

1. Úvod

Rada úloh identifikace, odhadování, rozpoznávání, klasifikace, plánování, filtrace, extrémální regulace vede na úlohu určení extrému funkce při neúplné informaci. K řešení těchto úloh se používají různé rekurentní algoritmy.

V přehledovém článku (1) jsou vylíčeny hlavní etapy rozvoje rekurentních algoritků optimalizace při apriorní neurčitosti a uvedeny způsoby optimalizace samotných algoritků. Je konstatováno, že optimální rekurentní algoritmy jsou velmi citlivé na nesplnění předpokladů, za kterých byly konstruovány, nebo-li nejsou robustní a uvažují se možnosti odstranění této nerobustnosti na základě apriorní informace jak o pozorováních tak i o řešení.

Poznámka. Metodu odhadování, pomocí níž odhad θ_n vypočítáme na základě minulého odhadu θ_{n-1} a nového pozorování Y_n podle vztahu

$$\theta_n = \varphi(n-1, \theta_{n-1}, Y_n), \quad n=1, 2, \dots,$$

kde φ je známá funkce, nazýváme rekurentní metoda a příslušné odhady $\theta_1, \theta_2, \dots, \theta_n$ rekurentní odhady.

Rekurentní algoritmy mají řadu předností: jednoduchost výpočtů, malé požadavky na rozsah paměti počítače, možnost zpracování dat v reálném čase, univerzálnost a rozmanitost.

Na příkladu regresní úlohy ukážeme, jak je možné využít apriorní informaci k zvýšení rychlosti konvergence rekurentních algoritků a učinit je robustními. Vyložený přístup umožní vybrat z různých možných rekurentních algoritků ten, který je ve vztahu k apriorní informaci v určitém smyslu optimální (2).

2. Formulace problému; základní předpoklady.

Uvažujeme obecný lineární regresní model

$$(2.1) \quad y_N = X_N \theta^\circ + u_N, \quad N = 1, 2, \dots,$$

kde

$\theta^\circ = (\theta_1^\circ, \theta_2^\circ, \dots, \theta_p^\circ)^T \in R^p$ je neznámý vektor regresních koeficientů;

$X_N = [x_{ij}]_{i=1, 2, \dots, N}^{j=1, 2, \dots, p}$ je matice typu $(N \times p)$ s řádky $x_i^T = (x_{i1}, x_{i2}, \dots, x_{ip}) \in R^p$, $i=1, 2, \dots, N$;

$u_N = (u_1, u_2, \dots, u_N)^T \in R^N$ je vektor náhodných chyb pozorování, u_i jsou nezávislé stejně rozdělené náhodné veličiny s distribuční funkcí F ;

$y_N = (y_1, y_2, \dots, y_N)^T \in R^N$ je vektor nezávislých pozorování takový, že $y_i, i=1, 2, \dots, N$ má distribuční funkci $F(y - x_i^T \theta)$.

Úloha: Odhadnout na základě naměřených y_i a naměřených nebo daných $x_i, i=1, 2, \dots, N$ neznámý vektor θ° .

Řešení: Je možné použít jak nerekurentní metody (např. metodu maximální věrohodnosti (ML) nebo metodu nejmenších čtverců (LS)), tak metody rekurentní (např. typu stochastických aproximací (SA) nebo rekurentní metodu nejmenších čtverců (RLS)). Ve zmíněných rekurentních metodách se odhad θ_n vypočítá z předcházejícího odhadu θ_{n-1} na základě nové informace (to je veličin y_n, x_n).

Obecný přístup ke konstrukci rekurentních algoritků spočívá v následujícím. Zvolíme nějakou skalární funkci $\varphi(z)$ (ztrátová funkce) a hledáme minimum funkcionálu

$$(2.2) \quad J(\theta) = E \varphi(y - x^T \theta) = \min_{\theta}$$

Za určitých podmínek nabývá $J(\theta)$ svého minima v bodě θ° . K minimalizaci lze použít stochastickou gradientní metodu, která dává odhady ve tvaru

$$(2.3) \quad \theta_n = \theta_{n-1} + P_n x_n \psi(y_n - x_n^T \theta_{n-1}), \quad n=1,2,\dots,$$

kde $\psi = \beta'$, P_n je nějaká matice (koeficient zesílení). Četné konkrétní algoritmy se vzájemně liší způsobem výběru matice P_n a funkce ψ .

Ukážeme, že matici P_n a funkci ψ můžeme vybrat na základě apriorní informace, kterou máme k dispozici, v určitém smyslu optimálním způsobem.

Poznámka. Odhad $\hat{\theta}$ neznámého θ^* určený posloupností θ_n generovanou algoritmem (2.3) nazýváme odhad odpovídající funkcí ψ .

Zabýváme se tedy modelem (2.1) a algoritmem (2.3). Uděláme následující předpoklady.

A1. Distribuční funkce $F(z)$ je symetrická kolem nuly ($F(z) = 1 - F(-z)$).

A2. Vektory $x_i, i=1,2,\dots,N$ jsou nezávislé (mezi sebou i s u_1) náhodně stejně rozdělené. Kovarianční matice $\Sigma = E x x^T$ je konečná a pozitivně definitní ($\Sigma > 0$).

A3. Matice P_n mají tvar $P_n = \gamma_n P$, kde P je symetrická, $P > 0$, skalární činitele $\gamma_n > 0$, $\sum_{n=1}^{\infty} \gamma_n = \infty$, $\sum_{n=1}^{\infty} \gamma_n^2 < \infty$.

A4. Funkce $\psi(z)$, $z \in R^1$ je lichá ($\psi(-z) = -\psi(z)$) a spojitá až na konečný počet bodů.

Označme
$$E_F \psi^2 = \int_{R^1} \psi^2(z) dF(z)$$

a předpokládáme
$$E_F \psi^2 < \infty.$$

A5. Buď A5₁. Funkce ψ je neklesající; existuje z_0 takové, že pro každé $\varepsilon > 0$ platí

$$\begin{aligned} \psi(z_0 + \varepsilon) &> \psi(z_0 - \varepsilon) \\ F(z_0 + \varepsilon) &> F(z_0 - \varepsilon); \end{aligned}$$

nebo A5₂. Existuje ryze unimodální hustota $f(z) = dF(z)/dz$, $f(0) < \infty$ ($f(z_1) < f(z_2)$ pro $|z_1| > |z_2|$).

A6. Buď A6₁. $|\psi(z)| \leq \lambda(1 + |z|)$, λ konstanta, $\int_{R^1} z^2 dF(z) < \infty$, $E \|x_i\|^2 < \infty$;

nebo A6₂. $|\psi(z)| \leq \lambda$ pro všechna $z \in R^1$.

A7. Funkce

$$m(\alpha) = - \int_{R^1} \psi(z-\alpha) dF(z)$$

je derivovatelná v nule a $0 < m'(0) < \infty$.

Nyní jsme připraveni zformulovat základní výsledky.

3. Konvergence a rychlost konvergence.

Níže uvedený výsledek ukazuje univerzálnost rekurentních algoritmů tvaru (2.3). Konvergují s pravděpodobností 1 ke skutečné hodnotě θ^* za dosti obecných předpokladů o náhodných chybách, o funkci ψ a matici P_n .

VĚTA 1. ("konzistence") Za předpokladů A1-A6 odhad θ_n v algoritmu (2.3) konverguje k θ^* skoro jistě.

Důkaz. Opírá se o obecné výsledky teorie stochastických aproximací Robbinsa-Monroova typu(3).

Skutečnost, že algoritmus konverguje, nestačí pro rozhodnutí, který algoritmus použít. Rozhodující je rychlost konvergence. Zde vzniká řada obtíží (4). Za prvé existují různé charakteristiky rychlosti konvergence (je možné odhadovat $E \|\theta_n - \theta^*\|^2$, $E(J(\theta_n) - J(\theta^*))$, $P(\|\theta_n - \theta^*\| \leq \varepsilon_n)$, $P(J(\theta_n) \leq J(\theta^*) + \varepsilon_n)$ atd.). Za druhé, obvykle se podaří získat pouze odhady shora těchto charakteristik. Přitom není jasné, do jaké míry je možné důvěřovat podobným odhadům a vybírat algoritmus na jejich základě. Tyto těžkosti odpadají, jestliže uvažujeme asymptotickou rychlost konvergence. Dokážeme-li, že pro odhad pořízený rekurentním vztahem je $\sqrt{n}(\theta_n - \theta^*)$ asymptoticky normální s nulovým středem a asymptotickou kovarianční maticí K (dále ASKO), pak K není pouze odhadem shora, ale adekvátní charakteristikou chování rekurentního procesu pro velká n . Tato charakteristika je vyčerpávající, neboť jak známo, normální rozdělení je úplně popsáno středem a kovarianční maticí. Jestliže se ukáže, že pro jeden algoritmus je K menší (v maticovém

smyslu) než pro jiný, pak i libovolná rozumná skalární charakteristika přesnosti pro první algoritmus je lepší. To co bylo řešeno, objasňuje, proč další závěry a srovnání algoritmů je založeno na asymptotické rychlosti konvergence.

VĚTA 2. ("asymptotická normalita"). Nechť jsou splněny předpoklady A1-A7, $\gamma_n = 1/n$ a matice $A = -m'(0)P\Sigma + 1/2 I$ je stabilní (I je jednotková matice), pak $\sqrt{n}(\theta_n - \theta^*)$ má asymptoticky p-rozměrné normální rozdělení s nulovým středem a ASKO K , to je

$$\mathcal{L}(\sqrt{n}(\theta_n - \theta^*)) \xrightarrow{n \rightarrow \infty} N_p(0, K),$$

kde K je řešení maticové rovnice

$$(3.1) \quad AK + KA^T = -E_P \psi^2 P \Sigma P.$$

Důkaz: Je založen na obecných výsledcích o asymptotické normalitě odhadů stochastických aproximací Robbins-Monroova typu (3).

Tudíž asymptotická rychlost konvergence algoritmu měřená maticí K , závisí na vlastnostech chyb pozorování, na funkci ψ a na matici P_n .

Ve dvou speciálních případech může být maticová rovnice (3.1) vyřešena explicitně.

i) Nechť $P = \gamma_M \Sigma^{-1}$, to je $P_n = (\gamma_M/n) \Sigma^{-1}$ - maticový koeficient zesílení. Potom pro $2 \cdot m'(0) \gamma_M > 1$ je matice A stabilní a

$$K_M = \frac{\gamma^2 E_P \psi^2}{2 \gamma m'(0) - 1} \Sigma^{-1}.$$

ii) Nechť $P = \gamma_S I$, to je $P_n = (\gamma_S/n) I$ - skalární koeficient zesílení. Jestliže $2 m'(0) \cdot k \cdot \gamma_S > 1$, $k = (\|\Sigma^{-1}\|)^{-1}$, pak A je stabilní a

$$K_S = \frac{\gamma E_P \psi^2}{4 m'(0)} (2I - A^{-1}).$$

Dříve než přistoupíme k otázce výběru algoritmu optimálního ve smyslu největší asymptotické rychlosti konvergence, uvedeme příklady rekurentních algoritmů odpovídajících různým funkcím ψ .

Předpokládejme, že jsou splněny předpoklady A1-A3. Zavedeme označení $\Delta_n = (y_n - x_n^T \theta_{n-1})$.

A) Proporcionální algoritmus: $\psi(z) = z$.

$$\theta_n = \theta_{n-1} + P_n x_n \Delta_n.$$

Jestliže má náhodná chyba omezený rozptyl $E u^2 = \sigma^2$, $E \|x\|^4 < \infty$, pak $\theta_n \xrightarrow{a.s.} \theta^*$. Pro $P_n = (\gamma/n) \Sigma^{-1}$, $\gamma > 1/2$ je $K = \sigma^2 \gamma^2 (2\gamma - 1)^{-1} \Sigma^{-1}$.

B) Znaménkový algoritmus: $\psi(z) = \text{sign}(z)$.

$$\theta_n = \theta_{n-1} + P_n x_n \text{sign}(\Delta_n).$$

Myšlenka znaménkového algoritmu připadá Fabiánovi, konvergenci dokázal Poljak.

C) Releový algoritmus s intervalem necitlivosti:

$$\psi(z) = \begin{cases} 0, & |z| \leq d, \\ \text{sign}(z), & |z| > d. \end{cases}$$

Jestliže d je bodem růstu funkce $F(z)$, pak podle věty 1 $\theta_n \rightarrow \theta^*$ skoro jistě. Jestliže v okolí bodu d existuje $f(z)$ spojitá v d , $f(d) > 0$, pak lze použít větu 2.

D) Proporcionální algoritmus s nasycením:

$$\psi(z) = \begin{cases} z, & |z| \leq d, \\ d \cdot \text{sign}(z), & |z| > d. \end{cases}$$

Jestliže $F(z)$ není konstantní v intervalu $\langle -d, d \rangle$, pak platí věta 1. Je-li $F(z)$ spojitá v d , pak je použitelná věta 2.

E) Proporcionální algoritmus s useknutím:

$$\psi(z) = \begin{cases} z, & |z| \leq d, \\ 0, & |z| > d. \end{cases}$$

Funkce ψ není rostoucí. Za předpokladu A5₂ algoritmus konverguje. Jestliže $f(z)$ je spojitá

v 4, pak platí věta 2.

4. Optimální algoritmus.

Vzhledem k tomu, co bylo řečeno v odstavci 3 zvolíme přístup k optimalitě algoritmu (2.3) založený na asymptotické rychlosti konvergence. Matice K , která je mírou asymptotické rychlosti konvergence algoritmu tvaru (2.3), závisí na distribuční funkci F , na funkci ψ a na matici P_n .

i) Výběr matice P_n .

Lze dokázat, že mezi všemi algoritmy (2.3) s danou funkcí ψ a F má minimální ASKO ten algoritmus, ve kterém

$$P_n = (n \cdot m'(0) \Sigma)^{-1},$$

tudíž volba $\gamma = (m'(0))^{-1}$ vede na algoritmus s ASKO

$$(4.1) \quad K(\psi, P) = \frac{E_P \psi^2}{(m'(0))^2} \Sigma^{-1} = \sigma^2(\psi, P) \Sigma^{-1}.$$

Asymptotická kovarianční matice daná (4.1) je shodná s asymptotickou kovarianční maticí M-odhadů.

ii) Výběr funkce ψ .

Abychom mohli vybrat funkci ψ , která minimalizuje matici $K(\psi, P)$ musíme udělat některé doplňující předpoklady.

A8. Existuje hustota $f(z) = dF(z)/dz$, je absolutně spojitá a má konečnou Fisherovu informaci

$$J(f) = \int_{\mathcal{R}^k} (f'(z)/f(z))^2 f(z) dz < \infty.$$

$$A9. \quad m'(0) = - \int_{\mathcal{R}^k} \psi f'(z) dz.$$

VĚTA 3. ("asymptotická optimalita odhadů"). Nechť jsou splněny A1-A2, existuje absolutně spojitá hustota $f(z) = dF(z)/dz$, přičemž $\psi = -f'/f$ vyhovuje předpokladům A4- A7 a je splněn A9, pak algoritmus

$$(4.3) \quad \theta_n = \theta_{n-1} + (n J(f) \Sigma)^{-1} x_n \psi(\Delta_n),$$

$$\psi(z) = -(\log f(z))'$$

je asymptoticky eficientní.

Důkaz. Závislost $K(\psi, f)$ na ψ je obsažena ve skalárním činiteli $\sigma^2(\psi, f)$ před pozitivně definitní maticí Σ v (4.1). Pomocí Cauchy-Buniakovského nerovnosti dostaneme

$$\sigma^2(\psi, f) = \frac{\int_{\mathcal{R}^k} \psi^2 f dz}{(\int_{\mathcal{R}^k} \psi f' dz)^2} \geq \frac{1}{\int_{\mathcal{R}^k} (f'/f)^2 f dz} = (J(f))^{-1}$$

a rovnost nastane právě tehdy, když $\psi = k \cdot f'/f$, k je libovolná konstanta. Tudíž pro $\psi = -(\log f)'$ dostaneme

$$K(\psi, f) = (J(f) \Sigma)^{-1},$$

což je Rao-Cramerova dolní hranice pro nestranné odhady a tudíž odhad (4.3) je asymptoticky eficientní.

Jinými slovy věta 3 říká, že rekurentní odhad (4.3) je asymptoticky optimální nejen mezi všemi rekurentními odhady, ale mezi všemi nestrannými odhady (tím i nerekurentními).

Speciálně použijeme-li k řešení úlohy odhadu vektoru θ^0 v modelu (2.1) metodu ML, to je

$$\theta_n = \arg \min_{\theta} \sum_{i=1}^n \varphi(y_i - x_i^T \theta), \quad \varphi(z) = -\log f(z),$$

nedostaneme lepší odhady. Algoritmus (4.3) je možné považovat za rekurentní verzi metody ML.

V matici zesílení P_n vystupuje kovarianční matice Σ . Jestliže informace o rozdělení vstupních veličin chybí, pak tato matice není známá. V tom případě můžeme matici Σ nahradit její výběrovou hodnotou $\Sigma_n = n^{-1} \sum_{i=1}^n x_i x_i^T$ a maticový koeficient zesílení můžeme vypočítat

rekurentně

$$P_n = (n J(r) \Sigma_n)^{-1} = (J(r) \sum_{i=1}^n x_i x_i^T)^{-1} = (J(r) \sum_{i=1}^{n-1} x_i x_i^T + J(r) x_n x_n^T)^{-1} =$$

$$P_{n-1} - \frac{P_{n-1} x_n x_n^T P_{n-1}}{(J(r))^{-1} + x_n^T P_{n-1} x_n}$$

Asymptoticky optimální algoritmus má pak tvar

$$(4.4) \quad \theta_n = \theta_{n-1} + P_n x_n \psi(\Delta_n), \quad \psi(z) = -(\log f(z))',$$

$$P_n = P_{n-1} - \frac{P_{n-1} x_n x_n^T P_{n-1}}{(J(r))^{-1} + x_n^T P_{n-1} x_n}$$

Dá se dokázat, že při libovolném počátečním P_0 jsou asymptotické vlastnosti (4.3) a (4.4) stejné.

Pro normální rozdělení chyb přechází algoritmus (4.4) v rekurentní metodu LS.

Asymptoticky optimální rekurentní algoritmus představuje spojení dvou algoritmů: algoritmu odhadu řešení úlohy a algoritmu pro matici zesílení. První závisí na pozorováních, druhý pouze na vstupních veličinách.

5. Robustní optimální rekurentní algoritmus

Abychom mohli použít optimální algoritmus (4.4), musíme znát hustotu pravděpodobnosti rozdělení náhodných chyb. Malé odchylky od předpokladů, za kterých byly optimální algoritmy konstruovány, vedou k velkým odchylkám odhadů od optimálního řešení. Jinými slovy tyto algoritmy nejsou robustní.

Příčinou nerobustnosti asymptoticky optimálních algoritmů je jejich informační neurčitost: vycházíme totiž z toho, že všechny předpoklady, na jejichž základě byl algoritmus zkonstruován, jsou přesné, zatím co mají pouze přibližný charakter. Proto při konstrukci optimálních algoritmů je třeba brát v úvahu apriorní informaci, která je k dispozici (i když většinou obecného charakteru) a tak odstranit informační neurčitost. Apriorní informace o pozorováních zahrnuje apriorní informaci o náhodných chybách pozorování a apriorní informaci o vstupních veličinách.

Apriorní informace může být vyjádřena zadáním nějaké třídy rozdělení, do které patří neznámé hustoty pravděpodobnosti. Typickými třídami je třída všech hustot rozdělení pravděpodobnosti, třída rozdělení s omezeným rozptylem, třída směsí dvou rozdělení, třída finitních rozdělení nebo třída přibližně finitních rozdělení.

Ke konstrukci asymptoticky optimálního robustního algoritmu použijeme Huberův přístup k robustnosti (5).

V souladu s Huberovou myšlenkou budeme předpokládat, že je známa pouze nějaká třída rozdělení náhodných chyb a optimálním rekurentním algoritmem budeme rozumět minimaxově optimální vzhledem ke třídě rozdělení. V tomto smyslu bude zkonstruovaný algoritmus robustní.

Předpokládáme, že je známa třída \mathcal{F} , do níž patří hustota náhodných chyb. Potom pro algoritmus (2.3) s $P_n = (n \cdot m'(0) \Sigma)^{-1}$ s libovolnou funkcí $\psi(z)$ za platnosti předpokladů věty 2 má ASKO tvar

$$K(\psi, r) = \frac{E \psi^2}{(\int \psi r' dz)^2} \Sigma^{-1} = \sigma^2(\psi, r) \Sigma^{-1},$$

které je pro dané ψ minimální.

Vybereme nyní $\psi = \psi_0$ optimálně ve smyslu minimaxu skalárního činitele $\sigma^2(\psi, r)$, to je pro optimální ψ bude platit

$$(5.1) \quad \sup_{f \in \mathcal{F}} \sigma^2(\psi, f) = \inf_{\psi} \sup_{f \in \mathcal{F}} \sigma^2(\psi, f).$$

Řešení této úlohy spočívá v nalezení "nejméně příznivé" hustoty třídy \mathcal{F} .

Definice: Nejméně příznivou hustotou třídy \mathcal{F} nazýváme hustotu f_0 , pro kterou platí

$$J(f_0) \leq J(f) < \infty \quad \text{pro všechna } f \in \mathcal{F}.$$

Odpověď na otázku, která volba ψ_0 vede ke splnění (5.1) dává následující věta.

VĚTA 4. ("minimaxovost odhadu"). Nechť \mathcal{F} je konvexní třída rozdělání majících absolutně spojitou hustotu f s $J(f) < \infty$ a nechť existuje $f_0 \in \mathcal{F}$ pro které $J(f_0) \leq J(f)$ pro všechna $f \in \mathcal{F}$. Nechť $f_0 \in \mathcal{F}_0$

$$\psi_0 = -f_0' / f_0 = -(\log f_0)'$$

splňují předpoklady A1, A2, A4-A7. Potom odhady dané algoritmem

$$(5.2) \quad \theta_n = \theta_{n-1} + (n \cdot m_F'(0) \Sigma)^{-1} x_n \psi_0(\Delta_n),$$

kde

$$m_F'(0) = - \int_{\mathcal{R}^k} \psi_0(z) f'(z) dz$$

jsou asymptoticky normální s ASKO

$$K(\psi_0, f) \leq K(\psi_0, f_0) = (J(f_0) \Sigma)^{-1}.$$

Jinými slovy pro libovolné $f \in \mathcal{F}$ je zaručena kvalita odhadu maticí $K(\psi_0, f_0)$.

Vzhledem k tomu, že pro $f = f_0$ je algoritmus (5.2) asymptoticky eficientní (věta 3), je

(5.2) asymptoticky optimální ve smyslu minimaxu vzhledem ke třídě \mathcal{F} .

6. Použití optimálního robustního rekurentního algoritmu.

Abychom mohli použít algoritmus (5.2) potřebujeme:

A) na základě apriorní informace určit třídu \mathcal{F} a najít $f_0 = \arg \min_{f \in \mathcal{F}} J(f)$

V článku (6) jsou nalezeny nejméně příznivé hustoty pro několik tříd;

B) vzhledem k tomu, že neznáme skutečnou hustotu f , aproximovat $m_F'(0)$ nahrazením f empirickou hustotou

$$m_F'(0) \approx n^{-1} \sum_{i=1}^n \psi_0'(y_i - x_i^T \theta^0) \approx n^{-1} \sum_{i=1}^n \psi_0'(y_i - x_i^T \theta_{i-1}) = n^{-1} \sum_{i=1}^n \psi_0'(\Delta_i);$$

C) v případě, že chybí informace o rozdělání vstupních veličin, nahradit neznámou maticí Σ její výběrovou hodnotou $\Sigma_n = n^{-1} \sum_{i=1}^n x_i x_i^T$

a odpovídající maticí P_n počítat rekurentně

$$P_n = P_{n-1} - \frac{P_{n-1} x_n x_n^T P_{n-1}}{(\psi_0'(\Delta_n))^{-1} + x_n^T P_{n-1} x_n};$$

D) znát počáteční podmínky P_0, θ_0 . Vhodně zvolené počáteční podmínky zvyšují rychlost konvergence. Je vhodné zvolit za počáteční odhad θ_0 v algoritmu (5.3) odhad metodou LS pro malé N_0 a položit $P_0 = (X_N^T X_N)^{-1}$.

Použitím identity

$$P_n x_n = \frac{P_{n-1} x_n}{(\psi_0'(\Delta_n))^{-1} + x_n^T P_{n-1} x_n}$$

s označení

$$v_n = P_{n-1} x_n, \quad r_n = x_n^T v_n,$$

můžeme algoritmus přepsat na tvar

$$(6.1) \quad \begin{aligned} \theta_n &= \theta_{n-1} + \frac{1}{(\psi'_0(\Delta_n))^{-1} + r_n} v_n \psi_0(\Delta_n), \\ P_n &= P_{n-1} - \frac{1}{(\psi'_0(\Delta_n))^{-1} + r_n} v_n v_n^T. \end{aligned}$$

V případě, že $\psi_0(\Delta) = 0$, položíme $\psi'_0(\Delta) = (1/\Delta)\psi_0(\Delta)$.

Příklady optimálních robustních algoritmů

A_R) $\mathcal{F}_1 = \{f: f(z) \text{ spojitá v } 0, f(0) \geq a > 0\}$ - třída libovolných nedegenerovaných rozdělání. Potom f_0 je hustota Laplaceova rozdělání, $\psi_0 = 2a \text{ sign}(z)$. Tudiž znaménkový algoritmus je robustní optimální algoritmus, jestliže chybí informace o rozdělání náhodných chyb pozorování.

B_R) $\mathcal{F}_2 = \{f: \int_{-\infty}^{\infty} z^2 f(z) dz \leq \sigma^2\}$. Pak f_0 je hustota normálního rozdělání $N(0, \sigma^2)$, $\psi_0 = z/\sigma^2$. Proporcionální algoritmus je optimální robustní ve třídě rozdělání s omezeným rozptylem.

C_R) $\mathcal{F}_3 = \{f: f = (1-\epsilon) f_N + \epsilon h, f_N \text{ je hustota } N(0, \sigma^2), h \text{ je libovolná}\}$ - třída ϵ -normálních rozdělání (normálních s ϵ -příměsí). Potom $\psi_0 = z/\sigma^2$ v intervalu $\langle -d, d \rangle$ a $\psi_0 = (d/\sigma^2) \text{ sign}(z)$ vně intervalu $\langle -d, d \rangle$. Jinými slovy proporcionální algoritmus s nasycením je optimální robustní ve třídě přibližně normálních rozdělání. Závislost d na ϵ a σ je uvedena např. v (6).

D_R) $\mathcal{F}_4 = \{f: f = (1-\epsilon) f_R + \epsilon h, f_R = 1/2d \text{ pro } |z| \leq d, f_R = 0 \text{ pro } |z| > d, h \text{ je libovolná hustota}\}$ - třída přibližně rovnoměrných rozdělání. Potom $\psi_0(z) = 0$ pro $|z| \leq d$, $\psi_0(z) = ((1-\epsilon)/(\epsilon d)) \text{ sign}(z)$ pro $|z| > d$; to znamená, že releový algoritmus s intervalem necitlivosti je optimální robustní pro tuto třídu.

7. Doplnující poznámky.

a) Výsledky o rychlosti konvergence a z nich udělané závěry mají asymptotický charakter. Optimální robustní algoritmy pracují dobře i při malých n , jak ukazují zkušenosti s jejich používáním.

b) Uvažovali jsme pouze lineární model. Všechny výsledky platí též pro model

$$y_n = \varphi(x_n)^T \theta^* + u_n, \quad n=1, 2, \dots$$

Pro nelineární model

$$y_n = g(x_n, \theta^*) + u_n, \quad n=1, 2, \dots$$

je možné použít stejný algoritmus jako (2.3)

$$\theta_n = \theta_{n-1} + P_n \nabla_{\theta} g(x_n, \theta_{n-1}) \psi(g(x_n, \theta_{n-1}) - y_n)$$

s doplňujícími předpoklady o funkci g

$$(g(x, \theta) - g(x, \theta^*)) (\nabla_{\theta} g(x, \theta))^T (\theta - \theta^*) \geq 0$$

pro všechna x, θ a také podmínky na růst funkce podle θ .

c) Uvažované algoritmy byly rekurentními modifikacemi odhadů typu

$$(7.1) \quad \theta_n = \arg \min_{\theta} \sum_{i=1}^n \rho(y_i - x_i^T \theta).$$

Speciálním případem (7.1) jsou odhady metodou ML a M-odhady Hubera. Konvergenci těchto odhadů je možné dokázat za slabších předpokladů než jejich rekurentní verze. Avšak z výpočetního hlediska jsou méně vhodné-explicitní tvar odhadu (7.1) dostaneme pouze ve vyjimečných případech a je třeba uchovávat v paměti počítače všechna $y_i, x_i, i=1, 2, \dots, n$.

Literature

- (1) Tsypkin Ya.Z. (1979). Adaptive optimization algorithms under a priori uncertainty. Automation and Remote Control, 6, 94-108.
- (2) Poljak B.T. and Tsypkin Ya.Z. (1980). Adaptive estimation Algorithms. Automation and Remote Control, 3, 71-84.
- (3) Nevel'son M.B. and Has'minskij R.Z. (1972). Stochastic approximation a recursive estimation, Moscow.
- (4) Poljak B.T. and Tsypkin Ya.Z. (1980). Optimal adaptive pseudogradient algorithms. Automation and Remote Control, 8, 74-84.
- (5) Huber P.J. (1964). Robust estimation of a location parameter. Ann. Math. Statist. 35, 73-101.
- (6) Tsypkin Ya.Z. and Poljak B.T. (1977). Robustized maximum likelihood method. In Dynamics of Systems, vol. 12, Gorkij University Press, 22-46.