Lineare Optimierung und Simplex-Algorithmus

Problemstellung

Beispiel 1: Unser Unternehmen verfügt über drei Maschinen A, B, C, mit denen zwei verschiedene Produkte P1, P2 hergestellt werden.

- Die Maschinen haben eine monatliche maximal Laufzeit von 170 (A), 150 (B) bzw. 180 (C) Stunden.
- Zur Herstellung von 1000 Stück von P1 braucht man eine Stunde Laufzeit der Maschine A und eine Stunde Laufzeit der Maschine B.
- Zur Herstellung von 1000 Stück von P2 braucht man zwei Stunden Laufzeit der Maschine A, eine Stunde Laufzeit der Maschine B und drei Stunden Laufzeit der Maschine C.

Wie viele Stück der beiden Produkte können wir monatlich unter den vorgegebenen Bedingungen herstellen? Wenn wir außerdem annehmen, dass sich mit 1000 Stück von P1 300 Euro Gewinn erzielen lassen, und mit 1000 Stück von P2 500 Euro, für welche mögliche Wahl wird der Gewinn maximiert?

Beispiel 2: Wir wollen aus verschiedenen Rohstoffen A, B, C, D ein neues Motoröl mischen. Damit das Öl die gewünschten Eigenschaften hat, müssen folgende Bedingungen erfüllt sein:

- Der Anteil von A muss mindestens doppelt so groß sein wie die Anteile von C und D zusammen.
- Die Summe der Anteile von B und C muss mindestens so groß sein wie das Doppelte des Anteils von D.
- Die Differenz der Anteile von A und D muss mindestens so groß sein wie die Differenz der Anteile von B und C.

Welche Zusammensetzungen genügen diesen Bedingungen? Wenn wir außerdem annehmen, dass die vier Rohstoffe 5, 4, 2, bzw. 1 Euro pro Liter kosten, welche Zusammensetzung muss man wählen, damit alle genannten Bedingngen erfüllt sind und die Kosten für die Rohstoffe minimal werden?

Die Bedingungen in den obigen Beispielen führen auf lineare Ungleichungssysteme. Dabei handelt es sich um Probleme der Form $Ax \geq b$ für eine gegebene Matrix $A \in \mathbb{R}^{n \times m}$ und einen gegebenen Vektor $b \in \mathbb{R}^n$ und einen unbekannten Vektor $x \in \mathbb{R}^m$. Mit "$\geq$" ist hier gemeint, dass jede Komponente des Vektors auf der linken Seite höchstens so groß sein darf wie der entsprechende Vektor auf der rechten Seite.
Für Beispiel 1 lautet das Ungleichungssystem

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -2 \\ -1 & -1 \\ 0 & -3 \end{pmatrix} \begin{pmatrix} P1 \\ P2 \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ -170 \\ -150 \\ -180 \end{pmatrix},$$

wobei die ersten beiden Zeilen die impliziten Bedingungen codieren, dass wir keine negative Stückzahlen produzieren können. Für Beispiel 2 erhalten wir

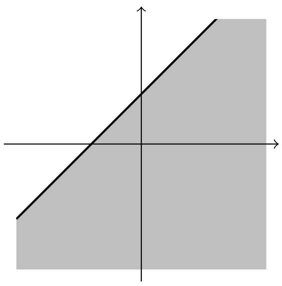
$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 \\ 2 & 0 & -1 & -1 \\ 0 & 1 & 1 & -2 \\ 1 & -1 & 1 & -1 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Dabei codieren die ersten vier Bedingungen, dass der Anteil eines Rohstoffs am Endprodukt nicht negativ sein kann, und die Bedingungen fünf und sechs erzwingen, dass sich die vier Anteile zu 1 aufaddieren.

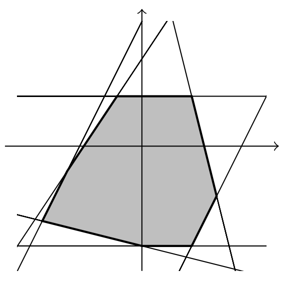
Die Lösungsmenge eines linearen Ungleichungssystems ist eine bestimmte Teilmenge von $\mathbb{R}^m$. Die Aufgabe gesteht darin, innerhalb der Lösungsmenge einen Punkt (oder mehrere Punkte) zu finden, an denen ein vorgegebenes Funktional $c^* \in (\mathbb{R}^m)^*$ maximal oder minimal wird. Im Beispiel 1 ist das Funktional $c^*\colon \mathbb{R}^2 \to \mathbb{R}$, $c(P1, P2) = 300P1 + 500P2$ zu maximieren, während wir im Beispiel 2 das Funktional $c^*\colon \mathbb{R}^4 \to \mathbb{R}$, $c(A, B, C, D) = 5A + 4B + 2C + D$ minimieren sollen. Es bietet sich an, die Funktionale als Vektoren zu schreiben, also $(300, 500)$ bzw. $(5, 4, 2, 1)$.

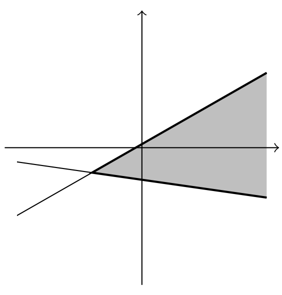
Geometrische Überlegungen

Wie sieht die Lösungsmenge eines linearen Ungleichungssystems aus? Die Lösungsmenge einer einzelnen Gleichung $a_1x_1 + \cdots + a_mx_m = b$ ist bekanntlich eine affine Hyperebene im $\mathbb{R}^m$. Diese Hyperebene trennt den Raum in zwei Halbräume. Sie besteht genau aus den Punkten, an denen das Funktional $a^*\colon \mathbb{R}^m \to \mathbb{R}$, $a^*(x) = a_1x_1 + \cdots + a_mx_m$ den Wert $b \in \mathbb{R}$ annimmt. Auf allen Punkten des einen Halbraums nimmt a^* einen kleineren Wert als b an, und auf den Punkten des anderen Halbraums einen größeren. Die Lösungsmenge einer einzelnen Ungleichung $a_1x_1 + \cdots + a_mx_m \geq b$ ist also ein Halbraum (incl. der Begrenzungsebene).

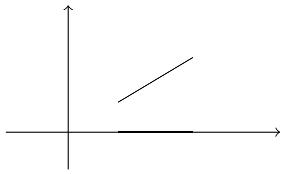


Die Lösungsmenge eines Ungleichungssystems ist ein Schnitt endlich vieler solcher Halbräume. Ein solches Gebilde nennt man ein *(konvexes) Polytop.* Dabei kann es sich um die leere Menge, um eine nicht-leere beschränkte Menge, oder um eine unbeschränkte Menge handeln.

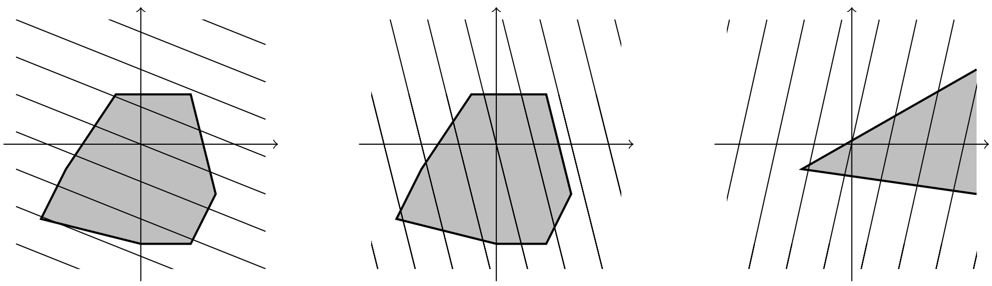




Die Aufgabe besteht darin, das Maximum oder Minimum einer linearen Funktion auf einem Polytop zu finden. Im eindimensionalen Fall überzeugt man sich leicht, dass für das Maximum und das Minimum nur Randpunkte des Polytops in Frage kommen.



Auch der zweidimensionale Fall ist noch recht anschaulich. Das Polytop ist hier eine Teilmenge der Ebene. Für jedes fixe $y \in \mathbb{R}$ ist die Menge aller $x \in \mathbb{R}^2$, für die $c^*(x) = y$ gilt, eine Gerade. All diese Geraden verlaufen parallel zu $\ker c^* \subseteq \mathbb{R}^2$. Um den Punkt x im Polytop zu finden, für den $c^*(x)$ maximal wird, schiebt man diese Gerade parallel über das Polytop, in Richtung steigender Werte von y. Die letzte Gerade, die einen Punkt des Polytops berührt, entspricht dann dem maximal möglichen Wert. Typischerweise ist das eine Ecke des Polytops. Es kann auch eine Kante sein. In diesem Fall sind alle Punkte dieser Kante optimale Lösungen, inclusive der beiden Ecken, durch die die Kante begrenzt wird. Wenn das Polytop unbeschränkt ist, gibt es als drittes noch die Möglichkeit, dass kein Maximum existiert.



Man kann zeigen, dass es sich in höheren Dimensionen genauso verhält: entweder das Funktional wird an einer Ecke des Polytops maximal, oder es gibt kein Maximum.

Normalisierung der Problemstellung

Bevor wir uns mit der Lösung des Problems beschäftigen, wollen wir die Problemstellung in eine etwas handlichere Form umwandeln. Dazu machen wir folgende Beobachtungen:

- O.B.d.A. genügt es, Maximierungsaufgaben zu betrachten, denn das Minimum eines Funktionals c^* ist das Maximum des Funktionals $-c^*$.
- O.B.d.A. können wir annehmen, dass jede Variable x_i einer Bedingung $x_i \geq 0$ unterworfen ist. Wenn eine solche Bedingung für eine bestimmte Variable x_i nicht vorgegeben ist, führen wir zwei neue Variablen x_i^+ und x_i^- ein, die den Bedingungen $x_i^+ \geq 0$ und $x_i^- \geq 0$ genügen sollen, und ersetzen die Variable x_i durch $x_i^+ - x_i^-$.
- O.B.d.A. können wir annehmen, dass alle Bedingungen, die nicht von der Form $x_i \geq 0$ sind, Gleichheiten sind, denn jede Ungleichung $a_1x_1 + \cdots + a_mx_m \geq b$ lässt sich ersetzen durch eine Gleichung $a_1x_1 + \cdots + a_mx_m - y = b$ für eine neue Variable y mit der Bedingung $y \geq 0$.

Insgesamt können wir also annehmen, dass wir $A \in \mathbb{R}^{n\times m}$, $b \in \mathbb{R}^n$, $c^* \in (\mathbb{R}^m)^*$ gegeben haben und einen Vektor $x \in \mathbb{R}^m$ suchen mit $Ax = b$ und $x \geq 0$ (koordinatenweise), für den $c^*(x)$ maximal wird. Durch Streichen redundanter Zeilen aus A und b können wir darüber hinaus annehmen, dass A vollen Rang hat. Insbesondere wird dann $m > n$ gelten (sofern es sich nicht um ein triviales Problem handet).

Lösungsstrategie

Aus der Konvexität des Polytops und der Linearität des Funktionals folgt, dass jedes lokale Maximum sofort auch ein globales Maximum ist. Deshalb genügt es, wenn wir einen Punkt im Polytop finden, in dessen unmittelbarer Umgebung es keine Punkte gibt, wo das Funktional einen höheren Wert annimmt. Wir wissen außerdem, dass es genügt, die Ecken des Polytops zu betrachten.

Der Simplex-Algorithmus startet an einer beliebigen Ecke. Entlang der Kanten, die an dieser Ecke zusammenlaufen, können wir uns zu einer anderen Ecke bewegen. Der Wert des Funktionals wird dabei abhängig von der gewählten Kante entweder steigen oder fallen oder gleich bleiben. Wir wählen eine Kante aus, entlang der der Wert des Funktionals steigt. Wenn es keine solche Kante gibt, dann haben wir das Maximum gefunden und sind fertig. Wenn es eine geeignete Kante gibt, dann gibt es zwei Möglichkeiten: wenn es entlang dieser Kante keine weitere Ecke gibt, dann ist das Polytop unbeschränkt und es gibt kein Maximum. Anderenfalls gehen wir bis zur Ecke und betrachten alle Kanten, die von jener Ecke abgehen. In jedem Schritt kommen wir so zu einer Ecke, bei der das zu maximierende Funktional einen höheren Wert annimmt. Da ein Polytop nur endlich viele Ecken hat, muss das Verfahren nach endlich vielen Iterationen terminieren.

Wie geht man rechnerisch von einer Ecke zur nächsten? Betrachten wir als Beispiel den Fall

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 3 & -5 \\ 0 & 1 & 0 & 0 & 3 & 4 \\ 0 & 0 & 1 & 0 & -2 & 6 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}, \qquad b = \begin{pmatrix} 6 \\ 37 \\ 36 \\ 7 \end{pmatrix}.$$

Da A schon eine Treppennormalform ist, lässt sich die Lösungsmenge des inhomogenen Gleichungssystems $Ax = b$ sofort ablesen. Sie lautet

$$\left\{ \begin{pmatrix} 6 \\ 37 \\ 36 \\ 7 \\ 0 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} -3 \\ -3 \\ 2 \\ -1 \\ 1 \\ 0 \end{pmatrix} + x_6 \begin{pmatrix} 5 \\ -4 \\ -6 \\ 0 \\ 0 \\ 1 \end{pmatrix} : x_5, x_6 \in \mathbb{R} \right\} \subseteq \mathbb{R}^6.$$

Es handelt sich um einen affinen Unterraum von $\mathbb{R}^4$. Für $x_5 = x_6 = 0$ bekommt man einen Vektor mit lauter nicht-negativen Einträgen, also einen Punkt des Polytops. Es handelt sich um eine Ecke, von der zwei Kanten ausgehen, nämlich eine, die dem Erhöhen von x_5 entspricht (wobei $x_6 = 0$ festgehalten wird) und eine, die dem Erhöhen von x_6 entspricht (wobei $x_5 = 0$ festgehalten wird). Wenn wir $x_6 = 0$ festhalten, können wir den Wert von x_5 bis maximal $x_5 = 2$ erhöhen, denn eine weitere Erhöhung würde zu einem Vektor mit negativen Komponenten führen, d.h. wir würden das Polytop verlassen. Aus dem gleichen Grund können wir, wenn wir $x_5 = 0$ festhalten, den Wert von x_6 bis maximal $x_6 = 6$ erhöhen.

Folgen wir der Kante von x_5 bis zum ihrem Ende, so landen wir an der Ecke $(0, 31, 40, 5, 2, 0)$. Welche Kanten von dort abgehen, lässt sich aus der obigen Darstellung des Gleichungssystems nicht mehr ohne weiteres ablesen. Besser geeignet wäre eine Darstellung, bei der die beiden freien Parameter x_1 und x_6 statt x_5 und x_6 sind. (Beachte: die erste und sechste Koordinate der neuen Ecke sind null.) Eine solche Darstellung kann man berechnen, indem man die erweiterte Matrix $(A|b)$ mit geeigneten elementaren Zeilenumformungen in eine Form bringt, bei der in der fünften Spalte der erste Einheitsvektor steht:

$$\left(\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 3 & -5 & 6 \\ 0 & 1 & 0 & 0 & 3 & 4 & 37 \\ 0 & 0 & 1 & 0 & -2 & 6 & 36 \\ 0 & 0 & 0 & 1 & 1 & 0 & 7 \end{array}\right) \begin{array}{l} \mid :3 \quad {\scriptstyle -3} \quad {\scriptstyle 2} \quad {\scriptstyle -1} \\ \longleftarrow {\scriptstyle +} \\ \longleftarrow {\scriptstyle +} \\ \longleftarrow {\scriptstyle +} \end{array} \quad \leftrightarrow \quad \left(\begin{array}{cccccc|c} 1/3 & 0 & 0 & 0 & 1 & -5/3 & 2 \\ -1 & 1 & 0 & 0 & 0 & 9 & 31 \\ 2/3 & 0 & 1 & 0 & 0 & 8/3 & 40 \\ -1/3 & 0 & 0 & 1 & 0 & 5/3 & 5 \end{array}\right)$$

Daraus folgt die alternative Darstellung

$$\left\{\begin{pmatrix}0\\31\\40\\5\\2\\0\end{pmatrix}+x_1\begin{pmatrix}1\\1\\-2/3\\-1/3\\1/3\\0\end{pmatrix}+x_6\begin{pmatrix}0\\-9\\-8/3\\-5/3\\5/3\\1\end{pmatrix}: x_1,x_6\in\mathbb{R}\right\}\subseteq\mathbb{R}^6$$

für die Lösungsmenge des inhomogenen Gleichungssystems $Ax=b$.

Die neue Matrix ist keine TNF, lässt sich aber durch Umordnung der Spalten in eine TNF überführen. Allerdings darf man Spalten bekanntlich nicht einfach vertauschen. Das Umordnen von Spalten entspricht einer Umordung der Koordinarten des Lösungsvektors. Wenn z.B. (x_1,x_2,x_3,x_4,x_5) ein Lösungsvektor des Systems $Ax=b$ ist, dann ist (x_5,x_2,x_3,x_1,x_5) ein Lösungsvektor für das System $\tilde{A}x=b$, bei dem $\tilde{A}$ aus A entsteht, indem die Spalten 1 und 5 miteinander vertauscht werden. Man kann sich die Umordnungen merken, indem man über jede Spalte die zugehörige Variable schreibt, und diese mit den Spalten mitvertauscht.

$$\begin{array}{cccccc|c}
x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & \\
\end{array}
\qquad\qquad
\begin{array}{cccccc|c}
x_5 & x_2 & x_3 & x_4 & x_1 & x_6 & \\
\end{array}$$

$$\left(\begin{array}{cccccc|c}
1&0&0&0&3&-5&6\\
0&1&0&0&3&4&37\\
0&0&1&0&-2&6&36\\
0&0&0&1&1&0&7
\end{array}\right)
\begin{array}{l}
|:3 \quad -3 \quad 2 \quad -1\\
+\\
+\\
+
\end{array}
\leftrightarrow
\left(\begin{array}{cccccc|c}
1&0&0&0&1/3&-5/3&2\\
0&1&0&0&-1&9&31\\
0&0&1&0&2/3&8/3&40\\
0&0&0&1&-1/3&5/3&5
\end{array}\right)$$

Wenn man die Spalten immer so vertauscht, dass im linken Teil der Matrix die Einheitsmatrix steht, dann braucht man diese nicht zu notieren (oder im Computer: zu speichern). Stattdessen kann man die Variablen auch vor die entsprechenden Zeilen schreiben:

$$\begin{array}{c|c}
 & \begin{array}{cc|c} x_1 & x_6 & \end{array}\\
\begin{array}{c}x_5\\x_2\\x_3\\x_4\end{array} &
\left(\begin{array}{cc|c}
1/3&-5/3&2\\
-1&9&31\\
2/3&8/3&40\\
-1/3&5/3&5
\end{array}\right)
\end{array}$$

Üblicherweise schreibt man unter die Spalten dieser Matrix noch die Werte des zu optimierenden Funktionals der entsprechenden Vektoren in der Darstellung des Lösungsraums. Wenn zum Beispiel $c^*\colon\mathbb{R}^6\to\mathbb{R}$ gegeben ist durch $c(x_1,x_2,x_3,x_4,x_5,x_6)=2x_5+x_6$, so würde man notieren

$$\begin{array}{c|c}
 & \begin{array}{cc|c} x_1 & x_6 & \end{array}\\
\begin{array}{c}x_5\\x_2\\x_3\\x_4\end{array} &
\left(\begin{array}{cc|c}
1/3&-5/3&2\\
-1&9&31\\
2/3&8/3&40\\
-1/3&5/3&5
\end{array}\right)\\
 & \begin{array}{cc|c} 2/3 & -13/3 & 4 \end{array}
\end{array},$$

weil $c^*(1,1,-2/3,-1/3,1/3,0)=2/3$, $c^*(0,-9,-8/3,-5/3,5/3,1)=13/3$, $c^*(0,31,40,5,2,0)=4$. Diese Darstellung bezeichnet man als ein Simplex-Tableau. Der Wert des Funktionals lässt sich erhöhen, solange es eine Variablenspalte gibt, für die der Wert von c^* positiv ist. Die Spalte mit dem maximalen Wert wählt man aus und bewegt sich in Richtung des Vektors in dieser Spalte zur nächsten Ecke.

Zusammenfassung: Aus der TNF der erweiterten Matrix $(A|b)$ kann man eine Darstellung der Lösungsmenge des inhomogenen Systems $Ax=b$ gewinnen. Zusammen mit der Einschränkung, dass die Variablen nur nichtnegative Werte annehmen sollen, lässt sich diese Darstelllung der Lösungsmenge als Ecke des Polytops zusammen mit Informationen über die Richtungen der abgehenden Kanten interpretieren. An den Koordinaten dieser Richtungsvektoren und des Startpunkts kann man ablesen, wie weit man sich entlang der Kante zum nächsten Eckpunkt bewegen kann. Man wählt die Kante, entlang der der Wert des Funktionals am stärksten steigt. Durch eine entsprechende Umordnung der Variablen und erneute TNF-Berechnung findet man eine Darstellung der Lösungsmenge mit der neuen Ecke als Startpunkt.

Die Startlösung

Mit dem oben beschriebenen Verfahren kann man von einer bekannten Ecke des Polytops zu einer besser bewerteten benachbarten Ecke wandern. Es bleibt zu überlegen, wie man überhaupt einmal eine erste Ecke findet. Im allgemeinen wird es nicht reichen, $(A|b)$ in eine TNF $(T|\tilde{b})$ zu überführen, weil nicht garantiert ist, dass der Vektor $\tilde{b}$ nur nichtnegative Koordinaten hat. (Wenn alle Koordinaten nichtnegativ sind, handelt es sich um eine Ecke.)

Eine Möglichkeit zum Auffinden einer Ecke besteht darin, zunächst ein modifiziertes Problem zu betrachten, für das man leicht eine Ecke finden kann. Statt $Ax = b$ betrachtet man $(A|I_n)\binom{x}{z} = \binom{b}{0}$ mit neuen Variablen $z = (z_1, \ldots, z_n)$. Mit geeigneten nichtnegativen Vielfachen der zusätzlichen Einheitsvektoren kann man etwaige negative Koordinaten eines Lösungspunkts ausgleichen und erhält so eine Ecke für das modifizierte Problem. Ausgehend von dieser Ecke minimiert man das Funktional $c^* \colon \mathbb{R}^{m+n} \to \mathbb{R}$, $c^*(x_1, \ldots, x_m, z_1, \ldots, z_m) = z_1 + \cdots + z_m$. Das ursprüngliche Polytop ist genau dann nicht leer, wenn das Minimum 0 ist. In diesem Fall gilt: Wenn das Minimum bei $(x_1, \ldots, x_m, z_1, \ldots, z_m)$ angenommen wird, dann ist $(x_1, \ldots, x_m)$ eine Ecke des ursprünglichen Polytops.