Sekventiell kvadratisk programmering

Problem: $\min f(x)$

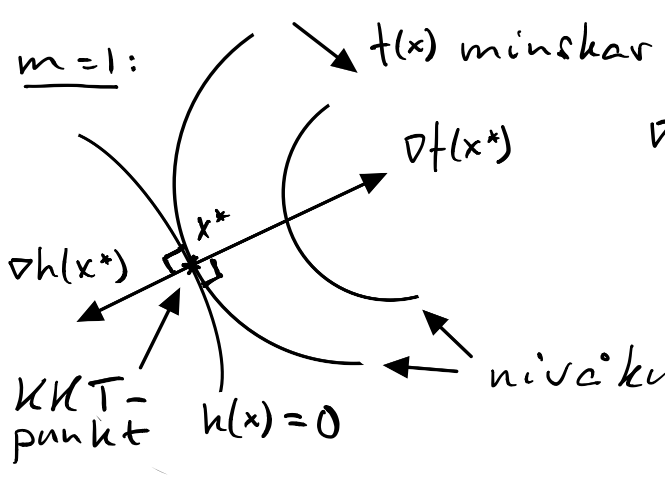
$$(P) \quad \text{då } h_i(x) = 0, \ i = 1, \dots, m \quad | \, v_i$$

(Olikhetsvillkor kan också hanteras.)
Ett globalt minimum till problemet finns bland dess KKT-punkter. (Givet att problemet uppfyller något lämpligt regularitetskrav.) Om problemet är icke-konvext så kan det bland KKT-punkterna även finnas punkter som är lokala, men inte globala, minima, och även punkter som inte ens är lokala minima (utan lokala maxima eller "sadelpunkter").

KKT-villkoren för (P):

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{m} v_i \nabla h_i(x) = 0 \\ h_i(x) = 0, \quad i = 1, \dots, m \end{cases}$$

<u>m = 1:</u>



$$\nabla f(x^*) + v^* \nabla h(x^*) = 0$$
$$h(x^*) = 0$$

Låt $h(x) = (h_1(x), \dots, h_m(x))^T$ och låt $v \in \mathbb{R}^m$ vara vektorn av Lagrange-multiplikatorer för villkoren $h(x) = 0$. Lagrange-funktionen är då $L(x,v) = f(x) + v^T h(x) = f(x) + \sum_{i=1}^m v_i h_i(x)$.

Men $\nabla f(x) + \sum_{i=1}^m v_i \nabla h_i(x) = \nabla_x L(x,v)$ och $h(x) = \nabla_v L(x,v)$, varför KKT-villkoren kan skrivas som

$$\begin{cases} \nabla_x L(x,v) = 0 \\ \nabla_v L(x,v) = 0 \end{cases} \quad \text{eller som } \nabla L(x,v) = 0.$$

Alltså sammanfaller KKT-punkter med stationära punkter till Lagrange-funktionen $L(x,v)$!

För att finna en stationär punkt till Lagrange-funktionen kan Newtons metod användas! Eftersom Lagrange-multiplikatorer även kan betraktas som dualvariabler så fås en <u>primal</u>-<u>dual</u> <u>metod</u>.

Givet en iterationspunkt (x^k, v^k). En andra ordningens Taylor-utveckling ger att

$$L(x,v) \approx L(x^k,v^k) + \nabla L(x^k,v^k)^T \begin{pmatrix} x - x^k \\ v - v^k \end{pmatrix} + \frac{1}{2} \begin{pmatrix} x - x^k \\ v - v^k \end{pmatrix}^T \nabla^2 L(x^k,v^k) \begin{pmatrix} x - x^k \\ v - v^k \end{pmatrix}$$

$$\Rightarrow \nabla L(x,v) \approx \nabla L(x^k,v^k) + \nabla^2 L(x^k,v^k)\begin{pmatrix} x-x^k \\ v-v^k\end{pmatrix}$$

$$HL = 0 \Rightarrow \nabla^2 L(x^k,v^k)\begin{pmatrix} x-x^k \\ v-v^k\end{pmatrix} = -\nabla L(x^k,v^k)$$

Linjärt ekvationssystem i x och v!

Här är $\nabla^2 L(x^k,v^k) = \begin{pmatrix} \nabla^2_{xx} L(x^k,v^k) & \nabla^2_{xv} L(x^k,v^k) \\ \nabla^2_{vx} L(x^k,v^k) & \nabla^2_{vv} L(x^k,v^k)\end{pmatrix}$

med

$$\nabla^2_{xx} L(x^k,v^k) = \nabla^2 f(x^k) + \sum_{i=1}^m v_i^k \nabla^2 h_i(x^k)$$

$$\nabla^2_{xv} L(x^k,v^k) = \nabla^2_{vx} L(x^k,v^k) = \nabla h(x^k)^T = (\nabla h_1(x^k), \ldots, \nabla h_m(x^k))$$

$$\nabla^2_{vv} L(x^k,v^k) = 0$$

och $\nabla L(x^k,v^k) = \begin{pmatrix} \nabla_x L(x^k,v^k) \\ \nabla_v L(x^k,v^k)\end{pmatrix} = \begin{pmatrix} \nabla f(x) + \sum v_i^k \nabla h_i(x^k) \\ h(x^k)\end{pmatrix}.$

Alltså fås systemet

$$\begin{cases} \nabla^2_{xx} L(x^k,v^k)(x-x^k) + \nabla h(x^k)^T(v-v^k) = -\left(\nabla f(x^k) + \sum_{i=1}^m v_i^k \nabla h_i(x^k)\right) \\ \nabla h(x^k)(x-x^k) = -h(x^k), \end{cases}$$

eller med $d = x - x^k$,

$$\begin{cases} \nabla^2_{xx} L(x^k,v^k)\, d + \nabla h(x^k)^T v = -\nabla f(x^k) \\ \nabla h(x^k)\, d = -h(x^k). \end{cases}$$

Låt (d^k, v^*) vara en lösning till detta system. Om $d^k = 0$ gäller så är (x^k, v^k) en KKT-punkt för (P). Annars fås nästa iterationspunkt i Newtons metod på problemet att finna e stationär punkt till $L(x,v)$ som $(x^{k+1}, v^{k+1}) = (x^k + d^k, v^*)$.

Genom att skriva systemet som

$$\begin{cases} \nabla f(x^k) + \nabla^2_{xx} L(x^k, v^k)\, d + \sum_{i=1}^{m} v_i \nabla h_i(x^k) = 0 \\ h_i(x^k) + \nabla h_i(x^k)^T d = 0, \quad i = 1, \dots, m \end{cases}$$

inses dock att det är KKT-villkoren för det kvadratiska optimeringsproblemet

$$\begin{array}{ll} & \min\ q(d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T \nabla^2_{xx} L(x^k, v^k)\, d \\ (QP) & \text{då } h_i(x^k) + \nabla h_i(x^k)^T d = 0, \quad i = 1, \dots, m, \quad | \; v_i \end{array}$$

varför nästa iterationspunkt alltså istället kan beräknas genom att lösa (QP). [Inkrementet d^k utgör primalt optimum till (QP), medan v^* är de optimala värdena på dualvariablerna för de linjäriserade villkoren i (QP).] Notera att målfunktionen i (QP) är kvadratisk medan villkoren är linjära, och vidare utgör linjärapproximationer (första ordningens Taylor-utvecklingar) av villkoren i (P). Vidare kan målfunktionen skrivas ut som

$$q(d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T \nabla^2 f(x^k) d + \frac{1}{2} d^T \Big(\sum_{i=1}^{m} v_i^k \nabla^2 h_i(x^k) \Big) d,$$

där de tre första termerna utgör en kvadratisk approximation av f i x^k (en andra ordningens Taylor-utveckling), medan den sista, kvadratiska, termen innehåller bivillkorsfunktionernas Hessianer i x^k, ihopviktade med v_i^k, $i=1,\dots,m$, och därmed beskriver bivillkorens kurvatur (krökning) i x^k. (QP) innehåller alltså information från andra ordningens approximationer av både målfunktion och bivillkor, men bidragen från andra ordningens termer hanteras på olika sätt.

Newtons metod på problemet att finna en stationär punkt till $L(x,v)$ konvergerar till en KKT-punkt om startpunkten (x^0,v^0) ligger tillräckligt nära KKT-punkten. Nära KKT-punkten fås kvadratisk konvergens.

En möjlig komplikation i metoden är att (QP) kan bli otillåtet. Detta kan dock enkelt åtgärdas genom att i (P) införa artificiella variabler med mycket höga kostnader, varigenom (QP) alltid kommer att ha en tillåten lösning.

Om (P) innehåller olikhetsvillkor så kommer de i (QP) approximeras med linjäriserade olikhetsvillkor (och motsvarande dualvariabler kommer att bli teckenbegränsade).

Hur garantera *global* konvergens till en KKT-punkt, dvs konvergens till en KKT-punkt *oavsett* *startpunkt*? Introducera en *straffunktion* som mäter både målfunktionsvärde och total otillåtenhet, samt lägg till *linjesökningar* map denna!

Låt l_1-straffunktion $F_E : \mathbb{R}^n \to \mathbb{R}$ ges av

$$F_E(x) = f(x) + \mu \sum_{i=1}^{m} |h_i(x)|$$

där $\mu > 0$ och tillräckligt stor.
Observera att F_E *inte* är differentierbar i punkter där $h_i(x) = 0$ gäller för något i.
Observera också att F_E beror av endast de primala variablerna, x.

Givet den primala iterationspunkten x^k, görs linjesökningen $\min_{\lambda \geq 0} F_E(x^k + \lambda d^k)$, som ger $\lambda = \lambda_k$. Man kan visa att d^k är en avtaganderiktning för F_E från x^k, varför $\lambda_k > 0$ gäller. Nästa primala iterationspunkt ges av $x^{k+1} = x^k + \lambda_k d^k$, med $F_E(x^{k+1}) < F_E(x^k)$. Nästa duala iterationspunkt fås som tidigare som $v^{k+1} = v^*$ [från duallösningen till (QP)].

Om (P) innehåller olikhetsvillkor, $g_i(x) \leq 0$, $i = 1, \ldots, m$, så ges l_1-straffunktionen istället av $F_E(x) = f(x) + \mu \sum_{i=1}^{m} \max\{0, g_i(x)\}$. (Man kan självklart även hantera en kombination av likhets- och olikhetsvillkor.)

Utvidgningar:

Problemet (P) är konvext om f är konvex och
i) för fallet likhetsvillkor: alla h_i är affina
ii) för fallet olikhetsvillkor: alla g_i är konvexa.
Då blir (QP) alltid ett konvext optimeringsproblem och det är typiskt måttligt beräkningskrävande. (Det är

väsentligen lika svårlöst som ett linjärprogrammeringsproblem av motsvarande storlek.) Om däremot (P) är ett icke-konvext problem (vilket är vanligt i tillämpningar) så kan matrisen $\nabla^2_{xx} L(x^k, v^k)$ bli indefinit, varigenom (QP) blir icke-konvext och beräkningskrävande. Ett vanligt sätt att åtgärda detta är att ersätta $\nabla^2_{xx} L(x^k, v^k)$ med $\nabla^2_{xx} L(x^k, v^k) + \nu I$, där $\nu > 0$ är en parameter som väljs större än beloppet av det mest negativa egenvärdet till $\nabla^2_{xx} L(x^k, v^k)$. Därigenom blir $\nabla^2_{xx} L(x^k, v^k) + \nu I$ positivt definit och (QP) konvext. (Jämför Newton-Marquardt metoden!)

Ofta är det beräkningskrävande att finna matrisen $\nabla^2_{xx} L(x^k, v^k)$. Det är då vanligt att den approximeras och uppdateras enligt en kvasi-Newton strategi, som t.ex. BFGS.