KTH Matematik

5B1817 Tillämpad ickelinjär optimering

Föreläsning 9

Sekvensiell kvadratisk programmering

Optimalitetsvillkor för ickelinjära programmeringsproblem

Betrakta ett ickelinjärt programmeringsproblem med likhetsbivillkor

$$(P_=) \quad \begin{array}{ll} \min & f(x) \\ \text{då} & g(x) = 0, \end{array} \qquad \text{där } f, g \in C^2,\ g : I\!R^n \to I\!R^m.$$

Om vi definierar lagrangefunktionen $\mathcal{L}(x,\lambda) = f(x) - \lambda^T g(x)$ är första ordningens optimalitetsvillkor $\nabla \mathcal{L}(x,\lambda) = 0$. Vi skriver dem som

$$\begin{pmatrix} \nabla_x \mathcal{L}(x,\lambda) \\ -\nabla_\lambda \mathcal{L}(x,\lambda) \end{pmatrix} = \begin{pmatrix} \nabla f(x) - A(x)^T \lambda \\ g(x) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

där $A(x)^T = \begin{pmatrix} \nabla g_1(x) & \nabla g_2(x) & \cdots & \nabla g_m(x) \end{pmatrix}$.

Första ordningens optimalitetsvillkor som ett ekvationssystem

Första ordningens nödvändiga optimalitetsvillkor kan alltså ses som ett ickelinjärt ekvationssystem med $n+m$ ekvationer och $n+m$ obekanta, x och λ, enligt

$$\begin{pmatrix} \nabla f(x) - A(x)^T\lambda \\ g(x) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

En Newtoniteration tar formen $\begin{pmatrix} x^+ \\ \lambda^+ \end{pmatrix} = \begin{pmatrix} x \\ \lambda \end{pmatrix} + \begin{pmatrix} p \\ \nu \end{pmatrix}$, där

$$\begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x,\lambda) & -A(x)^T \\ A(x) & 0 \end{pmatrix} \begin{pmatrix} p \\ \nu \end{pmatrix} = \begin{pmatrix} -\nabla f(x) + A(x)^T\lambda \\ -g(x) \end{pmatrix}.$$

Första ordningens optimalitetsvillkor som ett ekvationssystem, forts.

Newtonsystemet kan ekvivalent skrivas om som

$$\begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x,\lambda) & -A(x)^T \\ A(x) & 0 \end{pmatrix} \begin{pmatrix} p \\ \lambda + \nu \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ -g(x) \end{pmatrix},$$

alternativt

$$\begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x,\lambda) & A(x)^T \\ A(x) & 0 \end{pmatrix} \begin{pmatrix} p \\ -\lambda^+ \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ -g(x) \end{pmatrix}.$$

Vi föredrar formen med λ^+, då den direkt kan generaliseras till problem med olikhetsbivillkor.

Kvadratisk programmering med likhetsbivillkor

Jämför kvadratiskt programmeringsproblem

$$(EQP) \quad \begin{array}{ll} \text{min} & \frac{1}{2} p^T H p + c^T p \\ \text{då} & Ap = b, \\ & p \in I\!R^n, \end{array}$$

där optimallösningen p och multiplikatorvektorn λ^+ ges unikt av

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p \\ -\lambda^+ \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix},$$

om $Z^T H Z \succ 0$ och A har full radrang.

Newtoniteration och kvadratisk programmering med likhetsbivillkor

Jämför
$$\begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x,\lambda) & A(x)^T \\ A(x) & 0 \end{pmatrix} \begin{pmatrix} p \\ -\lambda^+ \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ -g(x) \end{pmatrix}$$

med
$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p \\ -\lambda^+ \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix}.$$

Identifiera:
$$\begin{aligned} \nabla^2_{xx}\mathcal{L}(x,\lambda) &\longleftrightarrow H \\ \nabla f(x) &\longleftrightarrow c \\ A(x) &\longleftrightarrow A \\ -g(x) &\longleftrightarrow b. \end{aligned}$$

Newtoniteration som QP-problem

En Newtoniteration för att lösa första ordningens nödvändiga optimalitetsvillkor till $(P_=)$ kan uppfattas som att lösa QP-problemet

$$
(QP_=) \quad \begin{array}{ll} \text{min} & \frac{1}{2} p^T \nabla^2_{xx} \mathcal{L}(x,\lambda) p + \nabla f(x)^T p \\ \text{då} & A(x)p = -g(x), \\ & p \in I\!R^n, \end{array}
$$

och låta $x^+ = x + p$ samt λ^+ ges av multiplikatorerna till $(QP_=)$.

Problemet $(QP_=)$ är väldefinierat med unik optimallösning p och multiplikatorvektor λ^+ om $Z(x)^T \nabla^2_{xx} \mathcal{L}(x,\lambda) Z(x) \succ 0$ och $A(x)$ har full radrang, där $Z(x)$ är en matris vars kolumner bildar bas för null$(A(x))$.

En SQP-iteration för problem med likhetsbivillkor

Givet x, λ så att $Z(x)^T \nabla^2_{xx} \mathcal{L}(x,\lambda) Z(x) \succ 0$ och $A(x)$ har full radrang, tar en Newtoniteration följande form.

- Tag fram optimallösning p och multiplikatorvektor λ^+ till

$$
(QP_=) \qquad \begin{array}{ll} \text{min} & \frac{1}{2} p^T \nabla^2_{xx} \mathcal{L}(x,\lambda) p + \nabla f(x)^T p \\ \text{då} & A(x) p = -g(x), \\ & p \in I\!R^n, \end{array}
$$

- $x \leftarrow x + p, \qquad \lambda \leftarrow \lambda^+$.

Vi kallar denna metod *sekvensiell kvadratisk programmering* (SQP).

OBS! $(QP_=)$ löses genom att lösa ett linjärt ekvationssystem.

OBS! x och λ har givna numeriska värden i $(QP_=)$.

Konvergenshastighet för SQP-metod för problem med likhetsbivillkor

Sats. *Antag att* $f \in C^3$, $g \in C^3$ *på* R^n *och att* x^* *är en reguljär punkt som uppfyller andra ordningens tillräckliga optimalitetsvillkor till* $(P_=)$. *Om SQP-metoden (med steglängd ett) startas i en punkt tillräckligt nära* x^*, λ^* *så är den väldefinierad och konvergerar mot* x^*, λ^* *med konvergenshastighet minst två.*

Bevis. I en omgivning av x^*, λ^* är $Z(x)^T \nabla^2_{xx} \mathcal{L}(x, \lambda) Z(x) \succ 0$ och $\begin{pmatrix} \nabla^2_{xx} \mathcal{L}(x, \lambda) & A(x)^T \\ A(x) & 0 \end{pmatrix}$ ickesingulär. Subproblemet $(QP_=)$ blir därmed väldefinierat och resultatet följer från kvadratisk konvergenshastighet av Newtons metod. $\square$

OBS! Iterationspunkterna är normalt *inte* tillåtna till $(P_=)$, men kommvergerar mot en optimal punkt.

SQP-metod för problem med likhetsbivillkor

Vi har hittills diskuterat SQP för $(P_=)$ i "idealt" fall. Några kommentarer:

- Om $Z(x)^T\nabla^2_{xx}\mathcal{L}(x,\lambda)Z(x) \not\succ 0$ kan vi ersätta $\nabla^2_{xx}\mathcal{L}(x,\lambda)$ med B i $(QP_=)$ där B är en symmetrisk approximation av $\nabla^2_{xx}\mathcal{L}(x,\lambda)$ som uppfyller $Z(x)^TBZ(x) \succ 0$.
- Man kan använda kvasi-Newtonapproximation B av $\nabla^2_{xx}\mathcal{L}(x,\lambda)$.
- Om $A(x)$ inte har full radrang kan $A(x)p = -g(x)$ sakna lösningar. Detta kan exempelvis hanteras genom att införa "elastiska" variabler. Vi behandlar inte detta här.
- Vi har visat lokala konvergensegenskaper. För att få konvergens från godtycklig startpunkt kan vi införa en *meritfunktion* och utföra linjesökning.

Exempel på meritfunktion för SQP på $(P_=)$

En meritfunktion består typiskt av en viktning av optimalitet och tillåtenhet. Exempel är den utökade lagrangefunktionen $M_\mu(x) = f(x) - \lambda(x)^T g(x) + \frac{1}{2\mu} g(x)^T g(x)$, där μ är en positiv parameter och $\lambda(x) = (A(x)A(x)^T)^{-1}A(x)\nabla f(x)$. (Vektorn $\lambda(x)$ är här minstakvadratlösningen till $A(x)^T\lambda = \nabla f(x)$.)

Då blir SQP-riktningen p en descentriktning till M_μ i x om μ är tillräkligt nära noll och $Z(x)^T B Z(x) \succ 0$, se Nocedal och Wright sid 545–546.

Vi kan därför göra linjesökning på M_μ i x-led och definiera $\lambda(x) = (A(x)A(x)^T)^{-1}A(x)^T\nabla f(x)$.

Idealt väljs steglängden $\alpha = 1$.

Vi tänker oss den "rena" metoden, där $\alpha = 1$ och λ^+ ges av $(QP_=)$.

SQP för problem med olikhetsbivillkor

I SQP-suproblemet $(QP_=)$ approximeras bivillkoren med en linearisering kring x, dvs kravet på p blir $g_i(x) + \nabla g_i(x)^T p = 0$, $i = 1, \ldots, m$.

För ett olikhetsbivillkor $g_i(x) \geq 0$ generaliseras detta till kravet $g_i(x) + \nabla g_i(x)^T p \geq 0$.

En SQP-metod ger här i varje iteration en prediktion av de aktiva bivillkoren till (P) genom de bivillkor som är aktiva i QP-subproblemet.

QP-subproblemet ger ickenegativa multiplikatorer för olikhetsbivillkoren.

SQP-subproblem för ickelinjärt programmeringsproblem

Problemet

$$(P) \qquad \begin{array}{ll} \min & f(x) \\ \text{då} & g_i(x) \geq 0, \; i \in \mathcal{I}, \\ & g_i(x) = 0, \; i \in \mathcal{E}, \\ & x \in I\!R^n, \end{array} \qquad \text{där } f, g \in C^2, \; g : I\!R^n \to I\!R^m,$$

får i en viss punkt x, λ ett SQP-subproblem

$$(QP) \qquad \begin{array}{ll} \min & \frac{1}{2} p^T \nabla^2_{xx} \mathcal{L}(x, \lambda) p + \nabla f(x)^T p \\ \text{då} & \nabla g_i(x)^T p \geq -g_i(x), \; i \in \mathcal{I}, \\ & \nabla g_i(x)^T p = -g_i(x), \; i \in \mathcal{E}, \\ & p \in I\!R^n, \end{array}$$

vilket har optimallösning p och lagrangemultiplikatorvektor λ^+.

En SQP-iteration för ickelinjär programmering

Givet x, λ så att $\nabla^2_{xx}\mathcal{L}(x,\lambda) \succ 0$, tar en SQP-iteration för (P) följande form.

- Tag fram optimallösning p och multiplikatorvektor λ^+ till

$$
(QP) \quad
\begin{array}{ll}
\text{min} & \frac{1}{2}p^T\nabla^2_{xx}\mathcal{L}(x,\lambda)p + \nabla f(x)^T p \\
\text{då} & \nabla g_i(x)^T p \geq -g_i(x), \;\; i \in \mathcal{I}, \\
 & \nabla g_i(x)^T p = -g_i(x), \;\; i \in \mathcal{E}, \\
 & p \in I\!R^n.
\end{array}
$$

- $x \leftarrow x + p, \qquad \lambda \leftarrow \lambda^+.$

Observera att $\lambda_i \geq 0$, $i \in \mathcal{I}$, bibehålls tack vare att λ^+ är lagrangemultiplikatorer till (QP).

Konvergenshastighet för SQP-metod

Sats. *Antag att $f \in C^3$, $g \in C^3$ på R^n och att x^* är en reguljär punkt som uppfyller andra ordningens tillräckliga optimalitetsvillkor till (P) med strikt komplementaritet. Om SQP-metoden (med steglängd ett) startas i en punkt tillräckligt nära x^*, λ^* så är den väldefinierad, har samma bindande bivillkor som (P) har i x^* och konvergerar mot x^*, λ^* med konvergenshastighet minst två.*

Bevis. Idé: Om vi bortser från bivillkoren som inte är bindande i x^* konvergerar SQP-metoden för detta problem enligt diskussion från likhetsfallet. Därmed går $\|p\|$ mot noll med minst kvadratisk hastighet.

För bivillkor i får vi $g_i(x) + \nabla g_i(x)^T p \geq g_i(x) - \|\nabla g_i(x)\| \|p\|$. Om x är tillräckligt nära x^* och bivillkor i inte är bindande i x^* blir $g_i(x) > 0$, och om $\|p\|$ är tillräckligt liten blir därmed
$g_i(x) + \nabla g_i(x)^T p \geq g_i(x) - \|\nabla g_i(x)\| \|p\| > 0$. $\square$

SQP-metod för ickelinjär programmering

Vi har diskuterat "idealfallet". Några kommentarer:

- Om $\nabla^2_{xx}\mathcal{L}(x,\lambda) \not\succ 0$ kan vi ersätta $\nabla^2_{xx}\mathcal{L}(x,\lambda)$ med B i (QP) där B är en symmetrisk approximation av $\nabla^2_{xx}\mathcal{L}(x,\lambda)$ som uppfyller $B \succ 0$.

- Man kan använda kvasi-Newtonapproximation B av $\nabla^2_{xx}\mathcal{L}(x,\lambda)$.

- QP-subproblemet kan sakna tillåtna lösningar. Detta kan exempelvis hanteras genom att införa "elastiska" variabler. Vi behandlar inte detta här.

- Vi har visat lokala konvergensegenskaper. För att få konvergens från godtycklig startpunkt kan vi införa en *meritfunktion* och utföra linjesökning. Vi behandlar inte detta här för olikhetsproblemsfallet.