KTH Matematik

5B1817 Tillämpad ickelinjär optimering

Föreläsning 5

Metoder för problem utan bivillkor, forts.

Lösningar

För en given metod blir en *lösning* den bästa punkt metoden kan ge.

- Med en metod som använder förstaderivatsinformation blir en *lösning* typiskt en punkt som uppfyller $\nabla f(x^*) = 0$. Exempel: steepest descent.

- Med en metod som även använder andraderivatsinformation blir en *lösning* typiskt en punkt där $\nabla f(x^*) = 0$ och $\nabla^2 f(x^*) \succeq 0$. Exempel: Modifierad Newtonmetod. (Dock behöver man använda även negativa krökningsriktningar för att uppnå detta.)

OBS! Vi kan inte med någon av dessa metoder garantera att vi når en lokal minpunkt, om vi inte antar konvexitet eller något motsvarande. Vi kan dock testa om andra ordningens tillräckliga optimalitetsvillkor är uppfyllda om vi använder en andraderivatsmetod.

Gradientbaserade metoder för problem utan bivillkor

Betrakta återigen ett ickelinjärt programmeringsproblem utan bivillkor

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

Vi ska nu titta på metoder som använder förstaderivator, liksom steepest-descentmetoden, men på ett "bättre sätt".

Speciellt *kvasi-Newtonmetoder* och *konjugerade gradientmetoden*.

Båda har motivering från kvadratisk programmering.

Kvasi-Newtonuppdatering

En kvasi-Newtonmetod beräknar sökriktningen p_k ur $B_k p_k = -\nabla f(x_k)$, där $B_k = B_k^T \succ 0$ och låter $B_{k+1} = B_k + C_k$, där C_k är en symmetrisk matris av låg rang, typiskt rang ett eller rang två.

Antag att $f(x) = \frac{1}{2} x^T H x + c^T x$, där $H = H^T \succ 0$. Då blir $\nabla f(x_k + \alpha_k p_k) = \nabla f(x_k) + \alpha_k H p_k$. Alternativt kan vi skriva $H s_k = y_k$, där $s_k = \alpha_k p_k$ och $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$.

Motiverat av denna kvadratiska modell, krävs i en kvasi-Newtonmetod att B_{k+1} ska uppfylla *Kvasi-Newtonvillkoret* $B_{k+1} s_k = y_k$, där $s_k = \alpha_k p_k$ och $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$.

Kvasi-Newtonmetod

Iteration k i en linjesökande kvasi-Newtonmetod tar följande form i en given punkt x_k med Hessianapproximation B_k.

- Beräkna sökriktning p_k ur $B_k p_k = -\nabla f(x_k)$.
- Lös approximativt $\min_{\alpha \geq 0} f(x_k + \alpha p_k)$, vilket ger α_k.
- $x_{k+1} \leftarrow x_k + \alpha_k p_k$.
- $s_k \leftarrow x_{k+1} - x_k, \quad y_k \leftarrow \nabla f(x_{k+1}) - \nabla f(x_k)$
- $B_{k+1} \leftarrow B_k + C_k$, där $B_{k+1} s_k = y_k$ och $B_{k+1} \succ 0$ krävs.

Olika metoder baseras på olika val av C_k och α_k. Typiskt har C_k rang ett eller rang två.

Typiskt väljs $B_0 = I$.

Symmetriska rang-ett uppdateringen

Påstående. *Låt* B_k *vara en symmetrisk matris. Låt* $B_{k+1} = B_k + C_k$, *där* C_k *är symmetrisk med rang ett. Antag att* $B_{k+1}s_k = y_k$ *och* $(y_k - B_k s_k)^T s_k \neq 0$. *Då blir*

$$C_k = \frac{1}{(y_k - B_k s_k)^T s_k}(y_k - B_k s_k)(y_k - B_k s_k)^T.$$

Bevis. Se Nash och Sofer, Lemma 11.4. Vi kan skriva $B_{k+1} = B_k + \gamma_k w_k w_k^T$, där $\|w_k\| = 1$. Kvasi-Newtonvillkoret ger $B_k s_k + \gamma_k w_k w_k^T s_k = y_k$, dvs $w_k = \dfrac{1}{\gamma_k w_k^T s_k}(y_k - B_k s_k)$. Då $\|w_k\| = 1$ får vi $\dfrac{\gamma_k (y_k - B_k s_k)^T s_k}{\|y_k - B_k s_k\|} = \|y_k - B_k s_k\|$, vilket ger $\gamma_k = \dfrac{\|y_k - B_k s_k\|^2}{(y_k - B_k s_k)^T s_k}$.
Insättning ger $\gamma_k w_k w_k^T = C_k$. $\square$

Symmetriska rang-ett uppdateringen är unik. $B_{k+1} \succ 0$ ej garanterad.

BFGS-uppdateringen

Rang-två uppdateringen är inte unik. BFGS-uppdateringen ges av

$$B_{k+1} = B_k - \frac{1}{s_k^T B_k s_k} B_k s_k s_k^T B_k + \frac{1}{y_k^T s_k} y_k y_k^T.$$

Påstående. *Låt* $B_{=}B_k^T \succ 0$, *och låt* B_{k+1} *vara BFGS-uppdateringen av* B_k. *Då är* $B_{k+1} \succ 0$ *om och endast om* $y_k^T s_k > 0$.

Bevis. Se Nash och Sofer, Lemma 11.5. $\square$

Positiv definithet kan garanteras om Wolfe-villkor används i linjesökningen. Då får vi $|\nabla f(x_{k+1})^T s_k| \leq -\eta \nabla f(x_k)^T s_k$, vilket ger $y_k^T s_k \geq -(1-\eta)\nabla f(x_k)^T s_k > 0$.

Om R_k är Choleskyfaktor av B_k, dvs $B_k = R_k^T R_k$, kan BFGS-uppdateringen göras som en rang-ett uppdatering av R_k till R_{k+1}.

Kvasi-Newtonmetod på kvadratiskt problem

Antag att en kvasi-Newtonmetod appliceras på ett kvadratisk problem där $f(x) = \frac{1}{2}x^T Hx + c^T x$, $H = H^T \succ 0$. Antag också att exakt linjesökning utförs samt att $\{B_k\}$ är positivt definita. Då är problemet löst i högst n iterationer och $B_n = H$.

Detta innebär att en kvasi-Newtonmetod löser sitt "kvadratiska modellproblem" i ett ändligt antal iterationer, till skillnad från steepest-descentmetoden.

Man kan visa ovanstående genom att visa att en kvasi-Newtonmetod genererar *konjugerade* sökriktningar för ett kvadratiskt problem. På ett kvadratiskt problem blir metoden ekvivalent med *konjugerade gradientmetoden*.

Lösning av ekvationssystem med positivt definit matris

Antag att vi vill lösa $Hx = -c$, där $H = H^T \succ 0$.

Ekvivalent, minimera $f(x) = \frac{1}{2}x^T Hx + c^T x$.

Alternativ:

- Fatoriseringsmetod. Förslag: Choleskyfaktorisering, $H = R^T R$, R högtriangulär.
- Iterativ metod. Förslag: Konjugerade gradientmetoden.

Struktur hos H avgör vad som är mest effektivt.

Konjugerade riktningar

Konjugerade gradientmetoden är en iterativ metod som använder sig av *konjugerade* sökriktningar.

Definition. *En mängd vektorer i* $\mathbb{R}^n$, $\{p_i\}$, *sägs vara* konjugerade *med avseende på en positivt definit symmetrisk* $n \times n$*-matris* H *om* $p_i^T H p_j = 0$ *för* $i \neq j$.

Påstående. *Konjugerade riktningar är linjärt oberoende.*

Bevis. Låt p_i, $i = 1, \ldots, m$, vara konjugerade riktningar. Antag att $\sum_{i=1}^m \gamma_i p_i = 0$ för några γ_i, $i = 1, \ldots, m$. Då blir $\sum_{i=1}^m \gamma_i p_j^T H p_i = 0$ för en godtycklig vektor p_j bland de konjugerade riktningarna. Då riktningarna är konjugerade ger detta $\gamma_j p_j^T H p_j = 0$, och eftersom $H \succ 0$ och $p_j \neq 0$ får vi $\gamma_j = 0$. $\square$

Konjugerade riktningar, forts.

Låt $f(x) = \frac{1}{2}x^THx + c^Tx$, där $H = H^T \succ 0$.

Låt $y = \sum_{k=1}^{m} \alpha_k p_k$, där p_k, $k = 1, \dots, m$ är konjugerade. Då blir

$f(y) = \sum_{k=1}^{m} \left(\frac{\alpha_k^2}{2} p_k^T H p_k + \alpha_k c^T p_k \right)$, vilket ger $\alpha_k = -\dfrac{c^T p_k}{p_k^T H p_k}$.

Konjugerade riktningar ger alltså ett separabelt problem.

Linjära konjugerade gradientmetoden

Iteration k i konjugerade gradientmetoden tar följande form i en given punkt x_k med $g_k = Hx_k + c$ och p_{k-1} givna.

- $\beta_k \leftarrow \dfrac{g_k^T g_k}{g_{k-1}^T g_{k-1}}$ om $k \geq 1$.
- $p_k \leftarrow -g_k + \beta_k p_{k-1}$.
- $\alpha_k \leftarrow -\dfrac{g_k^T p_k}{p_k^T H p_k} = \dfrac{g_k^T g_k}{p_k^T H p_k}$.
- $x_{k+1} \leftarrow x_k + \alpha_k p_k$.
- $g_{k+1} \leftarrow g_k + \alpha_k H p_k$.

Initialt väljs $\beta_0 = 0$.

Då $g_k = 0$ är problemet löst. I praktiken $\|g_k\| < \epsilon$.

Linjära konjugerade gradientmetoden, forts.

Påstående. *Antag att* $\{p_k\}$ *och* $\{g_k\}$ *genereras med konjugerade gradientmetoden. Då gäller att* $g_k^T g_l = 0$, $g_k^T p_l = 0$ *och* $p_k^T H p_l = 0$ *för* $k > l$.

Bevis. Se Nash och Sofer, Theorem 12.1. $\square$

Detta innebär att sökriktningarna $\{p_k\}$ är konjugerade samt att gradienterna $\{g_k\}$ är inbördes ortogonala samt ortogonala mot alla tidigare sökriktningar.

En konsekvens är att linjära konjugerade gradientmetoden konvergerar i högst n iterationer.

I praktiken behöver H ofta *förkonditioneras* med en positivt definit matris M som "liknar" H men där det är "lätt" att lösa ekvationssystem $Mz_k = r_k$. I stället för att lösa $Hx = -c$ löser man $M^{-1}Hx = -M^{-1}c$.

Ickelinjära konjugerade gradientmetoder

Antag nu att $f(x)$ är en ickelinjär (ej nödvändigtvis kvadratisk) funktion.

Iteration k i en ickelinjär konjugerad gradientmetod i en punkt x_k kan skrivas på följande form.

- $\beta_k \leftarrow \dfrac{\nabla f(x_k)^T \nabla f(x_k)}{\nabla f(x_{k-1})^T \nabla f(x_{k-1})}$ om $k \geq 1$.
- $p_k \leftarrow -\nabla f(x_k) + \beta_k p_{k-1}$.
- Lös approximativt $\min_{\alpha \geq 0} f(x_k + \alpha p_k)$, vilket ger α_k.
- $x_{k+1} \leftarrow x_k + \alpha_k p_k$.

Initialt väljs $\beta_0 = 0$.

Då $\nabla f(x_k) = 0$ är problemet löst. I praktiken $\|\nabla f(x_k)\| < \epsilon$.

Idén är att försöka få bättre metod än steepest descent.