Sekventiell linjärprogrammering (SLP)

Iterativ lösningsmetod för olinjär optimering med bivillkor.

Problem: $\min f(x)$

$$(P) \quad \text{då } g_i(x) \leq 0, \quad i = 1, \ldots, m$$

där både $f \in C^1$ och alla $g_i \in C^1$.
(Likhetsvillkor kan också hanteras.)

Metodprincip:

- Linjärapproximera målfunktionen och bivillkorsfunktionerna med första ordningens Taylor-utvecklingar,
- Lös linjärapproximerade problem,
- inkludera explicita steglängdsrestriktioner för att förhindra att linjärapproximationerna används alltför långt bort från punkterna där de skapats (och där de därför kan vara opålitliga),
- använd en straffunktion för att mäta målfunktionsvärde och grad av otillåtenhet tillsammans

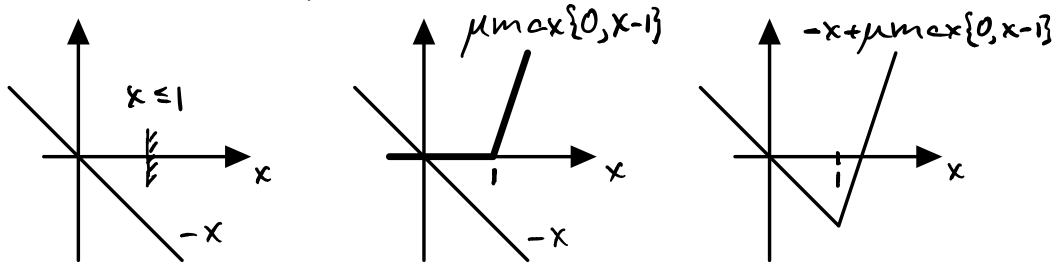
Låt $\mu > 0$ vara en straffparameter.

Betrakta den <u>exakta</u> straffunktionen
$F_E : R^n \to R$ med

$$F_E(x) = f(x) + \mu \sum_{i=1}^{m} \max\{0, g_i(x)\}.$$

Observera att $F_E(x)$ <u>inte</u> är differentierbar i punkter där $g_i(x) = 0$ gäller för något i.

Enkelt exempel: betrakta min $-x$ då $x \le 1$



Med ett tillräckligt stort värde på μ gäller att det obegränsade problemet $\min_{x \in R^n} F_E(x)$ är ekvivalent med (P).
(Om f och alla g_i är konvexa funktioner så gäller att μ är tillräckligt stor om μ är minst lika stor som den största optimala KKT-multiplikatorn.)
Problemet $\min_{x \in R^n} F_E(x)$ kan skrivas som

$$\begin{aligned} \min\ & f(x) + \mu \sum_{i=1}^{m} y_i \\ \text{då}\ & g_i(x) \le y_i, \quad i = 1, \dots, m \\ & y_i \ge 0, \quad i = 1, \dots, m, \end{aligned}$$

ty pga minimeringen fås $y_i = \max\{0, g_i(x)\}$.

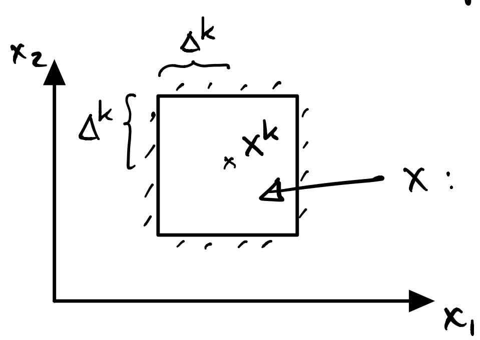
Betrakta en iterationspunkt x^k.
Linjärapproximera f och alla g_i i x^k och konstruera den linjäriserade straffunktionen $F_{EL} : \mathbb{R}^n \to \mathbb{R}$ med

$$F_{EL}(x) = f(x^k) + \nabla f(x^k)^T (x - x^k) + \mu \sum_{i=1}^{m} \max\{0, g_i(x^k) + \nabla g_i(x^k)^T (x - x^k)\}.$$

Approximationerna är tillräckligt tillförlitliga endast i en liten omgivning till x^k, varför en steglängdsbegränsning (eng. trust region) införs. Beräkningsmässigt är det lämpligast om begränsningen beskrivs av enkla linjära villkor, varför max-normen används:

$$\|x - x^k\|_\infty \le \Delta^k \iff -\Delta^k \le x_j - x_j^k \le \Delta^k, \ j = 1, \dots, n,$$

där $\Delta^k > 0$ är en parameter.



$x : x_j^k - \Delta^k \le x_j \le x_j^k + \Delta^k, \ j = 1, 2$

Den linjäriserade straffunktionen minimeras därefter över det område som definieras av steglängdsbegränsningen.

Analogt med omskrivningen ovan kan detta problem skrivas som

$$\begin{aligned}
&\min\ f(x^k)+\nabla f(x^k)^T(x-x^k)+\mu\sum_{i=1}^{m} y_i \\
&\text{då } g_i(x^k)+\nabla g_i(x^k)^T(x-x^k)\le y_i\ ,\ i=1,\dots,m \\
&(LP)\qquad\qquad -\Delta^k \le x_j - x_j^k \le \Delta_k\ ,\ j=1,\dots,n \\
&\qquad\qquad\qquad\qquad y_i \ge 0\ ,\ i=1,\dots,m,
\end{aligned}$$

vilket är ett linjärt optimeringsproblem som utgör en (lokal) approximation av (P). (LP) har alltid en tillåten lösning, eftersom y_i, $i=1,\dots,m$, fungerar som artificiella variabler, och ett begränsat optimum, pga steglängdsbegränsningen.

Om något $g_i(x)$ är icke-konvext så kan variablerna y_i, $i=1,\dots,m$, vara nödvändiga för att säkerställa att (LP) är tillåtet, även om (P) är tillåtet. Ett trivialt exempel på detta: $g(x)=1-x^2$ ger $\{x \mid g(x)\le 0\}= \{x \mid x\le -1 \text{ eller } x\ge 1\}$, men om $g(x)$ linjär-approximeras i $x=0$ så fås $\{x \mid g(0)+g'(0)(x-0)\le 0\}= = \{x \mid 1+0(0-0)\le 0\} = \emptyset$.

Om däremot alla g_i är konvexa så

gäller att $g_i(x^k)+\nabla g_i(x^k)^T(x-x^k) \leq g(x)$, $i=1,\ldots,m$, varför (LP) då alltid är tillåtet med alla $s_i=0$, givet att (P) är tillåtet. Dock behöver inte därför $s_i=0$, $i=1,\ldots,m$, gälla i ett optimum i (LP).

Man kan visa att om μ är tillräckligt stor så är x^k optimal i (LP) [dvs linjäringspunkten är själv optimal i det linjäriserade problemet] omm x^k är en KKT-punkt för (P).

Iterativ metod:

Låt $\bar{x}^k \neq x^k$ vara optimal i (LP).

Definiera $\Delta F_{E_k} = F_E(x^k) - F_E(\bar{x}^k)$ [= förbättring i värde på straffunktionen].

Definiera $\Delta F_{EL_k} = \underbrace{F_{E_k}(x^k)}_{=F_E(x^k)} - \underbrace{F_{E_k}(\bar{x}^k)}_{=\text{prediktion av } F_E(\bar{x}^k)}$ [= förbättring i värde på den linjäriserade straff-funktionen = en prediktion av ΔF_{E_k}].

Relationen mellan ΔF_{E_k} och ΔF_{EL_k} styr hur x^{k+1} väljs och Δ^k förändras.

Olika fall:

om ΔF_{E_k} avviker mycket från $\Delta F_{EL_k} \rightarrow$ låt $x^{k+1}=x^k$ och minska Δ^k

om ΔF_{E_k} avviker lite från ΔF_{EL_k} →
låt $x^{k+1} = \bar{x}^k$ och behåll Δ^k

om ΔF_{E_k} avviker mycket lite från ΔF_{EL_k} →
låt $x^{k+1} = \bar{x}^k$ och öka Δ^k

(Vad som betraktas som stora och små avvikelser mellan ΔF_{E_k} och ΔF_{EL_k} är algoritmparametrar och lämpliga val av värden på dem är problem-beroende. Värdet på Δ^0 och hur mycket Δ^k minskas eller ökas är också algoritmparametrar och val av lämplig värden på dem är också problemberoende. Värdet på Δ^k förändras typiskt mha en enkel strategi, som tex genom halveringar och fördubblingar.)