Definitioner

Vi kommer att studera matematiska problem på formen

$$\begin{array}{ll} & \min_x f(x) \\ \text{då} & x \in \mathcal{S} \end{array}$$

Funktionen f kallas *objektfunktion*. Mängden $\mathcal{S}$ kallas den *tillåtna mängden* och definieras vanligen med hjälp av en mängd av bivillkor

$$\mathcal{S} = \left\{ x : \left\{ \begin{array}{lll} g_i(x) = 0, & i \in \mathcal{E} & \text{(likhetsbivillkor)} \\ g_i(x) \geq 0, & i \in \mathcal{I} & \text{(olikhetsbivillkor)} \end{array} \right. \right\}.$$

Ofta anges $\mathcal{S}$ implicit på formen

$$\begin{array}{ll} \min_x & f(x) \\ \text{då} & g_i(x) = 0,\ i \in \mathcal{E} \\ & g_i(x) \geq 0\ i \in \mathcal{I} \end{array}$$

Alla optimeringsproblem och bivillkor går att skriva om på denna form.

En punkt x som uppfyller alla bivillkor kallas *tillåten* och tillhör då $\mathcal{S}$.

Definitioner, forts.

Vid en tillåten punkt $\bar{x}$ sägs ett olikhetsbivillkor $g_i(\bar{x}) \geq 0$ vara *bindande* eller *aktivt* om $g_i(\bar{x}) = 0$. Om $g_i(\bar{x}) > 0$ är bivillkoret *ej bindande* eller *inaktivt*. Likhetsbivillkor sägs alltid vara aktiva.

Punkten $\bar{x}$ sägs vara på *randen* av bivillkoret om $g_i(\bar{x}) = 0$ och vara en *inre punkt* till bivillkoret om $g_i(\bar{x}) > 0$. Likhetsbivillkor saknar inre punkter.

Den *aktiva bivillkorsmängden* vid en tillåten punkt är mänden av alla aktiva bivillkor.

En tillåten punkt med minst ett aktivt bivillkor tillhör *randen* av det tillåtna området. Övriga tillåtna punkter säga vara *inre punkter*.

Optimalitet, globala minimerare

Betrakta den n-dimensionella problemet

$$\min_{x \in \mathcal{S}} f(x).$$

Om en punkt x_* uppfyller

$$f(x_*) \leq f(x)\ \forall x \in \mathcal{S}$$

sägs x_* vara en *global minimerare* till f i $\mathcal{S}$. Punkten x_* kallas ofta *lösningspunkt*.

Om

$$f(x_*) < f(x)\ \forall x \in \mathcal{S}, x \neq x_*$$

kallas x_* *strikt global minimerare* och är unik.

Globala minimerare är önskvärda, men svåra att bestämma om inte f och $\mathcal{S}$ har speciella egenskaper.

Optimalitet, lokala minimerare

Många gånger får vi nöja oss med en *lokal minimerare*, dvs ett x_* sådant att

$$f(x_*) \leq f(x)\ \forall x \in \mathcal{S}, \|x - x_*\| < \epsilon, \epsilon \text{ “litet”}.$$

På samma sätt definieras en *strikt lokal minimerare* x_* som

$$f(x_*) < f(x)\ \forall x \in \mathcal{S}, x \neq x_*, \|x - x_*\| < \epsilon, \epsilon \text{ “litet”}.$$

Konvexitet

Ett specialfall då en lokal lösning också är en global är då objektfunktionen f och den tillåtna mänden $\mathcal{S}$ är *konvexa*.

En mängd $\mathcal{S}$ är *konvex* om $\forall x, y \in \mathcal{S}$ gäller att

$$\alpha x + (1-\alpha)y \in \mathcal{S},\ 0 \le \alpha \le 1,$$

dvs alla linjesegment mellan alla punkter i $\mathcal{S}$ ligger helt i $\mathcal{S}$, eller mängden "buktar ej inåt".

En funktion f är konvex på en konvex mängd $\mathcal{S}$ om det $\forall x, y \in \mathcal{S}$ gäller att

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha)f(y),\ 0 \le \alpha \le 1,$$

dvs f ligger under linjen från $x, f(x)$ till $y, f(y)$ eller linjen "buktar ej nedåt".

Konvexitet, forts.

Ett konvext programmeringsproblem är på formen

$$\min_{x \in \mathcal{S}} \ f(x)$$
$$\mathcal{S} \text{ konvex, } f \text{ konvex.}$$

Problemet

$$\min_{x \in \mathcal{S}} \ f(x)$$
$$\text{då} \quad g_i(x) \ge 0, \forall i$$

är konvext om f konvex och alla g_i konkava.

Om f är två ggr kont. deriverbar, så är f konvex omm

$$f''(x) \ge 0,\ \forall x \in \mathcal{S}.$$

För flerdimensionella problem $x \in \Re^m,\ f \in \Re^1$ gäller att Hessianen måste vara positivt semidefinit, dvs

$$y^T \nabla^2 f(x) y \ge 0,\ \forall y \in \Re^m,$$

där

$$\nabla^2 f(x)_{(ij)} = \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Den generella optimeringsalgoritmen, v I

- Bestäm startgissning x_0.
- Upprepa för $k = 0, 1, \ldots$
  - Om x_k optimal, avsluta.
  - Bestäm x_{k+1} som nästa approximation av lösningen.

Den generella optimeringsalgoritmen, v II

- Bestäm startgissning x_0.
- Upprepa för $k = 0, 1, \ldots$
  - Om x_k optimal, avsluta.
  - Bestäm *sökriktning* p_k.
  - Bestäm *steglängd* α_k.
  - $x_{k+1} = x_k + \alpha_k p_k$

Sökriktningen p_k

Ett vanligt krav på p_k är att funktionsvärdet ska minska i sökriktningen, dvs att p_k är *nedförsriktning* (*descent direction*) som uppfyller

$$f(x_k + \alpha p_k) < f(x_k),\ 0 < \alpha \leq \varepsilon$$

för något ε.

För problem *med* bivillkor krävs dessutom att riktningen är tillåten, dvs p_k måste dessutom uppfylla

$$x_k + \alpha p_k \in \mathcal{S}.$$

Om en tillåten nedförsriktning finns kan x_k ej vara en optimal punkt, ty $f(x_k + \alpha p_k) < f(x_k)$. Vi kommer att använda detta som test på optimalitet.

Steglängden α_k

Givet punkten x_k och en tillåten nedförsriktning p_k bestäms steglängden α_k genom att studera problemet

$$\min_{\alpha_k \geq 0} f(x_k + \alpha_k p_k).$$

Detta kallas *linjesökning*, då f studeras längs en linje.

Effektivitetsmått

För att kunna jämföra olika optimeringsalgoritmer med varandra måste vi ha ett mått på effektivitet.

Beräkningskomplexiteten, som används som mått på direkta (icke-iterativa) algoritmer går ej att använda, då vi ej känner antalet iterationer i förväg.

I stället används *konvergenshastigheten*.

Konvergenshastigheten

Antag vi har en serie $\{x_k\}$ som konvergerar mot en lösning x_*. Definiera felsekvensen

$$e_k = x_k - x_*.$$

Notera att

$$\lim_{k \to \infty} e_k = 0.$$

Sekvensen $\{x_k\}$ sägs *konvergera* mot x_* med hastighet r och hastighetskonstant C om

$$\lim_{k \to \infty} \frac{\|e_{k+1}\|}{\|e_k\|^r} = C$$

och $C < \infty$.

Konvergenshastighet, forts.

I praktiken förekommer tre olika konvergenshastigheter

- Linjär, då $r = 1$ och $0 < C < 1$.
- Kvadratiskt, då $r = 2$.
- Super-linjär, då $r = 1$ och $C = 0$.

Linjär konvergens, exempel

Om $r = 1$, $C = 0.1$ och $\|e_0\| = 1$ så blir normen på felsekvensen

$$\underbrace{1, 10^{-1}, 10^{-2}, \ldots, 10^{-7}}_{7 \text{ iterationer}}$$

Om i stället $C = 0.99$ blir motsvarande sekvens

$$\underbrace{1, 0.9, 0.9801, \ldots, 0.997 \cdot 10^{-7}}_{1604 \text{ iterationer}}$$

Som synes spelar konstanten C stor roll vid linjär konvergens. Notera också att om vi endast ställer kravet att ”felet ska bli mindre” för varje iteration kan vi få väldigt långsamma iterationsmetoder.

Kvadratisk konvergens, exempel

För $r = 2$, $C = 0.1$ och $\|e_0\| = 1$ blir sekvensen

$$\underbrace{1, 10^{-1}, 10^{-3}, 10^{-7}, \ldots}_{3 \text{ iterationer}}$$

För $r = 2$, $C = 3$ och $\|e_0\| = 1$ blir sekvensen

$$1, 3, 27, \ldots$$

Serien divergerar.

För $r = 2$, $C = 3$ och $\|e_0\| = 0.1$ blir sekvensen

$$0.1, 0.03, 0.0027, \ldots$$

Serien konvergerar, trots $C > 1$.

Här spelar alltså C mindre roll. I stället är det viktigt att man startar “tillräckligt nära” lösningen, dvs $\|e_0\|$ är liten.

Konvergenshastigheter, exempel

Studera sekvensen $x_k = 1 + 10^{-k}$

$$2, 1.1, 1.01, \ldots$$

Sekvensen konvergerar mot $x_* = 1$ och $e_k = x_k - x_* = 10^{-k}$. Alltså

$$\lim_{k\to\infty} \frac{\|e_{k+1}\|}{\|e_k\|} = \lim_{k\to\infty} \frac{10^{-(k+1)}}{10^{-k}} = \frac{1}{10},$$

dvs sekvensen konvergerar linjärt med hastighetskonstant $1/10$.

Studera sekvensen $x_{k+1} = \frac{x_k}{2} + \frac{2}{x_k}$

$$4, 2.5, 2.05, \ldots$$

$$\lim_{k\to\infty} x_k = 2$$

$$e_{k+1} = x_{k+1} - x_* = \cdots = \frac{1}{2x_k}(x_k - 2)^2 = \frac{1}{2x_k} e_k^2.$$

$$\lim_{k\to\infty} \frac{\|e_{k+1}\|}{\|e_k\|^2} = \frac{1}{2|x_*|} = \frac{1}{4},$$

dvs $r = 2$ och $C = 1/4$.

Taylorserier

Taylorserier är ett verktyg för att approximera en funktion f nära en punkt x_0.

Det är ett generellt verktyg:

- Möjligt uppskatta ett funktionsvärde nära en punkt.
- Approximationspolynomet kan lätt deriveras och integreras.
- Används för att härleda många minimeringsalgoritmer.

Definition: Låt x_0 vara specificerad punkt och $f : \Re \to \Re$ med n kontinuerliga derivator. Då är taylorapproximationen av grad n:

$$f(x_0+p) \approx f(x_0)+pf'(x_0)+\frac{p^2}{2}f''(x_0)+\cdots+\frac{p^n}{n!}f^{(n)}(x_0).$$

Taylorserier, exempel

För $f(x) = \sqrt{x},\ x_0 = 1$ är derivatorna

$$\begin{aligned} f(x_0) &= \sqrt{x_0} &&= 1 \\ f'(x_0) &= \tfrac{1}{2}x_0^{-1/2} &&= \tfrac{1}{4} \\ f''(x_0) &= -\tfrac{1}{4}x_0^{-3/2} &&= -\tfrac{1}{4} \\ f^{(3)}(x_0) &= \tfrac{3}{8}x_0^{-5/2} &&= \tfrac{3}{8} \end{aligned}$$

och approximationen blir då

$$\begin{aligned} \sqrt{1+p} &= f(x_0+p) \\ &\approx f(x_0)+pf'(x_0)+\tfrac{p^2}{2}f''(x_0)+\tfrac{p^3}{6}f^{(3)}(x_0) \\ &= 1+\tfrac{1}{2}p-\tfrac{1}{8}p^2+\tfrac{1}{16}p^3. \end{aligned}$$

Approximationen $\sqrt{1.6} = \sqrt{1+0.6} \approx 1+\frac{1}{2}\cdot 0.6 - \frac{1}{8}\cdot 0.6^2 + \frac{1}{16}\cdot 0.6^3 \approx \boxed{1.2685}$ ligger rätt nära det verkliga värdet $\sqrt{1.6} \approx \boxed{1.2649}$.

Flerdimensionella taylorserier

Det går också att härleda flerdimensionella taylorserier:

1-variabelfallet:

$$f(x_0+p) = f(x_0)+pf'(x_0)+\tfrac{1}{2}pf''(x_0)p+\cdots.$$

n-variabelfallet:

$$f(x_0+p) = f(x_0)+p^T\nabla f(x_0)+\tfrac{1}{2}p^T\nabla^2 f(x_0)p+\cdots.$$

$\nabla f(x_0)$ betecknar *gradienten* till f i x_0, dvs en kolumnvektor med element $\frac{\partial f}{\partial x_i}(x_0)$ och $\nabla^2 f(x_0)$ betecknar *hessianen* till f i x_0, dvs en matris med element $\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)$.

Flerdim. taylor, exempel

För funktionen $f(x_1,x_2) = x_1^3+5*x_1^2x_2+7x_1x_2^2+2x_2^3$ är gradienten

$$\nabla f(x) = \begin{bmatrix} 3x_1^2+10x_1x_2+7x_2^2 \\ 5x_1^2+14x_1x_2+6x_2^2 \end{bmatrix}$$

och hessianen

$$\nabla^2 f(x) = \begin{bmatrix} 6x_1+10x_2 & 10x_1+14x_2 \\ 10x_1+14x_2 & 14x_1+12x_2 \end{bmatrix}$$

I punkten $x_0 = [-2,3]^T$ blir dessa

$$\nabla f(x_0) = \begin{bmatrix} 15 \\ -10 \end{bmatrix} \text{och} \nabla^2 f(x_0) = \begin{bmatrix} 18 & 22 \\ 22 & 8 \end{bmatrix}.$$

Flerdim. taylor, exempel, forts.

För $p = [0.1, 0.2]^T$ blir

$$\begin{aligned} f(-1.9, 3.2) &= f(x_0 + p) \\ \approx \;& f(x_0) + p^T \nabla f(x_0) + \frac{1}{2} p^T \nabla^2 f(x_0) p \\ = \;& -20 + [0.1\ 0.2] \begin{bmatrix} 15 \\ -10 \end{bmatrix} + [0.1\ 0.2] \begin{bmatrix} 18 & 22 \\ 22 & 8 \end{bmatrix} \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix} \\ = \;& -20 - 0.5 + 0.69 = -19.81 \end{aligned}$$

Jämför det exakta värdet $f(-1.9, 3.2) = -19.755$.

Newtons metod

Metoden för att lösa det icke-linjära problemet

$$f(x) = 0$$

är att i stället lösa en *sekvens* av linjära problem. Dessa fås ur taylorapproximationen av f runt x_k:

$$f(x_k + p) \approx f(x_k) + p f'(x_k).$$

Om $f'(x_k) \neq 0$ kan vi lösa den linjära ekvationen

$$f(x_*) \approx f(x_k) + p f'(x_k) = 0$$

för p och får

$$p = -f(x_k)/f'(x_k),$$

vilket ger

$$x_{k+1} = x_k - f(x_k)/f'(x_k).$$

Newtons metod, konvergens

Newtons metod svarar mot att approximera (den icke-linjära) funktionen f med motsvarande linjära funktion (tangenten) runt x_k,

Konvergenshastigheten kommer bl a att bero på noggrannheten i den approximationen:

$$\begin{aligned} 0 = f(x_*) &= f(x_k - e_k) \\ &= f(x_k) - e_k f'(x_k) + \tfrac{1}{2} e_k^2 f''(\xi) \\ e_k - \frac{f(x_k)}{f'(x_k)} &= \frac{1}{2} e_k^2 \frac{f''(\xi)}{f'(x_k)} \\ \underbrace{x_k - \frac{f(x_k)}{f'(x_k)}}_{x_{k+1}} - x_* &= \frac{1}{2} e_k^2 \frac{f''(\xi)}{f'(x_k)} \\ e_{k+1} &= \frac{1}{2} e_k^2 \frac{f''(\xi)}{f'(x_k)} \\ \lim_{k \to \infty} \frac{\|e_{k+1}\|}{\|e_k\|^2} &= \frac{1}{2} \left| \frac{f''(x_*)}{f'(x_*)} \right| \end{aligned}$$

Newtons metod i flera dimensioner

Antag att vi har ett icke-linjärt problem

$$f(x) = 0,$$

där $f, x \in \Re^n$, dvs

$$\begin{aligned} f_1(x_1, \ldots, x_n) &= 0 \\ f_2(x_1, \ldots, x_n) &= 0 \\ &\vdots \\ f_n(x_1, \ldots, x_n) &= 0 \end{aligned}$$

Definiera gradientmatrisen $\nabla f(x)$ som

$$\begin{aligned} \nabla f(x) &= \begin{bmatrix} \nabla f_1(x) & \nabla f_2(x) & \ldots & \nabla f_n(x) \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \ldots & \frac{\partial f_n}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \ldots & \frac{\partial f_n}{\partial x_2} \\ \vdots & & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \ldots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} \end{aligned}$$

Newtons metod i flera dimensioner, forts.

Då blir den linjära approximationen

$$f(x_k + p) \approx f(x_k) + \nabla f(x_k)^T p$$

och om vi löser

$$f(x_*) \approx f(x_k) + \nabla f(x_k)^T p = 0$$

får vi lösningen

$$p = -\nabla f(x_k)^{-T} f(x_k),$$

vilket ger den nya approximationen

$$x_{k+1} = x_k + p = x_k - \nabla f(x_k)^{-T} f(x_k).$$