KTH Matematik

5B1817 Tillämpad ickelinjär optimering

Föreläsning 4

Metoder för problem utan bivillkor

Ickelinjära programmeringsproblem utan bivillkor

Betrakta idag ett ickelinjärt programmeringsproblem utan bivillkor

$$(P) \qquad \begin{array}{ll} \text{min} & f(x) \\ \text{då} & x \in \mathbb{R}^n, \end{array} \qquad \text{där } f \in C^2.$$

Önskvärt att hitta global minpunkt. I allmänhet för svårt.

Med förstaderivator: Hitta x^* så att $\nabla f(x^*) = 0$.

Med första- och andraderivator: Hitta x^* så att $\nabla f(x^*) = 0$ och $\nabla^2 f(x^*) \succeq 0$.

Hur görs detta?

Iterativa metoder

$$(P) \quad \begin{array}{ll} \text{min} & f(x) \\ \text{då} & x \in I\!\!R^n, \end{array} \qquad \text{där } f \in C^2.$$

En iterativ metod genererar x_0, x_1, x_2, $\ldots$ så att $\lim_{k\to\infty} x_k = x^*$, där x^* är en lösning till (P).

Avbryter då lämpligt avbrottskriterium är uppfyllt, tex $\|\nabla f(x_k)\| < \epsilon$.

Idealt ska x^* vara en global minpunkt. Om (P) är ett konvext problem kan vi i allmänhet uppnå detta.

För ickekonvexa problem kommer "lösning" i allmänhet att bero på vilken derivatainformation som används.

För ickekonvexa problem är startpunkten speciellt viktig.

Linjesökande metoder

En *linjesökande metod* genererar i varje iteration en *sökriktning* och utför *linjesökning* längs sökriktningen.

Iteration k tar följande form i punkten x_k.

- Beräkna sökriktning p_k så att $\nabla f(x_k)^T p_k < 0$.
- Lös approximativt $\min_{\alpha \geq 0} f(x_k + \alpha p_k)$, vilket ger α_k.
- $x_{k+1} \leftarrow x_k + \alpha_k p_k$.

Olika metoder baseras på olika val av p_k och α_k.

Några linjesökande metoder

Vi kommer idag initialt studera två fundamentala metoder.

- *Steepest-descentmetoden*, där $p_k = -\nabla f(x_k)$, och
- *Netwtons metod*, där $\nabla^2 f(x_k) p_k = -\nabla f(x_k)$.

Steepest descent:	$+$	Sökriktningen billig att beräkna,
	$-$	Långsam konvergens.
Newtons metod:	$-$	Sökriktningen tyngre att beräkna,
	$+$	Snabbare konvergens.

Vi kommer dessutom bland annat att titta på *kvasi-Newtonmetoder* som försöker efterlikna Newtons metod utan att räkna ut andraderivator.

Kvadratisk målfunktion

Titta på modellproblem med kvadratisk målfunktion,

$$(QP) \qquad \begin{array}{ll} \text{min} & f(x) = \frac{1}{2}x^T H x + c^T x \\ \text{då} & x \in I\!R^n. \end{array}$$

Påstående. *Följande gäller för (QP) beroende på H:*
(i) Om $H \succ 0$ har (QP) en unik global minpunkt x^* som ges av $Hx^* = -c$.
(ii) Om $H \succeq 0$, $H \not\succ 0$ är varje x^* som uppfyller $Hx^* = -c$ en global minpunkt till (QP). (Eventuellt finns inga sådana x^*).
(iii) Om $H \not\succeq 0$ är $f(x)$ obegränsad nedåt.

Bevis. Nödvändiga optimalitetsvillkor och konvexitet ger resultaten. □

Vi antar att $H \succ 0$ i diskussionen.

Linjesökning på kvadratisk målfunktion

Betrakta (QP) med $f(x) = \frac{1}{2}x^T Hx + c^T x$, där $H \succ 0$. Vi får:

x^* global minpunkt till $(QP) \iff 0 = \nabla f(x^*) = Hx^* + c$.

Antag att vi har sökriktning p_k så att $\nabla f(x_k)^T p_k < 0$.

Låt $\varphi(\alpha) = f(x_k + \alpha p_k) = f(x_k) + \alpha \nabla f(x_k)^T p_k + \frac{\alpha^2}{2} p_k^T H p_k$.

Då ger $\alpha_k = -\dfrac{\nabla f(x_k)^T p_k}{p_k^T H p_k}$ global minpunkt till $\min_{\alpha \geq 0} f(x_k + \alpha p_k)$ dvs vi kan utföra *exakt linjesökning*.

Steepest descent på kvadratisk målfunktion

Antag att $f(x) = \frac{1}{2}x^T H x + c^T x$, där $H \succ 0$. Vi får:

Antag också att steepest descent med exakt linjesökning används, dvs $p_k = -\nabla f(x_k)$ och $\alpha_k = -\dfrac{\nabla f(x_k)^T p_k}{p_k^T H p_k}$.

Det gäller då att $f(x_{k+1}) - f(x^*) \leq \left(\dfrac{\text{cond}(H) - 1}{\text{cond}(H) + 1}\right)^2 (f(x_k) - f(x^*))$, se Lemma 11.1 och Lemma 11.2 i Nash och Sofer.

$\text{cond}(H) \gg 1 \Rightarrow \dfrac{\text{cond}(H) - 1}{\text{cond}(H) + 1} \approx 1$, dvs långsam *linjär konvergens*.

För ickelinjär målfunktion får vi typiskt sådan linjär konvergens om exakt linjesökning utförs, där H ersätts med $\nabla^2 f(x_k)$.

Konvergenshastighet

Definition. *Antag att* $x_k \in I\!R^n, k = 0, 1, \ldots,$ *och att* $\lim_{k\to\infty} x_k = x^*$. *Vi säger att* $\{x_k\}_{k=0}^{\infty}$ *konvergerar mot* x^* *med konvergenshastighet* r *om*

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^r} = C, \quad \textit{där } C < \infty.$$

Vi får då $\|x_{k+1} - x^*\| \approx C \cdot \|x_k - x^*\|^r$.

Vi vill ha r stort (och C nära noll). Av intresse:

- $r = 1$, $0 < C < 1$, linjär konvergens. (Steepest descent.)
- $r = 1$, $C = 0$, superlinjär konvergens. (Kvasi-Newton.)
- $r = 2$, kvadratisk konvergens. (Newtons metod.)

Newtons metod

Vi har $f(x_k+p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k) p + o(\|p\|^2)$.

Newtons metod är väldefinierad då $\nabla^2 f(x_k) \succ 0$. Sökriktningen p_k ges som lösningen till $\min_{p \in I\!R^n} \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k) p$, dvs en kvadratiska approximation av målfunktionen minimeras.

Ekvivalent med att lösa $\nabla^2 f(x_k) p_k = -\nabla f(x_k)$.

Då p_k minimerar den kvadratiska funktionen till såväl riktning som belopp, är det naturligt att försöka välja $\alpha = 1$ i en linjesökning.

Eftersom $\nabla^2 f(x_k) \succ 0$ antagits får vi $0 < p_k^T \nabla^2 f(x_k) p_k = -\nabla f(x_k)^T p_k$ om $\nabla f(x_k) \neq 0$, dvs p_k är en descentriktning.

En *modifierad Newtonmetod* hanterar också $\nabla^2 f(x_k) \not\succ 0$.

Newtons metod löser ett kvadratiskt problem (med $H \succ 0$) i en iteration.

Konvergenshastighet för Newtons metod

Sats. *Antag att* $f \in C^3$ *på* R^n *och att* x^* *är en lokal minpunkt till* (P) *där* $\nabla^2 f(x^*) \succ 0$*. Om Newtons metod (med steglängd ett) startas i en punkt tillräckligt nära* x^* *så konvergerar den mot* x^* *med konvergenshastighet minst två.*

Bevis. Då $f \in C^3$, $\nabla f(x^*) = 0$ och $\nabla^2 f(x^*) \succ 0$ följer att det finns $\rho > 0$, $\beta_1 > 0$, $\beta_2 > 0$ så att $\|\nabla^2 f(x)^{-1}\| < \beta_1$ och $\|\nabla f(x) + \nabla^2 f(x)(x^* - x)\| < \beta_2 \|x - x^*\|^2$ för alla x där $\|x - x^*\| < \rho$.

Välj ρ och x_k så att $\|x_k - x^*\| < \rho$ och $\beta_1 \beta_2 \rho < C$ för något $C < 1$. Då får vi

$$
\begin{aligned}
x_{k+1} - x^* &= x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k) - x^* \\
&= -\nabla^2 f(x_k)^{-1} (\nabla f(x_k) + \nabla^2 f(x_k)(x^* - x_k)),
\end{aligned}
$$

vilket ger $\|x_{k+1} - x^*\| \le \beta_1 \beta_2 \|x_k - x^*\|^2 \le C \|x_k - x^*\|$. $\square$

Modifierad Newtonmetod

Antag att $\nabla f(x_k) \neq 0$. Om vi definierar p_k ur $B_k p_k = -\nabla f(x_k)$, där $B_k = B_k^T \succ 0$, blir p_k en descentriktning.

$B_k = I \Rightarrow$ steepest descent,
$B_k = \nabla^2 f(x_k) \Rightarrow$ Newtons metod då $\nabla^2 f(x_k) \succ 0$.

I en modifierad Newtonmetod låter man $B_k = \nabla^2 f(x_k)$ då $\nabla^2 f(x_k)$ är "tillräckligt positivt definit", och låter B_k vara en positivt definit approximation av $\nabla^2 f(x_k)$ om så inte är fallet.

Exempelvis kan man definiera en *modifierad Choleskyfaktorisering* så att $\nabla^2 f(x_k) + E_k = L_k D_k L_k^T$, där L_k är lågtriangulär med ettor på diagonalen, D_k är diagonal och positivt definit, samt $E_k = 0$ om $\nabla^2 f(x_k)$ är tillräckligt positivt definit.

Choleskyfaktorisering

Låt $B = B^T$. Partitionera $B = \begin{pmatrix} a & b^T \\ b & C \end{pmatrix}$. Antag att $a > 0$. Då blir

$$\begin{aligned} B &= \begin{pmatrix} 1 & 0 \\ \frac{1}{a}b & I \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & C - \frac{1}{a}bb^T \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{a}b^T \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ \frac{1}{a}b \end{pmatrix} a \begin{pmatrix} 1 & \frac{1}{a}b^T \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & C - \frac{1}{a}bb^T \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{a} \\ \frac{1}{\sqrt{a}}b \end{pmatrix} \begin{pmatrix} \sqrt{a} & \frac{1}{\sqrt{a}}b^T \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & C - \frac{1}{a}bb^T \end{pmatrix}. \end{aligned}$$

Choleskyfaktorisering, forts.

Antag att $B = B^T \succ 0$. Då är $a > 0$. Dessutom $C - \frac{1}{a}bb^T \succ 0$, eftersom

$$C - \frac{1}{a}bb^T = \begin{pmatrix} -\frac{1}{a}b & I \end{pmatrix} \begin{pmatrix} a & b^T \\ b & C \end{pmatrix} \begin{pmatrix} -\frac{1}{a}b^T \\ I \end{pmatrix}.$$

Alltså, om $B = B^T \succ 0$ kan vi skriva $B = LDL^T$, där L är lågtriangulär med ettor på diagonalen och D är diagonal med positiva diagonalelement.

Ekvivalent, låt $R = D^{1/2}L^T$. Då blir $B = R^T R$, där R är högtriangulär med positiva diagonalelement. Vi kallar R *Choleskyfaktorn* av B.

Choleskyfaktorisering, forts.

Antag att $B = B^T \succ 0$ och att vi vill lösa $Bp = -\nabla f(x)$.

- Bilda $B = R^T R$.
- Lös $R^T u = -\nabla f(x)$.
- Lös $Rp = u$.

Då R är triangulär är det enkelt att lösa med R och R^T.

Om $B = B^T \not\succ 0$ kan vi bilda $B + E = R^T R$, där E är diagonal med ickenegativa diagonalelement så att $B + E = R^T R$. Detta kallas *modifierad Choleskyfaktorisering*.

Vi kan då lösa $(B + E)p = -\nabla f(x)$, vilket leder till en form av *modifierad Newtonmetod*.

Tillräcklig descentriktning

För att en sökriktning p_k ska garantera konvergens krävs att den ska vara en *tillräcklig descentriktning*. Typiskt villkor är

$$-\frac{\nabla f(x_k)^T p_k}{\|\nabla f(x_k)\| \|p_k\|} \geq \sigma, \quad \text{där } \sigma \text{ är en positiv konstant.}$$

Detta innebär att p_k ska uppföra sig "tillräckligt mycket" som minusgradienten.

För sökriktning p_k ur $B_k p_k = -\nabla f(x_k)$ åstadkoms detta genom att se till att $\|B_k\| \leq M_1$ och $\|B_k^{-1}\| \leq M_2$, där M_1 och M_2 är positiva konstanter.

För en modifierad Newtonmetod kan modifikation av $\nabla^2 f(x_k)$ göras i modifierad Choleskyfaktorisering.

Linjesökning

I linjesökningen bestäms α_k som en approximativ lösning till $\min_{\alpha \geq 0} \varphi(\alpha)$, där $\varphi(\alpha) = f(x_k + \alpha p_k)$. Vi vill ha $f(x_{k+1}) < f(x_k)$, dvs $\varphi(\alpha_k) < \varphi(0)$. Detta villkor räcker dock ej för att visa konvergens.

Exempel på krav för att steget inte blir för långt:

$\varphi(\alpha) \leq \varphi(0) + \mu\alpha\varphi'(0)$, dvs (Armijo-villkor)

$f(x_k + \alpha p_k) \leq f(x_k) + \mu\alpha\nabla f(x_k)^T p_k$,

där $\mu \in (0, \frac{1}{2})$.

Exempel på krav för att steget inte blir för kort:

$|\varphi'(\alpha)| \leq -\eta\varphi'(0)$, dvs (Wolfe-villkor)

$|\nabla f(x_k + \alpha p_k)^T p_k| \leq -\eta\nabla f(x_k)^T p_k$,

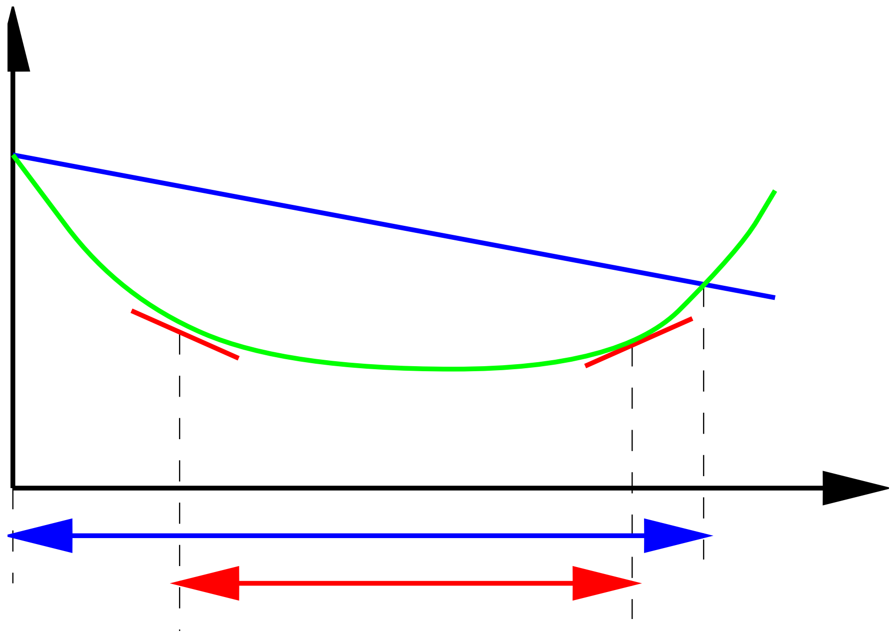
där $\eta \in (\mu, 1)$. Alternativt krav för att steget inte blir för kort:

Tag minsta ickenegativa heltal i sådant att

$f(x_k + 2^{-i} p_k) \leq f(x_k) + \mu 2^{-i}\nabla f(x_k)^T p_k$. ("Backtracking")

Illustration av linjesökningsvillkoren

Linjesökningsvillkoren av Wolfe-Armijotyp illustreras i nedanstående figur.



Linjesökningsvillkor

För att hitta $\bar{\alpha}$ så att Wolfe och Armijo-villkoren är uppfyllda kan vi betrakta $\hat{\varphi}(\alpha) = \varphi(\alpha) - \varphi(0) - \mu\alpha\varphi'(0)$.

Då blir $\hat{\varphi}(0) = 0$ och $\hat{\varphi}'(0) < 0$. Dessutom måste det finnas $\bar{\alpha} > 0$ så att $\hat{\varphi}(\bar{\alpha}) = 0$, annars blir φ obegränsad nedåt.

Alltså finns enligt medelvärdessatsen $\hat{\alpha} \in (0, \bar{\alpha})$ så att $\hat{\varphi}(\hat{\alpha}) < 0$ och $\hat{\varphi}'(\hat{\alpha}) = 0$.

Då $\mu < \eta$ får vi därmed $\varphi(\alpha) \le \varphi(0) + \mu\alpha\varphi'(0)$ och $|\varphi'(\alpha)| \le -\eta\varphi'(0)$ för α i en omgivning av $\hat{\alpha}$.

Exempelvis kan intervallhalvering kombinerat med polynomiell interpolation användas på $\hat{\varphi}$ för att hitta lämpligt α.