Grzegorz Mika

SVD i obcięte SVD

1. Wstęp

Niech H, G będą przestrzeniami Hilberta, $A\colon H \to G$ niech będzie liniowym i ograniczonym operatorem między tymi przestrzeniami. Naszym zadaniem jest, mając dany $g \in G$, znaleźć taki $f \in H$, że

$$Af = g.$$

W dalszym ciągu będziemy zakładać, że problem jest źle postawiony i rozwiązanie poprzez odwrócenie operatora A nie jest stabilne.

2. Operatory zwarte

Rozważmy liniowy operator ograniczony A między dwoma przestrzeniami Hilberta H, G. Założymy, że $D(A) = \{f \in H : \exists_{g \in G}\ Af = g\} = H$.

Twierdzenie 1. *Niech $A \in L(H, G)$. Wtedy*
1. $KerA = (RangeA^*)^\perp$ oraz $\overline{RangeA} = (KerA^*)^\perp$,
2. jeśli A jest iniektywny, to A^*A też,
3. $A^*A \in L(H)$ oraz A^*A jest dodatni i samosprzężony.
4. $H = KerA \oplus KerA^\perp = KerA \oplus \overline{RangeA^*}$,
5. $G = \overline{RangeA} \oplus RangeA^\perp = \overline{RangeA} \oplus KerA^*$.

Definicja 1. *Operator $A\colon H \to G$ nazywamy zwartym (compact), jeżeli dla każdego ograniczonego zbioru w H, jego obraz przez operator A jest względnie zwarty w G, czyli jego domknięcie jest zwarte w G. Przez $K(H, G)$ będziemy oznaczać zbiór operatorów zwartych między przestrzeniami H i G.*

Uwaga 1. *Jeżeli $A \in K(H, G)$ oraz $dimH = \infty$ to operator A^{-1} jest nieograniczony.*

Twierdzenie 2. ***(Reprezentacja operatora zwartego według wartości osobliwych)*** *Niech $A\colon H \to G$ będzie operatorem zwartym na przestrzeniach Hilberta H, G. Wtedy istnieją skończony lub zbieżny do zera ciąg liczb dodatnich $\{b_n\}_{n\in I}$ oraz układy ortonormalne $\{v_n\}_{n\in I} \subset H$, $\{u_n\}_{n\in I} \subset G$ takie, że*
— $\overline{KerA^\perp} = \overline{span\{v_n,\ n \in I\}}$,
— $\overline{RangeA} = \overline{span\{u_n,\ n \in I\}}$,
— $Af = \sum_n b_n \langle f, v_n\rangle u_n$ oraz $A^*g = \sum_n b_n \langle g, u_n\rangle v_n$.
Ponadto $g \in RangeA$ wtedy i tylko wtedy, gdy spełniony jest tzw. warunek Picarda

$$\sum_n b_n^{-2} |\langle g, u_n\rangle|^2 < \infty \text{ oraz } g = \sum_n \langle g, u_n\rangle u_n$$

Wtedy rozwiązania równania $Af = g$ mają postać

$$f = f_0 + \sum_n b_n^{-1} \langle g, u_n\rangle v_n$$

przy czym $f_0 \in KerA$ jest dowolne.

Układ (u_n, v_n, b_n) nazywamy układem singularnym operatora A a jego reprezentację w postaci $Af = \sum_n \lambda_n \langle f, v_n\rangle u_n$ nazywamy dekompozycją według wartości osobliwych (singular value decomposition– SVD) operatora A.

Dowód. Dowód twierdzenia opiera się na wykorzystaniu twierdzenia spektralnego do operatora A^*A.

Operator A^*A jest samosprzężony, zwarty i dodatni, a zatem istnieją liczby $b_1^2 \geqslant b_2^2 \geqslant \cdots \geqslant 0$ oraz funkcje ortonormalne v_n takie, że $A^*Av_n = b_n^2 v_n$. Niech $I = \{n \colon b_n > 0\}$ oraz przez u_n oznaczmy znormalizowane obrazy wektorów v_n, czyli $u_n = b_n^{-1}Av_n$ dla $n \in I$. Zauważmy, że $\langle u_k, u_l\rangle = b_k^{-1}b_l^{-1}\langle Av_k, Av_l\rangle = b_k^{-1}b_l^{-1}\langle v_k, A^*Av_l\rangle = b_k^{-1}b_l^{-1}\langle v_k, b_l^2 v_l\rangle = \delta_{kl}$.

Korzystając w wykazanego wcześniej twierdzenia dostajemy, że $KerA^{\perp} = (KerA^*A)^{\perp} = \overline{RangeA^*A} = \overline{span\{v_n,\ n \in I\}}$.

Analogicznie rozpatrując operator AA^* z rozkładem spektralnym $AA^*u_n = b_n^2 u_n$ dostajemy, że $\overline{RangeA} = \overline{span\{u_n,\ n \in I\}}$.

Tożsamości $Af = \sum_n b_n\langle f, v_n\rangle u_n$ oraz $A^*g = \sum_n b_n\langle g, u_n\rangle v_n$ otrzymujemy, zauważając, że $Af = \sum_n \langle Af, u_n\rangle u_n = \sum_n \langle Af, b_n^{-1}Av_n\rangle u_n = \sum_n \langle f, b_n^{-1}A^*Av_n\rangle u_n = \sum_n \langle f, b_n^{-1}b_n^2 v_n\rangle u_n = \sum_n b_n\langle f, v_n\rangle u_n$ oraz drugą analogicznie.

Z nierówności Bessela dostajemy, że $\sum_n |\langle f, v_n\rangle|^2 < \infty$, bo $f \in H$ a stąd $\sum_n |\langle f, v_n\rangle|^2 = \sum_n b_n^{-4}|\langle f, b_n^2 v_n\rangle|^2 = \sum_n b_n^{-4}|\langle f, A^*Av_n\rangle|^2 = \sum_n b_n^{-4}|\langle Af, Av_n\rangle|^2 = \sum_n b_n^{-2}|\langle g, b_n^{-1}Av_n\rangle|^2 = \sum_n b_n^{-2}|\langle g, u_n\rangle|^2 < \infty$. W drugą stronę wnioskujemy, że jeśli spełniony jest warunek Picarda to możemy wypisać jawny wzór na rozwiązanie, gdyż odpowiedni szereg norm współczynników jest zbieżny i g jest sumą swojego szeregu Fouriera..

Ostatecznie możemy wnioskować, że $f = f_0 + \sum_n b_n^{-1}\langle g, u_n\rangle v_n$, gdzie $f_0 \in KerA$. □

Udało nam się zaprezentować działanie zwartego operatora w postaci jego rozwinięcia według wartości osobliwych w postaci $Af = \sum_n b_n\langle f, v_n\rangle u_n$ oraz uzyskać postać szukanych rozwiązań w postaci $f = f_0 + \sum_n b_n^{-1}\langle g, u_n\rangle v_n$. Jednak takie rozwiązanie sytuacji stawia przed nami nowe problemy. Po pierwsze zauważmy, że jeżeli tylko g posiada niezerowe składowe w przestrzeni ortogonalnej do domknięcia obrazu operatora A równanie $Af = g$ nie może być spełnione dokładnie. Niech $P \colon G \to \overline{RangeA}$ będzie rzutem ortogonalnym, czyli $\forall_{g \in G}\ Pg = \sum_n \langle g, u_n\rangle u_n$. Wtedy dla dowolnego elementu $f \in H$ mamy, że $||Af - g||^2 = ||Af - Pg||^2 + ||(1-P)g||^2 \geqslant ||(1-P)g||^2$.

Drugi problem związany jest ze zbieżnością szeregu w warunku Picarda. Z twierdzenia o reprezentacji spektralnej operatora zwartego samosprzężonego wiemy, że liczby $b_n \to 0$ gdy $n \to \infty$ a zatem liczby $b_n^{-2} \to \infty$ gdy $n \to \infty$ a nie mamy żadnej gwarancji, że liczby $\langle g, u_n\rangle$ zbiegają do zera odpowiednio szybko by zrównoważyć ten przyrost szczególnie w przypadku zaburzonej wartości y.

3. Stochastyczny problem odwrotny

Powróćmy do sformułowania problemu z zaburzonymi pomiarami wyrażonymi w języku stochastyki, czyli niech $A \colon H \to G$ będzie zwartym operatorem między dwoma przestrzeniami Hilberta. Obserwując pewną zaburzoną informację Y naszym zadaniem jest poznać $f \in H$ według modelu

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

gdzie ϵ oznacza wielkość szumu.

Podamy teraz założenia jakie będzie musiał spełniać losowy szum.

Definicja 2. *Stochastycznym błędem ξ nazwiemy proces na przestrzeni Hilberta, czyli ograniczony liniowy operator $\xi \colon G \to L^2(\Omega, \mathcal{F}, \mathbb{P})$ taki, że dla dowolnych elementów $g_1, g_2 \in G$ mamy zdefiniowane zmienne losowe $\langle \xi, g_i\rangle$ takie, że $\mathbb{E}\langle \xi, g_i\rangle = 0$ oraz możemy zdefiniować kowariancję Cov_ξ jako ograniczony liniowy operator ($||Cov_\xi|| \leqslant 1$) z przestrzeni G w przestrzeń G taki, że $\langle Cov_\xi g_1, g_2\rangle = Cov(\langle \xi, g_1\rangle, \langle \xi, g_2\rangle)$.*

Często wykorzystywanym modelem będącym idealizacją pewnych innych modeli jest model białego szumu.

Definicja 3. *Powiemy, że losowy błąd ξ jest białym szumem, jeśli $Cov_\xi = I$ oraz indukowane zmienne losowe są gaussowskie, czyli dla dowolnych elementów $g_1, g_2 \in G$ mamy, że $\langle \xi, g_i \rangle \sim \mathcal{N}(0, ||g_i||^2)$, $(\langle \xi, g_1 \rangle, \langle \xi, g_2 \rangle, \ldots, \langle \xi, g_k \rangle) \sim \mathcal{N}_k(0, \Sigma)$ oraz $Cov(\langle \xi, g_1 \rangle, \langle \xi, g_2 \rangle) = \langle g_1, g_2 \rangle$.*

Lemat 1. *Niech ξ będzie białym szumem w przestrzeni G oraz niech $\{u_n\}$ będzie ortonormalną bazą tej przestrzeni. Oznaczając $\xi_k = \langle \xi, u_k \rangle$ dostajemy, że $\{\xi_n\}$ niezależnymi zmiennymi losowymi o tym samym standardowym rozkładzie gaussowskim.*

Zauważmy, że gdy ξ jest białym szumem, Y nie jest elementem przestrzeni G a staje się operatorem działającym na przestrzeni G w następujący sposób

$$\forall_{g \in G} \ \langle Y, g \rangle = \langle Af, u_n \rangle + \epsilon \langle \xi, g \rangle$$

gdzie $\langle \xi, g \rangle \sim \mathcal{N}(0, ||g||^2)$.

Rozważmy teraz układ singularny (u_n, v_n, b_n) operatora zwartego A oraz niech ξ będzie białym szumem. Możemy wtedy zapisać rozpatrując projekcję Y na układ $\{u_n\}$, że

$$\langle Y, u_n \rangle = \langle Af, u_n \rangle + \epsilon \langle \xi, u_n \rangle = \langle Af, b_n^{-1} A v_n \rangle + \epsilon \xi_n = b_n^{-1} \langle A^* A f, v_n \rangle + \epsilon \xi_n =$$

$$b_n^{-1} \langle \sum_k b_k^2 \langle f, v_k \rangle v_k, v_n \rangle + \epsilon \xi_n = b_n \theta_n + \epsilon \xi_n$$

gdzie $\theta_n = \langle f, v_n \rangle$ są współczynnikami w rozwinięciu Fouriera funkcji f w bazie $\{v_n\}$.

Oznaczając przez $y_n = \langle Y, u_n \rangle$ możemy wyjściowy problem $Y = Af + \epsilon \xi$ zapisać w równoważnej postaci sequence space model jako

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

W tej postaci widać dokładnie trudności związane ze stochastycznymi problemami odwrotnymi. Jako że b_n są wartościami osobliwymi operatora zwartego mamy, że $b_n \to 0$ gdy $n \to \infty$, czyli widać, że wraz ze wzrostem n sygnał $b_n \theta_n$ staje się coraz słabszy i coraz trudniej estymować θ_n. Dodatkową trudnością jest fakt, że naszym celem jest estymacja współczynników θ_n a nie współczynników $b_n \theta_n$, dlatego możemy zapisać równoważną postać problemu

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

gdzie $x_n = y_n / b_n$ oraz $\sigma_n = b_n^{-1}$, czyli $\sigma_n \to \infty$ gdy $n \to \infty$. Widzimy zatem, że wraz ze wzrostem n szum zaczyna dominować nad sygnałem czyniąc estymację θ_n trudną.

4. Obcięte SVD

Na początek rozważmy problem estymacji w nieparametrycznym modelu regresji

$$y_n = f(x_n) + \sigma \epsilon_n.$$

Naszym celem jest znalezienie funkcji f. W tym celu możemy posłużyć się metodą rzutowania na pewną bazę. Funkcję f możemy wtedy zapisać w postaci szeregu $f = \sum_{n=1}^{\infty} a_n \phi_n$ i wtedy zadanie estymacji sprowadzi się do znalezienia współczynników rozwinięcia a_n. Możemy zastosować następującą metodę: pierwsze N współczynników oszacujemy na podstawie posiadanych danych natomiast pozostałe współczynniki oszacujemy przez 0. Metoda ta znajduje swoje uzasadnienie w tym, że w przypadku gładkich funkcji f o jej kształcie decydują początkowe współczynniki, natomiast pozostałe stają się zaniedbywalne.

Podobną metodologię możemy spróbować zastosować w przypadku stochastycznych problemów odwrotnych z operatorami zwartymi posiadającymi dekompozycję według wartości osobliwych.

Rozważmy problem w postaci

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

Wtedy możemy zaproponować następujący estymator dla współczynników θ_n

$$\hat{\theta}(N) = \begin{cases} x_k, & k \leqslant N \\ 0, & k > N \end{cases}.$$

Wtedy estymatorem elementu f staje się $\hat{f} = \sum_{k=1}^{N} x_k v_k$.

Problemem jaki pozostał jest dobranie odpowiedniego poziomu obcięcia k.

W celu oceny jakości estymatora posłużymy się błędem średniokwadratowym, czyli $R(f, \hat{f}) = \mathbb{E}(||f - \hat{f}||^2)$. Mając do dyspozycji układ wartości osobliwych możemy zapisać estymator uzyskany metodą TSVD w postaci $\hat{f} = \sum_{n=1}^{k} x_n v_n$. Dzięki temu możemy zauważyć, że jest to estymator liniowy z wektorem wag posiadających jedynki na pierwszych k pozycjach i zerach na pozostałych. Ryzyko estymatora możemy wtedy zapisać jako $R(f, \hat{f}) = \mathbb{E}(||f - \hat{f}||^2) = \mathbb{E}(\sum_n (\hat{f}_k - f_k)^2) = \sum_{n=k+1}^{\infty} \theta_n^2 + \sum_{n=1}^{k} \epsilon^2\sigma_n^2$, czyli ryzyko estymatora TSVD wyraża się w bardzo prosty sposób.

Możemy teraz zastanowić się jak wybór k będzie wpływał na ryzyko estymatora TSVD

$$R(f, \hat{f}) = \sum_{n=k+1}^{\infty} \theta_n^2 + \sum_{n=1}^{k} \epsilon\sigma_n.$$

Z uzyskanego wzoru widzimy, że wraz ze wzrostem k zmniejsza się się obciążenie estymatora (ubywa pominiętych współrzędnych), ale rośnie wariancja, odwrotny skutek obserwujemy zmniejszając k- rośnie obciążenie, ale maleje wariancja. Optymalny wybór k powinien prowadzić do zbalansowania tych dwóch przeciwstawnych tendencji. Ogólnie wiadomo jednak, że wybór optymalnego poziomu odcięcia wymaga znajomości pewnych parametrów poszukiwanej funkcji (gładkość).

Literatura

[1] L. Cavalier, *Inverse Problems in Ststistics* in P. Alquier et al., *Inverse problems and high- dimensional estimation*, Springer, 2011,

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

[3] Z. Szkutnik, *Statystyczne problemy odwrotne*, notatki do wykładu,

[4] L. Wasserman, *All of nonparmetric statistics*, Springer, 2006.