Z. Wahrscheinlichkeitstheorie verw. Geb. 5, 167—172 (1966)

Über die Optimierung einer bedingt konvexen Funktion unter linearen Nebenbedingungen

PETER KALL

Eingegangen am 1. Februar 1965

Zusammenfassung. Es zeigt sich, daß auch für die allgemeinere Klasse der bedingt konvexen Funktionen der von A. ORDEN für quadratische Zielfunktionen formulierte Satz gültig bleibt. Ferner lassen sich aus diesem Satz auch für das Programmierungsproblem ganz ähnliche Schlüsse ziehen wie im quadratischen Fall. Es ist damit gezeigt, daß derartige nicht-konvexe Programmierungsprobleme grundsätzlich lösbar sind. In Zukunft wird es darauf ankommen, bessere Verfahren zur Lösung dieser Probleme zu finden.

Im Zusammenhang mit der Unternehmensforschung ist in den letzten Jahren die Optimierung konvexer Funktionen unter Nebenbedingungen ausführlich untersucht worden. Dabei sind besonders die linearen Funktionen und Funktionen, die sich additiv aus linearen und quadratischen Formen zusammensetzen, behandelt worden, wobei zunächst von den quadratischen Formen Definitheit verlangt wurde. ORDEN [*1*] hat dann auf die Definitheit verzichtet, dafür aber nur lineare Gleichungen als Nebenbedingungen zugelassen. Es ergab sich, daß für die Lösbarkeit eines solchen Problems zwei Bedingungen notwendig und hinreichend sind: 1. Im Optimum ist der Gradient der Zielfunktion orthogonal zu dem durch die Nebenbedingungen definierten linearen Raum. 2. Die quadratische Zielfunktion $f(x)$ ist konvex über diesem Raum, sofern das Optimum ein Minimum ist, andernfalls ist sie konkav.

Daß man sich im Falle einer nichtlinearen Zielfunktion bisher hauptsächlich mit den quadratischen Funktionen befaßt hat, mag zwei Gründe haben. Zum einen sind gerade die quadratischen Funktionen dem Mathematiker schon lange vertraut, und zum andern haben vielfach die in der Praxis auftretenden Funktionen (Kosten- oder Gewinnfunktionen) Eigenschaften, die den Versuch nahelegen, sie durch quadratische Funktionen zu approximieren. Diejenigen dieser Eigenschaften, die in dieser Arbeit verlangt werden, sind die folgenden:

1. Die Funktion $f(x)$ ist mindestens einmal stetig differenzierbar.
2. Die Funktion $f(x)$ ist bedingt konvex.

Definition. Die Funktion $f(x)$ heißt *bedingt konvex,* wenn daraus, daß für $x_1 \neq x_2$ und für ein λ_0 mit $0 < \lambda_0 < 1$

$$f(\lambda_0 x_1 + (1-\lambda_0)x_2) < \lambda_0 f(x_1) + (1-\lambda_0) f(x_2)$$

gilt, folgt:

Für

$$\left.\begin{aligned} x_3 &= x_1 + \mu(x_2 - x_1) \\ x_4 &= x_1 + \nu(x_2 - x_1) \end{aligned}\right\} \mu, \nu \in R$$

gilt

$$f(\lambda x_3 + (1-\lambda)x_4) \leqq \lambda f(x_3) + (1-\lambda) f(x_4)$$

für jedes $\lambda \in R$ mit $0 < \lambda < 1$.

Für eine Approximation durch quadratische Funktionen sind sicher noch weitere Eigenschaften zu verlangen — z.B. daß, wenn die Funktion über einem Strahl konvex ist, das gleiche für jeden dazu parallelen Strahl gelten muß (was offenbar mit der bedingten Konvexität nicht verlangt wird) — jedoch werden wir mit den beiden erwähnten auskommen.

Die Definition der bedingten Konvexität bedeutet anschaulich, daß eine bedingt konvexe Funktion, die über einem beliebigen Abschnitt eines Strahles unterhalb der Sehne zwischen den beiden Endwerten verläuft, über dem ganzen Strahl konvex ist, d.h., daraus, daß von drei Funktionswerten über einer Geraden der mittlere unter der Verbindungsgeraden der beiden äußeren liegt, schließt man bereits auf die Konvexität der Funktion über der betrachteten Geraden. Aus der Definition folgt sofort, wenn wir „bedingt konkav" analog zu „bedingt konvex" definieren, das

Lemma 1. *Eine Funktion, die bedingt konvex ist, ist auch bedingt konkav.*

Beweis. Zu zeigen ist, daß aus

$$f(\lambda_0 x_1 + (1-\lambda_0)x_2) > \lambda_0 f(x_1) + (1-\lambda_0) f(x_2) \tag{1}$$

mit $x_1 \neq x_2$; $0 < \lambda_0 < 1$ folgt, daß $f(x)$ längs des ganzen durch x_1, x_2 bestimmten Strahles konkav ist, d.h.

$$f(\lambda x + (1-\lambda)y) \geqq \lambda f(x) + (1-\lambda) f(y) \tag{2}$$

für

$$\left.\begin{aligned} x &= x_1 + \mu(x_2 - x_1) \\ y &= x_1 + \nu(x_2 - x_1) \end{aligned}\right\} \begin{aligned} &\mu, \nu \in R \\ &\mu \neq \nu \end{aligned}$$

und alle λ mit $0 < \lambda < 1$.

Wäre die Relation (2) für irgend eine Wahl von λ, μ, ν falsch, so würde folgen, daß $f(x)$ auf dem von x, y bestimmten Strahl — auf dem natürlich x_1, x_2 liegen — konvex ist, also (1) nicht gelten kann. Daraus folgt die Behauptung.

Für späteren Gebrauch benötigen wir noch

Lemma 2. *Ist die Funktion $\varphi(x)$ konkav über dem Strahl $x + \mu(y - x)$; $\mu \in R$, dann gilt*

$$\varphi(x + \nu(y-x)) \leqq \varphi(x) + \nu(\varphi(y) - \varphi(x)) \quad \textit{für} \quad \nu > 1\,.$$

Beweis. Nehmen wir an, für $\nu_0 > 1$ gelte

$$\varphi(x + \nu_0(y-x)) > \varphi(x) + \nu_0(\varphi(y) - \varphi(x))\,; \tag{3}$$

dann ist mit $z = x + \nu_0(y - x)$

$$y = \mu_0 x + (1-\mu_0) z\,, \quad \text{wobei } \nu_0(1-\mu_0) = 1\,.$$

Da φ konkav ist über dem betrachteten Strahl und (3) gelten soll, folgt

$$\begin{aligned}\varphi(y) &= \varphi(\mu_0 x + (1-\mu_0) z) \geqq \mu_0 \varphi(x) + (1-\mu_0)\varphi(z) \\ &> \mu_0\varphi(x) + (1-\mu_0)\varphi(x) + \nu_0(1-\mu_0)[\varphi(y) - \varphi(x)] = \varphi(y),\end{aligned}$$

also $\varphi(y) > \varphi(y)$.

Aus diesem Widerspruch folgt das Lemma.

Nunmehr wollen wir uns der folgenden Frage zuwenden: Unter welchen Bedingungen existiert ein Minimum einer bedingt konvexen Funktion unter linearen Nebenbedingungen? Präzise lautet das Problem folgendermaßen:

Sei $x \in R^n$, und $f(x)$ sei einmal stetig differenzierbar und bedingt konvex. Gesucht ist

$$(4) \qquad \begin{cases} \operatorname{Min} f(x) \\ \text{unter den Bedingungen} \\ A x = b \end{cases}$$

A ist eine $(m \times n)$-Matrix vom Rang r und b ein m-Vektor.

Nun beweist man leicht den folgenden

Satz 1. *Notwendig und hinreichend dafür, daß das Problem* (4) *eine Lösung* x_0 *besitzt, sind die beiden folgenden Bedingungen:*

$$\text{I.} \qquad \begin{aligned} N' \cdot \nabla f(x_0) &= 0 \\ A x_0 &= b \end{aligned}$$

II. $f(x)$ *ist konvex über dem linearen Raum* $L = \{x \mid Ax = b\}$. *Dabei ist* $\nabla f(x_0)$ *der Gradient von* $f(x)$ *im Punkte* x_0, *und* N *ist eine* $n \times (n-r)$-*Matrix, deren Spalten eine Basis des Vektorraumes* $V = \{x \mid Ax = 0\}$ *bilden.*

Beweis. Daß die Bedingungen notwendig sind, kann man folgendermaßen zeigen: Mit den Lagrange-Multiplikatoren u_i bildet man die Funktion

$$\Phi(x, u) = f(x) + u'(Ax - b).$$

Ist x_0 der gesuchte Optimalpunkt, so muß bekanntlich gelten

$$(5) \qquad \nabla_x \Phi(x_0, u) = \nabla f(x_0) + A' u = 0$$

$$(6) \qquad \nabla_u \Phi(x_0, u) = A x_0 - b = 0.$$

(6) ist identisch mit dem zweiten Teil von I.

Multipliziert man (5) mit N', so folgt

$$N' \nabla f(x_0) + N' A' u = 0.$$

Da $AN = 0$ ist, folgt also der erste Teil von I.

Um die Notwendigkeit von II zu beweisen, zeigen wir, daß $f(x)$ über L gegen $-\infty$ strebt, sofern diese Bedingung verletzt ist. Wir nehmen also an, es gebe zwei Punkte $x_1, x_2 \in L$ und einen Wert λ_1 mit $0 < \lambda_1 < 1$ derart, daß

$$(7) \qquad f(\lambda_1 x_1 + (1-\lambda_1) x_2) > \lambda_1 f(x_1) + (1-\lambda_1) f(x_2).$$

Ohne Einschränkung der Allgemeinheit können wir annehmen, daß

$$(8) \qquad f(x_1) \geqq f(x_2)$$

gilt. Setzen wir $x_3 = \lambda_1 x_1 + (1-\lambda_1) x_2$, so gilt $A x_3 = b$, d.h. $x_3 \in L$.

Nach Lemma 1 ist $f(x)$ konkav über dem Strahl

$$x_1 + \nu(x_2 - x_3); \quad \nu \in R.$$

Ferner gilt nach (7) und (8)

$$(9) \qquad f(x_3) > \lambda_1 f(x_1) + (1 - \lambda_1) f(x_2) \geqq f(x_2)$$

Da nun $f(x)$ konkav ist über dem Strahl $x_1 + \nu(x_2 - x_3)$ und folglich über $x_3 + \nu(x_2 - x_3)$, folgt nach Lemma 2

$$(10) \qquad f(x_3 + \nu(x_2 - x_3)) \leqq f(x_3) + \nu[f(x_2) - f(x_3)] \quad \text{für} \quad \nu > 1.$$

Da

$$[x_3 + \nu(x_2 - x_3)] \in L \quad \text{für jedes} \quad \nu \in R$$

und da nach (9) $f(x_2) - f(x_3) < 0$, strebt also $f(x)$ längs des Strahles $x_3 + \nu(x_2 - x_3)$ gegen $-\infty$. Also ist Bedingung II notwendig für die Existenz einer Lösung $x_0 \in L$. Daß die Bedingungen hinreichend sind, sieht man sofort: Für konvexe Funktionen gilt bekanntlich

$$f(x_2) - f(x_1) \geqq (x_2 - x_1)' \nabla f(x_1).$$

Haben wir also einen Punkt x_0, der den Bedingungen I genügt, und ist II erfüllt, so gilt für jedes $x \in L$

$$f(x) - f(x_0) \geqq (x - x_0)' \nabla f(x_0)$$

also, da $x - x_0 = Nz$ mit einem eindeutig bestimmten Vektor z,

$$f(x) - f(x_0) \geqq z' N' \nabla f(x_0) = 0 \quad \text{nach I}.$$

Folglich liefert x_0 ein globales Minimum.

Man sieht sofort, daß der Satz von ORDEN [*1*] in diesem Satz vollständig enthalten ist. Ferner ist es natürlich möglich, aus diesem Satz für das Programmierungsproblem

$$(11) \qquad \begin{cases} & \operatorname{Min} f(x) \\ \text{unter} & Ax = b \\ & x \geq 0 \end{cases}$$

ähnliche Schlußfolgerungen zu ziehen, wie das im quadratischen Fall (siehe [*2*]) geschehen ist.

Wir nehmen dazu an, daß (11) eine optimale Lösung x_0 hat, für die $x_0 > 0$ gilt. Für eine derartige Lösung gilt der folgende

Satz 2. *Ist $x_0 > 0$ eine optimale zulässige Lösung von* (11), *so gilt notwendigerweise*

I'. $N' \nabla f(x_0) = 0$;

II'. *$f(x)$ ist über jedem Strahl $x_0 + Nz$ innerhalb des zulässigen Bereiches konvex. N hat dieselbe Bedeutung wie früher.*

Beweis. Da x_0 innerer Punkt des zulässigen Bereiches ist, folgt Bedingung I' genauso wie Bedingung I in Satz 1. Wäre Bedingung II' verletzt, so würde aus Lemma 2 folgen, daß es ein zulässiges x_1 gibt mit $f(x_1) < f(x_0)$, im Widerspruch dazu, daß x_0 optimal ist. Damit ist der Satz bewiesen.

Nehmen wir nun an, der zulässige Bereich $\{x \mid Ax = b,\ x \geqq 0\}$ sei beschränkt. Dann folgt aus Satz 2 ähnlich wie in [2] ein kombinatorisches Verfahren zur Lösung von (11).

Zuerst sucht man einen Punkt x_0, der den Bedingungen

$$\left\{\begin{aligned} N' \nabla f(x_0) &= 0 \\ A x_0 &= b \\ x_0 &\geqq 0 \end{aligned}\right. \tag{12}$$

genügt. Findet man einen solchen Punkt, so prüft man nach, ob Bedingung II′ erfüllt ist. Ist das der Fall, so hat man offensichtlich das Optimum gefunden. Ist Bedingung II′ nicht erfüllt oder hat (12) keine zulässige Lösung, so geht man zu Teilbereichen des zulässigen Bereiches von (11) über, indem man sukzessive Komponenten von x gleich Null setzt (d.h. die entsprechenden Spalten von A und die entsprechenden Komponenten von $\nabla f(x)$ streicht; dann muß man allerdings für die aus A entstandene Matrix $\tilde{A}$ eine neue Matrix $\tilde{N}$ berechnen). In jedem dieser Teilbereiche sucht man wieder eine Lösung zu

$$\left\{\begin{aligned} \tilde{N} \nabla \tilde{f}(\tilde{x}_0) &= 0 \\ \tilde{A} \tilde{x}_0 &= b \\ \tilde{x}_0 &\geqq 0\,. \end{aligned}\right. \tag{13}$$

$\tilde{N}$, $\nabla \tilde{f}$ und $\tilde{x}_0$ sollen angeben, daß die Rechnung in einem Teilbereich durchgeführt wird. Findet man eine Lösung, die II′ genügt (wobei II′ nur noch über dem entsprechenden Teilbereich verlangt wird), so notiert man sie. Da es nur endlich viele solcher Teilbereiche gibt (diejenigen mit der kleinsten Dimension sind offenbar die Eckpunkte des Ausgangsbereiches, die offensichtlich den Bedingungen (13) und II′ genügen), hat man schließlich endlich viele zulässige Lösungen notiert, aus denen das Optimum auszuwählen ist.

Um zu zeigen, daß dieses Verfahren zum Ziel führen muß, müssen wir das folgende Lemma beweisen:

Lemma 3. *Existiert eine Lösung x_0 von* (12), *die der Bedingung* II′ *nicht genügt, so gibt es keinen Punkt x_1 mit $Ax_1 = b$, $x_1 > 0$, der optimal ist.*

Beweis. Nehmen wir an, es gebe einen solchen Optimalpunkt x_1. Nach Satz 2 ist dann $f(x)$ über der Geraden, die durch x_0 und x_1 geht, konvex. Also gilt, da $x_1 - x_0 = N z_1$ ist, wobei z_1 eindeutig bestimmt ist,

$$f(x_1) - f(x_0) \geqq (x_1 - x_0)' \nabla f(x_0) = z_1' N' \nabla f(x_0) = 0 \quad \text{oder} \quad f(x_1) \geqq f(x_0)\,. \tag{14}$$

Da in x_0 Bedingung II′ verletzt ist, gibt es aber einen Punkt x_2 des zulässigen Bereiches derart, daß

$$f(x_2) < f(x_0)\,. \tag{15}$$

Aus (14) und (15) folgt

$$f(x_2) < f(x_1)\,. \tag{16}$$

Folglich ist x_1 entgegen der Annahme nicht optimal. Aus diesem Widerspruch folgt das Lemma.

Dieses Lemma zeigt, daß das Verfahren in endlich vielen Schritten in der angegebenen Weise zum Ziele führt (wobei das Lemma natürlich auf die Teilbereiche sinngemäß anzuwenden ist).

Literatur

[1] ORDEN, A.: Minimization of indefinite quadratic functions with linear constraints. In: Recent advances in mathematical programming. Edited by R. L. Graves and P. Wolfe. New York: McGraw-Hill Book Company, Inc. 1963.

[2] KALL, P., H. P. KÜNZI u. W. OETTLI: Ein Beitrag zur indefiniten quadratischen Programmierung. Vorgetragen anläßlich der Joint European Conference of the Econometric Society and The Institute of Management Sciences (Zürich, 7.—11. 9. 1964).

Rechenzentrum der Universität
Zürich