2006–05–15

Optimering, fortsättning

en dimension, utan bivillkor:

Problem:

$$\min_{x\in\mathbb{R}} f(x), \quad f: \mathbb{R}\rightarrow\mathbb{R}$$

Metoder:

1. Gyllene snittet.

2. Polynomapproximation. Populär i modern programvara.

Konvergerar om f är konvex. Illustration ⟨fig24⟩. p_2 är interpolationspolynom av grad 2, genom $f(x_1)$, $f(x_2)$, $f(x_3)$. Minimera $p_2(x)$ är enkelt. Låt x_4 vara $p_2(x)$:s minpunkt och byt x_4 mot någon gammal punkt så att mimimum ligger i intervallet som bestäms av de tre punkterna. I exemplet i figuren fotsätter vi alltså med x_2, x_4, x_3.

3. Newtons metod för att lösa $f'(x)=0$, dvs $x_{k+1}=x_k-\frac{f'(x_k)}{f''(x_k)}$, $k=0,1,\ldots$ Kräver andraderivator.

4. Modifierad Newton:

$$x_{k+1}=x_k-\frac{f'(x_k)}{f''(x_0)}, k=0,1,\ldots$$

2:a derivatan eller en approximation av densamma beräknas *bara* i startpunkten.

5. Sekantmetoden:

$$x_{k+1}=x_k-\frac{f'(x_k)\,(x_k-x_{k-1})}{f'(x_k)-f'(x_{k-1})}, k=1,2,\ldots$$

Endast förstaderivator, två startapproximationer.

Flerdimensionell optimering utan bivillkor

$$\min f(\boldsymbol{x}): \quad \boldsymbol{f}: \mathbb{R}^n\rightarrow\mathbb{R}$$

Sökmetoder: $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k\,\boldsymbol{d}_k$ där α_k är en steglängd och $\boldsymbol{d}_k$ är sökriktning. Anta först att $\boldsymbol{d}_k$ är bestämd (hur kommer vi tala om senare). Vi vill då bestämma α_k, dvs hur långt ska vi gå lägns sökriktningen?

Att bestämma α_k är *en*-dimensionell optimering, s.k. linjesökning. Linjesökningsproblemet:

$$(P_L) \quad \min_{\alpha} f(\boldsymbol{x}_k+\alpha\,\boldsymbol{d}_k)$$

Tidigare metoder 1–5 för endimensionell optimering kan användas.

I specialfallet f kvadratisk, dvs

$$f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^T H\,\boldsymbol{x}+\boldsymbol{c}^T\boldsymbol{x}$$

kan vi visa formeln för optimal steglängd:

$$\alpha_k = -\frac{\nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_k}{\boldsymbol{d}_k^T H \boldsymbol{d}_k}$$

Bevis.

$$\min_\alpha f(\boldsymbol{x}_k + \alpha\, \boldsymbol{d}_k) = \min_\alpha \ g(\alpha)$$

$$g(\alpha) = f(\boldsymbol{x}_k + \alpha\, \boldsymbol{d}_k)$$

$$g'(\alpha) = \nabla f(\boldsymbol{x}_k + \alpha\, \boldsymbol{d}_k)^T \boldsymbol{d}_k$$

$$f \text{ kvadratisk } \Longrightarrow \nabla f(\boldsymbol{x}_k + \alpha\, \boldsymbol{d}_k) = H(\boldsymbol{x}_k + \alpha\, \boldsymbol{d}_k) + \boldsymbol{c} =$$

$$= H\, \boldsymbol{x}_k + \boldsymbol{c} + \alpha\, H\, \boldsymbol{d}_k = \nabla f(\boldsymbol{x}_k) + \alpha\, H\, \boldsymbol{d}_k = \nabla f(\boldsymbol{x}_k) + \alpha\, H\, \boldsymbol{d}_k$$

$$g'(\alpha) = 0 \Longrightarrow (\nabla f(\boldsymbol{x}_k) + \alpha\, H\, \boldsymbol{d}_k)^T \boldsymbol{d}_k = 0$$

Lös ut α:

$$\alpha = -\frac{\nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_k}{\boldsymbol{d}_k^T H\, \boldsymbol{d}_k}$$

□

Val av sökriktning $\boldsymbol{d}_k$:

1. Steepest descent, $\boldsymbol{d}_k = -\nabla f(\boldsymbol{x}_k)$.
2. Newtons metod på $\nabla f(\boldsymbol{x}) = \boldsymbol{0}$, dvs $\boldsymbol{d}_k$ som lösning till ekvationssystemet

$$H(\boldsymbol{x}_k)\, \boldsymbol{d}_k = -\nabla f(\boldsymbol{x}_k)$$

3. Kvasi-Newton:

$$B_k\, \boldsymbol{d}_k = -\nabla f(\boldsymbol{x}_k)$$

   där B_k är approximation av $H(\boldsymbol{x}_k)$, exempelvist Broydens uppdatering.

4. Konjugerad gradient (ej på tentan)

$$\boldsymbol{d}_k = -\nabla f(\boldsymbol{x}_k) + \beta_k\, \boldsymbol{d}_k$$

   "metod med minne".

$$\beta_k = \frac{\|\nabla f(\boldsymbol{x}_k)\|_2^2}{\|\nabla f(\boldsymbol{x}_{k-1})\|_2^2}$$

   Egenskap: Om f är kvadratisk så blir sökriktningarna konjungerat ortogonala:

$$\boldsymbol{d}_k^T H\, \boldsymbol{d}_j = 0, \quad k \neq j$$

EXEMPEL:

$$\min_{\boldsymbol{x} \in \mathbb{R}^2} 5\, x_1^2 + x_1\, x_2 + \frac{1}{2}\, x_2^2 - x_1$$

$$\nabla f = \begin{pmatrix} 10x_1 + x_2 - 1 \\ x_1 + x_2 \end{pmatrix}, \quad H = \begin{pmatrix} 10 & 1 \\ 1 & 1 \end{pmatrix}$$

Steepest descent:

$$\boldsymbol{x}^{[0]} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \nabla f\left(\boldsymbol{x}^{[0]}\right) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad \boldsymbol{d}^{[0]} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\alpha_0 = -\frac{\nabla f\left(\boldsymbol{x}^{[0]}\right)^T \boldsymbol{d}^{[0]}}{\left(\boldsymbol{d}^{[0]}\right)^T H\, \boldsymbol{d}^{[0]}} = \frac{1}{10}$$

$$\boldsymbol{x}^{[1]} = \boldsymbol{x}^{[0]} + \alpha_0\, \boldsymbol{d}^{[0]} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \frac{1}{10} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.1 \\ 0 \end{pmatrix}$$

$$\nabla f\left(\boldsymbol{x}^{[1]}\right) = \begin{pmatrix} 0 \\ 0.1 \end{pmatrix}, \quad \boldsymbol{d}^{[1]} = \begin{pmatrix} 0 \\ -0.1 \end{pmatrix}$$

$$\alpha_1 = -\frac{\nabla f\left(\boldsymbol{x}^{[1]}\right)^T \boldsymbol{d}^{[1]}}{\left(\boldsymbol{d}^{[1]}\right)^T H\, \boldsymbol{d}^{[1]}} = \ldots = 1$$

$$\boldsymbol{x}^{[2]} = \boldsymbol{x}^{[1]} + \alpha\, \boldsymbol{d}^{[1]} = \begin{pmatrix} 0.1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -0.1 \end{pmatrix} = \begin{pmatrix} 0.1 \\ -0.1 \end{pmatrix}$$

Konjungerad gradient

$$\boldsymbol{x}^{[0]} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{x}^{[1]} = \begin{pmatrix} 0.1 \\ 0 \end{pmatrix}, \quad \nabla f\left(\boldsymbol{x}^{[1]}\right) = \begin{pmatrix} 0 \\ 0.1 \end{pmatrix}$$

$$\beta_1 = \frac{\left\|\nabla f\left(\boldsymbol{x}^{[1]}\right)\right\|^2}{\left\|\nabla f\left(\boldsymbol{x}^{[0]}\right)\right\|^2} = 0.01$$

$$\boldsymbol{d}^{[1]} = -\nabla f\left(\boldsymbol{x}^{[1]}\right) + \beta_1\, \boldsymbol{d}^{[0]} = \begin{pmatrix} 0 \\ -0.1 \end{pmatrix} + 0.01 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.01 \\ -0.1 \end{pmatrix}$$

$$\alpha_1 = \frac{-\nabla f\left(\boldsymbol{x}^{[1]}\right)^T \boldsymbol{d}^{[1]}}{\left(\boldsymbol{d}^{[1]}\right)^T H\, \boldsymbol{d}^{[1]}} = \frac{10}{9}$$

$$\boldsymbol{x}^{[2]} = \boldsymbol{x}^{[1]} + \alpha_1\, \boldsymbol{d}^{[1]} = \ldots = \frac{1}{9} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \boldsymbol{x}^*$$

Icke-linjära minsta-kvadrat-problem, Heath 6.4

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}; \quad \boldsymbol{f}: \mathbb{R}^m \rightarrow \mathbb{R}^n, m > n$$

Vi har ett överbestämt system av icke-linjära ekvationer. Minsta-kvadrat-lösning:

$$\min_{\boldsymbol{x}} \|\boldsymbol{f}(\boldsymbol{x})\|_2$$

Stegvis linjärisering som vid Newtons metod. Taylorutveckling kring en approximation $\boldsymbol{x}_k$:

$$\boldsymbol{f}(\boldsymbol{x}) \approx \boldsymbol{f}(\boldsymbol{x}_k) + \mathcal{J}(\boldsymbol{x}_k)\,(\boldsymbol{x} - \boldsymbol{x}_k)$$

Linjär modell:

$$\min_{\boldsymbol{x}} \|\boldsymbol{f}(\boldsymbol{x})\|_2 \approx \min_{\boldsymbol{x}} \|\boldsymbol{f}(\boldsymbol{x}_k) + \mathcal{J}(\boldsymbol{x}_k)\,(\boldsymbol{x} - \boldsymbol{x}_k)\|_2$$

Detta är ett *linjärt* minsta-kvadrat-problem som alltså ska lösas i varje iteration.

Gauss-Newtons metod: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{d}_k$, $\mathcal{J}(\boldsymbol{x}_k)\boldsymbol{d}_k = -\,\boldsymbol{f}(\boldsymbol{x}_k)$. Ser ut som Newtons metod, men innebörden i det ekvationssystemet är helt annorlunda: lösning av linjärt minsta-kvadrat-problem medan i vanlig Newton är det lösning av ett linjärt ekvationssystem.

Typisk tillämpning Anpassning av icke-linjär modell till mätdata.

EXEMPEL $\psi(\boldsymbol{x}, t) = x_1 + x_2\,\mathrm{e}^{-x_3 t}$, (t_i, ψ_i) uppmätta värden, $i = 1, ..., m$. Överbestämt system om $m > 3$. $f_i = \psi(\boldsymbol{x}, t_i) - \psi_i$ med

$$\boldsymbol{f} = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}, \quad \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}, \min_{\boldsymbol{x}} \|\boldsymbol{f}(\boldsymbol{x})\|_2$$

EXEMPEL. Byråkratisk tillväxtmodell, dagsproduktion av EU-papper (kilo A4).

$t =$	0	1	2.5	3	3.5	4	5	6
$Y =$	4	8	54	76	78	66	75	130

Modell: $Y = a\,e^{qt} + b\,\mathrm{e}^{-(t-t_0)^2}$: exponentiell tillväxt med superponerad topp vid extrem händelse (Maastricht-avtalet).

a) Anta $q = 0.5$ och $t_0 = 3$. Bestäm a och b med linjär minsta-kvadrat.

Lösning:

$$A = \begin{pmatrix} | & | \\ \mathrm{e}^{0.5}\,t_i & \mathrm{e}^{-(t_i-3)^2} \\ | & | \end{pmatrix}, \quad \boldsymbol{Y} = \begin{pmatrix} 4 \\ 8 \\ \vdots \\ 130 \end{pmatrix}$$

$$A \begin{pmatrix} a \\ b \end{pmatrix} = \boldsymbol{Y}$$

b) Icke-linjär minsta-kvadrat alla 4 parametrarna, sätt upp problemet med Gauss-Newtons metod. Residual:

$$f_i = a\,\mathrm{e}^{q t_i} + b\,\mathrm{e}^{-(t_i-t_0)^2} - Y_i$$

Jakobianen:

$$\mathcal{J} = \begin{pmatrix} | & | & | & | \\ \mathrm{e}^{q t_i} & t_i\,a\,\mathrm{e}^{q t_i} & \mathrm{e}^{-(t_i-t_0)^2} & 2\,(t_i - t_0)\,b\,\mathrm{e}^{-(t_i-t_0)^2} \\ | & | & | & | \end{pmatrix}$$

$$\boldsymbol{x} = \begin{pmatrix} a \\ q \\ b \\ t_0 \end{pmatrix}$$

Gauss-Newton

$$\begin{cases} \boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{d}_k \\ \mathcal{J}(\boldsymbol{x}_k)\,\boldsymbol{d}_k = -\,\boldsymbol{f}(\boldsymbol{x}_k) \end{cases}$$

$\boldsymbol{x}_0$ kan tas från linjära modellen.