Další přístupy k identifikaci

doc. Ing. Petr Blaha, PhD.

Komplexní inovace studijních programů a zvyšování kvality výuky na FEKT VUT v Brně
OP VK CZ.1.07/2.2.00/28.0193

evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Identifikace v uzavřené smyčce

Motivace

V některých případech je nutné provádět identifikaci v uzavřené smyčce

- systém je zapojen do zpětnovazebního obvodu, ve kterém pracuje
- v otevřené smyčce je systém nestabilní
- zpětná vazba je vyžadována z bezpečnostních důvodů

Používané metody

Pro identifikaci v uzavřené smyčce se používají následující metody

- přímý přístup - stejně jako při identifikaci v otevřené smyčce, nepoužívá se žádaná hodnota $w(k)$, jenom vstup $u(k)$ a výstup $y(k)$
- nepřímý přístup - identifikuje se přenos uzavřené smyčky a na základě znalosti přenosu regulátoru se vypočítá přenos systému

Motivační příklad - přímý přístup

Uvažujme systém prvního řádu s neznámými parametry a a b

$$y(k) + ay(k-1) = bu(k-1) + e(k)$$

s lineární zpětnou vazbou

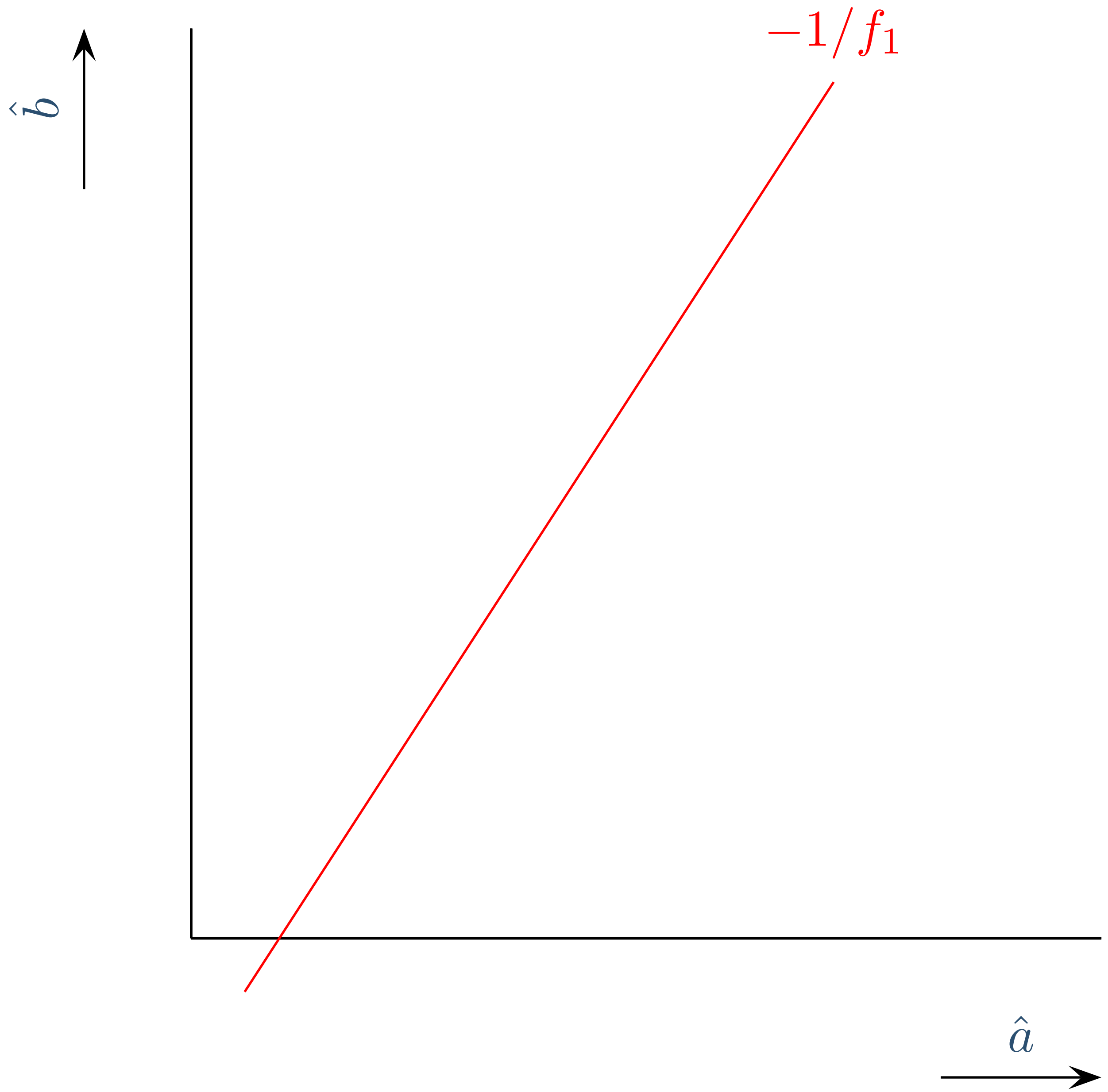
$$u(k) = -fy(k)$$

Zavedení zpětné vazby způsobí

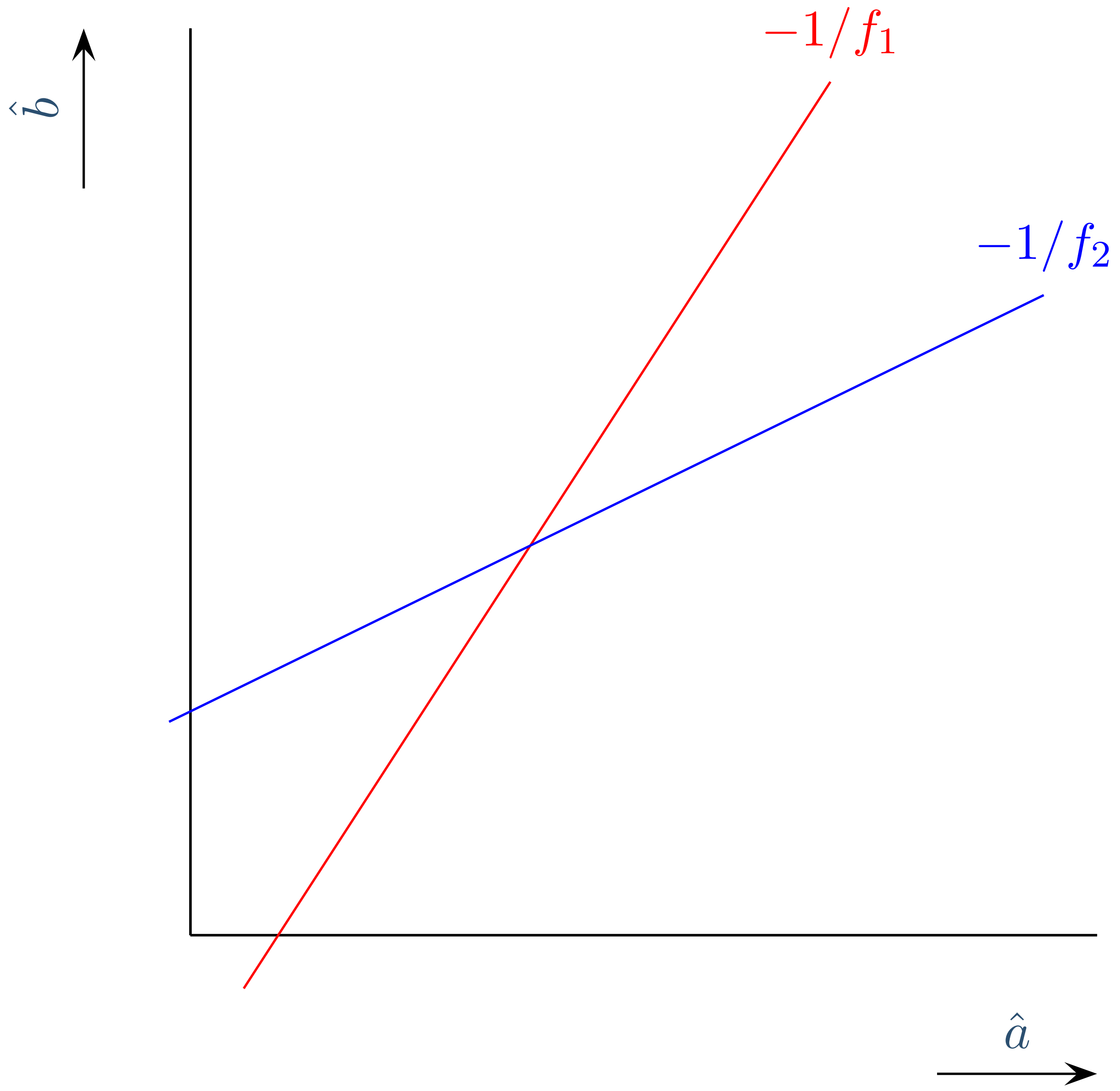
$$y(k) + (a + bf)y(k-1) = e(k)$$

Všechny modely $\hat{a} + \hat{b}f = c$ jsou řešením.
Identifikací získáme přímku $\hat{b} = -\frac{\hat{a}}{f} + \frac{c}{f}$, na které leží parametry identifikovaného systému - nedostaneme jednoznačnou dvojici parametrů. Nepomůže nám ani znalost zesílení f.

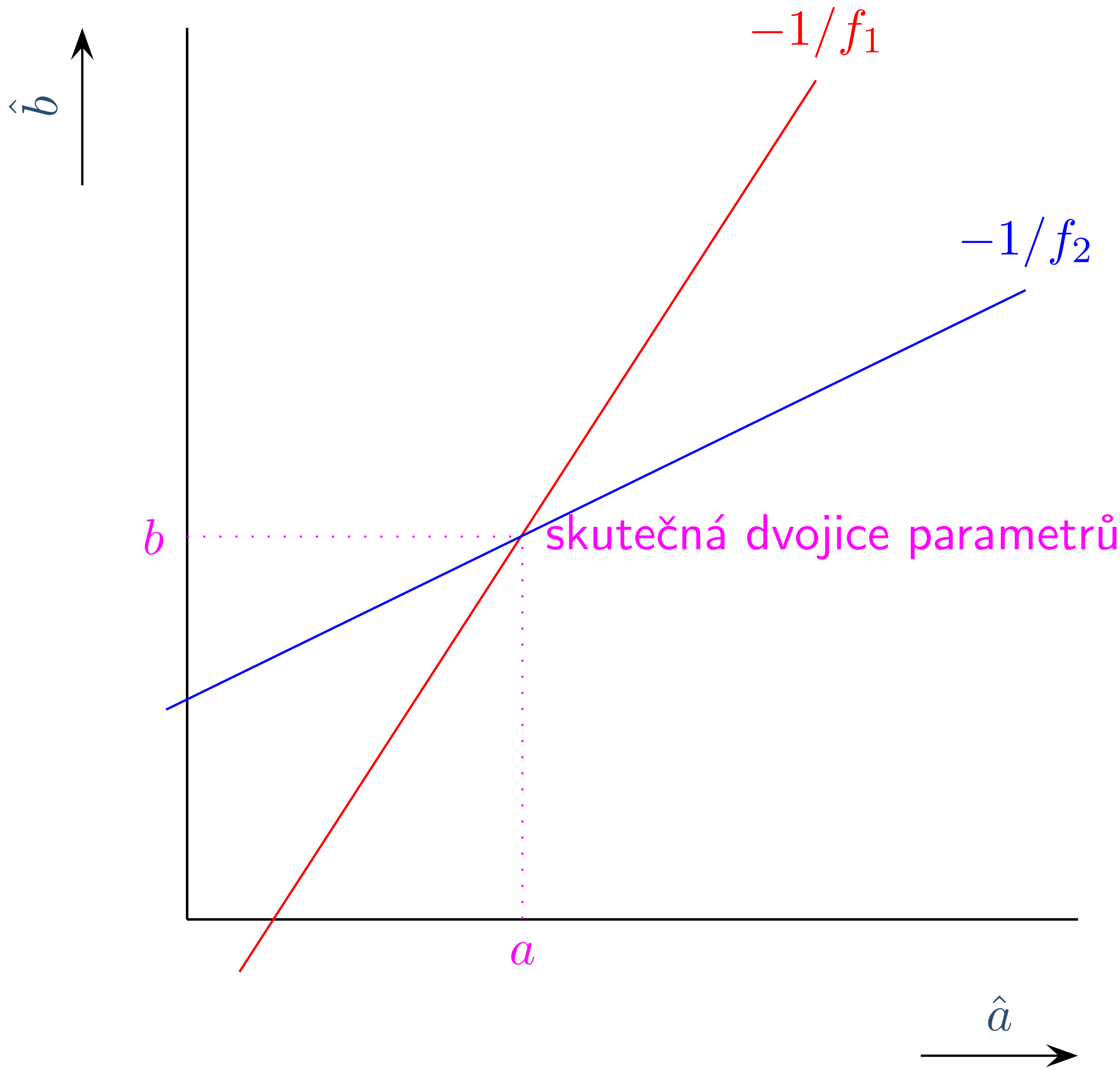
Ilustrace ztráty jednoznačnosti



Ilustrace ztráty jednoznačnosti



Ilustrace ztráty jednoznačnosti



Řešení problému

Použije se lineární zpětná vazba dostatečně vysokého řádu. Sloupce matice Φ pak přestanou být lineárně závislé. Pro předchozí příklad stačí třeba

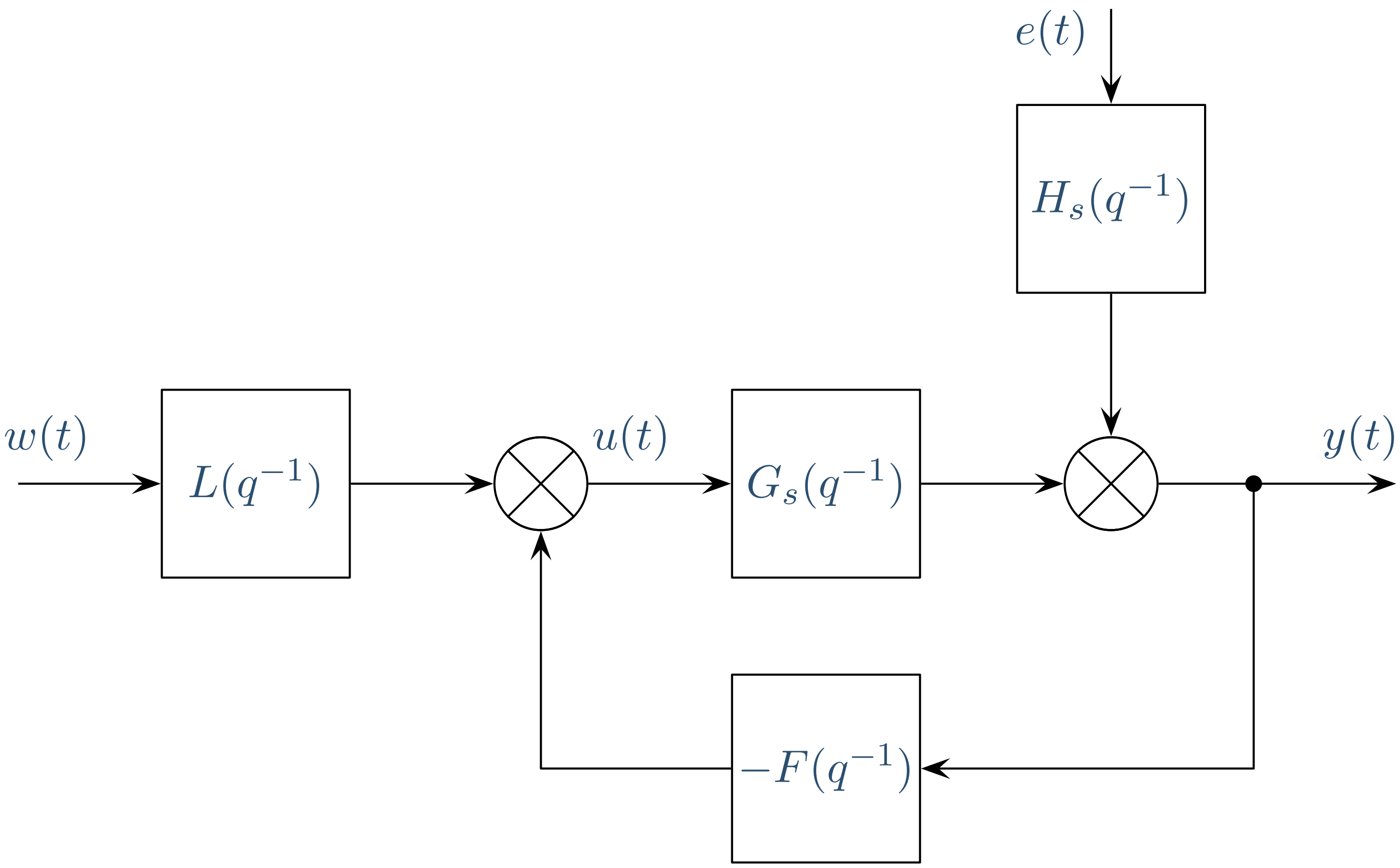
$$u(k) = -f_1 y(k) - f_2 y(k-1)$$

nebo se použije časově proměnná zpětná vazba

$$u(k) = -f(k)y(k)$$

V našem případě stačí použít dvě různé hodnoty zesílení. Pro dvě zesílení dostaneme dvě přímky, výsledná dvojice parametrů odpovídá průsečíku přímek.

Zpětnovazební zapojení



Popis

Mějme systém

$$
\begin{aligned}
y(k) &= G_s(q^{-1})u(k) + H_s(q^{-1})e(k) \\
u(k) &= -F(q^{-1})y(k) + L(q^{-1})w(k)
\end{aligned}
\tag{1}
$$

- $w(k)$ může být žádaná hodnota nebo šum vstupující do regulátoru
- $F(q^{-1})$ a $L(q^{-1})$ jsou regulátory

Cílem identifikace je určení $G_s(q^{-1})$ a $H_s(q^{-1})$. V některých případech také $F(q^{-1})$.

Předpoklady

Předpokládáme, že

- systém $G_s(q^{-1})$ neobsahuje přímou vazbu na výstup - zpožďuje minimálně o jeden krok (požadavek proti vzniku algebraické smyčky)
- subsystémy $L(q^{-1})$, $H_s(q^{-1})$ a $[I + G_s(q^{-1})F(q^{-1})]^{-1}$ jsou asymptoticky stabilní a nemají skryté nestabilní kořeny
- vstup $w(k)$ je budicím signálem postačujícího řádu.
- vstup $w(k)$ a porucha $e(k)$ jsou vzájemně nezávislé

Problém

- lineární zpětná vazba způsobí vznik lineárně závislých sloupců v matici Φ

$$\Phi = \begin{pmatrix} -y(n) & -y(n-1) & \cdots & -y(1) & u(n) & u(n-1) & \cdots & u(1) \\ -y(n+1) & -y(n) & \cdots & -y(2) & u(n+1) & u(n) & \cdots & u(2) \\ & & & & \vdots & & & \\ -y(N-1) & -y(N-2) & \cdots & -y(N-n) & u(N-1) & u(N-2) & \cdots & u(N-n) \end{pmatrix}$$

- parametry potom nemohou být určeny jednoznačně
- spektrální analýza funguje v případě nulové poruchy a nenulového vstupu $w(k)$, jinak lze v krajním případě obdržet jako výsledek identifikace $\hat{G}(e^{-\jmath\omega}) = -\frac{1}{F(e^{-\jmath\omega})}$

IVM

Uvažujme skalární systém

$$A(q^{-1})y(k) = B(q^{-1})u(k) + v(k) \tag{2}$$

kde je vstup $u(k)$ daný zpětnou vazbou RST regulátoru

$$R(q^{-1})u(k) = -S(q^{-1})y(k) + T(q^{-1})w(k) \tag{3}$$

Srovnání s obecným tvarem (1)

$$G_S(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})}$$

$$F(q^{-1}) = \frac{S(q^{-1})}{R(q^{-1})}$$

$$L(q^{-1}) = \frac{T(q^{-1})}{R(q^{-1})}$$

Sloučením rovnic (2) a (3) dostaneme (u polynomů jsou vynechány (q^{-1}))

$$[AR + BS]y(k) = BTw(k) + Rv(k)$$

$$[AR + BS]u(k) = ATw(k) - Sv(k)$$

Vyfiltrovaný vstup $\tilde{u}(k)$ a výstup $\tilde{y}(k)$

$$\tilde{u}(k) = \frac{AT}{AR + BS} w(k)$$

$$\tilde{y}(k) = \frac{BT}{AR + BS} w(k) = \frac{B}{A} \tilde{u}(k)$$

Pokračování

Pro vektor měřených veličin $\varphi(k)$ platí

$$\varphi(k) = \begin{pmatrix} -y(k-1) & \cdots & -y(k-n_a) & u(k-1) & \cdots & u(k-n_b) \end{pmatrix}^T$$

Vektor pomocných proměnných

$$\zeta(k) = \begin{pmatrix} -\tilde{y}(k-1) & \cdots & -\tilde{y}(k-n_a) & \tilde{u}(k-1) & \cdots & \tilde{u}(k-n_b) \end{pmatrix}^T$$

K realizaci algoritmu pomocné proměnné systému se zpětnou vazbou potřebujeme znát nejen přibližné parametry systému, ale také regulátor.

evropský sociální fond v ČR EVROPSKÁ UNIE MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Nepřímá identifikace

Vstup $w(t)$ musí být znám a musí být persistentně budicím signálem. Postup se skládá ze dvou kroků:

- Indentifikace přenosu uzavřené smyčky na základě znalosti $w(t)$ a $y(t)$
- Určení přenosu soustavy na základě znalosti zpětné vazby (regulátoru)

Metody přímého hledání extrému

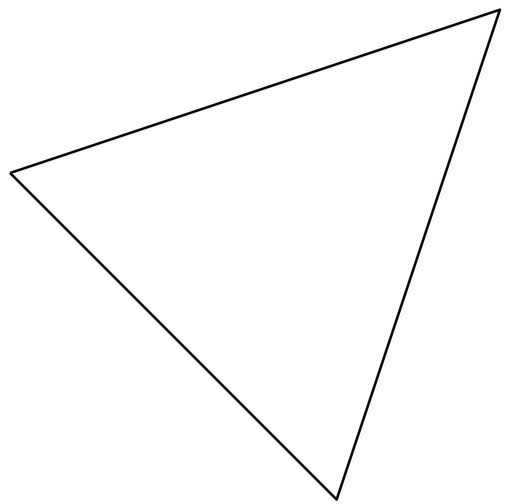
Simplexová metoda

Nelder-Mead - 1965

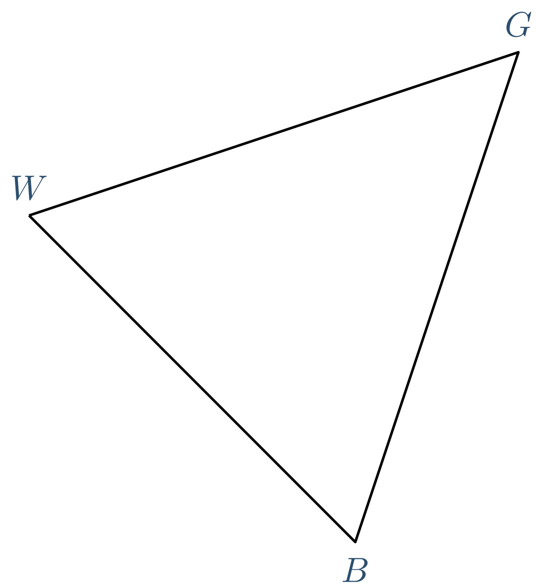
Simplex (či n-simplex) je n-rozměrným zobecněním trojúhelníku. Jedná se o konvexní obal $n+1$ afinně nezávislých bodů umístěný v euklidovském prostoru dimenze n či vyšší. Metoda pracuje s $n+1$ body v n rozměrném prostoru. Během jedné iterace může dojít k

- převrácení (reflexion)
- protažení (expansion)
- zkrácení (contraction)
- sražení (shrink)

Vysvětlení algoritmu

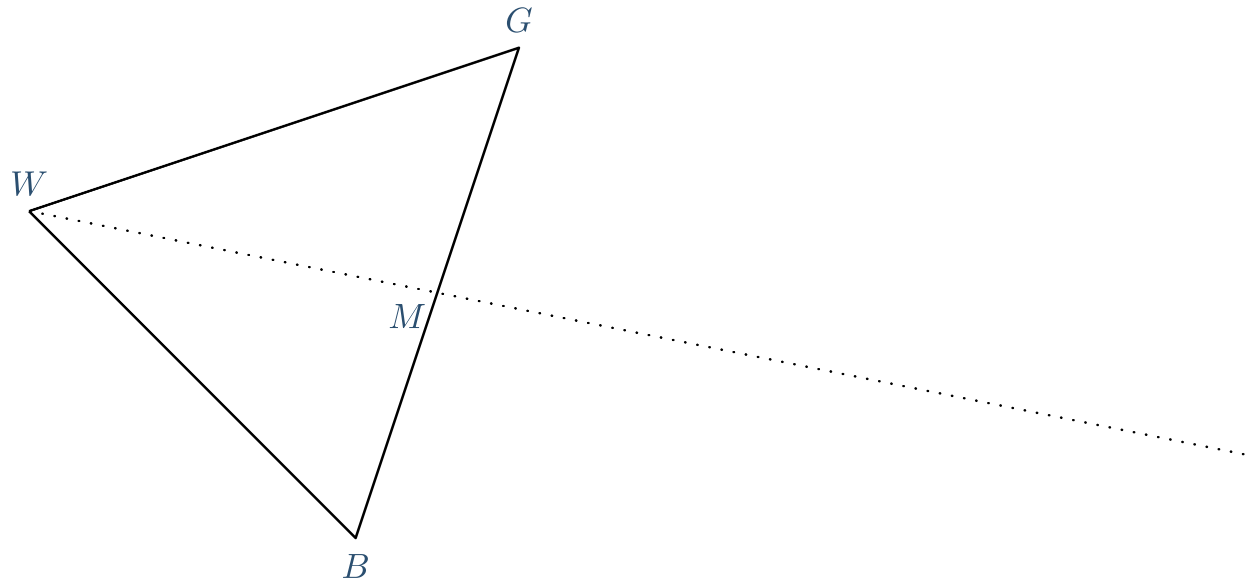


Při určování dvou neznámých parametrů (řešení ve dvourozměrném prostoru), máme tři body, které tvoří trojúhelník.



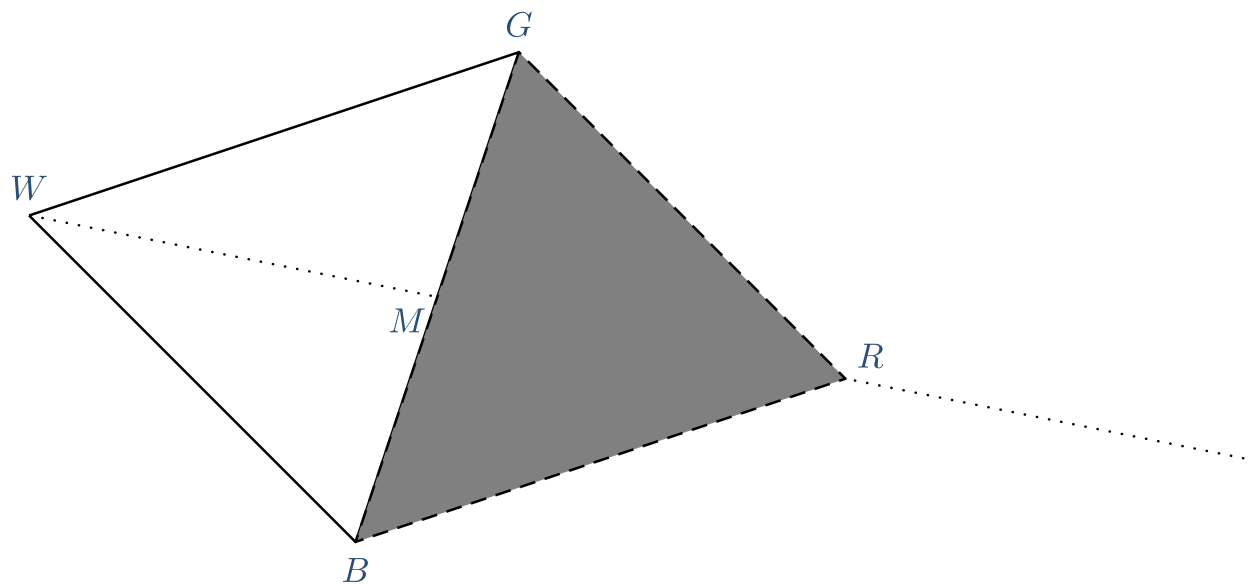
V každém vrcholu určíme hodnotu kriteriální funkce $V(x_k, y_k)$, kde $k = 1, 2, 3$ a seřadíme je od nejmenší hodnoty $B = (x_1, y_1)$ -Best, přes prostřední $G = (x_2, y_2)$ -Good, po největší $W = (x_3, y_3)$ -Worst

Vysvětlení algoritmu

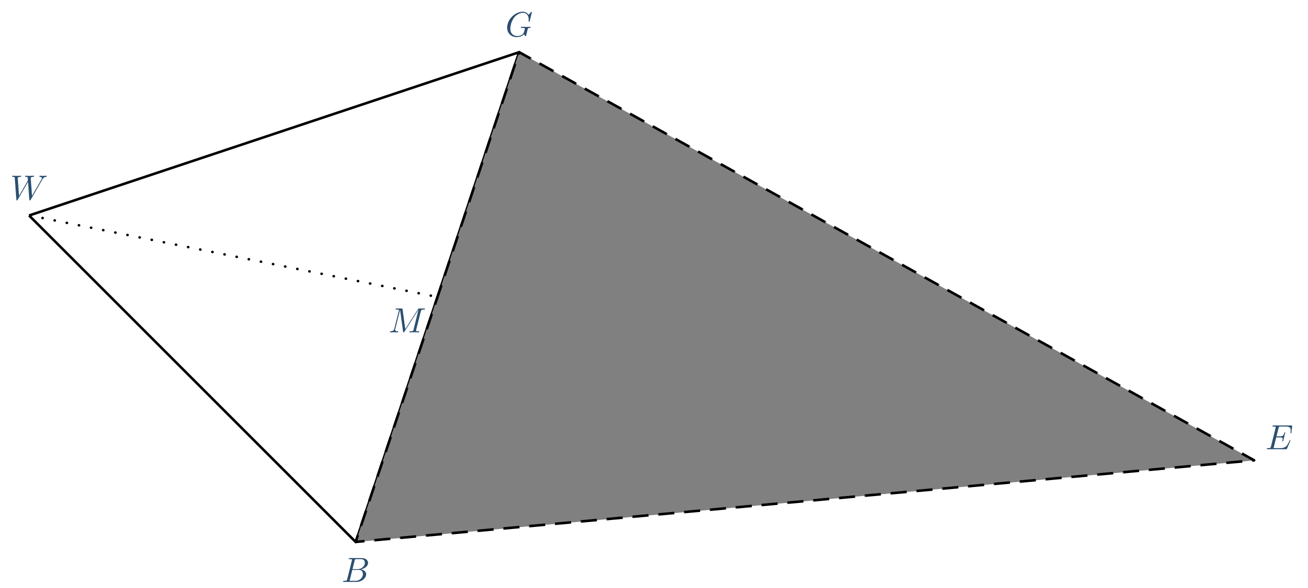


Vytvoříme střed úsečky $M = \frac{1}{2}(B + G)$ mezi body s nejnižší hodnotou kritéria a vedeme jím a bodem W přímku.

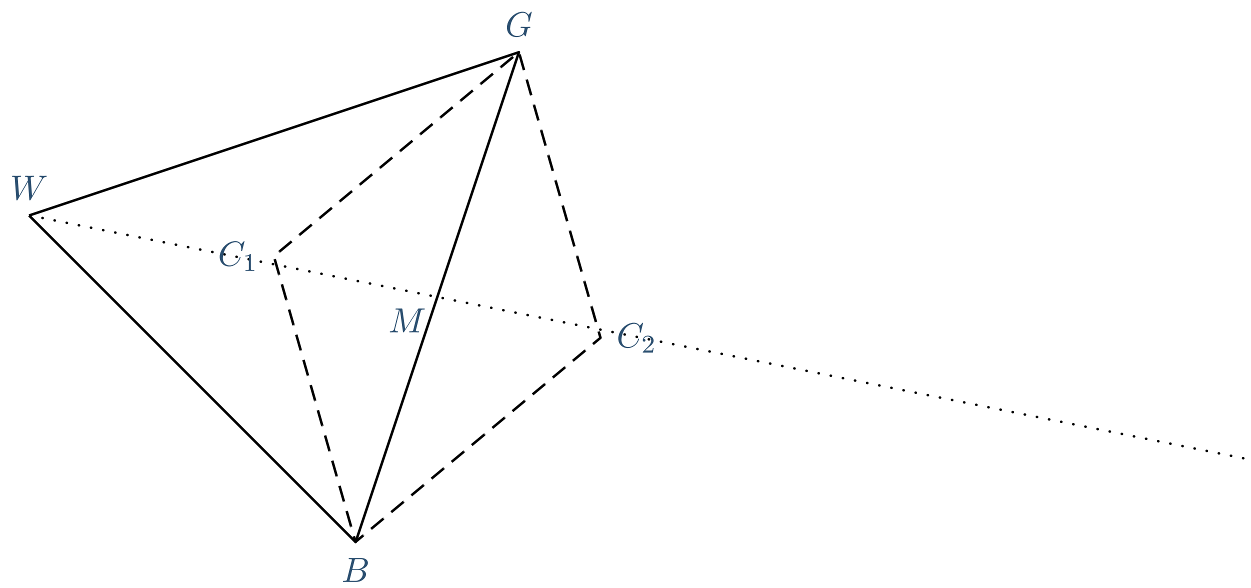
Vysvětlení algoritmu



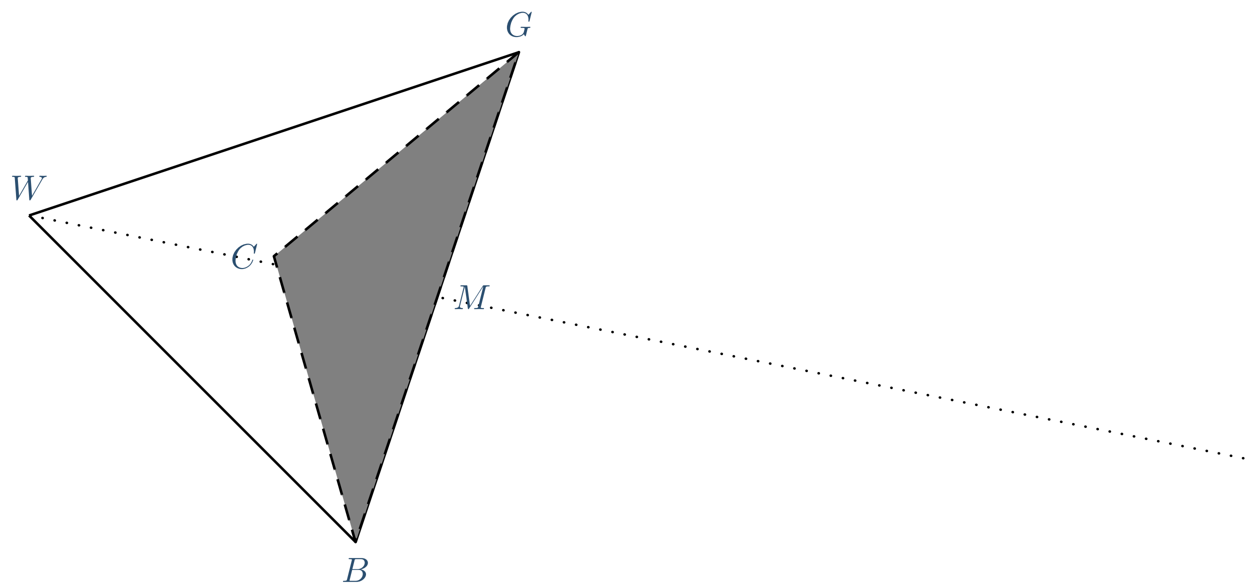
Provedeme převrácení trojúhelníka kolem úsečky BG. V nově vzniklém bodě $R = M + (M - W) = 2M - W$ otestujeme, zda je hodnota kritéria nižší než v W a větší než B, nahradíme nejhoršího a vracíme se na začátek.



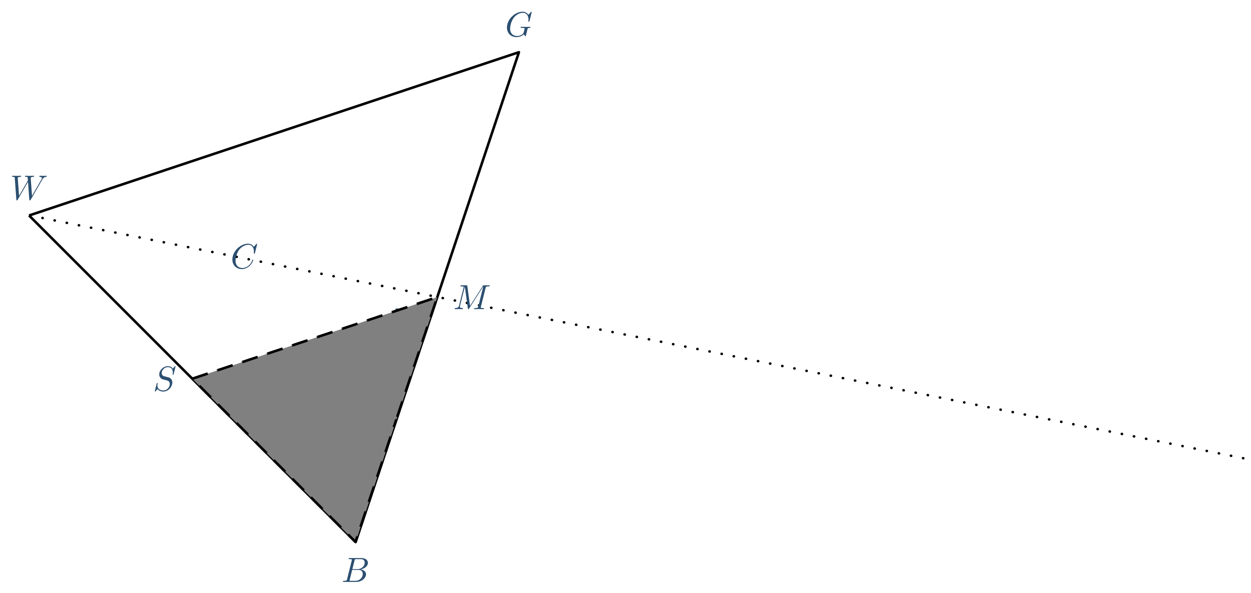
Jestliže je hodnota kritéria v R menší než B, přesunuli jsme se správným směrem k minimu. Vyzkoušíme, jestli není minimum ještě dál v bodě $E = R + (R - M) = 2R - M$. Srovnáme kritérium v bodech R a E. Bod s menší hodnotou je novým vrcholem trojúhelníka a vracíme se na začátek.



Jestliže je hodnota kritéria v R větší, zkusíme provést zkrácení. Nabízí se dva body C_1 a C_2.



Vybereme bod C, ve kterém je hodnota kritéria menší a zároveň menší než v bodě W.



Pokud je hodnota kritéria v bodě C větší než v bodě W, provedeme sražení trojúhelníka. Celý algoritmus se opakuje s nově vzniklým trojúhelníkem.

Zhodnocení algoritmu

- metoda je efektivní - na každou iteraci připadá v průměru 1-2 vyhodnocení kriteriální funkce
- v praxi může proběhnout velké množství iterací
- v některých případech vyžaduje restart
- výhodou je jednoduchá implementace, nevyžaduje počítat derivace, nestará se o hladkost funkce

Příklad

Pomocí příkazu fminsearch v Matlabu určete parametry systému k, T_1, T_2, T_3 s předpokládaným přenosem

$$F(p) = \frac{k}{(T_1p + 1)(T_2p + 1)(T_3p + 1)}$$

ze změřených vstup výstupních dat.
Nejprve vygenerujeme data

```
>> global t u y
>> t=0:0.1:100-0.1;
>> u=randn(1000,1);
>> f=zpk([],[-1 -1/2 -1/3],1/3);
>> %k=2, T1=1, T2=2, T3=3
>> y=lsim(f,u,t);
```

Nadefinujeme funkci, která počítá součet kvadrátů odchylek

```
function krit=sumkv(x)
global u t y
T1 = x(2);
T2 = x(3);
T3 = x(4);
k = x(1)/T1/T2/T3;
fm = zpk([],[-1/T1 -1/T2 -1/T3],k);
ym = lsim(fm,u,t);
krit = (y-ym)'*(y-ym)
```

Nakonec zavoláme funkci fminsearch s nastaveným počátečním odhadem parametrů

```
>>  param = fminsearch('sumkv',[1,1,1,1])
param =
    1.9975    2.8295    0.9536    2.1962
```

Metoda Hooke-Jeeves

Metoda Hooke-Jeeves

Prochází n-rozměrný prostor v n nezávislých pevných směrech (obvykle ortogonální, např. podél souřadnicových os).
Vychází se z počátečního bodu θ_0, v jednom ze směrů s daným krokem. Pokud je hodnota kritéria větší, zkusím opačný směr. Pokud je i v tomto bodě hodnota kritéria větší, zkrátíme krok. Stejným způsobem probíhá vyhledávání i ve všech ostatních směrech. Tím získáme nový bod θ_1 a postup se opakuje.
Dobrá volba vyhledávání extrému je ve směru $\theta_k - \theta_{k-1}$ (aproximace gradientu v k-té iteraci). Jestliže je hodnota kritéria v takto získaném bodě nižší, pak ho bereme jako nový bod pro hledání, jinak postupujeme podle předchozího postupu hledání ve všech směrech.

Gradientní metody

- Line search
- Gradientní (spádová) metoda (Steepest descent)
- Newtonova metoda
- Modifikovaná Newtonova metoda (Quasi-Newton method)
- Conjugate gradient method

Obecný princip gradientních metod je provést změnu vektoru parametrů θ_{k-1} s vahou ν_{k-1} ve směru vektoru p_{k-1}, který je dán směrem gradientu g_{k-1}, který může být natočen, prodloužen případně zkrácen maticí R_{k-1}.

$$\theta_k = \theta_{k-1} - \nu_{k-1} p_{k-1} \qquad \text{kde} \qquad p_{k-1} = R_{k-1} g_{k-1} \tag{4}$$

Gradient se pro případ kriteriální funkce dané součtem kvadrátů odchylek dá vyjádřit jako

$$g = \frac{\partial J(\theta)}{\partial \theta} = \frac{\partial \frac{1}{2} \sum_{i=1}^{N} e^2(i)}{\partial \theta} = \frac{1}{2} \sum_{i=1}^{N} \frac{\partial e^2(i)}{\partial \theta} = \sum_{i=1}^{N} e(i) \frac{\partial e(i)}{\partial \theta} \tag{5}$$

kde $e(i) = y(i) - \hat{y}(i)$.

Obecný popis gradientních metod - pokračování

Jedna složka gradientu v bodě θ ve směru i se dá aproximovat jako

$$g_i(\theta) \approx \frac{J(\theta + \Delta\theta_i) - J(\theta)}{\Delta\theta_i} \tag{6}$$

Výsledný gradient je dán vyčíslením g_i ve všech směrech $i = 1 \dots n$.
Pro takovéto vyčíslení gradientu potřebejeme n vyčíslení kriteriální funkce $J(\theta)$.
Podobným způsobem se dá určit také Hesián. i-tý sloupec h_i Hesiánu H se určí jako

$$h_i(\theta) \approx \frac{g(\theta + \Delta\theta_i) - g(\theta)}{\Delta\theta_i} \tag{7}$$

Poznámky

Výpočetně velmi náročné, spíš teoretické, použití pouze v případě, kdy se dá vyjádřit analyticky.
V praxi se řeší volba $\Delta\theta_i$. Co nejmenší z důvodu přesnosti x dost velký na to, aby se neprojevily kvantizační chyby při výpočtu.

Spádová metoda

Matice $R = I$, potom

$$\theta_k = \theta_{k-1} - \nu_{k-1} g_{k-1} \tag{8}$$

Směr hledání je tedy v opačném směru gradientu.

Newtonova metoda

Matice R se bere jako inverze Hesiánu H_{k-1}^{-1} ztrátové funkce v bodě θ_{k-1}

$$\theta_k = \theta_{k-1} - \nu_{k-1} H_{k-1}^{-1} g_{k-1} \tag{9}$$

Hlavní nevýhodou je požadavek na znalost druhých derivací a na inverzi matice.

Modifikovaná Newtonova metoda

Výpočet inverze Hesiánu může být příliš složitý (velký počet parametrů)
Modifikovaná Newtonova metoda nahrazuje Hesián nebo inverzi Hesiánu jeho/její aproximací.

$$\theta_k = \theta_{k-1} - \nu_{k-1} H_{k-1}^{-1} g_{k-1} \qquad (10)$$

$$H_k = H_{k-1} + Q_{k-1} \qquad \text{nebo} \qquad H_k^{-1} = H_{k-1}^{-1} + \widetilde{Q}_{k-1} \qquad (11)$$

Hlavní nevýhodou je požadavek na znalost druhých derivací a na inverzi matice.

Nejčastěji se řeší s H_k^{-1}.
Existuje spousta variant: Davidon–Fletcher–Powell metoda, Broyden–Fletcher–Goldfarb–Shanno (BFGS) methoda, Broydenova metoda, Symmetric Rank 1 (SR1) metoda.
Nejlepší výsledky dává většinou BFGS

$$\Delta\theta_{k-1} = \theta_k - \theta_{k-1} \qquad \Delta g_{k-1} = g_k - g_{k-1} \tag{12}$$

$$H_k^{-1} = \left(I - \frac{\Delta\theta_{k-1}\Delta g_{k-1}^T}{\Delta\theta_{k-1}^T\Delta g_{k-1}}\right) H_{k-1}^{-1} \left(I - \frac{\Delta\theta_{k-1}\Delta g_{k-1}^T}{\Delta\theta_{k-1}^T\Delta g_{k-1}}\right)^T + \frac{\Delta\theta_{k-1}\Delta\theta_{k-1}^T}{\Delta\theta_{k-1}^T\Delta g_{k-1}} \tag{13}$$

Modifikovaná Newtonova metoda - vlastnosti

- nevyžaduje výpočet druhých derivací
- kvadratická výpočetní náročnost z důvodu maticového násobení
- kvadratické požadavky na paměť z důvodu uložení Hesiánu
- velmi rychlá konvergence
- vhodné na problémy se střední složitostí (kolem 100 parametrů)

Nelineární nejmenší čtverce

Nelineární nejmenší čtverce

Uvažujeme kriteriální funkci

$$J(\theta) = \sum_{i=1}^{N} f^2(i,\theta) = \mathbf{f}^T\mathbf{f} \tag{14}$$

kde $\mathbf{f} = (f(1,\theta), f(2,\theta), \ldots, f(N,\theta))^T$
Potom j-tá složka gradientu:

$$g_j = 2\frac{\partial J(\theta)}{\partial \theta_j} = 2\sum_{i=1}^{N} f(i,\theta)\frac{\partial f(i,\theta)}{\partial \theta_j} \tag{15}$$

Jakobián je možné napsat jako

$$\mathbf{J} = \begin{pmatrix} \frac{\partial f(1)}{\partial \theta_1} & \cdots & \frac{\partial f(1)}{\partial \theta_N} \\ \vdots & & \vdots \\ \frac{\partial f(N)}{\partial \theta_1} & \cdots & \frac{\partial f(N)}{\partial \theta_N} \end{pmatrix} \tag{16}$$

NLS

Gradient se pak dá naspat jako

$$\mathbf{g} = 2\mathbf{J}^T\mathbf{f} \tag{17}$$

Hesián

$$\mathbf{H} = 2\mathbf{J}^T\mathbf{J} + 2\mathbf{S} \tag{18}$$

$\mathbf{S}$ se v některých případech vynechává.

Gauss-Newtonova metoda

$$\theta_k = \theta_{k-1} - \eta_{k-1}\left(\mathbf{J}_{k-1}^T\mathbf{J}_{k-1}\right)^{-1}\mathbf{J}_{k-1}^T\mathbf{f}_{k-1} \tag{19}$$

Levenberg-Marquardt metoda

$$\theta_k = \theta_{k-1} - \eta_{k-1}\left(\mathbf{J}_{k-1}^T\mathbf{J}_{k-1} + \alpha_{k-1}\mathbf{I}\right)^{-1}\mathbf{J}_{k-1}^T\mathbf{f}_{k-1} \tag{20}$$

Předchozí metody neberou v úvahu tvar kriteriální funkce, jsou obecné. Většinou se kriteriální funkce volí jako váhovaný součet kvadrátů odchylek. Pro tento případ se dá odvodit tvar Hesiánu, jeho výpočet je jednodušší a dají se použít specializované metody

- Gauss-Newtonova metoda
- Levenberg-Marquardtova metoda

Modely bez zpětné vazby

Modely bez zpětné vazby

Jak bylo řečeno dříve, problémy s posunutím odhadu jsou způsobeny zpětnou vazbou od výstupu použitou v modelu. Modely, které pracují bez této zpětné vazby neduhem posunutého odhadu netrpí. Jejich problémem však často bývá požadavek na větší počet určovaných parametrů.

- modely s konečnou impulsní charakteristikou (FIR)
- modely s ortonormálními bázovými funkcemi (OBF)
  - Laguerre filtry
  - Kautz filtry

Laguerre filtry

Jsou založeny na Laguerre-ových polynomech, které jsou ortonormálních jak v časové oblasti, tak v obraze.

$$F_{i,k}(p) = \sqrt{2p_i}\frac{(p-p_i)^{k-1}}{(p+p_i)^k} \tag{21}$$

Dají se reprezentovat žebříčkovou strukturou. Připomíná systém prvního řádu s aproximací dopravního zpoždění do Padého rozvoje.

Kautz filtry

Laguerre filtry slouží pro modelování tlumených systémů. Pro modelování kmitavých systémů se hodí spíš Kautz filtry, protože zahrnují dvojici komplexně sdružených pólů

$$F_{2i-1}(p) = \frac{\sqrt{(1-a^2)(1-b^2)}}{p^2 + a(b-1)p - b} g(a,b,p,i) \tag{22}$$

$$F_{2i}(p) = \frac{\sqrt{(1-b^2)(p-a)}}{p^2 + a(b-1)p - b} g(a,b,p,i) \tag{23}$$

$$g(a,b,p,i) = \left(\frac{-bp^2 + a(b-a)p + 1}{p^2 + a(b-1)p - b} \right)^{i-1} \tag{24}$$