Wstęp do metod numerycznych

Singular Value Decomposition

Metody iteracyjne

P. F. Góra

http://th-www.if.uj.edu.pl/zfs/gora/

2015

Singular Value Decomposition

Twierdzenie 1. *Dla każdej macierzy* $\mathbf{A} \in \mathbb{R}^{M \times N}$, $M \geqslant N$, *istnieje rozkład*

$$\mathbf{A} = \mathbf{U}\,[\mathrm{diag}(w_i)]\,\mathbf{V}^T, \tag{1}$$

gdzie $\mathbf{U} \in \mathbb{R}^{M \times N}$ *jest macierzą kolumnowo ortogonalną,* $\mathbf{V} \in \mathbb{R}^{N \times N}$ *jest macierzą ortogonalną oraz* $w_i \in \mathbb{R}$, $i = 1, \ldots, N$. *Rozkład ten nazywamy* rozkładem względem wartości osobliwych *(Singular Value Decomposition, SVD). Jeżeli* $M = N$, *macierz* $\mathbf{U}$ *jest macierzą ortogonalną.*

Jądro i zasięg operatora

Niech $\mathbf{A} \in \mathbb{R}^{M \times N}$. *Jądrem operatora* $\mathbf{A}$ nazywam

$$\text{Ker}\, \mathbf{A} = \{\mathbf{x} \in \mathbb{R}^N : \mathbf{A}\mathbf{x} = 0\}\,. \tag{2}$$

Zasięgiem operatora $\mathbf{A}$ nazywam

$$\text{Range}\, \mathbf{A} = \{\mathbf{y} \in \mathbb{R}^M : \exists \mathbf{x} \in \mathbb{R}^N : \mathbf{A}\mathbf{x} = \mathbf{y}\}\,. \tag{3}$$

Jądro i zasięg operatora są przestrzeniami liniowymi. Jeśli $M = N < \infty$, $\dim(\text{Ker}\, \mathbf{A}) + \dim(\text{Range}\, \mathbf{A}) = N$.

Sens *SVD*

Sens *SVD* najlepiej widać w przypadku, w którym co najmniej jedna z wartości $w_i = 0$. Dla ustalenia uwagi przyjmijmy $w_1 = 0$, $w_{i\neq 1} \neq 0$.
Po pierwsze, co to jest $\mathbf{z} = [z_1, z_2, \ldots, z_n]^T = \mathbf{V}^T\mathbf{x}$? Ponieważ $\mathbf{V}$ jest macierzą ortogonalną, $\mathbf{z}$ jest rozkładem wektora $\mathbf{x}$ w bazie kolumn macierzy $\mathbf{V}$. Korzystając z (1), dostajemy

$$\mathbf{A}\mathbf{x} = \mathbf{U}\left[\mathrm{diag}(w_i)\right]\mathbf{V}^T\mathbf{x} = \mathbf{U}\left[\mathrm{diag}(0, w_2, \ldots, w_N)\right]\mathbf{z} = \mathbf{U}\begin{bmatrix} 0 \\ w_2 z_2 \\ \vdots \\ w_N z_N \end{bmatrix}. \qquad (4)$$

Wynikiem ostatniego mnożenia będzie pewien wektor z przestrzeni $\mathbb{R}^M$. Ponieważ pierwszym elementem wektora $[0, w_2 z_2, \ldots, w_N z_N]^T$ jest zero, wynik ten nie zależy od pierwszej kolumny macierzy $\mathbf{U}$. Widzimy zatem, że *kolumny macierzy $\mathbf{U}$, odpowiadające niezerowym współczynnikom w_i, stanowią bazę w zasięgu operatora $\mathbf{A}$*.
Co by zaś się stało, gdyby $\mathbf{x}$ był równoległy do wektora stanowiącego pierwszą kolumnę $\mathbf{V}$? Wówczas $\mathbf{z} = 0$, a wobec tego $\mathbf{A}\mathbf{x} = 0$. Ostatecznie więc widzimy, że *kolumny macierzy $\mathbf{V}$, odpowiadające zerowym współczynnikom w_i, stanowią bazę w jądrze operatora $\mathbf{A}$.*

SVD i odwrotność macierzy

Niech $\mathbf{A} \in \mathbb{R}^{N\times N}$. Zauważmy, że $|\det \mathbf{A}| = \prod_{i=1}^{N} w_i$, a zatem $\det \mathbf{A} = 0$ wtedy i tylko wtedy, gdy co najmniej jeden $w_i = 0$. Niech $\det \mathbf{A} \neq 0$. Wówczas równanie $\mathbf{A}\mathbf{x} = \mathbf{b}$ ma rozwiązanie postaci

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \mathbf{V}\left[\mathrm{diag}(w_i^{-1})\right]\mathbf{U}^T\mathbf{b}\,. \tag{5}$$

Niech teraz $\det \mathbf{A} = 0$. Równanie $\mathbf{A}\mathbf{x} = \mathbf{b}$ *także* ma rozwiązanie, o ile tylko $\mathbf{b} \in$ Range $\mathbf{A}$. Rozwiązanie to ma postać $\mathbf{x} = \tilde{\mathbf{A}}^{-1}\mathbf{b}$, gdzie

$$\tilde{\mathbf{A}}^{-1} = \mathbf{V}\left[\mathrm{diag}(\tilde{w}_i^{-1})\right]\mathbf{U}^T\,. \tag{6a}$$

gdzie

$$\tilde{w}_i^{-1} = \begin{cases} w_i^{-1} & \text{gdy } w_i \neq 0\,, \\ 0 & \text{gdy } w_i = 0\,. \end{cases} \tag{6b}$$

SVD i macierze osobliwe

Wróćmy jeszcze raz do problemu osobliwych (z zerowym wyznacznikiem głównym) układów równań, wspomnianego już na stronie 5. Jeżeli $\det \mathbf{A} = 0$, układ równań z całą pewnością nie ma *jednoznacznego* rozwiązania. Może jednak mieć rozwiązanie (a nawet nieskończenie wiele rozwiązań), jeżeli prawa strona *należy do zasięgu* $\mathbf{A}$. Jest to równoważne warunkowi, że wszystkie wyznaczniki poboczne we wzorach Cramera zerują się. Wówczas **rozwiązaniem** układu równań jest każdy wektor postaci

$$\mathbf{x} = \tilde{\mathbf{A}}^{-1}\mathbf{b} + \mathbf{x}_0 \,, \tag{7}$$

gdzie $\tilde{\mathbf{A}}^{-1}$ jest pseudoodwrotnością daną przez (6), zaś $\mathbf{x}_0 \in \mathrm{Ker}\mathbf{A}$ jest dowolnym wektorem należącym do jądra. Rozwiązanie z $\mathbf{x}_0 = 0$ ma spośród nich najmniejszą normę. Zauważmy, że na wektory należące do zasięgu, pseudoodwrotność działa jak zwykła odwrotność macierzy.

Jeżeli $\mathbf{b}$ *nie* należy do zasięgu, wyrażenie (7) z $\mathbf{x}_0 = 0$ daje rozwiązanie przybliżone i najlepsze w sensie najmniejszych kwadratów, co niekiedy jest bardzo użyteczne.

SVD i współczynnik uwarunkowania

Twierdzenie 2. *Jeżeli macierz* $\mathbf{A} \in \mathbb{R}^{N\times N}$ *posiada rozkład* (1) *oraz* $\det \mathbf{A} \neq 0$*, jej współczynnik uwarunkowania spełnia*

$$\kappa = \frac{\max\limits_i |w_i|}{\min\limits_i |w_i|}. \tag{8}$$

Jeśli macierz jest źle uwarunkowana, ale *formalnie* odwracalna, numeryczne rozwiązanie równania $\mathbf{Ax} = \mathbf{b}$ może być zdominowane przez wzmocniony błąd zaokrąglenia. Aby tego uniknąć, często zamiast (bezużytecznego!) rozwiązania dokładnego (5), używa się *przybliżonego* (i użytecznego!) rozwiązania w postaci (6) z następującą modyfikacją

$$\tilde{w}_i^{-1} = \begin{cases} w_i^{-1} & \text{gdy } |w_i| > \tau\,, \\ 0 & \text{gdy } |w_i| \leqslant \tau\,, \end{cases} \tag{9}$$

gdzie τ jest pewną zadaną tolerancją.

Przykład

Mamy rozwiązać następujące dwa układy równań:

$$\begin{bmatrix} 0.666666667 & -0.166666666 & -0.333333333 \\ -0.166666666 & 0.166666667 & -0.166666666 \\ -0.333333333 & -0.166666666 & 0.666666667 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1.284457048 \\ -0.577350273 \\ -0.129756514 \end{bmatrix} \quad (10a)$$

$$\begin{bmatrix} 0.666666667 & -0.166666666 & -0.333333333 \\ -0.166666666 & 0.166666667 & -0.166666666 \\ -0.333333333 & -0.166666666 & 0.666666667 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1.284457052 \\ -0.577350265 \\ -0.129756510 \end{bmatrix} \quad (10b)$$

Równania te różnią się tylko prawymi stronami, przy czym norma różnicy prawych stron jest rzędu 10^{-8}. Błąd takich rozmiarów łatwo może pojawić się w wyniku jakichś poprzednich obliczeń lub na skutek niepewności danych "zewnętrznych", z którymi pracujemy.

Macierz w układach równań (10) jest symetryczna i dodatnio określona, a jej czynnik Cholesky'ego wynosi

$$\begin{bmatrix} 0.816496581 & & \\ -0.204124145 & 0.353553392 & \\ -0.408248290 & -0.707106777 & 0.000077460 \end{bmatrix} \tag{11}$$

Rozwiązania równań (10) za pomocą faktoryzacji Cholesky'ego mają postać

$$\mathbf{x}_a = \begin{bmatrix} -0.179434106 \\ -5.237183389 \\ -1.593647668 \end{bmatrix}, \qquad \mathbf{x}_b = \begin{bmatrix} 3.903048668 \\ 2.927782158 \\ 2.488835105 \end{bmatrix}. \tag{12}$$

Różnica rozwiązań jest, co do normy, rzędu 10, czyli jest rzędu 10^9 razy większa, niż różnica prawych stron.

Faktoryzacja *SVD* macierzy z układów (10) pokazuje, że wartości szczególne tej macierzy są w przybliżeniu równe $1, \frac{1}{2}, 10^{-9}$. Jeśli do rozwiązania układów równań (10) zastosować pseudoodwrotność (9) ($\tau = 10^{-8}$), *w obu wypadkach* otrzymamy

$$\mathbf{x} = \begin{bmatrix} 1.861807320 \\ -1.154700538 \\ 0.447593757 \end{bmatrix}. \tag{13}$$

(13) jest jedynie *przybliżonym* rozwiązaniem równań (10). Jest ono jednak bardziej użyteczne, niż "ścisłe" rozwiązania (12). Te dwa ostatnie najwyraźniej są zdominowane przez błąd, jaki wystąpił wzdłuż kierunku odpowiadającego najmniejszej wartości szczególnej macierzy. Nie wiemy — i nie mamy sposobu, aby to stwierdzić — które z dwu rozwiązań (12) jest "poprawne". Przybliżone rozwiązanie (13) po prostu ignoruje wpływ tego kierunku, a więc i zaburzeń wzdłuż niego występujących.

Nadokreślone układy równań

Niech $\mathbf{A} \in \mathbb{R}^{M \times N}$, $M > N$, $\mathbf{b} \in \mathbb{R}^M$, $\mathbf{x} \in \mathbf{R}^N$. Wówczas układ równań

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{14}$$

ma więcej równań, niż niewiadomych. Układ taki, zwany nadokreślonym, w ogólności nie ma rozwiązań. Za pomocą SVD można jednak znaleźć jego rozwiązanie przybliżone. Mianowicie

$$\left\| \mathbf{A}\left(\tilde{\mathbf{A}}^{-1}\mathbf{b}\right) - \mathbf{b} \right\|_2 = \text{minimum}\,, \tag{15}$$

gdzie $\tilde{\mathbf{A}}^{-1}$ jest pseudoodwrotnością (6). Widzimy, że $\tilde{\mathbf{A}}^{-1}\mathbf{b}$ jest przybliżonym, najlepszym w sensie najmniejszych kwadratów rozwiązaniem układu (14). Metoda ta jest *powszechnie* używana w liniowym zagadnieniu najmniejszych kwadratów.

Metody iteracyjne

Rozwiązanie układu równań liniowych, uzyskane za pomocą którejś z dotąd poznanych metod, byłoby dokładne (ścisłe), gdyby nie błędy zaokrąglenia (które, dodajmy, dla układów źle uwarunkowanych mogą być *znaczne*). Dlatego metody te nazywa się *metodami dokładnymi*.

W metodach iteracyjnych rozwiązanie dokładne otrzymuje się, teoretycznie, w granicy nieskończenie wielu kroków — w praktyce liczymy na to, że po skończonej (i niewielkiej) liczbie kroków zbliżymy się do wyniku ścisłego w granicach błędu zaokrąglenia.

Rozpatrzmy układ równań:

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \quad (16a)$$
$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2 \quad (16b)$$
$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3 \quad (16c)$$

Przepiszmy ten układ w postaci

$$x_1 = (b_1 - a_{12}x_2 - a_{13}x_3)/a_{11} \quad (17a)$$
$$x_2 = (b_2 - a_{21}x_1 - a_{23}x_3)/a_{22} \quad (17b)$$
$$x_3 = (b_3 - a_{31}x_1 - a_{32}x_2)/a_{33} \quad (17c)$$

Gdyby po prawej stronie (17) były “stare” elementy x_j, a po lewej “nowe”, dostalibyśmy metodę iteracyjną

$$x_i^{(k+1)} = \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k)} - \sum_{j=i+1}^{N} a_{ij} x_j^{(k)} \right) \Big/ a_{ii} \qquad (18)$$

Górny indeks $x^{(k)}$ oznacza, ze jest to przybliżenie w k-tym kroku. Jest to tak zwana *metoda Jacobiego*.

Zauważmy, że w metodzie (18) nie wykorzystuje się najnowszych przybliżeń: Powiedzmy, obliczając $x_2^{(k+1)}$ korzystamy z $x_1^{(k)}$, mimo iż znane jest już wówczas $x_1^{(k+1)}$. **Za to metodę tę łatwo można zrównoleglić.** Sugeruje to następujące ulepszenie:

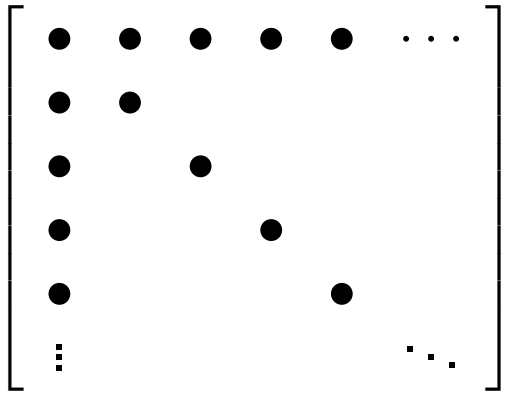
$$x_i^{(k+1)} = \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^{N} a_{ij} x_j^{(k)} \right) \Big/ a_{ii} \qquad (19)$$

Jest to tak zwana metoda Gaussa-Seidela.

Jeżeli macierz $\mathbf{A} = \{a_{ij}\}$ jest rzadka, obie te metody iteracyjne będą efektywne *tylko i wyłącznie* wówczas, gdy we wzorach (18), (19) uwzględni się ich strukturę, to jest uniknie redundantnych mnożeń przez zera.

Powtórzmy: Dla numerycznej efektywności metod iteracyjnych jest **niesłychanie** ważne, aby metodę zaprogramować w ten sposób, aby uwzględniać strukturę macierzy rzadkiej.

Przykład: Niech macierz $\mathbf{A} \in \mathbb{R}^{N \times N}$ ma strukturę



(20)

Metoda Gaussa-Seidela dla macierzy o strukturze (20) ma postać

$$
\begin{aligned}
x_1^{(k+1)} &= \left(b_1 - \sum_{j=2}^{N} a_{1j} x_j^{(k)} \right) / a_{11} \\
x_2^{(k+1)} &= \left(b_2 - a_{21} x_1^{(k+1)} \right) / a_{22} \\
x_3^{(k+1)} &= \left(b_3 - a_{31} x_1^{(k+1)} \right) / a_{33}
\end{aligned}
\tag{21}
$$

....

$$
x_N^{(k+1)} = \left(b_N - a_{N1} x_1^{(k+1)} \right) / a_{NN}
$$

Widać, że jedek krok (*sweep*) algorytmu (21) odbywa się w czasie proporcjonalnym do N.

Trochę teorii

Metody Jacobiego i Gaussa-Seidela należą do ogólnej kategorii

$$\mathbf{M}\mathbf{x}^{(k+1)} = \mathbf{N}\mathbf{x}^{(k)} + \mathbf{b} \tag{22}$$

gdzie $\mathbf{A} = \mathbf{M} - \mathbf{N}$ jest *podziałem* (*splitting*) macierzy. Dla metody Jacobiego $\mathbf{M} = \mathbf{D}$ (część diagonalna), $\mathbf{N} = -(\mathbf{L} + \mathbf{U})$ (części pod- i ponaddiagonalne, bez przekątnej). Dla metody Gaussa-Seidela $\mathbf{M} = \mathbf{D} + \mathbf{L}$, $\mathbf{N} = -\mathbf{U}$. Rozwiązanie równania $\mathbf{A}\mathbf{x} = \mathbf{b}$ jest punktem stałym iteracji (22).

Definicja *Promieniem spektralnym* (diagonalizowalnej) macierzy $\mathbf{G}$ nazywam

$$\rho(\mathbf{G}) = \max\{|\lambda| : \exists \mathbf{y} \neq 0 : \mathbf{G}\mathbf{y} = \lambda\mathbf{y}\} \tag{23}$$

Twierdzenie 3. *Iteracja* (22) *jest zbieżna jeśli* $\det \mathbf{M} \neq 0$ *oraz* $\rho(\mathbf{M}^{-1}\mathbf{N}) < 1$.

Dowód. Przy tych założeniach iteracja (22) jest odwzorowaniem zwężającym. □

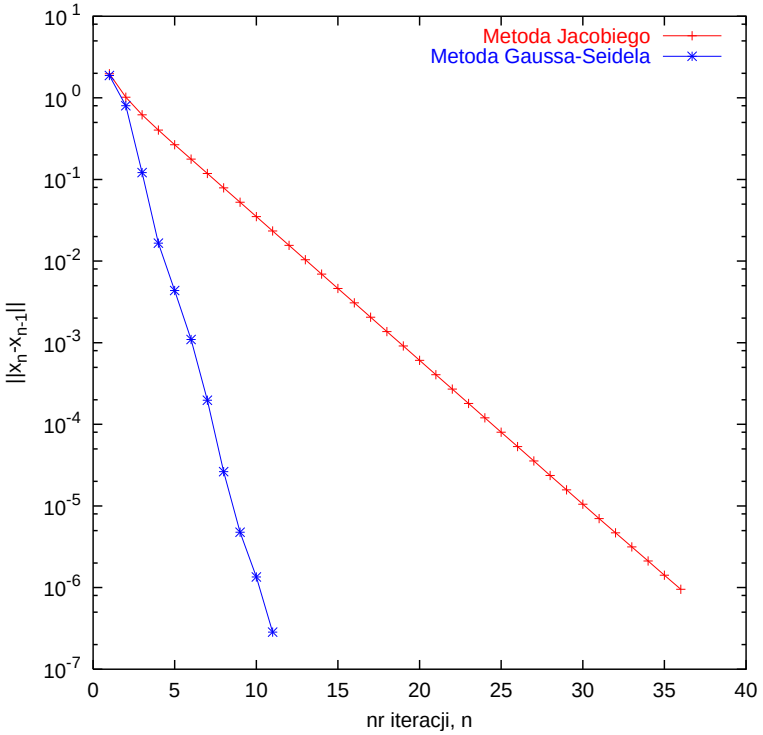
Twierdzenie 4. *Metoda Jacobiego jest zbieżna jeśli macierz* $\mathbf{A}$ *jest silnie diagonalnie dominująca.*

Twierdzenie 5. *Metoda Gaussa-Seidela jest zbieżna jeśli macierz* $\mathbf{A}$ *jest symetryczna i dodatnio określona.*

Przykład

Rozwiązujemy układ równań:

$$\begin{aligned} 3x &+ y &+ z &= 1 \\ x &+ 3y &+ z &= 1 \\ x &+ y &+ 3z &= 1 \end{aligned}$$



Inny przykład

Dla macierzy o wymiarach 128×128

$$\begin{bmatrix} 128 & 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & & & & \\ 1 & & 2 & & & \\ 1 & & & 2 & & \\ \vdots & & & & \ddots & \\ 1 & & & & & 2 \end{bmatrix} \tag{24}$$

(niezaznaczone elementy są zerami)

zbieżność z dokładnością do 10^{-12} w metodzie Gaussa-Seidela, według algorytmu (21), uzyskuje się w ~ 42 iteracjach.

SOR

Jeśli $\rho(\mathbf{M}^{-1}\mathbf{N})$ w metodzie Gaussa-Seidela jest bliskie jedności, zbieżność metody jest bardzo wolna. Można próbować ją poprawić:

$$x_i^{(k+1)} = w\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{N} a_{ij}x_j^{(k)}\right) \Big/ a_{ii} + (1-w)x_i^{(k)}\,, \tag{25}$$

gdzie $w \in \mathbb{R}$ jest *parametrem relaksacji*. Metoda ta zwana jest *succesive over-relaxation*, SOR. W postaci macierzowej

$$\mathbf{M}_w\mathbf{x}^{(k+1)} = \mathbf{N}_w\mathbf{x}^{(k)} + w\mathbf{b} \tag{26}$$

$\mathbf{M}_w = \mathbf{D} + w\mathbf{L}$, $\mathbf{N}_w = (1-w)\mathbf{D} - w\mathbf{U}$. *Teoretycznie* należy dobrać takie w, aby zminimalizować $\rho(\mathbf{M}_w^{-1}\mathbf{N}_w)$.