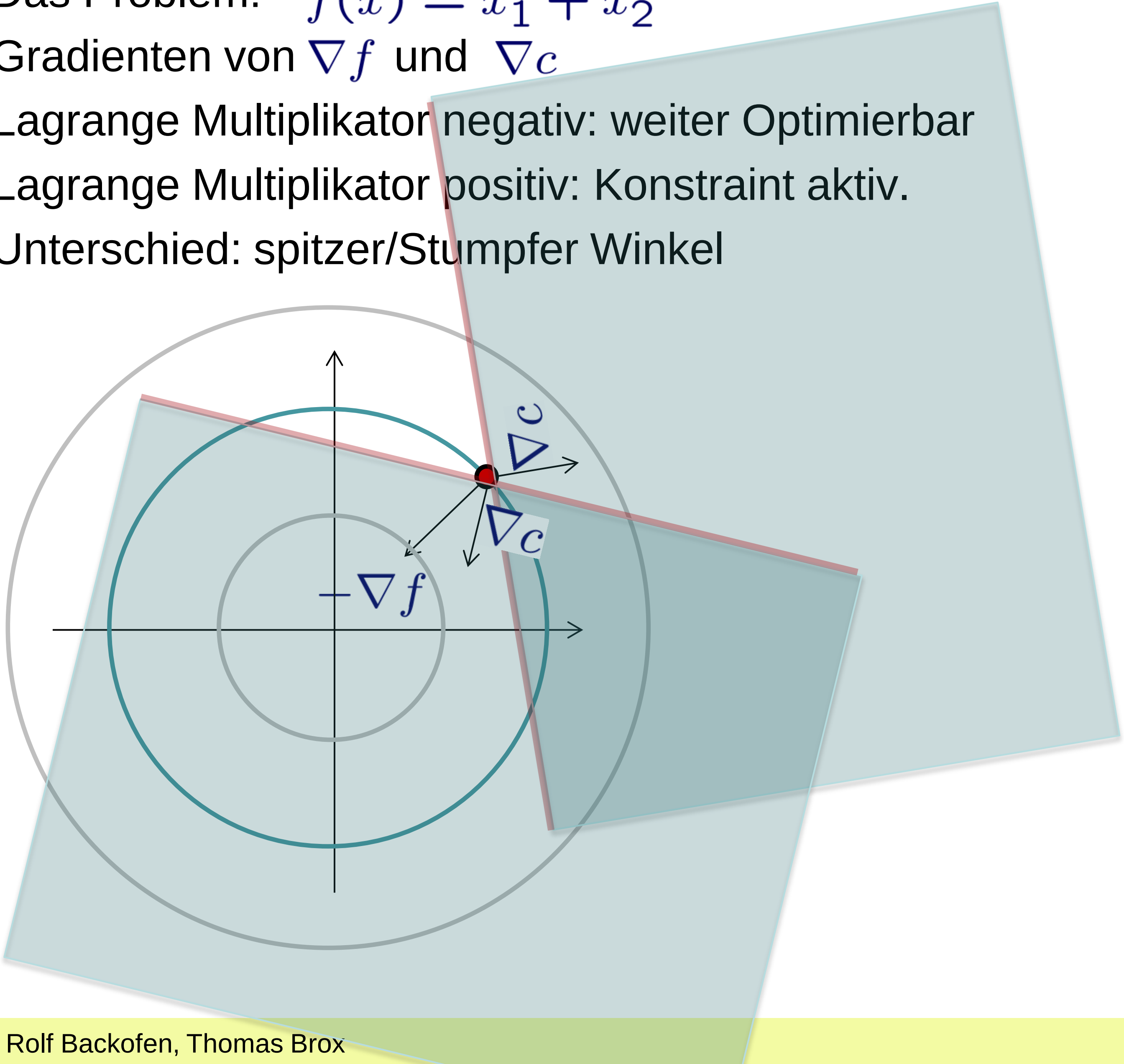
lässt sich nicht in eine solche Form bringen.



Unbekannt wie viele Bedingungen aktiv sind (Lösung muss keine Ecke sein)

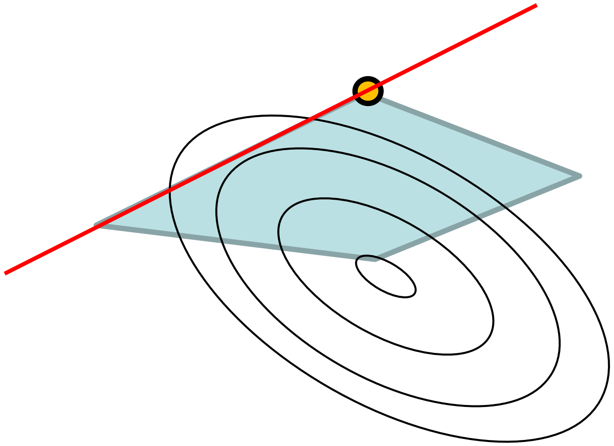
Negative Lagrange-Multiplikatoren

- Das Problem: $f(x) = x_1^2 + x_2^2$
- Gradienten von ∇f und ∇c
- Lagrange Multiplikator negativ: weiter Optimierbar
- Lagrange Multiplikator positiv: Konstraint aktiv.
- Unterschied: spitzer/Stumpfer Winkel



Arbeitsmenge und iteratives Verfahren

- Um die Ungleichheitsbedingungen in Gleichheitsbedingungen zu konvertiere führen wir eine **Arbeitsmenge** $\mathcal{W}$ ein.



- In ihr können nur Bedingungen enthalten sein, die in einem aktuellen Punkt x^k aktiv sind.
- Es müssen jedoch nicht alle aktiven Bedingungen enthalten sein.
- Die Arbeitsmenge kann auch leer sein.
- Die Bedingungen in $\mathcal{W}$ müssen linear unabhängig sein.
- Die Bedingungen in der Arbeitsmenge werden in x^k zu Gleichheitsbedingungen -> entspricht Zeilen von $Ax \geq b$ für die gilt:

  $a_i^T x = b_i$ (a_i ist die i-te Zeile von A)

Iteratives Verfahren

- Wir können nun einen Schritt d machen, der $f(x)$ unter den Gleichheitsbedingungen in $\mathcal{W}$ minimiert, d.h. wir minimieren

$$\min_x \frac{1}{2} x^\top Q x + c^\top x, \quad a_i^T x = b_i, \ \forall i \in \mathcal{W}$$

Wir wissen bereits, wie wir dieses Problem lösen können (z.B. mit der Schur-Komplement-Method)

- Das Minimierungsproblem kann auch in d ausgedrückt werden:

$$\min_d \frac{1}{2} d^\top Q d + g_k^\top d, \quad a_i^T d = 0, \ \forall i \in \mathcal{W}$$

mit

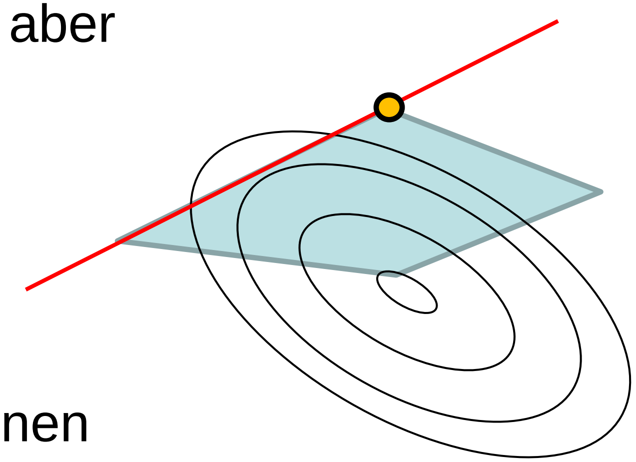
$$d = x - x^k, \quad g_k = Q x^k + c$$

- Wir können also den optimalen Schritt d finden, der $f(x)$ unter den Bedingungen in $\mathcal{W}$ minimiert.

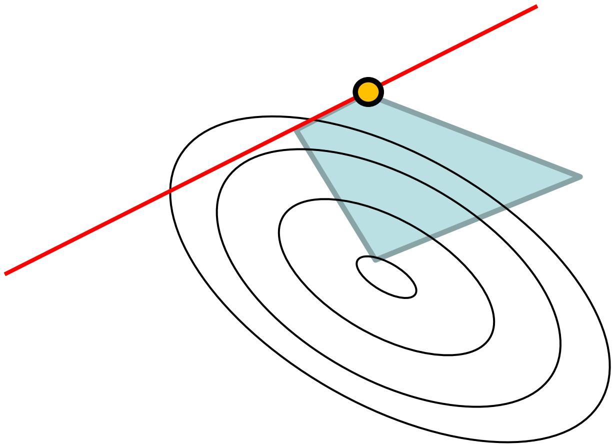
- Wäre $\mathcal{W}$ die aktive Menge im Optimum, wären wir nun fertig.

Anpassung der Arbeitsmenge

- Der Schritt d hält die Bedingungen in $\mathcal{W}$ ein, aber
  1. Diese Bedingungen sollten vielleicht nicht alle aktiv sein
  2. Es ist nicht sichergestellt, dass die anderen Bedingungen eingehalten werden



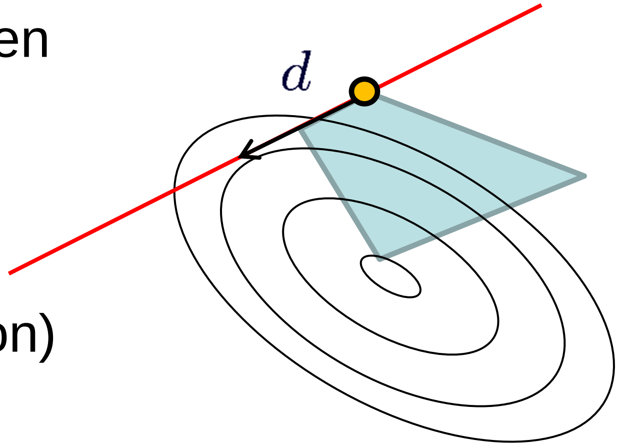
- Wir müssen also Bedingungen aus $\mathcal{W}$ entfernen (welche?)



- Außerdem müssen wir sicherstellen, dass der Schritt d nicht aus der gültigen Menge herausläuft (wie?)

- Die verletzte (**blockierende**) Bedingung wird in die Arbeitsmenge aufgenommen.

Einhaltung der Bedingungen



- Wir können unseren Schritt in der Länge verkürzen

$$x^{k+1} = x^k + \alpha d$$

- Ohne verletzte Bedingung wäre $\alpha = 1$ optimal (Newton-Verfahren, quadratische Funktion)

- Bedingungen $i \notin \mathcal{W}$ mit $a_i^\top d \geq 0$ sind unkritisch, da in dem Fall

$$a_i^\top (x^k + d) \geq a_i^\top x^k \geq b_i$$

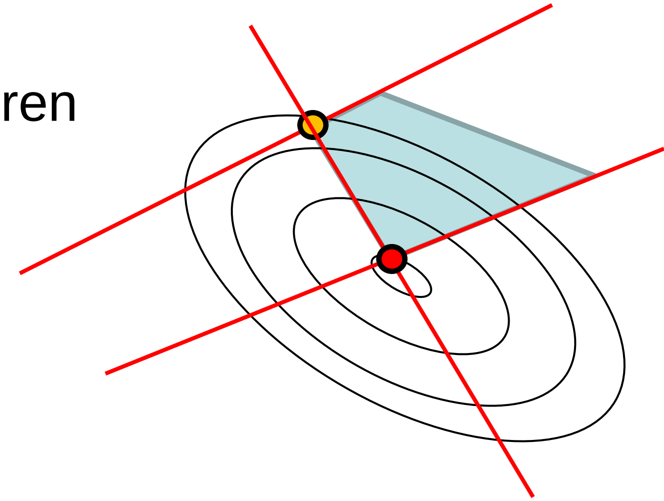
- Wir müssen also α so wählen, dass auch für alle Bedingungen $i \notin \mathcal{W}$ mit $a_i^\top d < 0$ gilt:

$$a_i^\top (x^k + \alpha d) \geq b_i$$

$$\rightarrow \alpha \leq \frac{b_i - a_i^\top x^k}{a_i^\top d} \qquad \rightarrow \alpha = \min\left(1, \min_{i \notin \mathcal{W}, a_i^\top d < 0} \left(\frac{b_i - a_i^\top x^k}{a_i^\top d}\right)\right)$$

Entfernen von Bedingungen

- In unserem Beispiel können wir offenbar $f(x)$ mit der aktuellen Menge $\mathcal{W}$ nicht weiter optimieren



- Sind die KKT-Bedingungen erfüllt?

$$Qx - A^\top \lambda = -c$$

$$Ax \geq b$$

$$\lambda^\top (Ax - b) = 0$$

$$\lambda \geq 0$$

- Falls $\lambda \geq 0$ haben wir das Minimum gefunden, ansonsten zeigen die negativen λ_i an, welche Bedingungen aus $\mathcal{W}$ entfernt werden müssen. (siehe Vorlesung 4, Folie 9 zu Lagrange-Mult. bei Ungleichungen)

- Danach lässt sich $f(x)$ weiter optimieren.

Beispiel

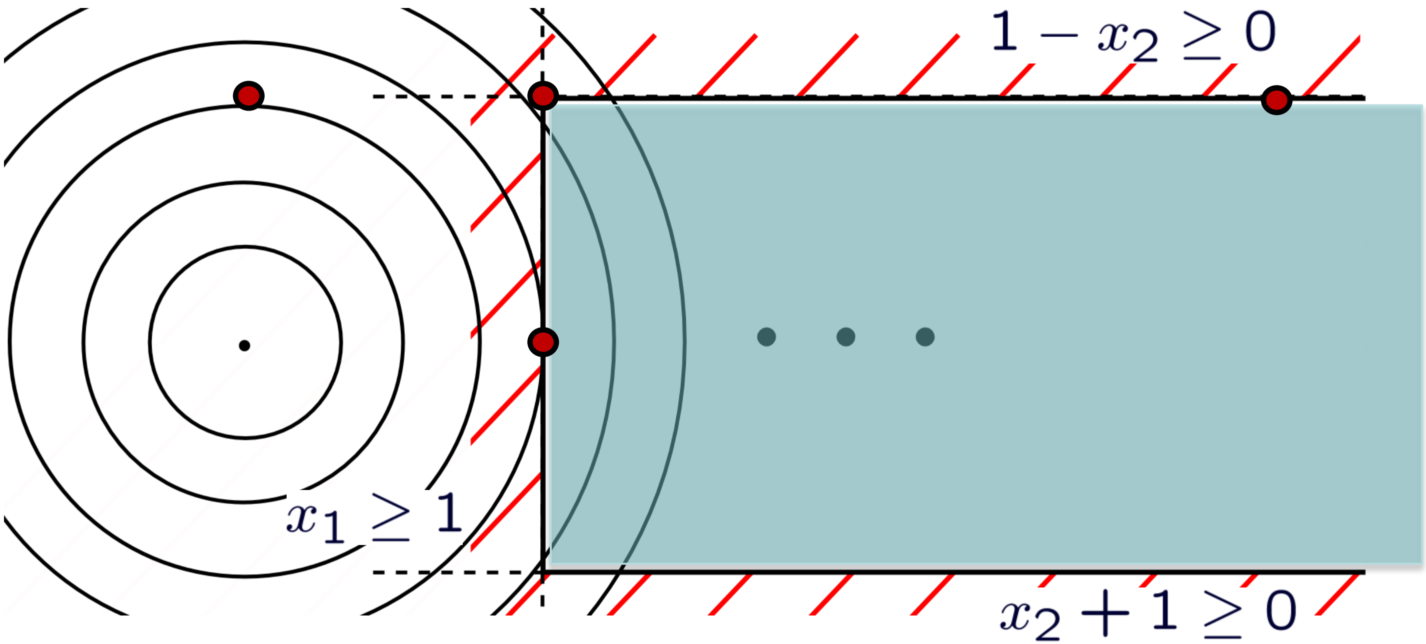
- Problem:

$$\min ||x||_2^2$$

$$\text{subject to} \quad x_1 \geq 1 \qquad (1)$$

$$x_2 + 1 \geq 0 \qquad (2)$$

$$1 - x_2 \geq 0 \qquad (3)$$



- Active Set: $\{3\} \;\to \{3, 1\} \;\to \{1\}$

Finden einer gültigen Startlösung

- Die Startlösung x^0 muss in der gültigen Menge liegen. Wie beim Simplex-Verfahren kann eine solche Lösung mit einem linearen Programm bestimmt werden (siehe Vorlesung 5):

$$\min_{x,z} e^\top z, \qquad a_i^\top x + z_i \geq b_i \; z \geq 0$$

$$e := (1, ..., 1)^\top$$

- Für beliebige Wahl von $x = x_s$ ist

  $$z_i = \max(b_i - a_i^\top x_s, 0)$$

  eine gültige Lösung

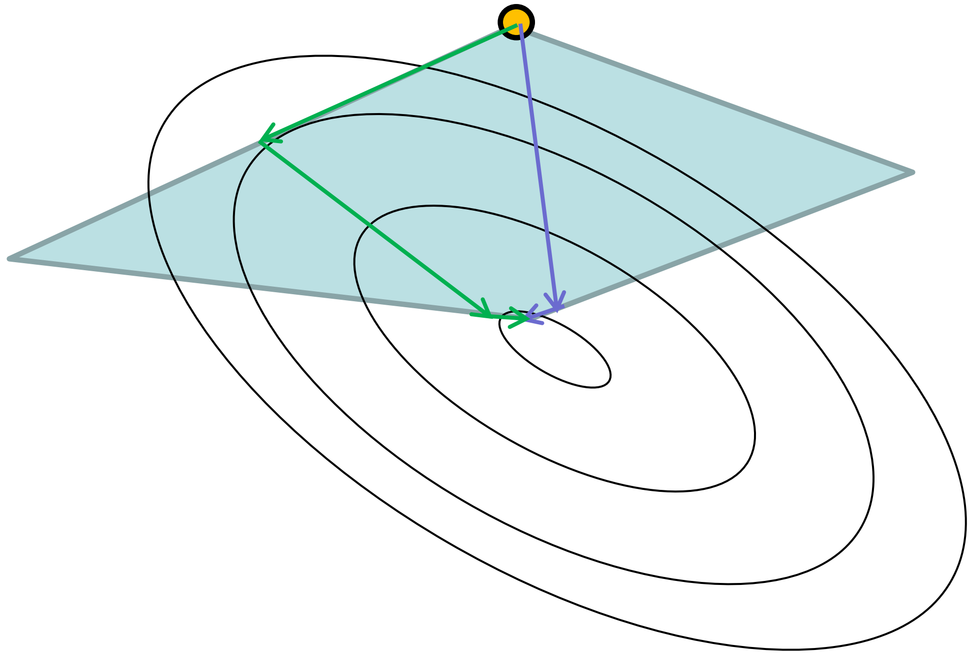
- Durch Minimierung von $e^\top z$ erhält man $z = 0$ und ein x, das in der gültigen Menge des quadratischen Programms liegt.

Wahl der ersten Arbeitsmenge

- Als erste Arbeitsmenge kann eine linear unabhängige Untermenge der aktiven Bedingungen gewählt werden.

- Bei einer großen Anzahl von Bedingungen kann auch $\mathcal{W} = \emptyset$ eine gute Wahl sein.

- Die Wahl der Arbeitsmenge beeinflusst den Weg zum Minimum.



Komplexität einer Iteration

- Dimension von x : n
  Anzahl linear unabhängiger Nebenbedingungen: $m \leq n$
  Anzahl Nebenbedingungen: M

- In jedem Schritt müssen wir ein quadratisches Programm mit Gleichheitsbedingungen (Arbeitsmenge) lösen.
  - Invertierung von Q : $\mathcal{O}(n^3)$
    (eventuell günstiger, falls Q z.B. eine Diagonalmatrix oder dünn besetzt ist)
  - Lösen des Gleichungssystems
    $$AQ^{-1}A^\top\lambda = AQ^{-1}c + b$$
    hat allgemein Komplexität: $\mathcal{O}(mn^2)$

- Außerdem muss in jedem Schritt die Gültigkeit aller Nebenbedingungen überprüft werden: $\mathcal{O}(Mn)$

Anzahl der Iterationen, Zyklen

- Da wir in jeder Iteration maximal eine Bedingung hinzufügen oder entfernen, brauchen wir mindestens so viele Iterationen wie es Unterschiede in der Größe der initialen Arbeitsmenge und der optimalen Arbeitsmenge gibt.

- Um keine Zyklen zu bekommen, muss wie beim Simplex-Verfahren sichergestellt sein, dass die gleiche Arbeitsmenge nicht mehrfach besucht wird.

- Bei nicht degenerierten Problemen ist dies der Fall, da dann einer von drei Fällen eintreten muss:
  1. Nicht-trivialer Schritt, d.h. die Funktion wird weiter minimiert und wir entfernen uns vom Rand ehemaliger Bedingungen der Arbeitsmenge. Diese werden daher nie wieder aufgenommen.
  2. Eine neue Bedingung ist verletzt. Nach spätestens n Iterationen kann keine Bedingung mehr verletzt sein.
  3. Eine Bedingung muss entfernt werden, um weiter zu minimieren. Danach machen wir wieder einen nicht-trivialen Schritt.

Alternative Strategien

- Wie bei linearen Programmen gibt es die Möglichkeit, den Rand der gültigen Menge zu meiden statt ihn zu suchen.
- Diese **Innere-Punkt-Methoden** sind bei großen Problemen meist effizienter.
- Falls die Nebenbedingungen eine einfache Form haben, sind **Projektionsmethoden** oft eine effiziente Wahl.
- Beide Ansätze werden in der nächsten Vorlesung kurz vorgestellt.

Zusammenfassung

- Quadratische Programme können konvex oder nicht konvex sein, abhängig von Q
- Konvexe quadratische Programme können immer global optimiert werden.
- Es kann wie bei der linearen Programmierung ein Aktive-Set-Verfahren verwendet werden.
- Das Optimum liegt nicht zwangsläufig auf dem Rand der gültigen Menge.
- Das Verfahren verwendet eine dynamische Arbeitsmenge von Nebenbedingungen.