Olinjärt exempel

Det lilla nystartade bryggeriet Frikeller ska designa en ny öl, med en blandning av Amarillo-humle och Citra-humle.

Låt x_1 stå för andelen Amarillo och x_2 andelen Citra.

Efter omfattande undesökningar kommer man fram till att smaken blir bäst om man minimerar följande funktion: $f(x) = x_1^2 + 2x_2^2 - 2x_1 - x_2$.

Bivillkor: $0 \leq x_1 \leq 1$, $0 \leq x_2 \leq 1$, och $x_1 + x_2 \leq 1$.

Optimum?

När man har formulerat sin optimeringsmodell vill man lösa den.

Dvs. finna en optimal lösning, x^*, till modellen.

Nästan alltid: Sökmetoder:

Stå i en punkt, gå till en annan (bättre).

Upprepa, tills punkten är optimal.

Viktigt att kunna inse optimalitet, dvs. förstå om punkten man står i är optimal.

För då ska man sluta.

Ibland får man en lösning av någon annan, och vill veta om den är optimal.

Vi behöver därför **optimalitetsvillkor.**

Optimalitetsvillkor

En viss punkt är *inte* optimal om vi vill och får gå till en annan punkt,

dvs. om det finns någon punkt som är bättre (enligt målfunktionen) och tillåten (enligt bivillkoren).

En punkt är inte optimal om det finns en **tillåten förbättringsriktning**. (Lokalt kriterium.)

Gäller omvändningen?

Vi är närsynta! Vad händer utanför vårt synfält?

Vilka slutsatser kan man dra av lokala egenskaper?

Är lösningen lokalt optimal?

Är lösningen globalt optimal?

För att kunna säga något säkert behövs exakta matematiska definitioner.

Matematisk notation

1. Se sida 55 i boken.
2. Inga vektorstreck. Vektorer: kolumnvektorer. Matriser versaler.
3. $\forall$: för alla.
4. Skalärprodukt: $c^T x = \sum_j c_j x_j$.
5. A medför B: A $\Rightarrow$ B.
6. Element i S som inte är i T: $S \setminus T$.
7. Optimala värden på x: x^*.

Optimum?

Vi betraktar ett minimeringsproblem
(gammal svensk tradition: minimera eländet).

Globalt minimum: En punkt som är bäst. (rita)

Definition

En punkt x^* är **globalt** minimum i X om $f(x^*) \leq f(x)$ för alla $x \in X$.

Lokalt minimum: En punkt som är bäst för närsynta. (rita)

Definition

En punkt x^* är **lokalt** minimum i X om $f(x^*) \leq f(x)$ för alla $x \in X$ som är nära x^*. (T.ex. för alla $x \in X : \|x - x^*\| \leq \delta$.)

Konvexitet

Konvex mängd: Innehåller alla linjesegment mellan punkter i mängden. (rita)

Definition

En mängd X är **konvex** om $\lambda x^{(1)} + (1 - \lambda)x^{(2)} \in X$ för alla $x^{(1)} \in X,\ x^{(2)} \in X,\ 0 \leq \lambda \leq 1$.

"Man kan se alla punkterna i en konvex mängd från vilken punkt som helst."

Exempel: Halvrum. Ett linjärt bivillkor.

Kombinerade mängder: Snittet av konvexa mängder är konvext. (rita)

Sats

Om X_i är konvex för alla i så är $X = \cap_i X_i$ konvex.

Exempel: Flera halvrum. Flera linjära bivillkor.

Extrempunkt

Konvexkombination: En punkt "innanför" de andra. (rita)

Definition

$\hat{x}$ är en **konvexkombination** av punkterna $x^{(k)}, k = 1, \ldots, p$, om
$\hat{x} = \sum_k \lambda_k x^{(k)}$, där $\sum_k \lambda_k = 1$ och $\lambda_k \geq 0$ för alla k.

Exempel: Mittpunkt. $\lambda_k = 1/p$.

Extrempunkt: En punkt som inte är en konvexkombination av två andra tillåtna punkter. (rita)

Definition

$\hat{x}$ är en **extrempunkt** i X om $\hat{x} = \lambda x^{(1)} + (1 - \lambda)x^{(2)}$, där $x^{(1)} \in X,\ x^{(2)} \in X$ och $0 < \lambda < 1$, endast är möjligt om $x^{(1)} = x^{(2)}$.

Exempel: Origo (under bivillkoren $x_i \geq 0$ för alla i).

Konvexa höljet

Konvexa höljet: Alla konvexkombinationer. (rita)

Sats/Definition

För en begränsad mängd S består conv(S) av alla konvexkombinationer av punkter i S.

Konvexa höljet av S är den *minsta konvexa mängden* som innehåller S.

Definition

Skärningen av alla konvexa mängder som innehåller en viss mängd S kallas **det konvexa höljet** av S och betecknas med conv(S).

Exempel: $S = \{(0,0), (1,0), (0,1)\}$.

$conv(S)$ ges av $x_1 \geq 0,\ x_2 \geq 0,\ x_1 + x_2 \leq 1$.

Polyeder

Mängd med "raka kanter".

Polyeder: (rita)

Definition

Skärningen av **ändligt** många halvrum kallas **polyeder**.

Exempel: $x_1 + x_2 \leq 1$, $x_1 \geq 0$, $x_2 \geq 0$.

Exempel: $x_1 + x_2 = 1$, $x_1 \geq 0$, $x_2 \geq 0$.

Exempel: $Ax \leq b$.

Exempel: $Ax = b$, $x \geq 0$.

En polyeder har **ändligt** många extrempunkter.

Konvex funktion

Konvex funktion: "Glad mun." Snäll funktion för minimering. (rita)
Linjesegmentet mellan två punkter ligger ovanför funktionen.

Definition

En funktion $f(x)$ är **konvex** om
$f(\lambda x^{(1)} + (1-\lambda)x^{(2)}) \leq \lambda f(x^{(1)}) + (1-\lambda)f(x^{(2)})$ för alla
$x^{(1)}$, $x^{(2)}$, $0 \leq \lambda \leq 1$.

Exempel: $f(x) = x_1^2$.

Exempel: $f(x) = 3x_1$.

Bivillkor definierade av en funktion: (rita)

Sats

Om $g(x)$ är en konvex funktion, så ger punkterna som uppfyller $g(x) \leq b$ en konvex mängd.

Exempel: $x_1^2 \leq 3$. $x_1^2 + x_2^2 \leq 17$.

Konvex funktion

Kombinera konvexa funktioner:

- $f_1(x) + f_2(x)$ är konvex om $f_1(x)$ och $f_2(x)$ är konvexa.
- $kf(x)$ är konvex om $f(x)$ är konvex och $k \geq 0$.

Tillsammans:

Sats

Om $f_i(x)$ är konvexa funktioner och $\alpha_i \geq 0$, så är $f(x) = \sum_i \alpha_i f_i(x)$ konvex.

Exempel: $f(x) = 2x_1^2 + 3x_2^2$.

Differentierbar konvex funktion

Egenvärden hos Hessianen: (jämför med andraderivator i endimensionella fallet)

Positivt semidefinit matris $\Leftrightarrow$ egenvärden ≥ 0.
Positivt definit matris $\Leftrightarrow$ egenvärden > 0.

Sats

Om $f(x)$ är en två gånger kontinuerligt differentierbar funktion, så är $f(x)$ konvex om dess hessian $H(x)$ är positivt semidefinit, och strikt konvex om $H(x)$ är positivt definit.

Beräkna egenvärden och kolla.

Metod med underdeterminanter:

Om varje ledande underdeterminant är positiv, är alla egenvärden positiva.

En "ledande underdeterminant", h_k, beräknas med de k första raderna och kolumnerna, dvs. alla kvadratiska delmatriser som börjar i övre vänstra hörnet.

Differentierbar konvex funktion

Exempel: min $x_1^2 + 2x_2^2 - 2x_1x_2$.

Gradienten är $\nabla f(x) = \begin{pmatrix} 2x_1 - 2x_2 \\ 4x_2 - 2x_1 \end{pmatrix}$.

Hessianen är $H(x) = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}$.

Dess egenvärden är $3 - \sqrt{5} \approx 0.7639$ och $3 + \sqrt{5} \approx 5.2361$, båda positiva, så Hessianen är positivt definit, vilket visar att funktionen $f(x)$ är konvex.

Metod med underdeterminanter:

$h_1 = 2 > 0$.

$h_2 = 2 * 4 - (-2) * (-2) = 8 - 4 = 4 > 0$.

Så alla ledande underdeterminanter är positiva, vilket betyder att alla egenvärden är positiva, Hessianen är positivt definit, och funktionen $f(x)$ är konvex.

Konvexitetskontroll

- Praktisk konvexitetskontroll av funktion:
  - Dela upp i konvexa bitar.
  - Använd satserna.
  - I värsta fall beräkna egenvärden.
- Kända konvexa funktioner:
  - $f(x) = a$ (konstant)
  - $f(x) = |x_j|$ (belopp)
  - $f(x) = x_j^2$ (kvadrat)
  - $f(x) = \sum_j a_{ij}x_j$ (linjär)
- Konvext problem (min): Konvex målfunktion, konvex tillåten mängd.
- Varför är detta intressant?
  Ett lokalt optimum i ett konvext problem är alltid ett globalt optimum!

Frikeller-exempel

min $f(x) = x_1^2 + 2x_2^2 - 2x_1 - x_2$

då $x_1 + x_2 \leq 1$ samt $0 \leq x_1 \leq 1$ och $0 \leq x_2 \leq 1$

Målfunktion: Summa av konvexa funktioner => Konvex.

Bivillkor: Alla linjära => Konvex tillåtet område.

=> Problemet konvext.

Mot optimum

Flervariabelanalyskunskap:

Gradienten, $\nabla f(x)$, anger lutningen av funktionen $f(x)$.

Den pekar i den riktning där funktionen $f(x)$ ökar snabbast.

Vill man maximera $f(x)$ är $\nabla f(x)$ en bra riktning att gå i.

(Vill man minimera $f(x)$ är $-\nabla f(x)$ en bra riktning att gå i.)

Exempel: min $x_1^2 + 2x_2^2$. Är punkten $\hat{x}_1 = 1$, $\hat{x}_2 = 1$ optimal?

Gradienten är $\nabla f(x) = \begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix}$, och i punkten $\nabla f(\hat{x}) = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$.

Alltså kan vi gå i riktningen $d = \begin{pmatrix} -2 \\ -4 \end{pmatrix}$ för att få bättre punkter.

Punkten $\hat{x}_1 = 1$, $\hat{x}_2 = 1$ är inte optimal, för det finns bättre punkter (i närheten).

Olinjär optimering utan bivillkor

$$\min f(x)$$

Gradienten $\nabla f(x)$ anger funktionens *lutning*.

Var ligger optimum?

$f(x)$ konvex, differentierbar: $\nabla f(\hat{x}) = 0 \Rightarrow \hat{x}$ är globalt minimum.

$f(x)$ ej konvex, men differentierbar: $\nabla f(\hat{x}) = 0 \Rightarrow$ Lutningen är noll.
$\hat{x}$ kan vara ett lokalt minimum, ett lokalt maximum eller en sadelpunkt.

Exempel: min $f(x) = x_1^2 + 2x_2^2$. Gradienten är $\nabla f(x) = \begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix}$.

$\nabla f(x) = 0$ ger $x_1 = 0$ och $x_2 = 0$.

Eftersom $f(x)$ är konvex, är detta globalt minimum.

Exempel: min $f(x) = -x_1^2 - 2x_2^2$. Gradienten är $\nabla f(x) = \begin{pmatrix} -2x_1 \\ -4x_2 \end{pmatrix}$.

$\nabla f(x) = 0$ ger $x_1 = 0$ och $x_2 = 0$.

Men $f(x)$ är inte konvex. (Detta är faktiskt maximum.)

Olinjär optimering utan bivillkor

I en *avtaganderiktning* minskar funktionsvärdet.
I en *ökanderiktning* ökar funktionsvärdet.

En riktning d är en avtaganderiktning för $f(x)$ i x om $\nabla f(x)^T d < 0$.
En riktning d är en ökanderiktning för $f(x)$ i x om $\nabla f(x)^T d > 0$.

En ökanderiktning pekar in i samma halvrum som gradienten,
en avtaganderiktning gör inte det.

Exempel: Är riktningen $d = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ en avtaganderiktning till funktionen $f(x) = x_1^2 + 2x_2^2$ i punkten $\hat{x} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$?

$\nabla f(x) = \begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix}$. $\nabla f(\hat{x}) = \begin{pmatrix} 4 \\ 12 \end{pmatrix}$. Så $\nabla f(\hat{x})^T d = 4 - 12 = -8 < 0$.

Ja, det är en avtaganderiktning.

Olinjär optimering med bivillkor

$$\begin{array}{ll} \min & f(x) \\ \text{då} & g_i(x) \leq 0 \quad i = 1, \ldots, m \end{array}$$

Var ligger optimum?

Inga aktiva bivillkor: Som utan bivillkor. $\nabla f(x) = 0$.

Ett eller flera aktiva bivillkor: Det får inte finnas någon tillåten förbättringsriktning.

$-\nabla f(x)$ är den mest önskade riktningen (brantaste lutningen).

Alla riktningar d med $\nabla f(x)^T d < 0$ är avtaganderiktningar.

Om bivillkoret $g_i(x) \leq 0$ är aktivt, får $g_i(x)$ inte öka.

$\nabla g_i(x)$ är då den mest förbjudna riktningen (utåtriktade normalen).

Alla riktningar d med $\nabla g_i(x)^T d > 0$ är förbjudna.

Så hur vet man om alla avtaganderiktningar är förbjudna?

Exempel

Exempel: min $x_1^2 + x_2^2 - x_1x_2 - x_1 - x_2$ då $x_1 + x_2 \leq 1$, $x_2 \leq 0.4$.

Skriv bivillkoren som $g_1(x) = x_1 + x_2 - 1 \leq 0$, $g_2(x) = x_2 - 0.4 \leq 0$.

Gradienter:
$\nabla f(x) = \begin{pmatrix} 2x_1 - x_2 - 1 \\ -x_1 + 2x_2 - 1 \end{pmatrix}$, $\nabla g_1(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\nabla g_2(x) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Är $\hat{x}_1 = 0.6$, $\hat{x}_2 = 0.4$ optimal?

Är punkten tillåten? Kolla:

$g_1(\hat{x}) = 0$. Tillåtet. Bivillkoret aktivt.

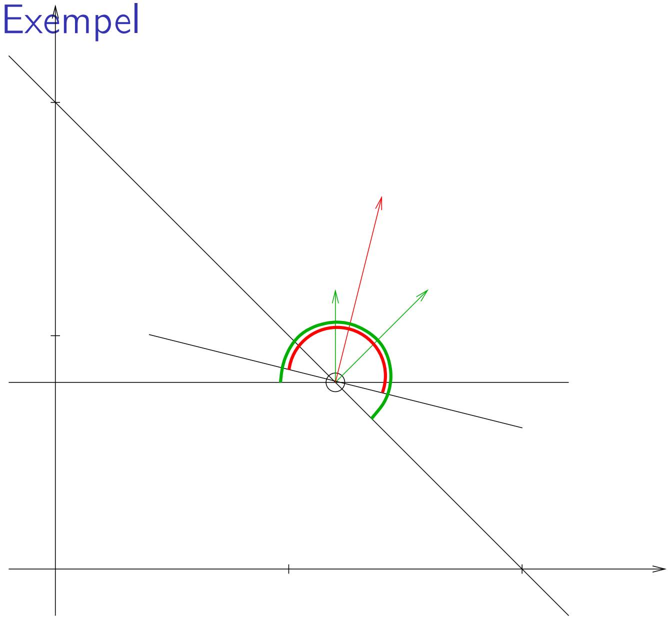
$g_2(\hat{x}) = 0$. Tillåtet. Bivillkoret aktivt.

Punkten är tillåten. Båda bivillkoren aktiva.

$\nabla f(\hat{x}) = \begin{pmatrix} -0.2 \\ -0.8 \end{pmatrix}$

Rita!

Exempel



Nej, förbättringsriktningarna täcks helt av de förbjudna riktningarna.

Olinjär optimering med bivillkor: Optimalitetsvillkor

Problemet:

$$\begin{array}{ll} \min & f(x) \\ \text{då} & g_i(x) \leq 0 \quad i = 1, \ldots, m \end{array}$$

Karush-Kuhn-Tuckervillkoren:

Punkten $\hat{x}$ är en KKT-punkt om följande är uppfyllt.

KKT1: $g_i(\hat{x}) \leq 0$ för alla i.

KKT2: $u_i g_i(\hat{x}) = 0$ för alla i.

KKT3: $\nabla f(\hat{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\hat{x}) = 0$.

KKT4: $u_i \geq 0$ för alla i.

En KKT-punkt är optimal om problemet är konvext.

Vad är detta???

Olinjär optimering med bivillkor: Optimalitetsvillkor

Analys av KKT-villkoren för problemet:

$$\begin{array}{ll} \min & f(x) \\ \text{då} & g_i(x) \leq 0 \quad i = 1, \ldots, m \end{array}$$

Vi ska kontrollera om punkten $\hat{x}$ är en KKT-punkt.

KKT1: $g_i(\hat{x}) \leq 0$ för alla i.

Först kontrollerar vi om punkten är tillåten.

KKT2: $u_i g_i(\hat{x}) = 0$ för alla i.

Vi ska göra en projektion av målfunktionen på de aktiva bivillkoren.

För att bara få med de aktiva bivillkoren, ser vi till att $u_i = 0$ för alla icke aktiva bivillkor, dvs. de som har $g_i(\hat{x}) < 0$.

(I två dimensioner kan man se detta grafiskt.)

Olinjär optimering med bivillkor: Optimalitetsvillkor

Analys av KKT-villkoren för problemet:

$$\begin{array}{ll} \min & f(x) \\ \text{då} & g_i(x) \leq 0 \quad i = 1, \ldots, m \end{array}$$

Vi har gjort KKT1 och KKT2, så vi vet att punkten är tillåten, och har sett till att $u_i = 0$ för alla icke aktiva bivillkor.

Nu är det dags att göra projektionen.

KKT3: $\nabla f(\hat{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\hat{x}) = 0$.

som är samma som $\quad -\nabla f(\hat{x}) = \sum_{i} u_i \nabla g_i(\hat{x})$.

Observera att $-\nabla f(\hat{x})$ är vår önskeriktning, medan $\nabla g_i(\hat{x})$ är utåtriktade normaler till bivillkoren, dvs. de mest förbjudna riktningarna.

Vi skriver önskeriktningen som en kombination av förbjudna riktningar.

Olinjär optimering med bivillkor: Optimalitetsvillkor

Vi har gjort projektionen $-\nabla f(\hat{x}) = \sum_i u_i \nabla g_i(\hat{x})$.

Påstående: Om $u_i \geq 0$ för alla i, så är alla förbättringsriktningar otillåtna.

Multiplicera med en valfri riktning d: $-\nabla f(\hat{x})^T d = \sum_i u_i \nabla g_i(\hat{x})^T d$.

Om d är en tillåten riktning, så är $\nabla g_i(\hat{x})^T d \leq 0$ för alla i.
Om $u_i \geq 0$ för alla i, så blir $\sum_i u_i \nabla g_i(\hat{x})^T d \leq 0$,

vilket betyder att $-\nabla f(\hat{x})^T d \leq 0$, dvs. $\nabla f(\hat{x})^T d \geq 0$,
vilket visar att d inte är en avtaganderiktning.

Å andra sidan, om något $u_i < 0$, så gäller inte detta.

Detta visar behovet av: KKT4: $u_i \geq 0$ för alla i.

Olinjär optimering med bivillkor: Optimalitetsvillkor

Vi vet nu att KKT4 tillsammans med KKT3, KKT2 och KKT1 bevisar att ingen tillåten förbättringsriktning finns i punkten $\hat{x}$,
så $\hat{x}$ kan vara optimal.

En sådan punkt kallas KKT-punkt.

För att en KKT-punkt säkert ska vara optimal, krävs konvexitet.
Annars kan det vara ett lokalt optimum eller en sadelpunkt (eller någon annan konstig punkt).

Detta är ett sätt att kontrollera om en viss *given* punkt är optimal.
Då är nämligen alla gradienter givna, och bara u behöver lösas ut.
Detta kan göras eftersom KKT3 då ger ett *linjärt* ekvationssystem.

Olinjär optimering med bivillkor: Optimalitetsvillkor

$$\begin{array}{ll} \min & f(x) \\ \text{då} & g_i(x) \leq 0 \quad i = 1, \dots, m \end{array}$$

Karush-Kuhn-Tuckervillkoren:

Punkten $\hat{x}$ är en KKT-punkt om följande är uppfyllt.
KKT1: $g_i(\hat{x}) \leq 0$ för alla i. (Tillåtenhet)
KKT2: $u_i g_i(\hat{x}) = 0$ för alla i. (Plocka bort inaktiva bivillkor)
KKT3: $\nabla f(\hat{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\hat{x}) = 0$. (Projektion)
KKT4: $u_i \geq 0$ för alla i. (Kolla riktning)

KKT3 är ett linjärt ekvationssystem i u som kan lösas metodiskt.

Tolkning av KKT-villkoren

KKT1 tar bort otillåtna punkter.

KKT2 tar bort inaktiva bivillkor.

KKT3 projicerar önskeriktningen $-\nabla f(x)$ på aktiva bivillkors normaler $\nabla g_i(x)$.

KKT4 kontrollerar att önskeriktningen $-\nabla f(x)$ ligger i konen som spänns upp av aktiva bivillkors normaler $\nabla g_i(x)$,
dvs. att alla förbättringsriktningar är otillåtna.

Hur gör man med likhetsbivillkor?
De är alltid aktiva, och $u_i < 0$ är tillåtet, så KKT2 och KKT4 faller bort för dessa bivillkor.

Vårt tidigare exempel

min $x_1^2 + x_2^2 - x_1 x_2 - x_1 - x_2$ då $x_1 + x_2 \leq 1$, $x_2 \leq 0.4$.

Är $\hat{x}_1 = 0.6$, $\hat{x}_2 = 0.4$ optimal?

$$\nabla f(\hat{x}) = \begin{pmatrix} -0.2 \\ -0.8 \end{pmatrix}, \nabla g_1(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \nabla g_2(x) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

KKT1: Är punkten tillåten?

$g_1(\hat{x}) = 0$. Tillåtet. Bivillkoret aktivt. KKT2: u_1 behöver inte vara noll.

$g_2(\hat{x}) = 0$. Tillåtet. Bivillkoret aktivt. KKT2: u_2 behöver inte vara noll.

Punkten är tillåten. KKT2: Inget av u_1 eller u_2 måste vara noll.

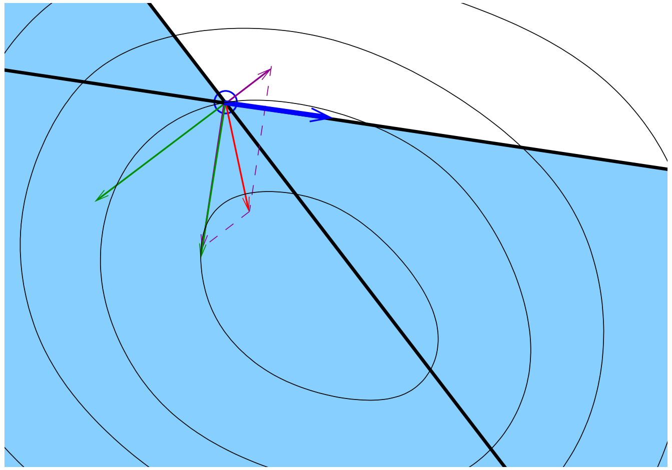
KKT3: $\begin{pmatrix} -0.2 \\ -0.8 \end{pmatrix} + u_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + u_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

Dvs. $u_1 = 0.2$, $u_1 + u_2 = 0.8$, vilket ger $u_1 = 0.2$, $u_2 = 0.6$.

KKT4 uppfyllt, $u_1 \geq 0$ och $u_2 \geq 0$.

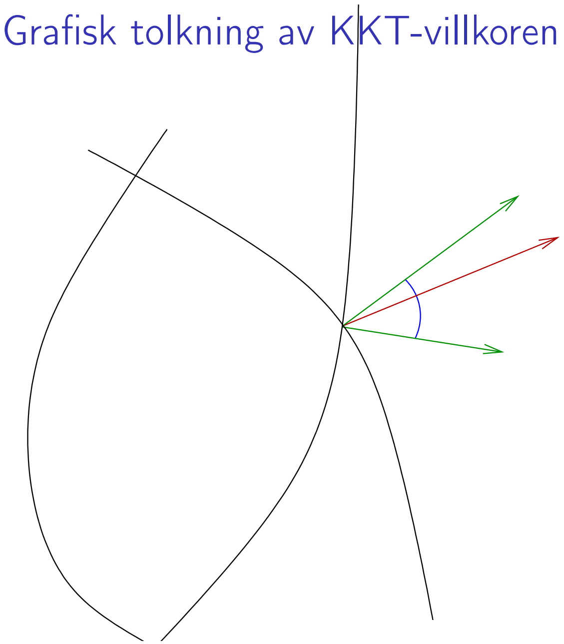
Punkten är optimal, ty den är en KKT-punkt och problemet är konvext.

Grafisk tolkning av KKT-villkoren



En tillåten förbättringsriktning finns.

Grafisk tolkning av KKT-villkoren



Slutsats: Om $-\nabla f(x)$ ligger i konen av $\nabla g_i(x)$ är det optimum.

KKT-villkoren och optimalitet

En KKT-punkt kan vara optimum.

Om problemet är *konvext*, så är en KKT-punkt ett *globalt optimum*.

Om problemet inte är konvext kan det finnas KKT-punkter som inte är optimala. Vissa KKT-punkter kan vara lokala optima.

Det finns vissa krav, kallade CQ (Constraint Qualification), för att optimum ska vara en KKT-punkt, bl.a.:

- Linjärt oberoende normaler för aktiva bivillkor.
- Konvext tillåtet område med inre punkt.

Om inget CQ är uppfyllt, kanske det inte finns någon KKT-punkt.

(CQ kan i denna kurs antas vara uppfyllda.)

Om något CQ är uppfyllt, och problemet är konvext, är alltså optimum och bara optimum är en KKT-punkt.

KKT-villkoren: Exempel

Exempel: min $f(x) = x_1^2 + 2x_2^2$ då $g(x) = x_1 + x_2 + 1 \leq 0$.

Är $\hat{x}_1 = -0.5$ och $\hat{x}_2 = -0.5$ optimal?

KKT1: Punkten är tillåten.

KKT2: Bivillkoret är aktivt. (Så u behöver inte vara 0.)

KKT3: $\nabla f(\hat{x}) = \begin{pmatrix} -1 \\ -2 \end{pmatrix}$, och $\nabla g(\hat{x}) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

KKT3 blir $\begin{pmatrix} -1 \\ -2 \end{pmatrix} + u \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$,

vilket betyder $u = 1$ och $u = 2$. Ej lösbart.

$\hat{x}$ är alltså ingen KKT-punkt.

KKT-villkoren: Exempel

Exempel: min $x_1^2 + 2x_2^2$ då $x_1 + x_2 \geq 1$, $x_1 \geq 0$, $x_2 \geq 0$.

Skriv bivillkoren som

$g_1(x) = -x_1 - x_2 + 1 \leq 0, \quad g_2(x) = -x_1 \leq 0, \quad g_3(x) = -x_2 \leq 0.$

Gradienter: $\nabla f(x) = \begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix}$, $\nabla g_1(x) = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$, $\nabla g_2(x) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$,

$\nabla g_3(x) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$.

KKT3: $\begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix} + u_1 \begin{pmatrix} -1 \\ -1 \end{pmatrix} + u_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + u_3 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

Sätt in de punkter som ska testas.

KKT-villkoren: Exempel

Är $\hat{x}_1 = 1$ och $\hat{x}_2 = 0$ optimal?

KKT1: Punkten är tillåten.

KKT2: $g_1(\hat{x}) = 0$. (u_1 behöver inte vara 0.)

KKT2: $g_2(\hat{x}) = -1 < 0$. ($u_2 = 0$.)

KKT2: $g_3(\hat{x}) = 0$. (u_3 behöver inte vara 0.)

KKT3: $\nabla f(\hat{x}) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$.

KKT3 blir $\begin{pmatrix} 2 \\ 0 \end{pmatrix} + u_1 \begin{pmatrix} -1 \\ -1 \end{pmatrix} + u_3 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$,

dvs. $2 - u_1 = 0$, $-u_1 - u_3 = 0$, vilket har lösningen $u_1 = 2$ och $u_3 = -2$.

$\hat{x}$ är ingen KKT-punkt, ty $u_3 < 0$.

KKT-villkoren: Exempel

Är $\hat{x}_1 = 0$ och $\hat{x}_2 = 1$ optimal?

KKT1: Punkten är tillåten.

KKT2: $g_1(\hat{x}) = 0$. (u_1 behöver inte vara 0.)

KKT2: $g_2(\hat{x}) = 0$. (u_2 behöver inte vara 0.)

KKT2: $g_3(\hat{x}) = -1 < 0$. ($u_3 = 0$.)

KKT3: $\nabla f(\hat{x}) = \begin{pmatrix} 0 \\ 4 \end{pmatrix}$.

KKT3 blir $\begin{pmatrix} 0 \\ 4 \end{pmatrix} + u_1 \begin{pmatrix} -1 \\ -1 \end{pmatrix} + u_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$,

dvs. $-u_1 - u_2 = 0$, $4 - u_1 = 0$, vilket har lösningen $u_1 = 4$ och $u_2 = -4$.

$\hat{x}$ är ingen KKT-punkt, ty $u_2 < 0$.

KKT-villkoren: Exempel

Istället för att ange en punkt kan man anta vilka bivillkor som är aktiva. (Detta fungerar bara ibland.)

Antag att $x_1 + x_2 = 1$ är aktivt och att $x_1 > 0$ och $x_2 > 0$.

Dvs. $g_1(\hat{x}) = 0$, $g_2(\hat{x}) < 0$ och $g_3(\hat{x}) < 0$.

KKT2 ger då $u_2 = 0$ och $u_3 = 0$.

KKT3 blir då $\begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix} + u_1 \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$,

vilket betyder $2x_1 - u_1 = 0$ och $4x_2 - u_1 = 0$.

Lös ut x i u: $x_1 = u_1/2$ och $x_2 = u_1/4$.

Antagandet $x_1 + x_2 = 1$ ger då $x_1 + x_2 = u_1/2 + u_1/4 = 1$,

vilket ger $u_1 = 4/3$, och $x_1 = 2/3$ och $x_2 = 1/3$.

Detta en KKT-punkt, ty $u_1 > 0$, och globalt optimum, ty problemet är konvext.

Frikeller-exempel

min $f(x) = x_1^2 + 2x_2^2 - 2x_1 - x_2$ då $x_1 + x_2 \leq 1$, $x_1 \geq 0$ och $x_2 \geq 0$.

(Bivillkoren $x_1 \leq 1$ och $x_2 \leq 1$ redundanta. Ta bort.)

Konvext.

Är det optimalt att bara använda en sort?

Kolla punkterna (1,0) och (0,1).

Först $\hat{x} = (1, 0)$:

KKT1: Punkten är tillåten.

KKT2: Bivillkor 1 och 3 aktiva, bivillkor 2 inte. => $u_2 = 0$.

KKT3: $\begin{pmatrix} 0 \\ -1 \end{pmatrix} + u_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + u_3 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$,

dvs. $u_1 = 0$, $u_1 - u_3 = 1$, vilket har lösningen $u_1 = 0$ och $u_3 = -1$.

$\hat{x}$ är ingen KKT-punkt, ty $u_3 < 0$.

Frikeller-exempel

min $f(x) = x_1^2 + 2x_2^2 - 2x_1 - x_2$ då $x_1 + x_2 \leq 1$, $x_1 \geq 0$ och $x_2 \geq 0$.

Nu $\hat{x} = (0, 1)$:

KKT1: Punkten är tillåten.

KKT2: Bivillkor 1 och 2 aktiva, bivillkor 3 inte. => $u_3 = 0$.

KKT3: $\begin{pmatrix} -2 \\ 3 \end{pmatrix} + u_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + u_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$,

dvs. $u_1 - u_2 = 2$, $u_1 = -3$, vilket har lösningen $u_1 = -3$ och $u_2 = -5$.

$\hat{x}$ är ingen KKT-punkt, ty $u_1 < 0$ och $u_2 < 0$.

Frikeller-exempel

min $f(x) = x_1^2 + 2x_2^2 - 2x_1 - x_2$ då $x_1 + x_2 \leq 1$, $x_1 \geq 0$ och $x_2 \geq 0$.

Antag istället att $0 < x_1 < 1$, $0 < x_2 < 1$ och $x_1 + x_2 = 1$.

KKT2: Bivillkor 1 aktivt, bivillkor 2 och 3 inte. => $u_2 = 0$, $u_3 = 0$.

KKT3: $\begin{pmatrix} 2x_1 - 2 \\ 4x_2 - 1 \end{pmatrix} + u_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$,

dvs. $2x_1 + u_1 = 2$, $4x_2 + u_1 = 1$, vilket ger $x_1 = (2 - u_1)/2$ och $x_2 = (1 - u_1)/4$.

KKT1: $x_1 + x_2 = 1$ ger $(2 - u_1)/2 + (1 - u_1)/4 = 1$, vilket ger $u_1 = 1/3$.

Det ger $x_1 = 5/6$ och $x_2 = 1/6$.

$x_1 \geq 0$ och $x_2 \geq 0$ så KKT1 är uppfyllt.

Detta är en KKT-punkt.

Optimum, ty problemet konvext.

Svar: Använd proportionerna 5/6 Amarillo och 1/6 Citra.

Olinjär optimering utan bivillkor: Metoder

$$\min f(x)$$

Generell sökmetod:

1. Finn en startpunkt, $x^{(k)}$. Sätt $k = 1$.
2. Beräkna en sökriktning, $d^{(k)}$.
3. Om $\|d^{(k)}\| \leq \varepsilon$ stopp.
4. Beräkna en steglängd, $t^{(k)}$.
5. Om $t^{(k)} \leq \varepsilon$ stopp.
6. Sätt $x^{(k+1)} = x^{(k)} + t^{(k)}d^{(k)}$. Sätt $k = k + 1$ och gå till 2.

Steglängd: För en given riktning bestäms t med **linjesökning** (en-dimensionell optimering):

Finn $t^{(k)}$ ur $\min_{t \geq 0} f(x^{(k)} + td^{(k)})$.

(I denna kurs analytiskt.)

Olinjär optimering utan bivillkor: Linjesökning

Exempel på linjesökning för funktionen $f(x) = 2x_1^2 + x_2^2 - 3x_1 + x_2$ i punkten $\hat{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ och riktningen $d = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$.

Vår nya punkt blir $x(t) = \hat{x} + td = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + t\begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1-t \\ 2-t \end{pmatrix}$.

Sätt in i funktionen:
$f(x(t)) = 2(1-t)^2 + (2-t)^2 - 3(1-t) + (2-t) = 2(1+t^2-2t) + (4+t^2-4t) - 3 + 3t + 2 - t = 2 + 2t^2 - 4t + 4 + t^2 - 4t - 3 + 3t + 2 - t = 3t^2 - 6t + 5$.

Derivera: $f'(x(t)) = 6t - 6$.

Sätt derivatan till noll (om funktionen är konvex): $f'(x(t)) = 6t - 6 = 0$ ger $t = 1$.

Sätt in i nya punkten: $x(1) = \begin{pmatrix} 1-1 \\ 2-1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Olinjär optimering utan bivillkor: Linjesökning

Linjesökning för funktionen $f(x) = 2x_1^2 + x_2^4 - 3x_1 - 2x_2$ i punkten $\hat{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ och riktningen $d = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

Ny punkt: $x(t) = \hat{x} + td = \begin{pmatrix} 1-t \\ t \end{pmatrix}$.

Sätt in i funktionen:
$f(x(t)) = 2(1-t)^2 + t^4 - 3(1-t) - 2t = t^4 + 2t^2 - 3t - 1$.

Derivera: $f'(x(t)) = 4t^3 + 4t - 3$.

Men $f'(x(t)) = 4t^3 + 4t - 3 = 0$ kan inte lösas.

Använd iterativ metod: Intervallhalvering:

Pröva olika värden på t.

Om $f'(x(t)) < 0$ så minskar funktionen, och vi vill gå längre.

Om $f'(x(t)) > 0$ så ökar funktionen, och vi har gått för långt.

Testa mittpunkten mellan dessa.

Olinjär optimering utan bivillkor: Linjesökning

Vi vill finna t som ger $f'(x(t)) = 4t^3 + 4t - 3 = 0$.

Metod: Intervallhalvering:

$t = 0$: $f'(x(0)) = -3 < 0$, funktionen minskar.

Testa med $t = 1$: $f'(x(1)) = 5 > 0$, funktionen ökar.

Minimum måste ligger mellan 0 och 1, testa mittpunkten, $t = 0.5$.

$f'(x(0.5)) = 0.5 + 2 - 3 = -0.5 < 0$, funktionen minskar.

Alltså ligger minimum mellan 0.5 och 1, inte mellan 0 och 0.5.

Testa mittpunkten, $t = 0.75$.

$f'(x(0.75)) = 2.7 > 0$, funktionen ökar.

Alltså ligger minimum mellan 0.5 och 0.75.

Upprepa.

Halva intervallet elimineras varje gång.

Får ej exakt lösning. Noggrannheten avgörs av hur länge man håller på.

Grundläggande lokala sökmetoder

Sökriktningen kan beräknas på flera olika sätt.

Gradienten $\nabla f(x)$ pekar i den riktning där funktionen $f(x)$ ökar snabbast. Vill man minimera $f(x)$ verkar $-\nabla f(x)$ vara en bra riktning att gå i.

Brantaste lutningsmetoden: Sätt $d^{(k)} = -\nabla f(x^{(k)})$.

Ta med andra ordningens information (dvs. sikta mot minimum av andra ordningens approximation av $f(x)$):

Newtons metod: Sätt $d^{(k)} = -H(x^{(k)})^{-1}\nabla f(x^{(k)})$.

Brantaste lutningsmetoden är robust, men kan sicksacka. (lab 1)

Newtons metod använder andra ordningens approximation och är generellt sett starkare men jobbigare och kan fallera utan konvexitet.

Lokala sökmetoder: Exempel

Exempel: min $3x_1^2 + 2x_2^2 - x_2x_2 - 3x_1 - 4x_2$. Gradienten är $\nabla f(x) = \begin{pmatrix} 6x_1 - x_2 - 3 \\ 4x_2 - x_1 - 4 \end{pmatrix}$. Hessianen är $H(x) = \begin{pmatrix} 6 & -1 \\ -1 & 4 \end{pmatrix}$. Dess egenvärden är 3.6 och 6.4, båda positiva, så funktionen $f(x)$ är konvex.

I punkten $\hat{x}_1 = 1$ och $\hat{x}_2 = 1$ har vi $\nabla f(x) = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$, så brantaste lutningsriktningen är $d_{br} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$.

Vi har $H^{-1} = \frac{1}{23}\begin{pmatrix} 4 & 1 \\ 1 & 6 \end{pmatrix}$, så Newtonriktningen blir

$$d_N = -H^{-1}\nabla f(\hat{x}) = -\frac{1}{23}\begin{pmatrix} 4 & 1 \\ 1 & 6 \end{pmatrix}\begin{pmatrix} 2 \\ -1 \end{pmatrix} = \frac{1}{23}\begin{pmatrix} -7 \\ 4 \end{pmatrix}$$

Newtonriktningen pekar precis mot optimum (eftersom funktionen är konvex och kvadratisk). Optimal steglängd blir då 1.

Lokala sökmetoder: Exempel

Exempel: min $-3x_1^2 - 2x_2^2 - x_2x_2 - 3x_1 - 4x_2$. Gradienten är $\nabla f(x) = \begin{pmatrix} -6x_1 - x_2 - 3 \\ -4x_2 - x_1 - 4 \end{pmatrix}$. Hessianen är $H(x) = \begin{pmatrix} -6 & -1 \\ -1 & -4 \end{pmatrix}$. Dess egenvärden är ung -6.4 och -3.6, så funktionen $f(x)$ är inte konvex.

I punkten $\hat{x}_1 = 1$ och $\hat{x}_2 = 1$ har vi $\nabla f(x) = \begin{pmatrix} -10 \\ -9 \end{pmatrix}$, så brantaste lutningsriktningen är $d_{br} = \begin{pmatrix} 10 \\ 9 \end{pmatrix}$.

Vi har $H^{-1} = \frac{1}{23}\begin{pmatrix} -4 & 1 \\ 1 & -6 \end{pmatrix}$, så Newtonriktningen blir

$$d_N = -H^{-1}\nabla f(\hat{x}) = -\frac{1}{23}\begin{pmatrix} -4 & 1 \\ 1 & -6 \end{pmatrix}\begin{pmatrix} -10 \\ -9 \end{pmatrix} = \frac{1}{23}\begin{pmatrix} -31 \\ -44 \end{pmatrix}$$

Newtonriktningen är inte en avtaganderiktning, utan en ökanderiktning. Linjesökning skulle ge steg noll.

Modifierade lokala sökmetoder

Försök åtgärda nackdelar:

Newtons metod: $d^{(k)} = -H(x^{(k)})^{-1}\nabla f(x^{(k)})$.

Newtons metod fungerar dåligt för ickekonvexa funktioner.
Kan t.o.m. ge ökningsriktning (istället för avtaganderiktning).

Konvexifiering:

Levenbergs modifiering av Newtons smetod: Ersätt H med $H + \nu I$.
Som att addera konvexa termen $\nu(x_1^2 + x_2^2 + \ldots + x_n^2)$.
$\nu = 0$ ger Newtons metod.
Ett mycket stort ν ger i princip brantaste lutningsmetoden.

Större ν ger "mer positiva" egenvärden, dvs. en "mer konvex" funktion.
Tillräckligt stort ν ger konvex funktion.
Jämför med mest negativa egenvärdet av H.

Modifierade lokala sökmetoder

Försök åtgärda nackdelar:

Newtons metod: $d^{(k)} = -H(x^{(k)})^{-1}\nabla f(x^{(k)})$.

Det är jobbigt att beräkna och invertera H i varje iteration i Newtons metod.

Förenkling av Newtons metod:

Kvasi-Newtonmetoder: Ersätt H^{-1} med en enklare beräknad approximation.
Approximationen uppdateras med enbart gradienter.

Flera varianter finns: DFP (Davidon, Fletcher, Powell), Broyden (Broyden), BFGS (Broyden, Fletcher, Goldfarb, Shanno) och SR1 (symmetric rank 1).

Krångliga formler för effektiv direkt uppdatering av inversapproximationen.

(Finns i Nileopt.)

Kvasi-Newtonmetoder

Formler enligt Wikipedia:

Method	$B_{k+1} =$	$H_{k+1} = B_{k+1}^{-1} =$
DFP	$\left(I - \frac{y_k \Delta x_k^T}{y_k^T \Delta x_k}\right) B_k \left(I - \frac{\Delta x_k y_k^T}{y_k^T \Delta x_k}\right) + \frac{y_k y_k^T}{y_k^T \Delta x_k}$	$H_k + \frac{\Delta x_k \Delta x_k^T}{y_k^T \Delta x_k} - \frac{H_k y_k y_k^T H_k^T}{y_k^T H_k y_k}$
BFGS	$B_k + \frac{y_k y_k^T}{y_k^T \Delta x_k} - \frac{B_k \Delta x_k (B_k \Delta x_k)^T}{\Delta x_k^T B_k \Delta x_k}$	$\left(I - \frac{y_k \Delta x_k^T}{y_k^T \Delta x_k}\right)^T H_k \left(I - \frac{y_k \Delta x_k^T}{y_k^T \Delta x_k}\right) + \frac{\Delta x_k \Delta x_k^T}{y_k^T \Delta x_k}$
Broyden	$B_k + \frac{y_k - B_k \Delta x_k}{\Delta x_k^T \Delta x_k} \Delta x_k^T$	$H_k + \frac{(\Delta x_k - H_k y_k) \Delta x_k^T H_k}{\Delta x_k^T H_k y_k}$
Broyden family	$(1-\varphi_k) B_{k+1}^{BFGS} + \varphi_k B_{k+1}^{DFP}, \quad \varphi \in [0,1]$	
SR1	$B_k + \frac{(y_k - B_k \Delta x_k)(y_k - B_k \Delta x_k)^T}{(y_k - B_k \Delta x_k)^T \Delta x_k}$	$H_k + \frac{(\Delta x_k - H_k y_k)(\Delta x_k - H_k y_k)^T}{(\Delta x_k - H_k y_k)^T y_k}$

Vi ska *inte* gå in på detaljer här.

Men de finns i Nileopt. I lab 1 kan ni jämföra dem.

Modifierade lokala sökmetoder

Försök åtgärda nackdelar:

Brantaste lutningsmetoden: $d^{(k)} = -\nabla f(x^{(k)})$.

Brantaste lutningsmetoden kan bli långsam p.g.a. sicksackning.
Problemet är att närliggande sökriktningar är ortogonala.

Förbättring av brantaste lutningsmetoden:

Konjugerade gradientmetoder: Modifiera riktningen för att undvika sicksackning, genom att addera in tidigare gradienter (på olika sätt).
Sätt $d^{(k)} = -\nabla f(x^{(k)}) + \beta_k d^{(k-1)}$.

Flera varianter (krångliga formler) för att beräkna β finns: FR (Fletcher-Reeves), HS (Hestenes-Stiefel), PR (Polak-Ribiere).

(Finns i Nileopt.)

Konjugerade gradientmetoder

Konjugerade gradientmetoder: Sätt $d^{(k)} = -\nabla f(x^{(k)}) + \beta_k d^{(k-1)}$.

Formler för att beräkna β:

Fletcher-Reeves: $\beta_k = \frac{\|\nabla f(x^{(k)})\|^2}{\|\nabla f(x^{(k-1)})\|^2}$

Polak-Ribiere: $\beta_k = \frac{\nabla f(x^{(k)})^T y^{(k)}}{\|\nabla f(x^{(k-1)})\|^2}$

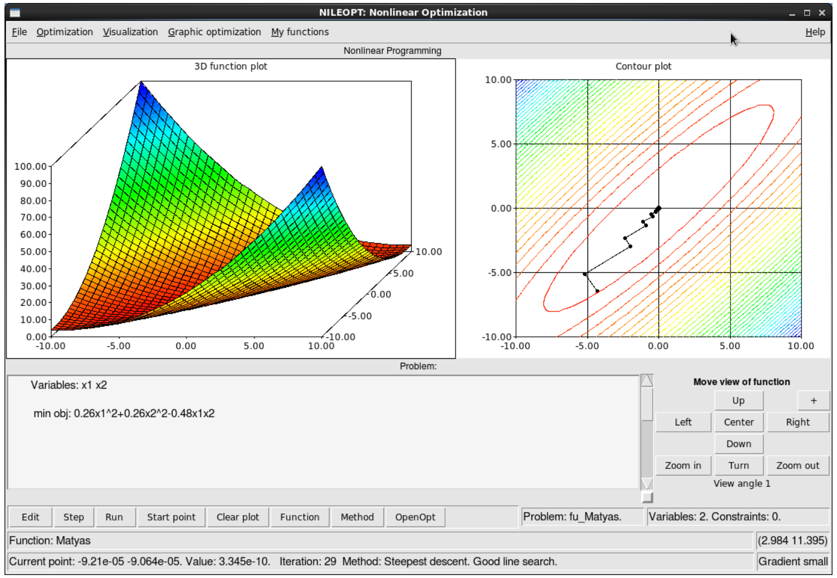
Hestenes-Stiefel: $\beta_k = \frac{\nabla f(x^{(k)})^T y^{(k)}}{d^{(k-1)^T} y^{(k)}}$

där $y^{(k)} = \nabla f(x^{(k)}) - \nabla f(x^{(k-1)})$.

(Vi ska inte gå in på varför.)

Vilken är bäst? Det beror på $f(x)$. Man kan pröva med Nileopt.

Nileopt



Brantaste lutningsmetoden, iteration 8-29.

Nileopt

Format för olinjära optimeringsproblem i Nileopt (GMPL,AMPL):

Problemet
$\min f(x) = x_1^2 + 2x_2^2 + x_1x_2 + 2x_1 - x_2$

skrivs in i Nileopt som

```
var x1;
var x2;
min obj: x1^2+2*x2^2+x1*x2+2*x1-1*x2;
```

För att få fram "upphöjt till" (^) krävs ett extra mellanslag.

Nileopt: Format

Problemet
$\min f(x) = 2x_1^2 - 2x_1x_2 + 3x_2$

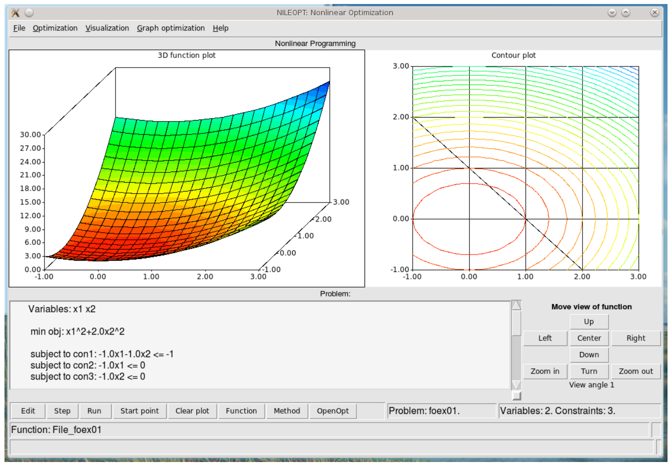
$$
\begin{array}{lll}
 & x_1^2 + x_2^2 & \leq 1 \\
\text{då} & x_1 + x_1x_2 & \leq 1 \\
 & x_1 + x_2 & \leq 2
\end{array}
$$

kan skrivas in i Nileopt som

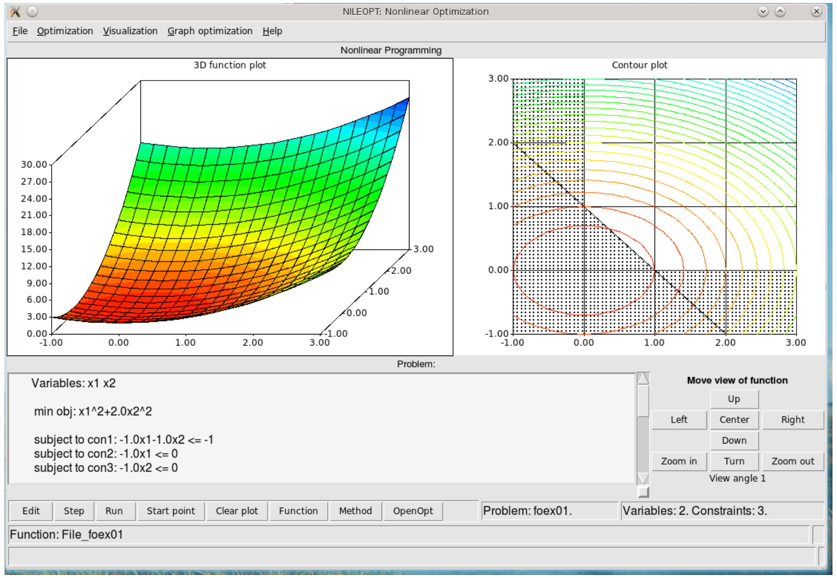
```
var x1;
var x2;

minimize obj:
 2*x1^2 - 2*x1*x2 + 3*x2;

subject to con1: x1^2 + x2^2 <= 1;
s.t. con2: x1 + x1*x2 <= 1;
st con3: x1+x2<=2;
```


Nileopt med bivillkor

$$
\begin{array}{llcccl}
\min & f(x) = & x_1^2 & + & 2x_2^2 & \\
\text{då} & & x_1 & + & x_2 & \geq 1 \\
 & & x_1 & & & \geq 0 \\
 & & & & x_2 & \geq 0
\end{array}
$$



Nileopt med bivillkor



Laboration 1

I laboration 1 ska ni:

- Skriva in olinjära optimeringsproblem enligt reglerna.
- Lösa olinjära problem med bivillkor med färdig programvara (Cobyla).
- Illustrera KKT-villkoren grafiskt.
- Testa olika metoder för olinjära problem utan bivillkor på olika funktioner.
- Intressanta frågor:

  Hur mycket sämre är brantaste lutningsmetoden än Newtons metod?

  Är sicksackning verkligen ett problem för brantaste lutningsmetoden?

  Är konjugerade gradientmetoder bättre än brantaste lutningsmetoden?

  Är kvasi-Newtonmetoder bättre än brantaste lutningsmetoden?

  Är kvasi-Newtonmetoder sämre än Newtons metod?

  Är icke-konvexa problem svårare än konvexa?