KTH Matematik

5B1817 Tillämpad ickelinjär optimering

Föreläsning 10

Inrepunktsmetoder. Straff- och barriärmetoder

Inrepunktsmetoder

- Inrepunktsmetoder används som ett sammanfattande namn för metoder av straff- och barriärtyp för ickelinjär programmering.
- Vi har tittat på inrepunktsmetoder för kvadratisk programmering tidigare. Tittar nu på generell ickelinjär programmering.
- Straff- och barriärmetoder i primal tappning är från 60-talet. De har några mindre bra egenskaper på grund av illakonditionering. Återupplivade 1984 för linjärprogrammering.
- Primal-duala inrepunktsmetoder är 90-talsmetoder. Har "bättre" uppförande.
- Vi tittar på straff- och barriärmetoder separat för enkelhets skull.

Straffunktion

Betrakta ett ickelinjärt programmeringsproblem med likhetsbivillkor

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

För en positiv parameter μ, bilda den kvadratiska straffunktionen $P_\mu(x) = f(x) + \dfrac{1}{2\mu} g(x)^T g(x)$.

Nödvändigt villkor för lokalt minimum av $P_\mu(x)$ är $\nabla P_\mu(x) = 0$, där $\nabla P_\mu(x) = \nabla f(x) + \dfrac{1}{\mu} A(x)^T g(x)$, med $A(x)^T = \begin{pmatrix} \nabla g_1(x) & \dots & \nabla g_m(x) \end{pmatrix}$.

Straffunktion, forts.

Om $x(\mu)$ är en lokal minpunkt till min $P_\mu(x)$ gäller att $\nabla f(x(\mu)) + \dfrac{1}{\mu} A(x(\mu))^T g(x(\mu)) = 0$.

Påstående. *Låt $x(\mu)$ beteckna en lokal minpunkt till* min $P_\mu(x)$. *Under lämpliga förutsättningar gäller*

$$\lim_{\mu \to 0} x(\mu) = x^*, \qquad \lim_{\mu \to 0} \frac{g(x(\mu))}{\mu} = -\lambda^*,$$

där x^* är en lokal minpunkt till $(P_=)$ och λ^* är motsvarande lagrangemultiplikatorvektor.

Straffunktion, forts.

Låt $\lambda(\mu) = -\dfrac{g(x(\mu))}{\mu}$.

Då blir $\nabla P_\mu(x(\mu)) = 0 \iff \nabla f(x(\mu)) - A(x(\mu))^T\lambda(\mu) = 0$.

Det vill säga $x(\mu)$ och $\lambda(\mu)$ löser ekvationssystemet

$$\begin{aligned} \nabla f(x) - A(x)^T\lambda &= 0, \\ \frac{g(x)}{\mu} + \lambda &= 0. \end{aligned}$$

Om andra blocket av ekvationer multipliceras med μ får vi ekvivalent

$$\begin{aligned} \nabla f(x) - A(x)^T\lambda &= 0, \\ g(x) + \mu\lambda &= 0. \end{aligned}$$

Störning av första ordningens nödvändiga optimalitetsvillkor. $\mu = 0$ ger optimalitetsvillkoren.

Straffunktion, exempel

$$\begin{array}{ll} \text{min} & -x_1x_2 - x_1x_3 - x_2x_3 \\ \text{då} & x_1 + x_2 + x_3 - 3 = 0. \end{array}$$

Det gäller att $x^* = (1\ 1\ 1)^T$ med $\lambda^* = -2$.

$$P_\mu(x) = -x_1x_2 - x_1x_3 - x_2x_3 + \frac{1}{2\mu}(x_1 + x_2 + x_3 - 3)^2.$$

$$\nabla P_\mu(x) = \begin{pmatrix} -x_2 - x_3 \\ -x_1 - x_3 \\ -x_1 - x_2 \end{pmatrix} + \frac{x_1 + x_2 + x_3 - 3}{\mu} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

$\nabla^2 P_\mu(x) = \left(\dfrac{1}{\mu} - 1\right) ee^T + I$, där $e = (1\ 1\ 1)^T$.

Straffunktion, exempel, forts.

Vi får $\nabla^2 P_\mu(x) \succ 0$ om $\mu < 3/2$.

Därmed har P_μ unik minpunkt för $\mu < 3/2$.

(P_μ saknar minpunkt för $\mu > 3/2$.)

Kravet $\nabla P_\mu(x(\mu)) = 0$ ger $x(\mu) = \dfrac{3}{3 - 2\mu} e$.

Därmed blir $\lambda(\mu) = -\dfrac{g(x(\mu))}{\mu} = \dfrac{6}{3 - 2\mu}$.

Insättning ger $\lim_{\mu \to 0} x(\mu) = e = x^*$ och $\lim_{\mu \to 0} \lambda(\mu) = 2 = -\lambda^*$.

Straffunktionsmetod

En straffunktionsmetod hittar approximationer till $x(\mu)$, $\lambda(\mu)$ för avtagande värden på μ.

En primal-dual metod tar Newtoniterationer på det primal-duala straffunktionsekvationssystemet

$$\begin{aligned} \nabla f(x) - A(x)^T\lambda &= 0, \\ g(x) + \mu\lambda &= 0. \end{aligned}$$

Newtoniterationen blir

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

Notera att $\mu = 0$ ger SQP-iteration.

Straffunktionsmetod, forts.

Notera att Δx löser

$$\left(\nabla^2_{xx}\mathcal{L}(x,\lambda)+\frac{1}{\mu}A(x)^TA(x)\right)\Delta x=-\left(\nabla f(x)+\frac{1}{\mu}A(x)^Tg(x)\right).$$

Om $\nabla^2_{xx}\mathcal{L}(x,\lambda)+\frac{1}{\mu}A(x)^TA(x)\succ 0$ blir Δx descentriktning till P_μ i x.

Om $\nabla^2_{xx}\mathcal{L}(x,\lambda)+\frac{1}{\mu}A(x)^TA(x)\not\succ 0$ kan vi ersätta $\nabla^2_{xx}\mathcal{L}(x,\lambda)$ med B så att $B+\frac{1}{\mu}A(x)^TA(x)\succ 0$ och lösa

$$\begin{pmatrix} B & A(x)^T \\ A(x) & -\mu I \end{pmatrix}\begin{pmatrix} \Delta x \\ -\Delta\lambda \end{pmatrix}=-\begin{pmatrix} \nabla f(x)-A(x)^T\lambda \\ g(x)+\mu\lambda \end{pmatrix}.$$

För konvexa problem är $g(x)=Ax-b$, och då väljs ofta $\mu=0$.

Vi väljer fortsättningsvis $\mu=0$ för linjära bivillkor.

Barriärfunktion

Betrakta ett ickelinjärt programmeringsproblem med olikhetsbivillkor

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

Vi antar att $\{x\in I\!R^n : g(x)>0\}\neq\emptyset$ och kräver $g(x)>0$ "implicit".

För en positiv parameter μ, bilda den logaritmiska barriärfunktionen $B_\mu(x)=f(x)-\mu\sum_{i=1}^m \ln g_i(x)$.

Nödvändigt villkor för lokalt minimum av $B_\mu(x)$ är $\nabla B_\mu(x)=0$, där $\nabla B_\mu(x)=\nabla f(x)-\mu\sum_{i=1}^m\frac{1}{g_i(x)}\nabla g_i(x)=\nabla f(x)-\mu A(x)^TG(x)^{-1}e$, med $G(x)=\text{diag}(g(x))$ och $e=(1\ 1\ \ldots 1)^T$.

Barriärfunktion, forts.

Om $x(\mu)$ är en lokal minpunkt till $\min_{x:g(x)>0} B_\mu(x)$ gäller att $\nabla f(x(\mu))-\mu A(x(\mu))^TG(x(\mu))^{-1}e=0$.

Påstående. *Låt $x(\mu)$ beteckna en lokal minpunkt till $\min_{x:g(x)>0} B_\mu(x)$. Under lämpliga förutsättningar gäller*

$$\lim_{\mu\to 0} x(\mu)=x^*,\quad \lim_{\mu\to 0}\mu G(x(\mu))^{-1}e=\lambda^*,$$

där x^* är en lokal minpunkt till $(P_\geq)$ och λ^* är motsvarande lagrangemultiplikatorvektor.

OBS! Det gäller att $g(x(\mu))>0$.

Barriärfunktion, forts.

Låt $\lambda(\mu)=\mu G(x(\mu))^{-1}e$, dvs $\lambda_i(\mu)=\dfrac{\mu}{g_i(x(\mu))}$, $i=1,\ldots,m$.

Då blir $\nabla B_\mu(x(\mu))=0 \iff \nabla f(x(\mu))-A(x(\mu))^T\lambda(\mu)=0$.

Det vill säga $x(\mu)$ och $\lambda(\mu)$ löser ekvationssystemet

$$\begin{aligned} \nabla f(x)-A(x)^T\lambda&=0, \\ \lambda_i-\frac{\mu}{g_i(x)}&=0,\quad i=1,\ldots,m, \end{aligned}$$

där vi också kräver $g(x)>0$ och $\lambda>0$. Om andra blocket av ekvationer multipliceras med $G(x)$ får vi ekvivalent

$$\begin{aligned} \nabla f(x)-A(x)^T\lambda&=0, \\ g_i(x)\lambda_i-\mu&=0,\quad i=1,\ldots,m. \end{aligned}$$

Störning av första ordningens nödvändiga optimalitetsvillkor.

Barriärfunktion, exempel

$$\begin{array}{ll} \text{min} & \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) \\ \text{då} & x_1 + x_2 + x_3 - 3 \geq 0. \end{array}$$

Det gäller att $x^* = (1\ 1\ 1)^T$ med $\lambda^* = 1$.

$B_\mu(x) = \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) - \mu \ln(x_1 + x_2 + x_3 - 3)$.

$$\nabla B_\mu(x) = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} - \frac{\mu}{x_1 + x_2 + x_3 - 3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

$\nabla^2 B_\mu(x) \succ 0$ då $x_1 + x_2 + x_3 - 3 > 0$.

Barriärfunktion, exempel, forts.

Eftersom $\nabla^2 B_\mu(x) \succ 0$ får B_μ unik minpunkt för alla $\mu > 0$.

Kravet $\nabla B_\mu(x(\mu)) = 0$ ger $x(\mu) = \left(\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{\mu}{3}}\right) e.$

Därmed blir $\lambda(\mu) = \dfrac{\mu}{g(x(\mu))} = \left(\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{\mu}{3}}\right).$

Insättning ger $\lim_{\mu \to 0} x(\mu) = e = x^*$ och $\lim_{\mu \to 0} \lambda(\mu) = 1 = \lambda^*$.

Barriärfunktionsmetod

En barriärfunktionsmetod hittar approximationer till $x(\mu)$, $\lambda(\mu)$ för avtagande värden på μ.

En primal-dual metod tar Newtoniterationer på det primal-duala barriärfunktionsekvationssystemet

$$\begin{aligned} \nabla f(x) - A(x)^T \lambda &= 0, \\ G(x)\lambda - \mu e &= 0. \end{aligned}$$

Newtonsteget Δx, $\Delta \lambda$ ges av

$$\begin{pmatrix} \nabla^2_{xx} \mathcal{L}(x, \lambda) & A(x)^T \\ \Lambda A(x) & -G(x) \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta \lambda \end{pmatrix} = - \begin{pmatrix} \nabla f(x) - A(x)^T \lambda \\ G(x)\lambda - \mu e \end{pmatrix},$$

där $\Lambda = \text{diag}(\lambda)$.

En iteration i en primal-dual barriärfunktionsmetod

En iteration i en primal-dual barriärfunktionsmetod tar följande form, givet $\mu > 0$, x så att $g(x) > 0$, och $\lambda > 0$.

- Beräkna Δx, $\Delta \lambda$ ur
$$\begin{pmatrix} \nabla^2_{xx} \mathcal{L}(x, \lambda) & A(x)^T \\ \Lambda A(x) & -G(x) \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta \lambda \end{pmatrix} = - \begin{pmatrix} \nabla f(x) - A(x)^T \lambda \\ G(x)\lambda - \mu e \end{pmatrix}.$$
- Välj "lämplig" steglängd α så att $g(x + \alpha \Delta x) > 0$, $\lambda + \alpha \Delta \lambda > 0$.
- $x \leftarrow x + \alpha \Delta x, \quad \lambda \leftarrow \lambda + \alpha \Delta \lambda$.
- Om (x, λ) "tillräckligt nära" $(x(\mu), \lambda(\mu))$, reducera μ.

Barriärfunktionsmetod, forts.

Notera att Δx löser

$$\left(\nabla^2_{xx}\mathcal{L}(x,\lambda) + A(x)^T\Lambda G(x)^{-1}A(x)\right)\Delta x = -\left(\nabla f(x) - \mu A(x)^T G(x)^{-1} e\right).$$

Om $\nabla^2_{xx}\mathcal{L}(x,\lambda) + A(x)^T\Lambda G(x)^{-1}A(x) \succ 0$ blir Δx descentriktning till B_μ i x.

Om $\nabla^2_{xx}\mathcal{L}(x,\lambda) + A(x)^T\Lambda G(x)^{-1}A(x) \not\succ 0$ kan vi ersätta $\nabla^2_{xx}\mathcal{L}(x,\lambda)$ med B så att $B + A(x)^T\Lambda G(x)^{-1}A(x) \succ 0$ och lösa

$$\begin{pmatrix} B & A(x)^T \\ \Lambda A(x) & -G(x) \end{pmatrix}\begin{pmatrix} \Delta x \\ -\Delta\lambda \end{pmatrix} = -\begin{pmatrix} \nabla f(x) - A(x)^T\lambda \\ G(x)\lambda - \mu e \end{pmatrix}.$$

Inrepunktsmetod

Barriärmetoder kallas *inrepunktsmetoder* då de generar iterationspunkter i det inre av det tillåtna området, dvs $g(x) > 0$ bibehålls i alla iterationspunkter.

Även straff/barriärmetoder räknas till inrepunktsmetoder.

Vi studerar speciellt *primal-duala inrepunktsmetoder*, som baseras på Newtoniterationer på det primal-duala ickelinjära ekvationssystemet.

Inrepunktsmetod, forts.

Som barriärmetoden framställts har vi krävt att $\{x \in I\!R^n : g(x) > 0\} \neq \emptyset$, och speciellt att initialpunkten uppfyller $g(x) > 0$.

Ett alternativ är att skriva om $(P_\geq)$ som

$$(P_\geq) \quad \begin{array}{ll} \min & f(x) \\ \text{då} & g(x) - s = 0, \quad s \geq 0. \end{array}$$

Vi kan då initialt välja $s > 0$, men kräver *inte* initialt att $g(x) - s = 0$.

Så gjorde vi då vi tog fram inrepunktsmetod för QP-problem.

Inrepunktsmetod, forts.

Om straffparametern till bivillkoret $g(x) - s = 0$ sätts till noll får vi det primal-duala ickelinjära ekvationssystemet

$$\begin{aligned} \nabla f(x) - A(x)^T\lambda &= 0, \\ g(x) - s &= 0, \\ S\lambda - \mu e &= 0. \end{aligned}$$

Om s elimineras genom att sätta $g(x) = s$ får vi det "ursprungliga" systemet.

Vi kräver dock inte $g(x) = s$, utan endast $s > 0$.

Inrepunktsmetod, forts.

Newtonsteget Δx, Δs, $\Delta\lambda$ ges av

$$\begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x,\lambda) & 0 & -A(x)^T \\ A(x) & -I & 0 \\ 0 & \Lambda & S \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta\lambda \end{pmatrix} = - \begin{pmatrix} \nabla f(x) - A(x)^T\lambda \\ g(x) - s \\ S\lambda - \mu e \end{pmatrix}.$$

Man kan eliminera Δs och få

$$\begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x,\lambda) & A(x)^T \\ \Lambda A(x) & -S \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta\lambda \end{pmatrix} = - \begin{pmatrix} \nabla f(x) - A(x)^T\lambda \\ G(x)\lambda - \mu e \end{pmatrix},$$

vilket har samma struktur som “ursprungliga” systemet för $(P_{\geq})$.

Inrepunktsmetod, forts.

Ekvationssystemet kan symmetriseras till formen

$$\begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x,\lambda) & A(x)^T \\ A(x) & -S\Lambda^{-1} \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta\lambda \end{pmatrix} = - \begin{pmatrix} \nabla f(x) - A(x)^T\lambda \\ G(x) - \mu\Lambda^{-1} e \end{pmatrix}.$$

Matrisen är här symmetrisk men blir illakonditionerad nära optimum. (Typiskt $s_i/\lambda_i \to 0$ eller $s_i/\lambda_i \to \infty$.)

Med en sådan symmetrisering kan man modifiera $\nabla^2_{xx}\mathcal{L}(x,\lambda)$ genom en faktorisering av den symmetriska matrisen.

Illakonditioneringen är i allmänhet “ofarlig”.

Inrepunktsmetod, forts.

- Inrepunktsmetoder är ofta effektiva på konvexa problem. Då används typiskt ingen straffparameter på likhetsbivillkoren. (De är linjära.)
- För LP och QP är antalet iterationer typiskt av storleksordningen 10-20. Varje iteration blir dock tyngre då problemstorleken växer.
- För ickekonvexa problem är situationen väsentligt mer komplex. Vi täcker inte detta i kursen mer än vad som diskuterats här.