





Metody nehladkých rovnic pro řešení úloh nelineárního programování

**Ladislav Lukšan,
Ctirad Matonoha, Jan Vlček**

Ústav informatiky AV ČR, v. v. i.
Pod Vodárenskou věží 2, 182 00 Praha 8

PANM 22, 23.-28.6.2024

-  Lukšan L.:
Numerické optimalizační metody. Učební text.
Ústav informatiky AV ČR, v.v.i., Praha, 2023.
-  Lukšan L., Matonoha C., Vlček J.:
Nonsmooth equation method for nonlinear nonconvex optimization.
In: Conjugate Gradient Algorithms and Finite Element Methods (M.Křížek,
P.Neittaanmäki, R.Glovinski, S.Korotov eds.). Springer Verlag, Berlin (2004), 131-145.
-  Wächter A., Biegler L.T.:
Line search filter methods for nonlinear programming - motivation and global convergence.
SIAM J. on Computation 16 (2005), 1-31.
-  Wächter A., Biegler L.T.:
Line search filter methods for nonlinear programming - local convergence.
SIAM J. on Computation 16 (2005), 32-48.

Úloha nelineárního programování

Úloha

Minimalizovat $F(x)$, pokud $c_k(x) \leq 0$, $1 \leq k \leq m$.

Podmínky optimality (Kuhn, Tucker)

$$\begin{aligned}g(x, u) &= 0, \\c(x) &\leq 0, \quad u \geq 0, \quad UC(x)e = 0,\end{aligned}$$

kde

$$g(x, u) = \nabla F(x) + \sum_{k=1}^n u_k \nabla c_k(x) = \nabla F(x) + A(x)u,$$

$$A(x) = [\nabla c_k(x), 1 \leq k \leq m].$$

Přitom $u \in R^m$ je vektor **Lagrangeových multiplikátorů**,

$$U = \text{diag}(u_k, 1 \leq k \leq m), \quad C(x) = \text{diag}(c_k(x), 1 \leq k \leq m)$$

a e je vektor s jednotkovými prvky.

Metody nelineárního programování

- Sekvenční kvadratické programování (SQP): Úloha

Minimalizovat $Q(d) = \frac{1}{2}d^T B d + g^T d$, pokud $A^T d + c \leq 0$.

- Metody vnitřních bodů (IP): Posloupnost úloh

Minimalizovat $F(x) - \mu e^T \log(S)e$, pokud $c(x) + s = 0$,

kde $S = \text{diag}(s_k, 1 \leq k \leq m) > 0$ a $\mu \rightarrow 0$.

- Metody nehladkých rovnic (NE): Úloha

Minimalizovat $F(x)$, pokud $h(x, u) = 0$.

Rovnice $h(x, u) = 0$ jsou polohladké.

Metoda nehladkých rovnic

Podmínky optimality se vyjádří ve tvaru

$$g(x, u) = 0, \quad h(x, u) = 0,$$

kde $h(x, u) = [\psi(-c_k(x), u_k), 1 \leq k \leq m]^T$.

Fisherova–Burmeisterova funkce

$$\psi(a, b) = \sqrt{a^2 + b^2} - (a + b)$$

- se rovná nule právě tehdy, když $a \geq 0$, $b \geq 0$, $ab = 0$;
- je spojitě diferencovatelná, pokud $|a| + |b| \neq 0$;
- je polohladká, pokud $|a| + |b| = 0$;
- funkce $\psi^2(a, b)$ je spojitě diferencovatelná v R^2 .

Zobrazení

$$f(z) = [g(z), h(z)]^T = [g(x, u), h(x, u)]^T$$

je polohladké v bodě z . Platí

$$f'(z, d) = Jd + o(\|d\|), \quad \text{pokud } \|d\| \rightarrow 0 \quad \text{a} \quad J \in \partial f(z + d),$$

takže

$$f(z + d) - f(z) = f'(z, d) + o(\|d\|) = Jd + o(\|d\|).$$

Určení směrového vektoru

Soustava rovnic $f(z + d) \approx f(z) + Jd = 0$, kde $z = (x, u)$ a $d = (d_x, d_u)$, se řeší nehladkou Newtonovou metodou. Pak

$$x_+ = x + d_x, \quad u_+ = u + d_u,$$

kde

$$\begin{bmatrix} B & A \\ R^{-1}(R + C)A^T & -R^{-1}(R - U) \end{bmatrix} \begin{bmatrix} d_x \\ d_u \end{bmatrix} = - \begin{bmatrix} g(x, u) \\ h(x, u) \end{bmatrix},$$

$$B \approx G(x, u) = \nabla^2 F(x) + \sum_{k=1}^m u_k \nabla^2 c_k(x),$$

$$R = \text{diag}(r_k, 1 \leq k \leq m), \quad r_k = \sqrt{c_k(x)^2 + u_k^2}.$$

Symetrizace

$$\begin{bmatrix} B & A \\ A^T & -M \end{bmatrix} \begin{bmatrix} d_x \\ d_u \end{bmatrix} = - \begin{bmatrix} g(x, u) \\ (R + C)^{-1} R h(x, u) \end{bmatrix},$$

kde $M = (R + C)^{-1}(R - U)$ je diagonální pozitivně semidefinitní matice.

Použití aktivních omezení

Omezení s indexem $1 \leq k \leq m$ nazveme aktivním, pokud

$$-\frac{\partial \psi_k}{\partial u_k} \leq \hat{\varepsilon} \frac{\partial \psi_k}{\partial c_k} \iff r_k - u_k \leq \hat{\varepsilon}(r_k + c_k),$$

kde $\hat{\varepsilon} > 0$ (obvykle $0.01 \leq \hat{\varepsilon} \leq 1$).

Aktivní veličiny: $\hat{c}_k, \hat{u}_k, \hat{r}_k, \hat{M}$; Neaktivní veličiny: $\check{c}_k, \check{u}_k, \check{r}_k, \check{M}$.

Eliminace neaktivních veličin: Platí

$$\check{d}_u = \check{M}^{-1}(\check{A}^T d_x + \check{c}) - \check{u}, \quad (1)$$

$$\begin{bmatrix} \hat{B} & \hat{A} \\ \hat{A}^T & -\hat{M} \end{bmatrix} \begin{bmatrix} d_x \\ d_u \end{bmatrix} + \begin{bmatrix} \hat{g}(x, u) \\ (\hat{R} + \hat{C})^{-1} \hat{R} \hat{h}(x, u) \end{bmatrix} = \begin{bmatrix} r_x \\ r_u \end{bmatrix}, \quad (2)$$

kde

$$\hat{B} = B + \check{A} \check{M}^{-1} \check{A}^T, \quad \hat{g}(x, u) = g(x, u) + \check{A} \check{M}^{-1} \check{c}.$$

Platí

$$\hat{M} \leq \hat{\varepsilon} I, \quad \check{M}^{-1} < 1/\hat{\varepsilon} I \quad \text{a} \quad \hat{M} \rightarrow 0, \quad \check{M}^{-1} \rightarrow 0, \quad \text{pokud} \quad x \rightarrow x_*.$$

Symetrická matice \hat{B} má omezenou normu a je snaha, aby byla pozitivně definitní.

Použití cenových funkcí

Cenová funkce

$$\varphi(t) = F_{1,2}(x + td_x) + \sigma P_{1,2}(x + td_x, u + td_u), \quad \sigma \geq 0,$$

$$F_1(x + td_x) = F(x + td_x),$$

$$F_2(x + td_x) = F(x + td_x) + (u + d_u)^T c(x + td_x),$$

$$P_1(x + td_x, u + td_u) = \|h(x + td_x, u + td_u)\|_1,$$

$$P_2(x + td_x, u + td_u) = \frac{1}{2} \|h(x + td_x, u + td_u)\|^2.$$

Je třeba, aby platilo $\varphi'(0) < 0$, a aby byla splněna Armijova podmínka

$$\varphi(t) - \varphi(0) \leq \varepsilon_1 t \varphi'(0), \quad \text{kde } 0 < \varepsilon_1 < 1/2.$$

Označení

$$F_1 : \chi(r) = d_x^T r_x - (\hat{u} + \hat{d}_u)^T \hat{r}_u, \quad F_2 : \chi(r) = d_x^T r_x,$$

$$F_1 : \gamma_0 = (u + d_u)^T M(u + d_u) - (u + d_u)^T c, \quad F_2 : \gamma_0 = 0,$$

$$P_1 : \gamma_1 = \|h\|_1 - \|(\hat{R} + \hat{C})\hat{R}^{-1}\hat{r}_u\|_1,$$

$$P_2 : \gamma_1 = \|h\|^2 - \hat{h}^T (\hat{R} + \hat{C}) \hat{R}^{-1} \hat{r}_u.$$

Přesnost řešení

Potřebujeme, aby platilo $\gamma_1 > 0$, což je splněno, pokud

$$P_1 : \|\hat{r}_u\|_1 \leq \frac{\omega_u}{2} \|h\|_1, \quad P_2 : \|\hat{r}_u\| \leq \frac{\omega_u}{2} \|h\|,$$

kde $0 \leq \omega_u < 1$.

Věta 1 Necht vektory d_x , \hat{d}_u jsou získány přibližným řešením soustavy rovnic (2) a vektor \check{d}_u je určen podle (1). Pak

$$\varphi'(0) = -d_x^T B d_x - \gamma_0 - \gamma_1 \sigma + \chi(r).$$

Pokud $\gamma_1 > 0$,

$$\sigma \geq \underline{\sigma} > -\frac{d_x^T B d_x + \gamma_0}{\gamma_1},$$

a soustava (2) je řešena s dostatečnou přesností, konkrétně když

$$\chi(r) < d_x^T B d_x + \gamma_0 + \gamma_1 \sigma,$$

platí $\varphi'(0) < 0$.

Globální konvergence

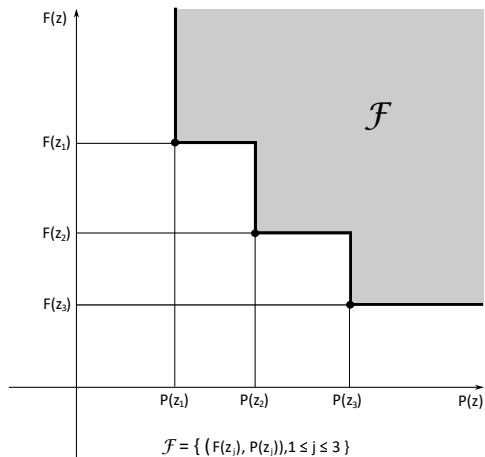
Globální konvergence: Iterační metoda konverguje z **libovolného bodu** ke **stacionárnímu bodu** (kde jsou splněny KKT podmínky).

Standardní předpoklady:

- Funkce $F(x)$ a $c_k(x)$, $1 \leq k \leq m$, jsou dvakrát spojitě diferencovatelné. Funkční hodnoty a derivace jsou omezené.
- Matice vystupující v (2) jsou stejnoměrně regulární.
- Matice B jsou stejnoměrně omezené a stejnoměrně pozitivně definitní.
- Platí $|u_k| + |c_k(x)| \geq \varepsilon$ (striktní komplementarita) a $r_k + c_k(x) \neq 0$.

Globální konvergenci lze dokázat pouze pokud cenová funkce $\varphi(t)$ obsahuje funkci $F_1(x + td_x) = F(x + td_x)$ a pokud parametr σ neklesá (neefektivní volba).

Geometrické znázornění filtru



$$P(z) = \frac{1}{2} \|h(z)\|^2$$

Kombinovaná metoda spádových směrů

Kombinace Armijovy podmínky a principu filtru.

Pokud platí

$$d^T \nabla F(z) < 0 \quad \text{a} \quad -d^T \nabla F(z)t > \delta_3 P^\nu(z), \quad (3)$$

kde $\delta_3 > 0$ a $\nu > 1$, používá se Armijova podmínka

$$F(z + td) - F(z) \leq \varepsilon_1 td^T \nabla F(z), \quad (4)$$

kde $0 < \varepsilon_1 < 1/2$. Pokud neplatí (3), používá se podmínka filtru

$$F(z + td) < F(z) - \delta_1 P(z) \quad \text{nebo} \quad P(z + td) < P(z) - \delta_2 P(z), \quad (5)$$

kde $\delta_1 > 0$ a $\delta_2 > 0$.

Algoritmus 1

- Vstup** Dvojice $z = (x, u)$, hodnoty $F(z)$, $P(z)$ a směrový vektor $d = (d_x, d_u)$ získaný řešením rovnic (1)–(2).
- Krok 1** Vypočteme minimální délku kroku $\underline{t} > 0$. Zvolíme počáteční délku kroku $t > 0$ (obvykle $t := 1$).
- Krok 2** Pokud $t < \underline{t}$ přejdeme na krok 6. Pokud $t \geq \underline{t}$, vypočteme nové hodnoty $F := F(z + td)$ a $P := P(z + td)$. Pokud $(F, P) \in \mathcal{F}$, přejdeme na krok 5.
- Krok 3** Pokud platí (3) a je splněna **Armijova podmínka** (4), přejdeme na krok 8. Pokud platí (3) a není splněna Armijova podmínka (4), přejdeme na krok 5.
- Krok 4** Pokud neplatí (3) a je splněna **podmínka filtru** (5), přejdeme na krok 7. Pokud neplatí (3) a není splněna podmínka filtru (5), přejdeme na krok 5.

Krok 5 Zkrátíme délku kroku $t > 0$ a přejdeme na krok 2.

Krok 6 *Obnovení přípustnosti.* Najdeme nový směrový vektor d a vhodnou délku kroku $t > 0$, pro kterou je splněna podmínka (5). Vypočteme hodnoty $F := F(z + td)$, $P := P(z + td)$ a přejdeme na krok 7.

Krok 7 Aktualizace filtru.

Krok 8 *Ukončení výpočtu.* Hodnota $t > 0$ je získaná délka kroku.

Minimální délka kroku

$$\begin{aligned} \underline{t} &= \delta_4 \min \left(\varepsilon_0, \frac{\delta_1 P(z)}{|d^T \nabla F(z)|}, \frac{\delta_3 P(z)^v}{|d^T \nabla F(z)|} \right), & d^T \nabla F(z) < 0, \\ \underline{t} &= \delta_4 \varepsilon_0, & d^T \nabla F(z) \geq 0, \end{aligned}$$

kde $0 < \delta_4 < 1$ a $\varepsilon_0 > 0$.

Věta 2 Uvažujme kombinovanou metodu spádových směrů, kde se délka kroku určuje algoritmem 1. Nechť jsou splněny standardní předpoklady pro globální konvergenci. Pak $h(z) \rightarrow 0$.

Věta 3 Uvažujme metodu nehladkých rovnic, kde se směrový vektor získává řešením rovnic (1)–(2) s přesností

$$d_x^T r_x \leq \bar{\omega}_x d_x^T \hat{B} d_x, \quad \|\hat{r}_u\| \leq \bar{\omega}_u \|\hat{c}(x)\|,$$

kde $0 \leq \bar{\omega}_x < 1$ a $0 \leq \bar{\omega}_u < 1$, a délka kroku se určuje algoritmem 1. Jsou-li splněny standardní předpoklady pro globální konvergenci (takže $h(z) \rightarrow 0$), je tato metoda **globálně konvergentní**.

Všechny výsledky jsou počítány pomocí systému UFO
– je to **modulární systém**:

- všechny metody používají stejné podprogramy
⇒ porovnání metod je věrohodné



Lukšan L., Tůma M., Matonoha C., Vlček J., Ramešová N., Šiška M., Hartman J.:
UFO 2022 - Interactive System for Universal Functional Optimization.

Technical report V-1289, ICS AS CR, Prague, 2022, 460 p.

<https://www.cs.cas.cz/luksan/ufo.html>

TEST21:

- Sada testovacích příkladů pro minimalizaci s omezeními ve tvaru nerovností.
- 18 řídkých úloh popsanych v



Lukšan L., Vlček J.:

Indefinitely preconditioned inexact Newton method for large sparse equality constrained nonlinear programming problems.

Numer. Linear Algebra Appl., 5 (1998), 219-247.

kde rovnosti jsou nahrazeny nerovnostmi.

- Například 1. úloha je modifikovaná Rosenbrockova funkce

$$F(x) = \sum_{i=1}^{n-1} \left[100(x_i^2 - x_{i+1})^2 + (x_i - 1)^2 \right]$$

$$c_k(x) = \sin(x_{k+1} - x_{k+2})\sin(x_{k+1} + x_{k+2}) - x_k \exp(x_k - x_{k+1}) \\ + 3x_{k+1}^3 + 4x_{k+1} + 2x_{k+2} - 8, \quad 1 \leq k \leq n - 2$$

Metody použité pro porovnání výsledků:

- NE-P – metoda nehladkých rovnic, použití cenové funkce
- NE-F – metoda nehladkých rovnic, kombinace cenové funkce a filtru
- IP-P – metoda vnitřních bodů, použití cenové funkce
- IP-F – metoda vnitřních bodů, použití filtru
- SQP – sekvenční kvadratické programování
- NE-MN – Newtonova metoda nehladkých rovnic (kombinovaný filtr)
- IP-MN – Newtonova metoda vnitřních bodů (cenová funkce)
- NE-VM – Metoda nehladkých rovnic s proměnnou metrikou (kombinovaný filtr)
- IP-VM – Metoda vnitřních bodů s proměnnou metrikou (cenová funkce)

Výkonnostní profily 1

Výkonnostní profily:

- užitečný nástroj k porovnávání optimalizačních metod,
- zohledňují výsledky odpovídající jednotlivým problémům.

Výkonnostní profil $\rho(\tau)$ je definován jako

$$\rho(\tau) = \frac{\text{počet problémů, pro něž } \log_2(\tau_{p,m}) < \tau}{\text{celkový počet problémů}}, \quad 0 \leq \tau \leq \bar{\tau}.$$

Pro $X = \text{čas, NIT, NFV apod.}$:

$$\tau_{p,m} = \frac{\text{hodnota } X \text{ potřebná k vyřešení problému } p \text{ metodou } m}{\text{hodnota } X_{\min} \text{ potřebná k vyřešení problému } p \text{ metodami } m},$$
$$\tau_{p,m} = \infty \quad \text{pokud se metodou } m \text{ nepodaří vyřešit problém } p.$$

Výkonnostní profily 2

Hodnota $\rho(\tau)$ udává:

- pro $\tau = 0$ procento testovacích úloh, pro které je metoda m nejlepší,
- pro dostatečně velké τ procento testovacích úloh, které metoda m dokáže vyřešit.

Relativní účinnost a spolehlivost jednotlivých metod lze přímo vyčíst z výkonnostních profilů:

Čím vyšší je konkrétní křivka, tím lepší je příslušná metoda.

Systém UFO vyrábí výkonnostní profily automaticky po zadání příslušného vstupního parametru.

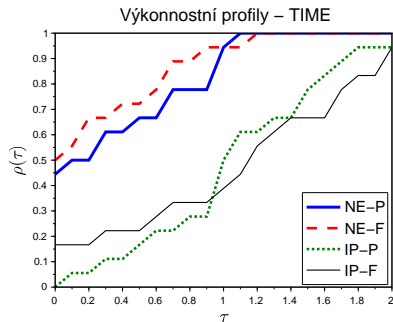
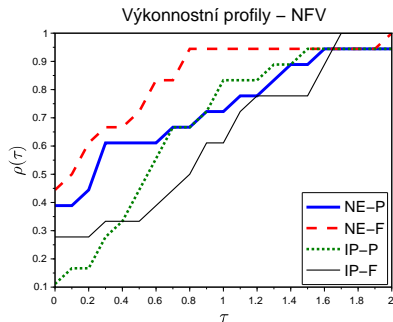


Dolan E.D., Moré J.J.:

Benchmarking optimization software with performance profiles.

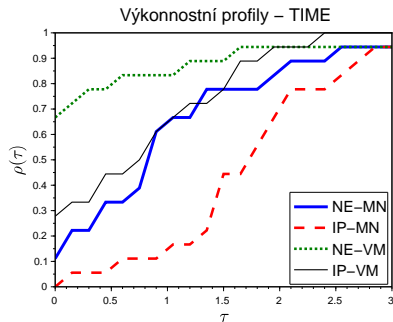
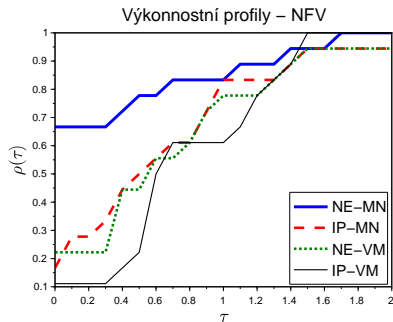
Mathematical Programming 91 (2002) 201-213.

Porovnání modifikací Newtonovy metody.



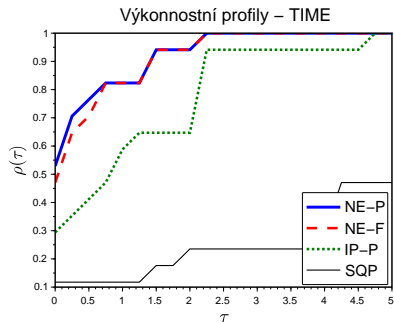
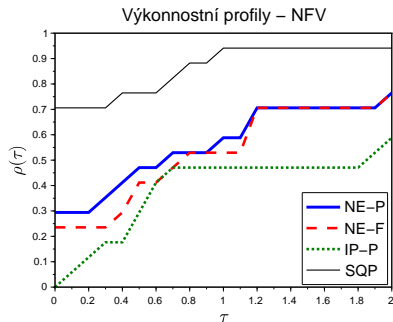
TEST21 - 18 úloh s 1000 proměnnými

Použití metod s proměnnou metrikou.



TEST21 - 18 úloh s 1000 proměnnými

Porovnání různých metod.



TEST21 - 17 úloh s 1000 proměnnými

Děkujeme za pozornost.