Václav Fabian

Přehled deterministických a stochastických aproximačních metod pro minimalizaci funkcí

Kybernetika, Vol. 1 (1965), No. 6, (499)--523

Persistent URL: http://dml.cz/dmlcz/125290

KYBERNETIKA ČÍSLO 6, ROČNÍK 1/1965

Přehled deterministických a stochastických aproximačních metod pro minimalizaci funkcí*

VÁCLAV FABIAN

Přehled základních výsledků a metod v oblasti minimalizace funkcí. Uvedeny a porovnány jsou metody gradientní a jim příbuzné metody, metody nelineárního programování a stochastické aproximace. Jsou popsány základní vlastnosti konvergence a její rychlosti a uvedeny otevřené otázky.

OBSAH

* Referát přednesený na *letním semináři o teorii informace a statistických metodách v teorii řízení*, který se konal v Praze ve dnech 25. května až 4. června 1965.

1. ÚVOD

Referát zahrnuje dosti široký obor. Deterministický a stochastický případ se většinou studují odděleně. Důvodem pro současné shrnutí bylo aplikační hledisko, podle něhož se zřetelně jedná o jednotnou nebo alespoň příbuznou problematiku, a příbuznost metod z obou oblastí. Vzhledem k šíři problematiky si nebudu moci všímat podrobností a omezím se na to, co považuji za hlavní myšlenky a výsledky.

Chtěl bych též podotknout, že shrnuji spíše mně známé výsledky než výsledky vůbec známé. Také seznam literatury, i když je poměrně bohatý, není ani zdaleka kompletní bibliografií. Z prací týkajících se stochastických aproximací jsou však zastoupeny všechny mě známé.

Existuje řada přehledných prací týkajících se různých částí naší problematiky. Schmetterer (1961) podal zasvěcený a do značné hloubky jdoucí přehled výsledků ve stochastických aproximacích, s důkazy a s některými novými výsledky. Starší přehled z tohoto oboru podal Derman (1956). Přehledy metod v deterministickém případě podávají Spang (1962), Wolfe (1962) a Saaty a Bram (1964, kap. 2 a 3). Poslední práce je nejpodrobnější, podává přesně formulované výsledky a většinou též jejich důkazy. Může být velmi užitečná množstvím velmi dobře vybraného materiálu. Uvádí však mnohé podobné výsledky paralelně vedle sebe bez poukázání na jejich často velmi těsnou souvislost; je to opět spíše přehled než monografie. Té by bylo velmi třeba zvláště proto, že zejména v oboru deterministické minimalizace je velmi mnoho prací rozdílné úrovně a mnoho paralelních výsledků bez jasných vzájemných vztahů.

2. TYP UVAŽOVANÝCH APROXIMAČNÍCH METOD

Budeme se zabývat aproximačními metodami pro získání bodu minima funkce f na nějaké množině A, části k-rozměrného euklidovského prostoru E_k. Situace, v níž $A = E_k$ bývá jednodušší než případ opačný, kdy předpokládáme, že $A = \{x; G(x) \leqq 0\}$ pro nějakou m-rozměrnou vektorovou funkci G na E_k. ($G(x) \leqq 0$ značí, že každá složka $G(x)$ je nekladná; i-tou složku vektoru y značíme $y^{(i)}$, i-tou složku G resp. $G(x)$ značíme $G^{(i)}$ resp. $G^{(i)}(x)$. Euklidovskou normu značíme $\| \ \|$).

V obou situacích, v situaci bez i s omezením ($A = E_k$, $A \neq E_k$) záleží na informaci, kterou můžeme o f získat. V deterministickém případě předpokládáme, že pro každé x z A můžeme určit hodnotu $f(x)$, případně hodnotu prvních nebo vyšších derivací této funkce. Ve stochastickém případě můžeme pro každé x získat pozorování hodnoty $f(x)$ zatížené náhodnou chybou. Uvažujeme aproximační metody, které z bodu x hledají lepší aproximaci v nějakém směru δ a vzdálenosti α. Ve spojitém případě je pak aproximační proces určen diferenciálními rovnicemi typu

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = \alpha_t \delta_t \; ;$$

v diskrétním případě je analogicky $x_{t+1} - x_t = \alpha_t \delta_t$. V tomto druhém případě je α_t buď voleno nějakým vhodným způsobem, většinou předem, nezávisle na x_t, a musí být většinou dostatečně malé. Budeme o těchto metodách mluvit jako o metodách s malým krokem.* α_t může být také voleno tak, aby buď minimalizovalo $f(x_t + \alpha_t \delta_t)$ při daných x_t a δ_t nebo – častěji – aby bylo α_t největším číslem, pro něž je $f(x_t + \tau\delta_t)$ klesající pro $0 \leqq \tau \leqq \alpha_t$. V tomto případě budeme říkat, že α_t je voleno metodou největšího spádu. (Též se mluví o optimální volbě, což je však název silně matoucí.) Budeme uvažovat gradientní a jim příbuzné metody. V gradientní metodě se volí δ_t rovno $-D(x_t)$, kde $D_h(x)$ značí vektor prvých derivací funkce h v bodě x. V případě $h = f$ vynecháváme označení funkce. Podobně značíme $H_h(x)$ matici druhých derivací funkce h v bodě x, $H = H_f$. V příbuzných metodách se volí $\delta_t = -BD(x_t)$, kde B je nějaká matice, obvykle pozitivně definitní, a např. blízká nebo rovna $H^{-1}(x_t)$. Jindy se takový směr jen odhaduje.

Zajímáme se o aproximační metody, které konvergují při každém výchozím bodu a pro všechna f z dosti bohaté třídy funkcí. Tyto vlastnosti nemají mnohé z aproximačních metod vykládaných ve standardních učebnicích.

Tak např. zobecnění Newtonovy metody pro náš problém vede ke vztahu

$$x_{n+1} = x_n - \alpha_n H^{-1}(x_n)\, D(x_n)$$

s $\alpha_n = 1$. Tato metoda konverguje jen za značně speciálních kvantitativních podmínek. Zvolíme-li však např. α_n metodou největšího spádu, dostaneme konvergenci za podmínek kvalitativních a podstatně obecnějších.

Hned zde lze říci, že úloha minimalizace dané funkce f (řekněme bez omezení; případ s omezením může být ještě obtížnější) může být velmi nesnadná. Potíže může působit velká dimenze (ta činí obtížným i řešení lineárních rovnic) i vzdorovitý tvar funkce, jako má např. funkce

$$f(u, v) = 100(v - u^2)^2 + (1 - u)^2 ,$$

kterou ke zkoušení své minimalizační metody použil Rosenbrock (1960) a po něm mnozí další.

Další potíží je, že aproximační metody nám aproximují bod lokálního minima, který nemusí být bodem absolutního minima. Samozřejmě, že pro některé funkce, např. konvexní, toto nebezpečí není, ale v ostatních případech je v podstatě aproximačních metod, že nemohou dosáhnout lepšího výsledku. Zde zbývá jen systematický průzkum definiční oblasti A funkce f, případně s aproximacemi použitými v jednotlivých jeho krocích. Při velké dimenzi je však tento postup většinou nemožný. Jedinou cestou zde asi je získat více informací o uvažované funkci a vymezit oblast, v níž již jiné lokální minimum kromě absolutního není.

Aproximační metody a to aspoň ty, o nichž mluvíme, jsou určeny pro řešení

* Názvy nejsou ustálené: Zoutendijk (1960) např. myslí spojitou metodu, mluví-li o metodě s malým krokem.

 úloh za obecných předpokladů. V případech speciálních lze použít speciálnějších metod s výhodnějšími vlastnostmi. Nemůžeme však zatím bohužel čekat, že speciálními postupy vyřešíme mnoho problémů.

3. VZTAH K ŘEŠENÍ ROVNIC

Minimalizačními metodami se řeší i soustavy rovnic. Je-li g vektorová funkce, lze řešit rovnici $g(x) = 0$ minimalizací funkce $f(x) = \|g(x)\|^2$. Jsou možné i jiné postupy. Je-li g derivací nějaké funkce f, $g = Df$, (obecné podmínky pro to podal Kerner (1933) a uvádí též Nashed (1964)) lze minimalizací f přijít ke stacionárnímu bodu x, v němž je $g(x) = 0$. Tento postup v Hilbertových prostorech studuje Nashed (1964, 1965).

4. APLIKACE

Okruh praktických problémů, které vedou k uvažovaným úlohám, je nesmírně rozsáhlý. Zahrnuje především úlohy optimalizace. V ekonomických oblastech tyto úlohy vedou často k jednoduchým f, složitým A (např. lineární a kvadratické programování) a velké dimenzi k; pro přehled mnoha aplikací viz např. Maňas (1965) a Arnoff a Sengupta (1961). V technických problémech bývá dimenze menší, ale f často složitější. Problémy experimentálního výzkumu jsou obvykle stochastického rázu. Aproximační metody pro minimalizaci funkcí se mohou stát strategiemi automatických optimalizátorů, viz např. Feldbaum (1958) a Stachovskij (1958), kteří navrhli takový optimalizátor na základě diskrétních gradientních metod, a Fabian (1961 a 1962), který použil jako strategie stochastické aproximační metody. Zpráva o poloprovozní zkoušce automatického optimalizátoru Opcon byla uveřejněna v Chem. Eng., 1959, str. 64 a 66 a v Chemische Industrie, 1959, Heft 5, str. 234; pro popis viz Kerstukos a Van Nice (1958) a Archer (1960). Na minimalizaci též vedou úlohy regrese, zejména nelineární, a tedy např. úlohy určování neznámých konstant v kinetických rovnicích a podobně. Zde je možné využít speciálních vlastností úlohy (Levenberg (1944), Marquardt (1963), Goldstein (1962) a Powell (1965)). Další aplikace jsou též v problémech adaptivní predikce, viz Hanš a Špaček (1960) a Gardner (1964).

5. DETERMINISTICKÝ PŘÍPAD BEZ OMEZENÍ

5.1 Přehled výsledků

Jako nejstarší bývá citována práce Cauchyho. Cauchy (1847) navrhl v krátkém sdělení zpřesňovat řešení systému rovnic $g(x) = 0$ ve směru $-Df$, kde $f(x) = \|g(x)\|^2$. Ukázal, že ke zlepšení dojde, zvolíme-li délku kroku buď dost malou nebo tak, aby byla minimalizována hodnota $f(x - \alpha\, Df(x))$. Tak mu je připisováno autorství gradientních metod s malým krokem i s krokem, voleným metodou největšího spádu. Dále se touto problematikou zabýval Schröder (1870). Mises a Pollaczek-Geiringer (1929) uvádějí aproximační schéma $x_{n+1} = x_n - cf(x_n)$ pro hledání kořene funkce f, která je záporná, resp. kladná vlevo resp. vpravo od kořene. Přitom c je číslo menší než

$$\sup_{x_1 \neq x_2} \frac{x_2 - x_1}{f(x_2) - f(x_1)}.$$

Postup zobecňují na systémy rovnic. Germansky (1934) použil tohoto postupu k aproximaci $x_{n+1} = x_n - c\,D(x_n)$ bodu minima funkce f. Curry (1944) podává, zdá se, prvý důkaz konvergenčních vlastností metody největšího spádu. Zřejmě nezávisle pak studuje Kantorovič (1945a, 1945b, 1948) vlastnosti metody největšího spádu pro kvadratické funkce definované na Hilbertově prostoru. Dostává také prvý výsledek o rychlosti konvergence: vzdálenost n-té aproximace od řešení klesá geometricky, tj. jako Ca^n, kde C a a jsou konstanty, $a = (\lambda_k - \lambda_1)/(\lambda_k + \lambda_1)$, λ_k a λ_1 největší a nejmenší charakteristické číslo matice $H(0)$ (Kantorovič (1947), viz též Faddeev a Faddeeva (1963, str. 70). Metodu největšího spádu pro řešení systémů nelineárních rovnic studuje též Both (1949), porovnává ji s metodou, v níž se na každém kroku mění jen jedna souřadnice a dochází k závěru, že je metoda největšího spádu k-násobně rychlejší (ovšem každý krok vyžaduje více výpočtů). (Pro řešení lineárních rovnic Householder a Bauer (1960) dostávají podobný výsledek s konstantou $4k$; přesněji řečeno, jedná se pouze o porovnání horních odhadů rychlosti.) Kantorovičův (1948) výsledek poněkud zobecňuje Fridman (1962) (ukazuje, že 0 nemusí být izolovaným bodem spektra operátoru). Crockett a Chernoff (1955) dokazují řadu výsledků týkajících se deterministických gradientních metod. Uvažují aproximační schémata tvaru $x_{n+1} = x_n - h_n B_n\,D(x_n)$ s pozitivně definitními maticemi B_n a zkoumají rychlost konvergence v okolí bodu minima. Při vhodné volbě délky kroků h_n je opět geometrická (jako v kvadratickém případě) s konstantou opět tím příznivější, čím je menší $(\lambda_m - \lambda_1)/(\lambda_1 + \lambda_m)$, kde λ_m a λ_1 jsou největší a nejmenší charakteristické číslo matice $B_n\,H(x_n)$. Nejvýhodnější je tedy „Newtonova" volba $B_n = H^{-1}(x_n)$, pokud $H(x_n)$ má inverzi. Autoři též uvažují případ, v němž se $H^{-1}(x_n)$ počítá „jen občas" v krocích n_1, n_2 atd. a $B_m = H(n_i)$ pro $n_i \leqq m < n_{i+1}$. Na tuto práci navázal až později Goldstein (1962), který studoval nelokální podmínky konvergence gradientních metod a uvedl znovu poněkud přehledněji výsledek o asymptotické rychlosti konvergence při používání malých kroků (zdá se, že pro kroky volené metodou největšího spádu takový výsledek není v obecném případě znám). Goldstein uvažuje též některé modifikace gradientní metody pro minimalizaci výrazu $f(x) = \|g(x)\|$, kde g je m-dimenzionální vektorová funkce. Zahrnuje též případ $m = k$ a volbu $\delta_n = -D_g^{-1}(x_n)\,g(x_n)$, kde $D_g^{(ij)}(x) = \partial g^{(i)}/\partial x^{(j)}$. K takto určenému směru dospíváme hledáním řešení rovnice $g(x_n) + D_g(x_n)(x - x_n) = 0$.

Poznámka o výhodnosti volby $B_n = H^{-1}(x_n)$ se vztahuje jen na x_n blízké bodu minima s maticí $H(x_n)$ pozitivně definitní, kdy tato volba je přibližně optimální.

Optimální volba B_n pro jiná x_n je, pokud vím, otevřeným problémem, i když se i zde zdá výhodnou volba $B_n = H_n^{-1}(x_n)$, je-li $H(x_n)$ pozitivně definitní.

Konvergenci gradientních metod v abstraktních prostorech uvažovali Vainberg (1960), Nashed (1964), Leżański (1963) a jiní.

5.2 Základní vlastnosti gradientních metod

Hlavní výsledky, o nichž jsme dosud referovali, shrnují následující 3 věty. Začínáme se spojitou aproximační metodou. Za silnějších předpokladů odvodil Rosenbloom (1956) podobný, o něco silnější výsledek, zahrnující odhad rychlosti konvergence (ta však asi ve spojitém případě není tak zajímavá jako v diskrétním, vzhledem k možnosti transformace času).

Věta 1. *Nechť f má spojitou derivaci D, která je pro nějaké kladné ε stejnoměrně spojitá na množině* $B = \{x; \|D(x)\| < \varepsilon\}$ *(k tomu stačí např. omezenost množiny B). Předpokládejme, že* x_t *vyhovuje rovnici*

$$\frac{dx_t}{dt} = -D(x_t) \tag{1}$$

(např. existuje-li spojitá H má (1) pro každou počáteční podmínku $x_0 = a$ *právě jedno řešení). Potom* $f(x_t) \searrow$ *a buď* $f(x_t) \searrow -\infty$ *nebo* $\|D(x_t)\| \to 0$ *pro* $t \to \infty$.

Důkaz: Je $\frac{d}{dt} f(x_t) = -D'(x_t) \frac{d}{dt} x_t = -\|D(x_t)\|^2$,

$$f(x_t) - f(x_0) = -\int_0^t [D(x_s)\|^2 \, ds \leqq 0 \, .$$

Odtud $f(x_t) \searrow$ a buď $f(x_t) \searrow -\infty$ nebo $\int_0^\infty \|D(x_s)\|^2 \, ds < +\infty$. V tomto druhém případu je $\liminf_{t\to\infty} \|D(x_t)\| = 0$; předpokládejme, že však není $\|D(x_t)\| \to 0$. Pak existuje takové kladné η menší než ε a taková posloupnost disjunktních intervalů (u_i, v_i), že $\|D(x_{u_i})\| \leqq \eta/2$, $\|D(x_{v_i})\| \geqq \eta$. Protože je $D(x_t)$ spojitá vzhledem k t, lze intervaly volit tak, aby byly částí množiny $T = \{t; \|D(x_t)\| \leqq \eta\}$. x_t je však stejnoměrně spojitá na T, $\{x_t, t \in T\} \subset B$ a tedy $D(x_t)$ je stejnoměrně spojitá na T, takže $\liminf_{i\to\infty} (v_i - u_i) > 0$, odkud ovšem ihned $\int_0^\infty \|D(x_t)\|^2 \, dt = +\infty$ a věta je dokázána.

Věta 2. *Nechť existují druhé derivace funkce f, buď* $x_0 \in E_k$, $S = \{x; f(x) \leqq f(x_0)\}$, *nechť H je omezená na S,* $\varrho = 2(\sup_{x\in S} \|H(x)\|)^{-1}$, *buď* $0 < \delta < \varrho$. *Nechť*

$$x_{n+1} = x_n - \alpha_n D(x_n) \, ; \tag{2.1}$$

pro každé n nechť je buď $\delta < \alpha_n < \varrho - \delta$ *nebo nechť je* α_n *voleno metodou největšího spádu.*

Pak při prvé volbě α_n je $f(x_{n+1}) - f(x_n) \leqq \delta\varrho^{-1}(\varrho - \delta) \|D(x_n)\|^2$; při druhé je $f(x_{n+1}) - f(x_n) \leqq (\varrho/4) \|D(x_n)\|^2$. Je $f(x_n) \searrow$ a buď $f(x_n) \searrow -\infty$ nebo $\|D(x_n)\| \to 0$.

Důkaz: Předpokládejme, že $x_0, x_1, \ldots, x_n$ již jsou dána a vyhovují tvrzení věty. Je-li $D(x_n) = 0$ je $x_m = x_n$ pro $m > n$ a vše platí. Buď $D(x_n) \neq 0$ a položme

$$\varphi(\alpha) = f(x_n - \alpha D(x_n)) - f(x_n)\,.$$

Je $\varphi(\alpha) = -\alpha\|D(x_n)\|^2 + \alpha^2/2 \,.\, D'(x_n)\, H(\eta(\alpha))\, D(x_n) \leqq -\alpha\|D(x_n)\|^2 (1 - \frac{1}{2}\alpha\|H(\eta(\alpha))\|)$, přičemž $\eta(\alpha)$ leží na úsečce $R(\alpha)$ spojující x_n a $x_n - \alpha\, D(x_n)$. Buď $\langle 0, \beta\rangle$ maximální interval, na němž je φ nekladná. Protože $\mathrm{d}/\mathrm{d}\alpha\ \varphi(0) = -\alpha\|D(x_n)\|^2 < 0$, je $\beta > 0$. Předpokládejme, že $\beta < \varrho$. Ze spojitosti f plyne, že $\varphi(\beta) = 0$, současně však $\eta(\beta)$, ležící na úsečce $R(\beta)$, je v S a tedy $\|H(\eta(\beta))\| \leqq 2\varrho^{-1}$. Odtud a z poslední nerovnosti pro $\varphi(\alpha)$ dostáváme $\varphi(\beta) \leqq -\beta\|D(x_n)\|^2 (1 - \beta/\varrho) < 0$, což je spor. Je tedy φ nekladné na $\langle 0, \varrho\rangle$ a v celém tomto intervalu je

$$\varphi(\alpha) \leqq -\alpha\|D(x_n)\|^2 \left(1 - \frac{\alpha}{\varrho}\right).$$

Minimum výrazu vpravo v tomto intervalu je v bodě $\alpha = \varrho/2$ a $\varphi(\varrho/2) \leqq -(\varrho/4)\,.$ $.\, \|D(x_n)\|^2$; maximum v intervalu $(\delta, \varrho - \delta)$ je v jeho krajních bodech, v nichž je

$$\varphi(\delta) \leqq \delta \frac{\varrho - \delta}{\varrho} \|D(x_n)\|^2 \geqq \varphi(\varrho - \delta)\,.$$

Tím dostáváme odhady pro $f(x_{n+1}) - f(x_n)$. Ostatní tvrzení z nich již bezprostředně plynou.

Další věta udává asymptotickou rychlost konvergence pro metodu malých kroků.

Věta 3. *Nechť platí předpoklady předešlé věty a navíc nechť z je hromadný bod posloupnosti x_n, H je spojitá v z a matice $H(z)$ nechť je pozitivně definitní s λ_1 a λ_k nejmenším a největším charakteristickým číslem. α_n nechť jsou volena v intervalu $\langle \delta, \varrho - \delta\rangle$. Potom $x_n \to z$ a*

$$\|x_{n+1} - z\| \leqq \|x_n - z\|\,(a_n + \varepsilon_n) \text{ s } a_n \leqq a < 1 \text{ a } \varepsilon_n \to 0\,; \tag{3.1}$$

podrobněji

$$a_n = \max\{1 - \alpha_n\lambda_1, \alpha_n\lambda_k - 1\}\,, \tag{3.2}$$

$$a\ = \max\{1 - \delta\lambda_1, 1 - 2\delta/\varrho\}\,. \tag{3.3}$$

Důkaz: Protože $D(x_n) \to 0$, je $D(z) = 0$ a $D(x) = H(\xi)\,(x - z) = [H(z) + \mathscr{E}(x)]\,(x - z)$, kde ξ je bod na úsečce mezi body x a z a $\mathscr{E}(x)$ je matice, která konverguje k nulové pro $x \to z$. Označme $\varepsilon_n = \alpha_n\|\mathscr{E}(x_n)\|$ a buď E jednotková matice. Je

$$\begin{aligned}\|x_{n+1} - z\| &= \|x_n - z - \alpha_n[H(z) + \mathscr{E}(x_n)]\,(x_n - z)\| \leqq \\ &\leqq \|(E - \alpha_n H(z))\,(x_n - z) - \alpha_n \mathscr{E}(x_n)\,(x_n - z)\,\| \leqq \\ &\leqq \|x_n - z\|\,(\|E - \alpha_n H(z)\| + \varepsilon_n)\,.\end{aligned}$$

Norma $a_n = \|E - \alpha_n H(z)\|$ je rovna největšímu vlastnímu číslu matice $E - \alpha_n H(z)$ a tedy $a_n = \max\{|1 - \alpha_n\lambda_1|, |1 - \alpha_n\lambda_k|\} = \max\{1 - \alpha_n\lambda_1, \alpha_n\lambda_k - 1\} \leqq \max\{1 - \delta\lambda_1, (\varrho - \delta)\lambda_k - 1\}$ a protože $\lambda_k \leqq \sup_{x\in S} \|H(x)\| = 2/\varrho$,

$$(\varrho - \delta)\lambda_k - 1 \leqq 2(\varrho - \delta)/\varrho - 1 = 1 - 2\delta/\varrho .$$

Platí tedy (3.1), (3.2) i (3.3) až na $\varepsilon_n \to 0$, což jsme ještě nedokázali. Existuje však $\eta > 0$, takové, že pro $\|x - z\| < \eta$ je $2\varrho\|\mathscr{E}(x)\| < 1 - a$. Existuje též n tak, že je $\|x_n - z\| < \eta$ a tedy $\varepsilon_n = \alpha_n\|\mathscr{E}(x_n)\| < (1 - a)/2$. Od tohoto n počínaje je pak podle (3.1) již $\|x_m - z\| \leqq \eta((1 + a)/2)^{m-n}$ pro $m > n$ a tedy $\varepsilon_n \to 0$.

Poznámka 1. Malý krok a krok největšího spádu

Délka kroku metodou největšího spádu maximalizuje úbytek funkční hodnoty v daném kroku, avšak toto lokálně optimální určení α_n není nejlepším z celkového hlediska. Zkušenosti ukazují lepší výsledky s malými kroky (viz Forsythe a Wasov (1959) str. 224–225 a Stiefel (1955); citováno jen podle Goldsteina (1962)). Nevýhodou je potíž ve volbě délky malého kroku, neznáme-li, což je častý případ, H. (Stochastické aproximační metody se této potíži vyhýbají tak, že volí $\alpha_n \to 0, \sum\alpha_n = +\infty$. Pro deterministický případ by to však asi nebylo vhodné.) Nevýhodou metody největšího spádu je, že na každém kroku řešíme jednorozměrný minimalizační problém. Neznám obecných výsledků o postačující přibližnosti při řešení této úlohy. Často se doporučuje přibližné řešení založené na prokládání polynomu (Brooks (1959), Powell (1964) a další). Minimaxové a asymptoticky minimaxové procedury pro odhad bodu minima v jednorozměrném případě navrhl Kiefer (1953, 1957). (Vícerozměrné analogie Kieferovy metody byly též uvažovány, viz Krolak a Cooper (1963) a Newman (1960, 1965).)

Poznámka 2. Volba souřadného systému a Newtonova metoda

Aplikujeme-li metodu z vět 2 a 3 na funkci $g(\xi) = f(A\xi)$, kde A je nesingulární matice a $B = AA'$, přechází vztah (2.1) v

$$\xi_{n+1} = \xi_n - \alpha_n D_g(\xi_n) = \xi_n - \alpha_n A'D(A\xi_n)$$

a vyjádříme-li ho v původních souřadnicích $x = A\xi$, ve vztah

$$x_{n+1} = x_n - \alpha_n A\,A'D(x_n) = x_n - \alpha_n BD(x_n) .$$

Protože $H_g(\xi) = A'H(A\xi)\,A$ je třeba volit α_n buď metodou největšího spádu nebo v intervalu $\langle\delta, \varrho_B - \delta\rangle$, kde $\varrho_B = 2(\sup_{x\in S}\|A'H(x)\,A\|)^{-1} = 2(\sup_{x\in S}\|BH(x)\|)^{-1}$. V odhadech věty 2 pro $f(x_{n+1}) - f(x_n)$ pak je třeba nahradit ϱ tímto ϱ_B a $\|D(x_n)\|^2$

číslem $\|A' D(x_n)\|^2$. Konvergenční vlastnosti zůstávají zachovány, neboť je $D(x_n) \to 0$ jakmile je $A'D(x_n) \to 0$. Ve větě 3 pak λ_1, λ_k jsou nejmenší a největší vlastní číslo matice $BH(z)$ a z (3.2) vidíme, že je výhodné volit B tak, aby $\lambda_1 \doteq \lambda_k$. To je splněno při tzv. Newtonově volbě, kdy volíme $B = B_n$ závislou na n a rovnu $H^{-1}(x_n)$, aspoň v blízkosti bodu z. Propracováním odhadů věty 3 v tomto případě bychom dostali výsledky, podobné výsledkům Crocketta a Chernoffa. Jak volit B, jsme-li ještě vzdáleni od řešení, je zatím otevřený problém. Také je patrné, že B_n musí zachovávat jisté rozumné předpoklady, např. omezenost $\|B_n^{-1}\|^{-1}$ a $\|B_n\|$, nemá-li dojít ke komplikacím a případně k porušení konvergenčních vlastností.

Poznámka 3. Praktické zkušenosti

Častá je taková zkušenost, že z počátku $f(x_n)$ silně klesá, pak se však konvergence velmi zpomalí a bývá i zastavena zaokrouhlovacími chybami, zejména používá-li se původní metody, tj. je-li $B = E$. Proto se intenzivně studují další příbuzné metody.

5.3 Metody spřažených gradientů

I v kvadratickém případě byly špatné zkušenosti s gradientními metodami. Zaručená geometrická konvergence Ca^n měla často koeficient velmi blízký 1. Použití Newtonova směru v tomto případě vede sice okamžitě k výsledku, avšak je nepraktické, protože potřebná inverze matice je problém složitější než původní problém. Snahy o dosažení lepších výsledků povzbuzovaly i některé anomalie bodů x_n, které byly pro velká n přibližně koplanární nebo kolineární, takže velmi jednoduché extrapolace podstatně zlepšovaly přibližnost řešení (Forsythe a Motzkin (1951), Forsythe a Forsythe (1952), Forsythe (1963), Akaike (1959)); asymptotickou kolinearitu při malé délce kroků dokázali za obecných podmínek Crockett a Chernoff (1955). Bylo jasné, že by velké zlepšení přinesla volba délek α_n tak, aby nikoliv jen $f(x_{n+1}) - f(x_n)$ bylo minimální, ale $f(x_{n+m}) - f(x_n)$ pro m větší než 1. Zdálo se, že tento požadavek, vedoucí na řešení soustavy lineárních rovnic, je asi stejně obtížný jako původní úloha, aspoň pro $m = k$. Avšak kolem roku 1951, pracujíce zprvu vzájemně nezávisle (viz Forsythe (1953)), objevili Hestens (1951), Lanszos (1952) a Stiefel (1952) jednoduchý přímý algoritmus, který vede v kvadratickém případu k postupnému určení $x_1, x_2, \ldots$ tak, že

$$x_i = x_0 + \alpha_{i,0}\, D(x_0) + \alpha_{i,1}\, D(x_1) + \ldots + \alpha_{i,i-1}\, D(x_{i-1})$$

a $\alpha_{i,j}$ minimalizují $f(x_i)$. V důsledku toho také lze ukázat, že algoritmus je konečný a nejpozději n-tý člen x_n je již řešením původní rovnice. V algoritmu, který se nazývá metodou sdružených (konjugovaných) gradientů, se neurčují konstanty $\alpha_{i,j}$, ale přímo se počítají x_i podle rekurentních vztahů. Při tomto postupu je délka směru volena metodou největšího spádu a směr δ_i v i-tém kroku je C-ortogonální ke všem dříve užitým směrům, tj. $\delta_i C \delta_j = 0$ pro $j < i$. Přitom $C = H(0)$ je matice druhých derivací (které jsou konstantní v kvadratickém případě, který nyní uvažujeme). Kdybychom přešli k souřadnicím, v nichž odpovídá matici $H(0)$ jednotková matice a funkci f funkce $g(\xi) = \xi'\xi$, zjistili bychom, že směry δ_i přecházejí ve směry vzájemně kolmé a tedy v k krocích dojdou k bodu minima funkce g s kulovými vrstevnicemi. Fletcher a Powell (1963/4) (algoritmus v ALGOLu 60 uveřejnil Wells (1965)) navrhli používat gradientní metody se směrem $B_n D(x_n)$ a délkou α_n určenou metodou největšího spádu, přičemž B_0 je nějaká pozitivně definitní matice a B_n se postupně konstruují ze vztahů

$$B_{n+1} = B_n + \frac{\sigma_n\sigma_n'}{\sigma_n' y_n} - \frac{B_n y_n y_n' B_n}{y_n B_n y_n},$$

kde

$$\sigma_n = x_{n+1} - x_n\,, \quad y_n = D(x_{n+1}) - D(x_n)\,.$$

Ukázali: 1. na několika příkladech podstatně rychlejší konvergenci než při $B_n = E$, 2. že B_n jsou pozitivně definitní (důkaz není v pořádku, ale tvrzení platí), 3. že v případě kvadratické funkce s $H = H(0)$ pozitivně definitní je x_n již rovno bodu minima a $B_n = H^{-1}$ (jako u Newtonovy metody).

Je však možné ukázat, že Fletcherova-Powellova metoda dává v kvadratickém případě a při $B_0 = E$ přesně touž posloupnost bodů $x_1, x_2, \ldots, x_n$ jako metoda sdružených gradientů. Numerická zkušenost, vlastnost 2. a vztah $D_n = M^{-1}$ zůstávají však stejně zajímavé, jako i jiná formulace metody sdružených gradientů, jako gradientní metody námi uvažovaného typu. Fletcher a Reeves (1964) – zřejmě bez znalostí vztahu mezi metodou konjugovaných gradientů a metodou Fletcherovou-Powellovou – experimentovali s jinými rekurentními vztahy metody konjugovaných gradientů (algoritmus v ALGOLu 60 uveřejnil Reeves (1964)). (Třídu obecnějších algoritmů, ale též jen pro kvadratický případ, studovali Shah, Buehler a Kempthorne (1964), jejichž práce též zobecňují článek Finkelův (1959).) Všechny tyto metody v kvadratickém případě totožné, mohou v nekvadratickém případě dávat různé výsledky. Všechny též dávají slibné výsledky, u žádné ovšem nebyly v obecném případě dokázány konvergenční vlastnosti. Powell (1964, 1965) uvažoval také zajímavé příbuzné postupy, v nichž není třeba počítat derivace (metody nepoužívající derivací navrhli již dříve Rosenbrock (1960), Hooke a Jeeves (1961); druhý článek jsem neviděl, cituje jej Spang (1962)). S Powellovou metodou (1964) porovnávali numericky svou metodu Nelder a Mead (1965), kterou bohužel navrhují bez důkazů vlastností.

5.4 Metoda založená na spojité aproximaci

Připomeneme, že již dříve se popisovaly spojité aproximační procesy diferenciálními rovnicemi. Tyto procesy lze realizovat buď na analogových počítačích (viz Ablow a Brigham (1955), Pyne (1959), DeLand (1959)), což má ovšem nevýhody v omezení na přesnosti i okruh problémů, nebo přibližně na číslicových počítačích. Pak ovšem je algoritmus specifikován, až když je specifikována metoda řešení diferenciální rovnice. Kizner ve značně speciálním případě (řešení systému nelineárních rovnic dimense $\leqq 2$) použil Rungeovy-Kuttovy metody a obdržel konvergenci řádu 5, tj. značí-li e_k nepřesnost k-té aproximace, je $e_{k+1}/e_k^5 = 0(1)$ pro $k \to \infty$. Kdyby šlo tento výsledek dostatečně zobecnit se zachováním rychlosti konvergence, bylo by možné na většinu toho, o čem jsme referovali dříve, zapomenout.

6. STOCHASTICKÝ PŘÍPAD BEZ OMEZENÍ

6.1 Přehled výsledků

Hotelling (1941) uvažuje stochastickou versi minimalizace funkce více proměnných. Jeho postup je nesekvenční. Funkce se aproximuje polynomem, jehož konstanty se odhadují z napozorovaných odhadů funkčních hodnot. Uvažuje chyby způsobené 1. odhady těchto konstant, 2. aproximací funkce polynomem daného stupně. Chyby druhé skupiny je ovšem velmi těžké analyzovat; zdá se že dnešní vývoj tímto směrem nejde.

Friedman a Savage (1947) se zabývají již sekvenčním vyhledáváním bodu minima. Uvažují postup, v němž se funkce minimalizuje postupně ve směru jednotlivých proměnných.

Box a Wilson (1951) navrhují jakousi směs kroků metodou největšího spádu a kroků, v nichž prokládají polynomy podobně jako u Hotellinga (1941). Tento článek byl sledován celou řadou dalších (Box (1954), Box a Youle (1955), Box (1957), Box a Hunter (1959)). Jejich postup byl často v praxi používán a přinesl úspěchy. Protože však je v něm ponecháno mnoho na subjektivním úsudku, nemůže být dost dobře studován matematickými prostředky. Nevím též nic o praktickém porovnávání se stochastickými aproximacemi.

Robbins a Monro (1951) popsali stochastickou aproximační metodu pro hledání řešení rovnice $f(x) = 0$ v jednorozměrném případě s funkcí f zápornou vlevo od z a kladnou vpravo od z. Aproximační formule zní $x_{n+1} = x_n - a_n y_n$, kde y_n je odhad hodnoty $f(x_n)$ a a_n jsou čísla splňující podmínku $c/n \leqq a_n \leqq C/n$ s kladnými c, C. Robbins a Monro dokázali za mírných dodatečných podmínek na f a y_n, že $x_n \to z$ podle pravděpodobnosti. Všimněme si analogie aproximací, kterou navrhli Mises a Pollaczek-Geiringer (1929), viz odst. 5.1. Odlišnost předpokladů znemožňuje zvolit ve stochastickém případě a_n rovny jedné konstantě c, avšak tím se vyhneme nesnázi, která spočívá v deterministickém případě v požadavku „dostatečně malého c“. Tato situace je stejná i v jiných případech analogických metod deterministických a stochastických.

Wolfowitz (1952) poněkud zobecnil původní podmínky Robbinsovy-Monroovy a Kiefer a Wolfowitz (1952) modifikovali Robbinsovu-Monroovu aproximační metodu tak, aby vyhledávala bod maxima funkce f v jednorozměrném případě. Zde $x_{n+1} = x_n + a_n[(y_{n,1} - y_{n,-1})/c_n]$, kde $y_{n,i}$ je odhad čísla $g(x_n + ic_n)$ a

$$c_n \to 0\,, \ \sum a_n = +\infty\,, \ \sum a_n^2 c_n^{-2} < +\infty\,, \ c_n > 0\,, \ a_n > 0\,.$$

 Proces připomíná deterministický proces, který navrhnul Germansky (1934) s rozdílem ve volbě konstant a s tím, že ve stochastické verzi se derivace odhaduje z diferencí.

Schmetterer (1953, 1954a, 1954b) asi jako prvý a dále Killianpur (1954), Hodges a Lehmann (1956) vyšetřovali rychlost konvergence Robbins-Monro metody; další výsledky o rychlosti konvergence a asymptotickém rozložení aproximujících x_n odvodil Chung (1954). Za jistých dodatečných předpokladů jsou x_n asymptoticky normální a označíme-li $e_n = \sqrt{\mathsf{E}(x_n - r)^2}$, kde E značí očekávanou hodnotu, je $e_n = 0(n^{-1/2})$.

Blum (1954a, 1954b) zobecnil podmínky metod Robbinsovy-Monroovy a Kieferovy-Wolfowitzovy a dokázal konvergenci s pravděpodobností 1 (dřívější výsledky se týkaly konvergence podle pravděpodobnosti). Jako prvý pak zobecnil tyto metody na vícerozměrný případ. Důležitější z obou zobecnění je metoda pro minimalizaci funkce. Je analogická gradientní metodě deterministické, avšak opět místo „vhodně malé konstanty" používá posloupnosti čísel konvergujících k nule a gradient aproximuje odhadovanými diferencemi.

Burkholder (1955, 1956), Derman (1956) a Dupač (1957, 1958) vzájemně nezávisle zobecňují Chungovu metodiku na studium rychlosti konvergence Kieferovy-Wolfowitzovy metody. Tato rychlost závisí na podmínkách, které splňuje f. Tyto výsledky byly převedeny i na vícerozměrný případ Sacksem (1958), který používá jiných důkazových prostředků než předchozí práce. Burkholder (1955, 1956) uvažoval též aproximační schémata zobecňující procesy Robbinsův-Monroův a Kieferův-Wolfowitzův a ukázal, že v případě Kieferova-Wolfowitzova postupu konvergují postupné aritmetické průměry odhadů $y_{n,1}$, $y_{n,-1}$ funkčních hodnot s pravděpodobností 1 k hodnotě funkce f v bodě minima.

Obecnou větu udávající vztah mezi deterministickými aproximacemi jistého typu a mezi stochastickými aproximacemi dokázal Dvoretzky (1956). Zahrnuje všechny předtím známé jednorozměrné konvergenční výsledky. Bohužel její vícerozměrná verse již asi není tak silná. Původní dost komplikovaný důkaz zjednodušil nejprve Wolfowitz (1956), ještě jednodušší důkaz podstatné části věty podali Derman a Sacks (1959). Zobecnění na obecnější prostory provedl Block (1957); jiné zobecnění, jehož cílem bylo zahrnout Blumovo (1958) zobecnění jednorozměrné Robbinsovy-Monroovy metody, podala Krasulina (1962). Jiné zobecnění podal Kitagawa (1959).

Ve snaze urychlit konvergenci navrhl Kesten (1958) pro Robbinsovu-Monroovu metodu, aby se místo původní posloupnosti a_n používalo jiné, která vznikne tak, že použijeme konstanty z předchozího kroku, jestliže $x_{n-1} - x_{n-2}$ a $x_n - x_{n-1}$ mají stejné znaménko; jinak použijeme dalšího (ještě nepoužitého) členu původní posloupnosti a_j. Podobná modifikace Kieferovy-Wolfowitzovy metody nebyla příliš úspěšná, také zobecnění na vícerozměrný případ Kesten (1958) neuvažoval. Později navrhl Fabian (1960) jinou modifikaci, stejně motivovanou, pro hledání minima ve vícerozměrném případě. Určuje-li původní metoda x_{n+1} jako $x_n + \Delta_n$, pak při této modifikaci se postupně získávají odhady hodnot $f(x)$ pro $x = x_n + j\Delta_n$, $j = 1, 2, \ldots$ Pokračuje se tak dlouho, pokud napozorované odhady tvoří klesající posloupnost. Předposlední z těchto hodnot je pak nové x_{n+1}. (Kdyby náhodné chyby neměly podstatný vliv, je délka kroku přibližně jako v metodě největšího spádu.). Za značně obecných podmínek zachovává tento postup konvergenci Blumovy vícerozměrné minimalizační metody i některých jejích modifikací.

V roce 1960 je též publikována celá řada výsledků o spojitých aproximačních metodách: Driml a Nedoma (1960) dokazují za jistých podmínek konvergenci spojité analogie Robbinsova-Monroova postupu pro řešení rovnice $r(x) = 0$. Aproximační postup je tvaru $dx_t/dt = -(1/t)\, y_t(x_t)$ pro $t \geqq 1$. Přitom je $y_t(x_t)$ odhad funkční hodnoty $r(x_t)$ v čase t, r je neklesající funkce. Driml a Hanš (1960) a Hanš a Špaček (1960) podávají stochastické verse aproximací fixních bodů kontrahujících transformací. Analogii Drimlovy-Nedomovy metody pro vícerozměrnou minimalizaci uvažoval Sakrison (1964), který též studoval aplikace na problémy optimalizace a adjstace filtrů (Sakrison (1962a, 1962b, 1963)). Ho (1962) ukázal na souvislosti s iterační metodou z teorie filtrace a predikce, Fabius (1959) publikoval informační článek, Gardner (1963, 1965) studuje aplikace stochastických aproximačních metod v adaptivní predikci. Blízké stochastickým aproximacím jsou též „Up and down" metody, pro něž viz Derman (1957), Brownlee a Hodges (1963) a Wetherill (1963). Hledá se v nich jen mezi celočíselnými body (nebo mezi body jiné sítě).

6.2 Blumova minimalizační metoda a její modifikace

Výsledek Blumovy (1954b) práce týkající se minimalizace lze shrnout větou, která je jen formální reformulací původní věty Blumovy.

Věta 1. *Nechť f má druhé derivace, $\sup_{x \in E_k} \|H(x)\|$ buď konečné a nechť pro každé $\varepsilon > 0$ je*

$$(1.1) \qquad f(z) = 0\,, \; \inf \{\|D(x)\|;\, \|x - z\| > \varepsilon\} > 0\,; \; \inf \{f(x);\, \|x - z\| > \varepsilon\} > 0\,.$$

pro nějaké $z \in E_k$ (takže z je bodem minima f). Buďte a_n, c_n kladná čísla,

$$(1.2) \qquad c_n \to 0\,, \; \sum_{n=1}^{\infty} a_n = +\infty\,, \; \sum_{n=1}^{\infty} a_n c_n < +\infty\,, \; \sum_{n=1}^{\infty} a_n^2 c_n^{-2} < +\infty\,,$$

X_n náhodné vektory splňující vztah

$$(1.3) \qquad X_{n+1} = X_n - a_n Y_n\,,$$

$$(1.4) \qquad \mathsf{E}_{X_1, X_2, \ldots, X_n} Y_n^{(i)} = c_n^{-1} [f(X_n + c_n e_i) - f(X_n)]$$

($\mathsf{E}_{X_1, \ldots, X_n}$ *je znak podmíněné očekávané hodnoty, $e_i \in E_k$, $e_i^{(j)} = 0$ pro $i \neq j$, $e_i^{(i)} = 1$.*)

$$(1.5) \qquad \mathsf{E}_{X_1, \ldots, X_n} (Y_n - E_{X_1, \ldots, X_n} Y_n)^2 \leqq \frac{\sigma^2}{c_n^2}\,.$$

$$\mathsf{P}\{X_n \to z\} = 1 .$$

Naznačíme hlavní kroky důkazu: Aritmetickými úpravami lze ukázat, že $f(X_n) = \zeta_n + \eta_n$, kde ζ_n je konvergentní martingal a η_n neklesající omezená posloupnost; odtud plyne konvergence $f(X_n)$. Pak se ukáže, že v případě $\mathsf{P}\{\|D(X_n)\| \geqq \varepsilon > 0\} > 0$ pro všechna n, musí být $\mathsf{E}\, f(X_n) \to -\infty$, což není možné. Proto je vzhledem k (1.1) s pravděpodobností 1 z hromadným bodem posloupnosti X_n a vzhledem k (1.1) $f(X_n) \to 0$ s pravděpodobností 1.

Sacks (1958) poznamenává, že podmínka (1.4), která umožňuje snížit počet odhadů na n-tém kroku na $k + 1$, vede k dominantní systematické odchylce X_n od řešení z, čímž se snižuje rychlost konvergence. Navrhl proto místo (1.4) použít vztahu

$$\mathsf{E}_{X_1,\ldots,X_n} Y_n^{(i)} = (2c_n)^{-1}[f(X_n + c_n e_i) - f(X_n - c_n e_i)] . \tag{1.6}$$

Dokazuje pak za řady doplňujících podmínek, že X_n jsou asymptoticky normální a vzdálenost X_n od z je $0(c_n^{-1} n^{-1/2})$. Přitom rychlost s jakou musí $c_n \to 0$, závisí na tvaru funkce. Pro názornost uvedeme největší možnou rychlost v jednorozměrném případě, jak ji vyšetřoval Dupač (1957, 1958) s omezením se na $a_n = n^{-\alpha}$, $c_n = n^{-\gamma}$. Je-li $K_0\|x - z\| \leqq \|D(x)\| \leqq K_1\|x - z\|$ pro kladné konstanty K_0, K_1, a volíme-li $a > 1/(2K_0)$, $a_n = a/n$, $c_n = c/n^{1/4}$, je $\sqrt{\mathsf{E}(X_n - z)^2} = 0(n^{-1/4})$. Je-li navíc třetí derivace funkce f omezena v absolutní hodnotě konstantou, pak je optimální volba $a > 1/(2K_0)$, $a_n = a/n$, $c_n = c/n^{1/6}$, pro níž $\sqrt{\mathsf{E}(X_n - z)^2} = 0(n^{-1/3})$. Obecně lze dosáhnout tím lepší rychlosti konvergence, čím je f symetričtější kolem bodu z (Dupač (1957, 1958), Sacks (1958), Schmetterer (1961)).

Fabian (1960) studuje jednotným způsobem konvergenci Robbinsovy-Monroovy, Kieferovy-Wolfowitzovy a Blumovy vícerozměrné aproximační metody a několika jejich modifikací. O jedné z nich, týkající se délky kroku jsme mluvili již výše. Druhá spočívá v tomto: V původní Blumově metodě je krok roven $-a_n Y_n$, kde Y_n je odhad gradientu v bodě x_n. Délka tohoto kroku je rovna $a_n\|Y_n\|$ a — přibližně — $a_n\|D(x_n)\|$. To není vhodné v obecném případě, v němž může být $\|D(x_n)\|$ velká v bodech x_n blízkých hledanému bodu a malá v bodech x_n vzdálených. Navrhuje se proto zaměnit Y_n vektorem se složkami c_n^{-1} sign $Y_n^{(i)}$ a dokazuje se, že konvergenční vlastnosti zůstanou zachovány za 1. zobecněných podmínek týkajících se f a za 2. poněkud zpřísněných, ale stále ještě velmi obecných požadavků na náhodné chyby. Zobecnění požadavků na f spočívá v tom, že není třeba vyloučit případy, v nichž x_n může divergovat vzhledem k příliš velkým $\|D(x_n)\|$, které vedou k neomezeně velkým krokům $a_n Y_n$. (Podobné modifikace uvažoval též Friedman (1963).) Konvergenční vlastnosti studuje autor za obecnějších předpokladů, které nezaručují existenci jenom jediného extrému. Výsledky ukazují, že, zhruba řečeno, x_n opět konverguje k bodu, v němž má f nulovou derivaci, avšak nepodařilo se dokázat, že to musí být lokální minimum. Později (Fabian (1962)) dokázal tuto silnější vlastnost v jednorozměrném případě pro modifikovanou aproximační metodu, v níž $a_n = 1/\sqrt{n}$ (a tedy $\sum a_n^2 = +\infty$ na rozdíl od dosavadních předpokladů). Asymptotická rychlost konvergence této metody je však pomalá (Vosiková (1964)).

6.3 Praktické zkušenosti

Guttman a Guttman (1959) podávají výsledky provedených aplikací. Později jiný takový příklad podává Fabian (1961) a dosti rozsáhlé Monte-Carlo studie chování aproximační Robbinsovy-Monroovy procedury provedl Wetherill (1963): Pro odhad α-kvantilu logistické křivky za obvyklých podmínek biologických zkoušek je zkušenost velmi dobrá u L_{50}, horší u L_{90}; to ovšem v této situaci je vůbec obtížnější problém. Wetherill také konstatuje, že je-li výchozí bod daleko od řešení, délka kroků a_n klesá příliš rychle: Tuto nevýhodu mají v mnohem menší míře modifikace navržené Kestenem (1958) a Fabianem (1960), o nichž jsme mluvili již výše.

Praktické zkušenosti autora svědčí o tom, že (při použití modifikací jím navržených) $f(x_n)$ zpočátku dosti rychle ubývá. Z teoretických výsledků je zřejmé, že později je konvergence velmi pomalá, avšak v experimentálních aplikacích často několik kroků stačí resp. není možné jich více provést. (Zde je ovšem otázka, do jaké míry jsou relevantní teoretické asymptotické výsledky pro takové aplikace.) Tyto dobré zkušenosti se týkají ovšem poměrně malé dimenze ($k = 4$, $k = 6$ a pod.) a také uvažované funkce nebyly asi vytvořeny přírodou s ďábelským úmyslem příliš složitě. Na rozdíl od toho požadavky na deterministické aproximační metody bývají podstatně náročnější (velká k, velký počet kroků a též požadavek, aby x_n bylo blízké bodu minima a nikoliv jen, aby se $f(x_n)$ zmenšilo ve srovnání s $f(x_0)$).

7. NELINEÁRNÍ PROGRAMOVÁNÍ

Nelineární programování je jen jiný název pro deterministický případ s omezením.

V této části předpokládáme, že existuje spojitá derivace D funkce f, že G je m-rozměrná vektorová funkce se spojitou derivací D_G ($D_G(x)$ je matice typu $k \times m$, $D_G^{(i,j)}(x) = D_{G^{(j)}}^{(i)}(x)$). Předpokládáme, že f a $G^{(i)}$ jsou konvexní funkce. Tento předpoklad lze v mnohých tvrzeních zeslabit (viz Zoutendijk (1960)); zjednodušuje však úvahy. Konvexita G implikuje konvexitu množiny $R = \{x; G(x) \leqq 0\}$. Předpokládáme též, že $R_0 = \{x; G^{(i)}(x) < 0$ pro všechna $i\}$ je neprázdná množina. Uvažovaným problémem je minimalizace f na R. Označíme dále $I(x)$ množinu $\{i; G^{(i)}(x) = 0\}$. Označení f, G, D_G, R, R_0, $I(x)$ budeme používat v celém tomto odstavci.

V oblasti nelineárního programování je mnoho výsledků paralelně dokazovaných podobnými prostředky. Většina je však důsledky několika málo základních. Jedním z nich je následující výsledek (Zoutendijk (1960)), patřící mezi ty, které charakterizují bod minima způsobem, na němž je možno založit metodu pro jeho hledání.

Věta 1. *Tyto 3 podmínky jsou vzájemně ekvivalentní:*

(i) *z není bodem minima f na R,*

(ii) *existuje s tak, že pro nějaké $\varepsilon_1 > 0$ a $\varepsilon_2 = 0$ platí*

$$s'D(z) \leqq -\varepsilon_1\,, \quad s'\,D_{G^{(i)}}(z) \leqq -\varepsilon_2 \quad \text{pro} \quad i \in I(z)\,; \tag{1.1}$$

(iii) *existuje s a $\varepsilon > 0$ tak, že platí (1.1) pro $\varepsilon_1 = \varepsilon_2 = \varepsilon$.*

Důkaz: Platí-li (iii), pak pro dostatečně malé α je $f(z + \alpha s) < f(z)$, $G^{(i)}(z + \alpha s) < G^{(i)}(z) = 0$ pro $i \in I(z)$, $G^{(i)}(z + \alpha s) < 0$ pro ostatní i a tedy platí (i). Platí-li (i), existuje $x \in R$ tak, že $f(x) < f(z)$. Z konvexity plyne (1.1) s kladným ε_1 a s $\varepsilon_2 = 0$, tj. platí (ii). Platí-li (ii) a položíme li $s_2 = x_2 - z$ pro nějaké $x_2 \in R_0$, pak směr $s_1 = s + \alpha s_2$ splňuje již pro dostatečně malé α podmínku (iii) a důkaz je skončen.

Věta 2. *z je bodem minima f na R tehdy a jen tehdy, existuje-li $\mu \in E_m$ tak, že*

$$D(z) = -D_G(z)\,\mu\,, \quad \mu \geqq 0\,, \; G'(z)\,\mu = 0\,. \tag{2.1}$$

(Poznamenáme, že $G'(z)\,\mu = 0$ lze zapsat též ve tvaru $\mu^{(i)} = 0$ pro $i \notin I(z)$, tj. pro $G^{(i)}(z) < 0$.)

Důkaz: Stačí dokázat, že (2.1) je ekvivalentní negaci podmínky (ii) z věty 1. Platí-li (2.1) a je-li $s'\,D_{G^{(i)}}(z) \leqq 0$ pro $i \in I(z)$, je podle (2.1) $s'\,D(z) = -\,s'\,D_G(z)\,\mu \geqq 0$, takže (ii) neplatí. Neplatí-li (ii), plyne existence μ splňujícího (2.1) z Farkasovy (1902) věty (viz též Zoutendijk (1960, § 2.2)), citované obvykle též v pracích o lineárním programování: Jsou-li a_i $(i = 0, 1, \ldots, q)$ k-rozměrné vektory a jestliže neexistuje s takové, že $s'a_0 < 0$, $s'a_i \leqq 0$ pro $i = 1, \ldots, q$, pak je $a_0 = \sum_{i=1}^{q} a_i n_i$ s $n_i \geqq 0$.

Věta 3. *(Věta o sedlovém bodu). K tomu, aby z bylo bodem minima f na R, je nutné a stačí, aby pro nějaké $\mu \in E_m$, $\mu \geqq 0$ bylo*

$$f(z) + \lambda'\,G(z) \leqq f(z) + \mu'\,G(z) \leqq f(x) + \mu'\,G(x) \tag{3.1}$$

pro všechna $x \in E_k$, $\lambda \in E_m$, $\lambda \geqq 0$.

Důkaz. Nechť platí (3.1). Z levé strany pro $\lambda = 0$ dostáváme $\mu'\,G(z) \geqq 0$. Kdyby $G^{(i)}(z) > 0$ pro nějaké i, pak při dostatečně velkém $\lambda^{(i)}$ by levá strana (3.1) nemohla platit. Je tedy $G(z) \leqq 0$, $z \in R$ a $\mu'\,G(z) = 0$ a z pravé části (3.1) $f(z) \leqq f(x) + \mu'\,G(x)$. Pro $x \in R$ je navíc $\mu'\,G(x) \leqq 0$ a tak je z bodem minima f na R.

Nechť je z bodem minima f na R. Podle věty 2 existuje μ takové, že $\mu \in E_m$, $D(z) = -D_G(z)\,\mu$, $\mu \geqq 0$, $G'(z)\,\mu = 0$. Pro $\lambda \geqq 0$ je $\lambda'\,G(z) \leqq 0$ neboť $G(z) \leqq 0$ a platí levá část (3.1). Pro naše pevné $\mu \geqq 0$ je funkce $f(x) + \mu'\,G(x)$ konvexní a má v bodě z minimum, neboť její derivace je rovna 0 v bodě z. Platí tedy pravá část nerovnosti (3.1).

Z věty 2 lze dále odvodit věty o dualitě (viz např. Mangasarian (1962); špatně citují Saaty a Bram (1964): v primárním problému nemá být podmínka $x \geqq 0$) a mnoho dalších ekvivalentních formulací úlohy nelineárního programování.

Podle mých znalostí první charakterizaci bodu minima (obecnější než věta 2, uvažuje nekonečně mnoho podmínek tvaru $G^{(i)}(x) \leqq 0$, nepředpokládá konvexitu) odvodil John (1948). Podmínku typu věty 2 odvodili Kuhn a Tucker (1950), později ji zobecnili Arrow a Hurwicz (1956). (Přímý důkaz ekvivalence podmínek typu věty 1 a věty 2 udal též Mangasarian (1963)). Charakterizaci stacionárních bodů v Banachových prostorech podal Altman (1964).

Brzy pak byly také navrženy a studovány spojité gradientní metody pro hledání sedlových bodů: Arrow a Hurwitz (1951), Kose (1956). Prvá z obou prací byla pak otištěna v upravené verzi spolu s mnohými dalšími články o této problematice ve sborníku Arrow, Hurwicz a Uzawa (1958). O mírně povzbudivých praktických zkušenostech s těmito metodami referují (podle Wolfeho (1962)) Manne (1953) a Marschak ve sborníku Arrow et al. (1958). Zoutendijk ((1960) věta 2, § 7.2) v poněkud jiné formulaci dokazuje ještě jednu možnost převedení problému s omezením na problém bez omezení: Je-li $\inf f(x) > -\infty$ pak existuje M tak, že bod minima $f(x) + M\,F(x)$, kde $F(x)$ je rovno maximu z kladných částí složek $G^{(i)}(x)$, $i = 1, 2, \ldots, m$, je bodem minima funkce f na množině $\{x; G(x) \leqq 0\}$. Tento výsledek není uváděn dalšími, kteří uvádějí podobné výsledky. Pietrzykowski (1961, 1962) dokazuje, že řešení původní úlohy je limitou bodů minima funkce $f(x)/\mu + \sum_{i=1}^{k} G_{+}^{(i)}(x)$ pro $\mu \to \infty$.*) Carroll (1961) navrhl minimalizovat funkci $f(x) + r \sum_{i=1}^{k} 1/[G_{+}^{(i)}(x)]$, $(r > 0)$, Fiacco a McCormick (1964a, 1964b) dokázali, že řešení této úlohy konverguje pro $r \to 0$ k bodu minima f na R.

Pro hledání bodu minima v deterministickém případě s omezením byly, zdá se, nejprve převážně studovány spojité aproximační procedury pro vyhledání sedlového bodu funkce $F(x, \lambda) = f(x) - \lambda\, G(x)$. Diskrétní gradientní metodu pro řešení této úlohy studoval Uzawa v kap. 10 sborníku Arrow, Hurwicz a Uzawa (1958).

Jiný směr představují metody, v nichž se minimalizuje přímo funkce $f(x)$ tak, že se z daného bodu x_n postupuje ve směru s_n, v němž f klesá, který však nevede ven z množiny $\{x; G(x) \leqq 0\}$. Za poněkud speciálnějších předpokladů navrhli takové metody Beale (1955, 1958) a Abrham (1958, 1961, 1962). Bealovu (1958) práci jsem neviděl, avšak Shah, Buehler a Kempthorne uvádějí, že je to dosti komplikovaná metoda, výhodná v těch případech, kdy počítání funkčních hodnot je relativně pomalé nebo nákladné ve srovnání s ostatními výpočty. Práce Abrhamova (1962) je upravená Frishova (1957, 1958) metoda lineárního programování. Za obecnějších podmínek navrhl metodu „možných směrů" Zoutendijk (1959, 1960). Zoutendijkova metoda používá jeden ze směrů, jejichž existence je – pokud již nejsme v bodu minima – zaručena podmínkou (iii) věty 1 a který vede dovnitř množiny R. Přitom vybírá ten směr, který splňuje podmínku (iii) s největším ε a současně vyhovuje některé z normalizačních podmínek. V podmínce (iii) se však přitom pro ta $G^{(i)}$, která jsou lineární, vyžaduje pouze $s' D_{G^{(i)}}(x) \leqq 0$. Nevýhodou je, že takový směr je třeba

* Symbol $_+$ zde značí nezápornou část funkce. Pietrzykowski uvádí poněkud jinou formulaci.

 určit řešením úlohy lineárního programování a tedy dosti pracně. Rosen (1960, 1961) používá jednodušeji určeného směru, splňujícího podmínku (ii) věty 1, po kterém se případně nemnoho – po tečně – vzdálíme od R a pak se opět vracíme. Wolfe (1962) referuje, že velmi dobrý přehled těchto prací podává Witzgall (1960) (tuto práci jsem neviděl). Později se objevila celá řada prací, které navrhují metody podobné metodě Zoutendijkově, aniž jeho práci citují (citují však Rosena (1960), který na Zoutendijkovu (1960) práci odkazuje): je to Ivanov (1962), Klingman a Himmelblau (používají diferencí místo derivací, neuvádějí žádné dokázané vlastnosti). Glass a a Cooper (1965) neudávají ani důkazy ani podrobnosti své navrhované metody.

Předpoklad konvexity umožňuje též převést problém přibližně na lineární programování. Původní problém se především převede na problém s lineární funkcí f, tak, že se přejde na E_{k+1}, položí se $\bar{f}(x) = x^{(k+1)}$ a přidá se omezení $G^{(m+1)}(x) = f(x) - x^{(k+1)} \leqq 0$.

Cheney a Goldstein (1959) (připisující svou myšlenku dozadu Remezovi (1934)) popisují algoritmus, v němž se minimalizuje lineární f na R a R se vyjádří ve tvaru $\{x;\ T(x) \leqq 0,\ T \in \mathsf{T}\}$, kde T je systém všech lineárních supportů původní množiny R. Téže myšlenky v jednodušším výkladu využívá Kelley (1960). Další práce v tomto směru se snaží řešit otázky efektivnosti tohoto přístupu: Wolfe (1960, 1961b), Hartley a Hocking (1963).

8. OTEVŘENÉ OTÁZKY

Jaká je vhodná volba B_n pro určení směru $-B_n\, D(x_n)$ pro x_n ne nutně blízké bodu minima? Jaké jsou vlastnosti metod uvažovaných z odst. 5.3 v nekvadratickém případě, lze zaručit aspoň konvergenci? Lze zobecnit Kiznerovu metodu z odst. 5.4? Jak modifikovat stochastické aproximační metody, aby v případě, že náhodné chyby jsou malé, nebyly horší než dobré deterministické metody? Lze zde též využívat odhadu inverse matice $H(X_n)$ ke zlepšení směru a případně i délky a_n směru v n-tém kroku? Nelze volit a_n, c_n v závislosti na průběhu aproximačního procesu tak, abychom dostali nejlepší možnou rychlost konvergence bez předběžné znalosti vlastností f, a abychom přitom zaručili i opuštění lokálního minima, ale nikoliv za cenu ztráty rychlosti? Jaká pravidla je možno udat pro zastavení? (Hrubé pravidlo pro konvexní f udává Zoutendijk (1960), prakticky asi nevýhodné pravidlo pro Robbinsovu-Monroovu metodu, dávající současně konfidenční interval pro řešení, navrhl Farrell (1962).)

(Došlo dne 26. července 1965.)

Poznámka: Během tisku byly dokončeny další práce: Dupač (1965a, 1965b) studuje stochastickou aproximaci za předpokladu, že se uvažovaná funkce mění průběhem času. Fabian (1965) navrhl stochastickou aproximační metodu pro případ s omezením. Do seznamu literatury byla též zahrnuta práce, v níž Gladyšev (1965) poněkud zobecňuje podmínky konvergence Robinsovy-Monroovy metody.

Zkratky časopisů:

AMS	The Annals of Mathematical Statistics
CACM	Communications of the Association for Computing Machinery
CZM	Чех. мат. журнал (Czech. Math. Journal)
DAN SSSR	Доклады Академии наук СССР
JACM	Journal of the Association for Computing Machinery
JASA	Journal of the American Statistical Association
JRSS	Journal of the Royal Statistical Society
JSIAM	Journal of the Soc. for Industrial and Applied Mathematics
MS	Management Science
OR	Operations Research
PRTR*i*	Transactions of the *i*-th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes. Czech. Ac. Sci., Prague

Ablow C. M. and G. Brigham (1955): An analog solution of programming problems. OR *3*, 388 to 394.

Abrham J. (1958): Přibližná metoda pro nelineární programování. Čas. pěst. mat. *83*, 425—439.

Abrham J. (1961): An Approximate Method for Convex Programming. Econometrica *29*, 700 to 703.

Abrham J. (1962): The Multiplex Method and its Application to Concave Programming. CZM *12*, 325—345.

Ackoff R. L. (1961): Progress in Operations Research, Vol. I. John Wiley & Sons. Inc., New York — London.

Akaike H. (1959): On successive transformation of probability distribution and its application to the analysis of the optimum gradient method. Ann. Inst. Statist. Math. Tokyo *11*, 1—16.

Altman M. (1964): Stationary points in non-linear programming. Bull. Acad. Polon. Sci. Ser. Sci. Math. Astronom. Phys. *12*, 29—35.

Archer D. H. (1960): An optimising control for the chemical process industries. Britisch Chemical Engineering, 88—94.

Arnoff E. L. and S. S. Sengupta (1961): Mathematical Programming. Viz sb. Ackoff (1961), 105—210.

Arrow K. J. and L. Hurwicz (1951): A gradient method for approximation of saddle points and conditionned maxima. Rep. 253 RAND Corporation, June *13*.

Arrow K. J. and L. Hurwicz (1956): Reduction of constrained maxima to saddle-point problems. Proc. 3rd Berkeley Symp. Math., Statist. Probability, vol. V, 1—20, Un. Calif. Press.

Arrow K. J. and L. Hurwicz, H. Uzawa (1958): Studies in Linear and Non-linear Programming. Stanford University Press, Stanford, California.

Beale E. M. L. (1955): On optimising a convex function subject to linear inequalities. JRSS *B 17*, 173—184.

Beale E. M. L. (1958): On an iterative method for finding a local minimum of a function of more than one variable. Statist. Tech. Research Group Techn. Report No. 25, Princeton Univ., Princeton, N. Y.

Beale E. M. L. (1959): On quadratic programming. Naval Research Logistic Quart. *6*, 227—243.

Bellman R. (1957): Dynamic Programming. Princeton University Press, Princeton, N. Y.

Block H. D. (1957): On stochastic approximation. ONR Report, Dept. of Mathematics. Cornell University, Ithaca, New York, 13 pp.

Blum J. R. (1954a): Approximation methods which converge with probability one. AMS *25*, 382 to 386.

Blum J. R. (1954b): Multidimensional stochastic approximation methods. AMS *25*, 463—483.

Blum J. R. (1958): A note on stochastic approximation. Proc. Am. Math. Soc., *9*, 404—407.

Both A. D. (1949): An application of the method of steepest descent to the solution of systems of non-linear simultaneous equations. Quar. J. Mech. Appl. Math. *2*, 460 —468.

Box G. E. P. and K. B. Wilson (1951): On the experimental attainment of optimum conditions, JRSS B *13*, 1—45.

Box G. E. P. (1954): The exploration and exploitation of response surfaces: some general considerations and examples. Biometrics *10*, 16—60.

Box G. E. P. and P. V. Youle (1955): The exploration and exploitation of response surfaces: an example of the link between the fitted surface and the basic mechanism of the system. Biometrics *11*, 287—323.

Box G. E. P. (1957): Evolutionary operation: a method for increasing industrial productivity. Applied Statistics *6*, 81—101.

Box G. E. P. and J. S. Hunter (1959): Condensed calculations for evolutionary operation programs. Technometrics, 77—95.

Brooks S. H. (1959): A comparison of maximum seeking methods J. Oper. Research Soc. *7*, 430—457.

Brownlee K. A., J. L. Hodges, Jr. and M. Rosenblatt (1963): The Up-and-Down method with small samples. JASA *48*, 262—277.

Burkholder D. L. (1956): On a certain class of stochastic approximation procedures. AMS *27*, 1044—1059.

Carroll Ch. W. (1961): The created response surface technique for non-linear restrained systems, OR *9*, 169—185.

Cauchy A. (1847): Méthode générale pour la résolution des systèms d'equations simultanées. Compte Rendu, Ac. Sci. Paris, *25*, 536—538.

Cheney E. W. and A. A. Goldstein (1959): Newton's method of convex programming and Tchebycheff approximation. Numerische Math. *1*, 253—268.

Chung K. L. (1954): On a stochastic approximation method. AMS *25*, 463—483.

Crockett J. B. and H. Chernoff (1955): Gradient methods of maximization. Pacif. J. Math. *5*, 33—50.

Curry H. B. (1944): The method of steepest descent for non-linear minimization problems. Quarterly Appl. Math. *2*, 258—261.

DeLand E. C. (1959): Continuous programming methods on an analog computer. Rand Corpor. Paper No. P-1815, Sept.

Derman C. (1956): An application of Chung's lemma to the Kiefer-Wolfowitz stochastic approximation procedure. AMS *27*, 532—536.

Derman C. (1956): Stochastic approximation. AMS *27*, 879—886.

Derman C. (1957): Non-parametric Up-and-Down experimentations. AMS *28*, 795—798.

Derman C. and J. Sacks (1959): On Dvoretzky's stochastic approximation theorem. AMS *30*, 601—606.

Dixon W. J. and A. M. Mood (1948): A method for obtaining and analyzing sensitivity data JASA *43*, 109—126.

Driml M. and O. Hanš (1960): On experience theory problems. PRTR2, 93—111.

Driml M. and O. Hanš (1960): Continuous stochastic approximation. PRTR2, 113—122.

Driml M. and J. Nedoma (1960): Stochastic approximation for continuous random processes. PRTR2, 145—158.

Dupač V. (1957): O Kiefer-Wolfowitzově aproximační metodě. Čas. pěst. mat. *82*, 47—75.

Dupač V. (1958): Notes on stochastic approximation methods. CZM *8* (83), 139—149.

Dupač V. (1965a): A dynamic stochastic approximation method. V tisku v AMS *36*.

Dupač V. (1965b): A dynamical stochastic approximation method. Referát na 4. pražské konferenci o teorii informace, statistických rozhodovacích funkcích a náhodných procesech.

Dvoretzky A. (1956): On stochastic approximation. Proc. Third. Berkeley Symp. Math. Statistics and Probability, Vol. I, 39—55. University of California Press.

Fabian V. (1960): Stochastic approximation methods, CZM *10* (85), 123—159.

Fabian V. (1961): Stochastická aproximační metoda pro hledání optimálních podmínek v experimentální práci a v samoadaptivních systémech. Aplikace matematiky, *6*, 162—183.

Fabian V. (1962): Blokové logické schema automatického optimalizátoru. Aplikace matematiky *7*, 426—440.

Fabian V. (1964): A new one-dimensional stochastic approximation method for finding a local minimum of a function. PRTR3, 85—105.

Fabian V. (1965): Stochastic approximation of constrained minima. Referát na 4. pražské konferenci o teorii informace, statistických rozhodovacích funkcích a náhodných procesech.

Fabius J. (1959): Stochastic approximation. Statistics Neerlandica *13*, 445—452.

Фаддеев Д. К., В. Н. Фаддеева (1963): Вычислительные методы линейной алгебры. Москва.

Farkas J. (1902): Über die Theorie der einfachen Ungleichungen. Zeit. Reine Angew. Math. *124*, 1—24.

Farrell R. H. (1962): Bounded length confidence intervals for the zero of a regression funkction. AMS *33*, 237—247.

Фельдбаум А. А. (1958): Автоматический оптимизатор. Автоматика и телемеханика *14*, 731—743.

Fiacco A. V. and G. P. McCormick (1964a): The sequential unconstrained minimization technique for nonlinear programming, a primal-dual method. Management Science *10*, 360—366.

Fiacco A. V. and G. P. McCormick (1964b): Computational algorithm for the sequential unconstrained minimization technique for nonlinear programming. Management Science *10*, 601—617.

Finkel R. W. (1959): The method of resultant descents for the minimization of an arbitrary function. Space Technology Lab. Tech. Report, Los Angeles.

Fletcher R. and M. J. D. Powell (1963/4): A rapidly convergent descent method for minimization. Comp. J. *6*, 163—168.

Forsythe A. I. and G. E. Forsythe (1952): Punched-cards experiments with accelerated gradient method for linear equations. NBS Report 1643, National Bureau of Standards, Los Angeles.

Forsythe G. E. and T. S. Motzkin (1951): Assymptotic properties of the optimum gradient method. Bull. Amer. Math. Soc. *57*, 73.

Forsythe G. E. (1953): Solving linear algebraic expressions can be interesting. Bull. Amer. Math. Soc. *59*, 299—329.

Forsythe G. E. and W. R. Wasow (1959): Finite-difference methods for partial difference equations. New York, Wiley & Sons.

Фридман В. М. (1962): О сходимости методов типа наискорейшего спуска. Усп. мат. наук *17*, 201—204.

Friedman M. and L. J. Savage (1947): Planning experiments seeking maxima. Select. Techn. of Statist. Analysis. McGraw-Hill, New York.

Friedman S. (1963): On stochastic approximations. AMS *34*, 343—346.

Frish R. (1957): The multiplex method for linear programming. Sankhya *18*, 329—362.

Frish R. (1958): The multiplex method for linear programming. Memorandum fra Socialøkonomisk Institut, Universitetet Oslo *12*.

Gardner L. A. Jr. (1963): Stochastic approximation and its application to problems of prediction and control synthesis. Internat. Symp. Nonlinear Differential Equations and Nonlinear Mech., Colorado Springs, New York—London, Acad. Press, 241—258.

Gardner L. A. Jr. (1964): Adaptive predictors. PRTR3.

Germansky B. (1934): Notiz über die Lösung von Extemaufgaben mittels Iteration. Z. Ang. Math. Mech. *14*, 187—187.

Гладышев Е. Г. О стохастической аппроксимации. Теория вероятностей и прим. *10*, 297—300.

Glass H. and L. Cooper (1965): Sequential search: a method for solving constrained optimization problems. JACM *12*, 71—82.

Goldstein A. A. (1962): Cauchy's method of minimization. Numer. Math. *4*, 146—150.

Goldstein A. A. and Ward Cheney (1958): A finite algorithm for the solutions of consistent linear equations and inequalities and for the Tchebycheff approximation of inconsistent linear equations. Pacific J. Math. *8*, 415—427.

Griffith R. E. and R. A. Stewart (1961): A nonlinear programming technique for the optimization of continuous processing systems. Management Science *7*, 379—392.

Guttman L. and R. Guttman (1959): An illustration of the use of stochastic approximation methods. Biometrics *15*, 551—559.

Hanš O. and A. Špaček (1960): Random fixed point approximation by differentiable trajectories. PRTR2, 203—213.

Hartley H. O. and R. R. Hocking (1963): Convex programming by tangential approximation. Management Science *9*, 600—612.

Hestens M. R. (1951): Iterative methods for solving linear equations. NAML Report 52—9, Nat. Bureau of Standards, Los Angeles, 19 pp.

Ho Yu Chi (1962): On the stochastic approximation method and optimal filtering theory. Journal of math. analysis and applications *6*, 152—154.

Hodges J. L., Jr. and E. L. Lehmann (1956): Two approximation to the Robbins Monro process. Proc. 3rd Berkeley Symp. on Math. Statistics and Prob., vol. *1*, 95—104.

Hooke R. and T. A. Jeeves (1961): Direct search solution of numerical and statistical problems. JACM *8*, 212—

Hotelling H. (1941): Experimental determination of the maximum of a function. AMS *12*, 20—46.

Householder A. S. and F. L. Bauer (1960): Certain iterative methods for solving linear systems. Numerische Mathematik *2*, 55—59.

Иванов В. В. (1962): Об алгоритмах быстрого спуска. ДАН СССР *143*, 775—778.

John F. (1964): Extremum problems with inequalities as subsidiary conditions. „Studies and Essays", Courant anniversary volume, pp. 187—204.

Kantorovič L. V. (1945): On an effective method of solving extremal problems for quadratic functionals. C. R. de l'Ac. Sci. de l'URSS XLVIII, no. *7*, 455—459.

Канторович Л. В. (1945): Об одном методе решения экстремальных задач для квадратных функционалов. ДАН СССР *48*, 483—487.

Канторович Л. В. (1947): О методе найскорейшего спуска. ДАН СССР *56*, 233—236

Канторович Л. В. (1948): Функциональный анализ и прокладная математика. Усп. мат. наук *3*, 89—185.

Kelley J. E. (1960): The cutting-plane method for solving convex programs. JSIAM *8*, 703—712.

Kerner M. (1933): Die Differentiale in der allegemeinen Analysis. Ann. of Math. *34*, 546—572.

Kerstukos A. and R. I. van Nice (1958): Optimizing process operations. Automation *5*.

Kesten H. (1958): Accelerated stochastic approximation. AMS *29*, 41—59.

Kiefer J. and J. Wolfowitz (1952): Stochastic estimation of the maximum of a regression function. AMS *23*, 462—466.

Kiefer J. (1953): Sequential minimax search for a maximum. Proc. Amer. Math. Soc. *4*, 502—506.

Kiefer J. (1957): Optimum sequential search and approximation methods under minimum regularity assumptions. JSIAM *5*, 105—136.

Killianpur G. (1954): A note on the Robbins-Monro stochastic approximation method. AMS *25*.

Kitagawa T. (1959): Successive processes of statistical controls (2). Met. Coll. Sci. Univ. Kyoto, Ser. A, Math. *13*, 1—16.

Kizner W. (1964): A numerical method for finding solutions of nonlinear equations. JSIAM *12*, 424—428.

Klingman W. R. and D. M. Himmelblau (1964): Nonlinear programming with the aid of a multiplegradient summation technique. JACM *11*, 400—415.

Kose T. (1956): Solutions of saddle value problems by differential equations. Econometrica *24*, 59—70.

Красулина Т. П. (1962): Замечания о некоторых процесах стохастической апроксимации. Теория вероят. к прим. *7*, 113—118.

Krolak P. and L. Cooper (1963): An extension of Fibonaccian search to several variables. CACM *6*, 639—641.

Kuhn H. W. and A. W. Tucker (1950): Nonlinear programming. Proc. 2nd Berkeley Symp. on Mathem. Statist. and Probability, Univ. of Calif. Pres., 481—492.

Lanczos C. (1952): Solution of systems of linear equations by minimized iterations. J. Res. Nat. Bureau of Standards *49*, 33—53.

Levenberg K. (1944): A method for the solution of certain non-linear problems in least squares. Qu. Appl. Math. *2*, 164—168.

Leżański T. (1963): Über das Minimumproblem für Funktionale in Banachschen Räumen. Math. Annalen *152*, 271—274.

Maňas M. (1965): Nelineární programování. Ek. mat. obzor *1*, 3—34.

Mangasarian O. I. (1962): Duality in nonlinear programming. Qu. Appl. Math. *20*, 300—302.

Mangasarian O. I. (1963): Equivalence in nonlinear programming. Naval. Res. Logist. Quart. *10*, 299—306.

Manne A. S. (1953): Concave programming for gasoline blends. Rand. Corp. Paper. No. P—383.

Marquardt D. W. (1963): An algorithm for least-squares estimation of nonlinear parameters. JSIAM *11*, 431—441.

Mises R. von and H. Pollaczek-Geiringer (1929): Praktische Verfahren der Gleichungsauflösung. Z. Angew. Math. Mech. *9*, 58—77.

Nashed M. Z. (1964): The convergence of the method of steepest descent for nonlinear equations with variational or quasi-variational operators. J. Math. Mech. *13*, 765—794.

Nashed M. Z. (1965): On general iterative methods of the solutions of a class of nonlinear operator equations. Math. of Computation *19*, 14—24.

Nelder J. A. and R. Mead (1965): A simplex method for function minimization. Comp. J. *7*, 308—313.

Newman D. J. (1960): Locating the maximum on a unimodal surface. Presented at the 18th National Meeting of Operations Res. Soc., October 1960, neuveřejněno, viz Spang (1962).

Newman D. J. (1965): Location of the maximum on unimodal surfaces. JACM *12*, 395—398.

Pietrzykowski T. (1961): On an iteration method for maximizing a concave function on a convex set. Práce ZAM, A *13*.

Pietrzykowski T. (1962): On a method of approximative finding conditional maximums. Algorytmy *1*, 9—15.

Powell M. J. D. (1964): An efficient method for finding the mininum of a function of several variables without calculating the derivatives. Computer Journal *7*, 155—162.

Powell M. J. D. (1965): A method for minimizing a sum of squares of non-linear functions without calculating derivatives. Comp. J. *7*, 303—307.

Reeves C. M. (1964): Algorithm 238, conjugate gradient method. CACM *7*, 481.

Remez E. (1934): Sur un procédé convergent d'approximations succesives pour determiner les polynomes d'approximation. C. R. Ac. Sci. Paris *198*, 2063—2065.

Robbins H., S. Monro (1951): A stochastic approximation method, AMS *22*, 400—407.

Rosen J. B. (1960): The gradient projection method for nonlinear programming. I. Linear constrainsts. JSIAM *8*, 181—217.

Rosen J. B. (1961): The gradient projection method for nonlinear programming. Part. II. Nonlinear constraints. JSIAM *9*, 514—532.

Rosenbloom P. C. (1956): The method of steepest descent. Proc. of Symp. in Applied Math. vol. 6, pp. 127—176.

Rosenbrock H. H. (1960): An automatic method for finding the largest or least value of a function. Computer Journal *3*, 175—184.

Saaty T. L. and J. Bram (1964): Nonlinear mathematics. McGraw-Hill Book Co., New York.

Sacks J. (1958): Asymptotic distributions of stochastic approximations. AMS *29*, 373—405.

Sakrison D. J. (1962a): Design of filters for non-mean-square error performance criteria by means of a continuous adjustment procedure. Quarterly Progress Report No. 66, 189—201 and No. 67, 119—126, Research Lab. of Electronics.

Sakrison D. J. (1962b): Application of stochastic approximation methods to system optimization. Techn. Rep. 391, July 10, Massachusetts Inst. Technology, Res. Lab. Electronics. Cambridge, Massachusetts.

Sakrison D. J. (1963): Iterative design of optimum filters for non-mean-square error criteria. Trans. of the IEEE Professional Group on Information Theory *9*, 161—167.

Sakrison D. J. (1964): A continuous Kiefer-Wolfowitz procedure for random processes. AMS *35*, 590—599.

Schmetterer L. (1953): Bemerkungen zum Verfahren der stochastischen Iteration. Österreichisches Ingenieur-Archiv VII, 111—117.

Schmetterer L. (1954a): Sur l'approximation stochastique. Bull. Inst. Internat. Statist. *24*. 203 to 206.

Schmetterer L. (1954b): Zum Sequentialverfahren von Robbins und Monro. Monatshefte für Mathematik *58*, 33—37.

Schmetterer L. (1958): Sur l'iteration stochastique. Colloques Internationaux du centre national de la recherch scientifique. *87*, 55—63.

Schmetterer L. (1961): Stochastic approximation. Proc. 4th Berkeley Sympos. Math. Statist. Probab. *1*, 587—609.

Schröder E. (1870): Über unendlich viele Algorithmen zur Auflösung der Gleichungen. Math. Ann. *2*, 317—365.

Shah B. V., R. J. Buehler and O. Kempthorne (1964): Some algorithms for minimizing a function of several variables. JSIAM *12*, 74—92.

Spang H. A. III (1962): A review of minimization techniques for nonlinear functions. SIAM Review *4*, 343—365.

Стаховский Р. И. (1958): Двухканальный автоматический оптимизатор. Автоматика и телемеханика, *14*, 744—756.

Stiefel E. (1952): Über einige Methoden der Relaxationsrechnung. Zeit. Angew. Math. Phys. *3*, 1—33.

Stiefel E. (1955): Relaxation Methoden bester Strategie zur Lösung linearer Gleichungssysteme. Comm. Math. Helv. *29*, 157—159.

Ваинберг М. М. (1960): О сходимости метода найскорейшего спуска для нелинейных уравнений. ДАН СССР, *130*, 9—12.

Vosiková M. (1964): Asymptotické vlastnosti Kiefer-Wolfowitzovy aproximační metody. Dipl. práce, Mat.-fyz. fak. KU.

Wells M. (1965): Algorithm 251, Function minimization [E 4]. CACM *8*, 169—170.

Wetherill G. B. (1963): Sequential estimation of quantal response curves (with discussion). JRS. *B 25*, 1—48.

Witzgall C. (1960): Gradient-projection methods for linear programming. Princeton Un. and IBM Corp. Rep. No. 2,

Wolfe P. (1960): Accelerating the cutting-plane method for nonlinear programming. Rand. Corp. Paper No. P-2010.
Wolfe P. (1961): Accelerating the cutting-plane method for nonlinear programming. JSIAM *9*, 481—488.
Wolfe P. (1962): Recent developments in nonlinear programming. Advance in computers, vol. 3, Acad. Press. New York, 155—187.
Wolfowitz (1952): On the stochastic approximation method of Robbins and Monro. AMS *23*, 457—561.
Wolfowitz (1956): On stochastic approximation methods. AMS *27*, 1151—1155.
Zoutendijk G. (1959): Maximizing a function in a convex region. JRSS *B 21*, 338—355.
Zoutendijk G. (1960): Methods of feasible directions. Elsevier Publ. Co., Amsterdam—London—New York—Princeton.

SUMMARY

A Review of Deterministic and Stochastic Approximation Methods for Function Minimization

VÁCLAV FABIAN

A review of basic results and methods in the field of function minimization. Gradient methods, methods of non-linear programming and stochastic approximations are included. Convergence properties, in particular the rate of convergence, are discussed and open questions are emphasized.

Dr. Václav Fabian, CSc., Ústav teorie informace a automatizace ČSAV, Vyšehradská 49, Praha 2.