Dekonwolucja zaburzonego sygnału

Podejście przez wyrocznie

Grzegorz Mika
Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie
g.w.mika@gmail.com

AGH
AGH University of Science and Technology

Streszczenie

Rozważony zostanie problem estymacji nieznanego elementu f na podstawie niebezpośrednich i zaburzonych obserwacji. Zaburzenie modelowane będzie przez pewien zwarty operator liniowy A. Niech Λ będzie pewnym skończonym zbiorem estymatorów liniowych. Celem będzie konstrukcja metody wyboru estymatora z rodziny Λ naśladującego estymator o minimalnym ryzyku w tej klasie. Okaże się, że można to osiągnąć poprzez minimalizację nieobciążonego estymatora ryzyka. W drugiej części zaprezentowany zostanie numeryczny przykład porównujący tę metodą z popularną metodą rozbieżności.

Wstęp

Przypuśćmy, że obserwujemy zjawisko, w którym interesuje nas uzyskanie informacji na temat elementu f należącego do pewnej przestrzeni Hilberta $\mathcal{H}$. Jednak w przeciwieństwie do klasycznych modeli nie obserwujemy sygnału f bezpośrednio z pewnym błędem, a mamy dostęp jedynie do przekształconej wersji z pewną niedokładnością. Zjawisko takie będzie modelowane w następujący sposób

$$Y = Af + \epsilon\xi,$$

gdzie Y jest obserwowane, A jest zwartym operatorem liniowym na $\mathcal{H}$, $\epsilon > 0$ reprezentuje poziom szumu, natomiast ξ jest gaussowskim białym szumem.
W rozważanym modelu, korzystając z reprezentacji singularnej SVD operatora A, możliwe jest zapisanie powyższego problemu w następującej postaci

$$y_n = b_n\theta_n + \epsilon\xi_n,\ n = 1, 2, \ldots,$$

lub równoważnie

$$x_n = \theta_n + \epsilon\sigma_n\xi_n,\ n = 1, 2, \ldots,$$

gdzie $\xi_n \sim \mathcal{N}(0, 1)$ są niezależnymi zmiennymi, θ_n są współczynnikami Fouriera funkcji f w pewnej bazie ortonormalnej związanej z operatorem A, $x_n = y_n/b_n$, przy czym b_n są wartościami singularnymi operatora A, natomiast ciąg $\sigma_n = 1/b_n \to \infty$ związany jest ze złym uwarunkowaniem zadania, które pojawia się za każdym razem, gdy operator A jest zwarty.
W ten sposób sprowadziliśmy zagadnienie do podobnego jak w przypadku regresji nieparametrycznej, czyli znalezienia dobrego estymatora dla współczynników θ_n, jednak tym razem z uwagi na czynniki σ_n sygnał "ginie" z czasem pośród szumu, który zaczyna dominować.

Teoria

Calem odzyskania informacji o sygnale f będziemy posługiwać się estymatorami liniowymi dla jego współczynników θ_n.
Niech $\lambda = (\lambda_1, \lambda_2, \ldots)$ będzie nielosowym ciągiem liczbowym. Estymatorem liniowym współczynników θ_n nazwiemy estymator $\theta(\lambda) = (\hat{\theta}_1, \hat{\theta}_2, \ldots)$, gdzie

$$\hat{\theta}_i = \lambda_i x_i,\ i = 1, 2, \ldots.$$

Ciąg λ nazywać będziemy filtrem lub wagami.
Jakość estymatora $\hat{f}$ elementu f mierzona będzie przy pomocy scałkowanego ryzyka średniokwadratowego

$$\mathcal{R}(\hat{f}, f) = \mathbb{E}_f||f - \hat{f}||^2.$$

W przypadku naszego modelu dostajemy, że

$$\mathcal{R}(\hat{f}, f) = \mathbb{E}_f||f - \hat{f}||^2 = \mathbb{E}_\theta \sum_n \left(\theta_n - \hat{\theta}_n\right)^2 = \mathbb{E}_\theta||\theta - \hat{\theta}||^2$$

Gdy mamy do czynienia z estymatorem liniowym wyrażenie to sprowadza się do postaci

$$\mathbb{E}_\theta||\theta - \hat{\theta}||^2 = \sum_{n=1}^{\infty}(1 - \lambda_n)^2\theta_n^2 + \epsilon^2\sum_{n=1}^{\infty}\sigma_n^2\lambda_n^2.$$

Rozważać będziemy teraz pewien skończony zbiór estymatorów liniowych Λ.
Zaproponowane podejście opiera się na znalezieniu takiego filtra, by minimalizował on powyższe ryzyko. Jednak element taki wymagałby znajomości estymowanego elementu, zatem nie jest on dostępny dla statystyka. Element taki nazywany jest wyrocznią. Zamiast minimalizować zatem bezpośrednio ryzyko będziemy minimalizować jego nieobciążony estymator, a dokładniej nieobciążony estymator wyrażeni $\mathcal{R}(\hat{\theta}, \theta) - \sum_{n=1}^{\infty}\theta_n^2$ postaci

$$U(\lambda, Y) = \sum_{n=1}^{\infty}(\lambda_n^2 - 2\lambda_n)x_n^2 + 2\epsilon^2\sum_{n=1}^{\infty}\lambda_n\sigma_n^2.$$

Oznaczając przez λ^* element minimalizujący $U(\lambda, Y)$ w zbiorze Λ i wprowadzając oznaczenia $\rho(\lambda) = \sup_n \sigma_n^2|\lambda_n| \left[\sum_{k=1}^{\infty}\sigma_k^4\lambda_k^4\right]^{-1/2}$, $\rho = \max_{\lambda\in\Lambda}\rho(\lambda)$, $S = \frac{\max_{\lambda\in\Lambda}\sup_n \sigma_n^2\lambda_n^2}{\min_{\lambda\in\Lambda}\sup_n \sigma_n^2\lambda_n^2}$, $M = \sum_{\lambda\in\Lambda}\exp\left(\frac{-1}{\rho(\lambda)}\right)$, $L_\lambda = \ln(NS) + \rho^2\ln^2(MS)$, gdzie N jest licznością zbioru Λ, otrzymujemy uzasadnienie zaproponowanej metody w postaci następującej nierówności wyroczni dla estymatora $\theta^* = \lambda^*Y$ przy odpowiednich założeniach

$$\mathbb{E}\,\|\theta^* - \theta\|^2 \leqslant (1 + \gamma_1 B^{-1})\min_{\lambda\in\Lambda}\mathcal{R}(\hat{\theta}, \theta) + \gamma_2 B\epsilon^2 L_\Lambda\omega(B^2L_\Lambda),$$

gdzie stałe $B > B_01 > 0, \gamma_1 > 0, \gamma_2 > 0$ niezależną od elementu f i rodziny Λ, a funkcja $\omega(x)$ jest postaci

$$\omega(x) = \max_{\lambda\in\Lambda}\sup_k\left[\sigma_k^2\lambda_k^2\mathbf{1}\left(\sum_{n=1}^{\infty}\sigma_n^2\lambda_n^2 \leqslant x\sup_k \sigma_k^2\lambda_k^2\right)\right],\ x > 0.$$

Ryzyko zaproponowanego estymatora pozwala się zatem porównać z ryzykiem wyroczni.
Co ważne nierówność ta jest nieasymptotyczna, dając kryterium wyboru estymatora dla dowolnej wielkości próby.

Praktyka

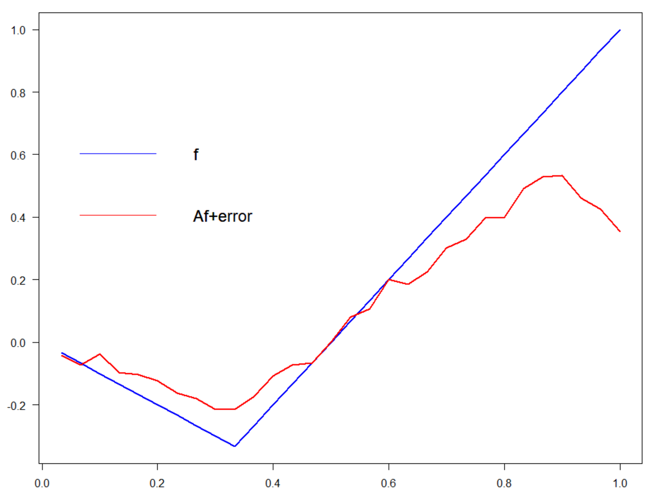
Rozważmy problem dekonwolucji funkcji f na odcinku $[0, 1]$ z zaburzonych obserwacji funkcji postaci

$$(Af)(s) = \int_0^1 \phi(s - t)f(t)dt,$$

gdzie jądro ϕ jest zdefiniowane następująco

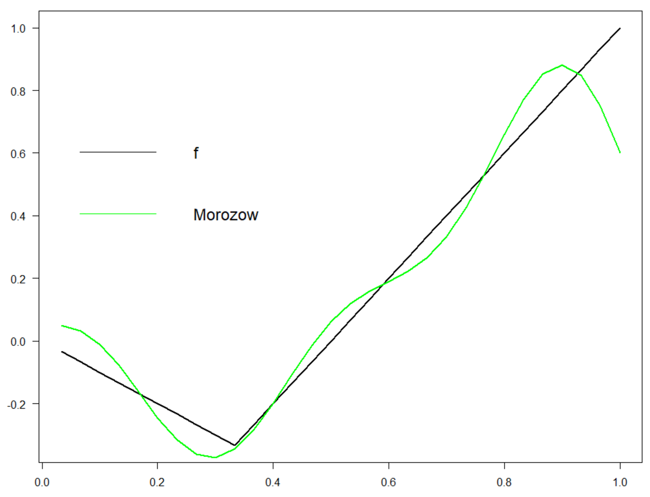
$$\phi(t) = \frac{1}{\sqrt{2\pi a^2}}e^{-\frac{x^2}{2a^2}},\ a \in \mathbb{R}.$$

Za testową funkcję przyjęto funkcję postaci $f(t) = \begin{cases} -t, & t<1/3 \\ 2x-1, & x>1/3 \end{cases}$, natomiast wartość parametru a wybrano na poziomie 0.1.



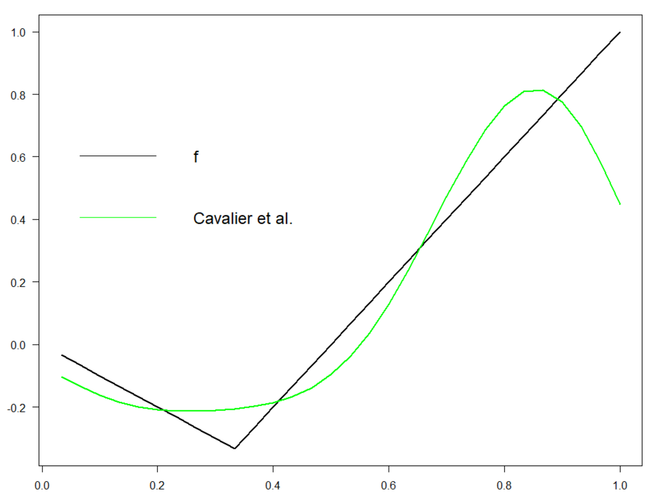
Rysunek 1: Funkcja f przed i po zniekształceniu

Błąd $\epsilon\xi_n$ był generowany jako realizacja niezależnych zmiennych losowych o rozkładzie normalnym ze średnią zero i odchyleniem standardowym w wysokości 5% maksymalnej wartości Af.
Rozważać będziemy teraz skończoną rodzinę estymatorów rzutowych z wagami postaci $\lambda_n = \mathbf{1}\{n \leqslant N\}$, dla których jakości kluczowy jest wybór poziomu obcięcia N.
Metodą, do której porównana zostanie zaproponowana, jest zasada rozbieżności Morozowa. Gdybyśmy znali prawdziwy poziom błędu $\|y - y_0\| = \epsilon$ optymalnym poziomem obcięcia byłby wtedy taki poziom k, że jest on najmniejszą taką liczbą, że zachodzi $||y - Af_k|| < \epsilon$, gdzie f_k jest rozwiązaniem uzyskanym z λ_k. Możemy zatem próbować estymować poziom szumu i na tej podstawie wybierać poziom obcięcia. Zastosowanie tej metody daje następujący rezultat



Rysunek 2: Zasada rozbieżności Morozowa

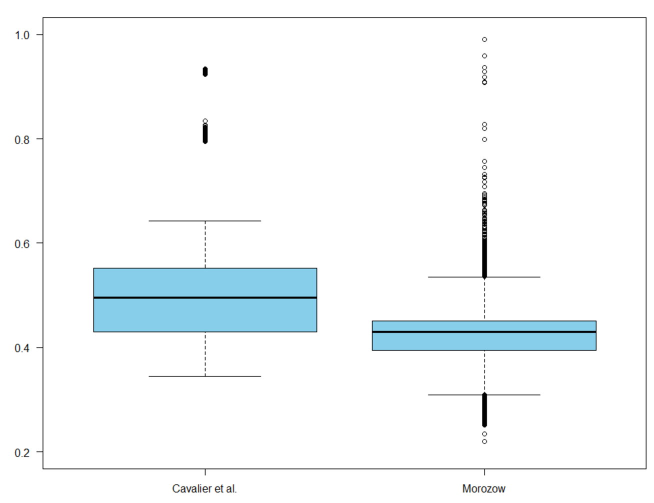
Natomiast w przypadku zastosowania kryterium zaproponowanego przez Cavaliera i in. otrzymano poniższy wynik



Rysunek 3: Cavalier et al.

Wnioski

W rozważanym eksperymencie metoda zaproponowana przez Cavalier i in. dała nieznacznie gorsze wyniki objawiające się wielkością średniego błędu rozwiązania $\mathbb{E}\,\|f - f_k\| \approx 0.58$ w porównaniu do błędu zasady Morozowa na poziomie 0.43 przy 10 000 powtórzeń eksperymentu. Różnica ta jednak jest nieistotna statystycznie. Na poniższym wykresie zaprezentowano wykresy rozkładu błędów dla obu metod



Rysunek 4: Podsumowanie wyników

Literatura

[1] P. Alquier, E. Gautier, G. Stoltz, Inverse Problems and High-Dimensional Estimation Springer-Verlag, 2011, wydanie zbiorowe,

[2] L. Cavalier, G. K. Golubev, D. Picard, A.B. Tsybakov, Oracle inequalities for inverse problems, The Annals of Statistics, Vol. 30, No. 3, 2002, pp. 843–874,

[3] J. Kaipo, E. Somersalo, Statistical and computational inverse problems, Springer, 2004.

Podziękowania

Chciałbym serdecznie podziękować mojemu Promotorowi Prof. Zbigniewowi Szkutnikowi za pomoc i pokierowanie moim rozwojem naukowym.