

Wstęp do metod numerycznych

Faktoryzacja Cholesky'ego i *QR*

P. F. Góra

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

2018

Faktoryzacja Cholesky'ego

Niech $\mathbf{A} \in \mathbb{R}^{N \times N}$ będzie symetryczna, $\mathbf{A}^T = \mathbf{A}$, i dodatnio określona:

$$\forall \mathbf{x} \in \mathbb{R}^N, \mathbf{x} \neq \mathbf{0}: \mathbf{x}^T \mathbf{A} \mathbf{x} > 0. \quad (1)$$

Wówczas istnieje alternatywa dla faktoryzacji LU : faktoryzacja postaci

$$\mathbf{A} = \mathbf{C} \mathbf{C}^T, \quad (2)$$

gdzie \mathbf{C} jest macierzą trójkątną dolną o elementach diagonalnych większych od zera. Znalezienie faktoryzacji Cholesky'ego jest mniej więcej o połowę szybsze, niż znalezienie faktoryzacji LU tej samej macierzy.

Najprostszy algorytm jest bardzo podobny do algorytmu Doolittle'a:

$$\underbrace{\begin{bmatrix} c_{11} & & & & \\ c_{21} & c_{22} & & & \\ c_{31} & c_{32} & c_{33} & & \\ c_{41} & c_{42} & c_{43} & c_{44} & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}}_C \underbrace{\begin{bmatrix} c_{11} & c_{21} & c_{31} & c_{41} & \dots \\ & c_{22} & c_{32} & c_{42} & \dots \\ & & c_{33} & c_{43} & \dots \\ & & & c_{44} & \dots \\ & & & & \ddots \end{bmatrix}}_{C^T} = \underbrace{\begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} & \dots \\ a_{21} & a_{22} & a_{32} & a_{42} & \dots \\ a_{31} & a_{32} & a_{33} & a_{43} & \dots \\ a_{41} & a_{42} & a_{43} & a_{44} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}}_A \quad (3)$$

Pierwsza kolumna macierzy A daje

$$\begin{aligned}
 c_{11}^2 &= a_{11} \\
 c_{21}c_{11} &= a_{21} \\
 c_{31}c_{11} &= a_{31} \\
 c_{41}c_{11} &= a_{41} \\
 &\dots\dots\dots
 \end{aligned} \quad (4)$$

Z pierwszego z tych równań obliczamy c_{11} , z kolejnych c_{21}, c_{31} itd.

Druga kolumna daje

$$\begin{aligned}c_{11}c_{21} &= a_{21} \\c_{21}^2 + c_{22}^2 &= a_{22} \\c_{31}c_{21} + c_{32}c_{22} &= a_{32} \\c_{41}c_{21} + c_{42}c_{22} &= a_{42} \\&\dots\dots\dots\end{aligned}\tag{5}$$

Pierwsze z równań (5) jest identyczne z drugim z równań (4). Drugie z równań (5) pozwala na wyliczenie c_{22} . Dalsze równania pozwalają wyliczyć c_{32} , c_{42} itd.

I tak dalej.

Z uwagi na symetrię problemu, przy obliczaniu faktoryzacji Cholesky'ego nie jest możliwy wybór elementów podstawowych. Z uwagi na kolejność obliczeń, obliczone czynniki Cholesky'ego można przechowywać w tym samym miejscu, co elementy pierwotnej macierzy A .

Faktoryzacja LDL

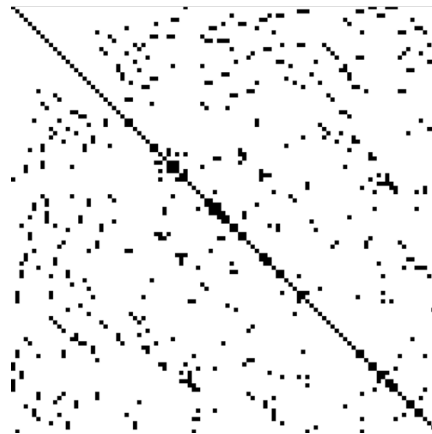
Jeżeli macierz spełnia założenia potrzebne do przeprowadzenia faktoryzacji Cholesky'ego, można także znaleźć jej inną faktoryzację:

$$\mathbf{A} = \mathbf{L} \mathbf{D} \mathbf{L}^T, \quad (6)$$

gdzie \mathbf{L} jest macierzą trójkątną dolną o tej własności, że $\forall i: l_{ii} = 1$, natomiast \mathbf{D} jest macierzą diagonalną o dodatnich elementach. Zaletą faktoryzacji LDL w stosunku do faktoryzacji Cholesky'ego jest to, iż do znalezienia LDL nie potrzeba pierwiastkowań.

Macierze rzadkie

W wielu praktycznych zastosowaniach występują **macierze rzadkie**, to znaczy takie, w których liczba elementów niezerowych rośnie wolniej niż N^2 , gdzie N jest wymiarem macierzy. Na przykład w macierzy trójdzielnej liczba niezerowych elementów skaluje się jak $O(3N)$, a w macierzy pasmowej o P dodatkowych diagonalach jak $O((2P + 1)N)$. Możliwe są także inne struktury macierzy rzadkich.



Macierze rzadkie i efektywność numeryczna

Dla efektywności numerycznej jest niesłuchanie ważne, aby zastosowany **algorytm** uwzględniał strukturę macierzy, tak, aby nie trzeba było wykonywać redundantnych mnożeń przez zero i dodawań zera, a nawet żeby nie przechodzić przez zerowe elementy.

- Dla macierzy trójdzielnej faktoryzacji LU dokonujemy w czasie liniowym, $O(N)$, ale za to niemożliwy jest wybór elementu podstawowego.
- Jeżeli możliwa jest faktoryzacja Cholesky'ego macierzy M -diagonalnej, także jej czynnik Cholesky'ego będzie M -diagonalny. Może jednak pojawić się niekorzystne zjawisko, zwane **wypełnieniem**: Jeżeli sama macierz ma zera “wewnątrz” pasma, jej czynnik Cholesky'ego nie musi ich mieć, co może bardzo niekorzystnie wpłynąć na wydajność numeryczną.

Przykład

Czynnik Cholesky'ego następującej macierzy *rzadkiej*

$$\begin{bmatrix} \bullet & \bullet & \bullet & \bullet & \bullet & \dots \\ \bullet & \bullet & & & & \\ \bullet & & \bullet & & & \\ \bullet & & & \bullet & & \\ \bullet & & & & \bullet & \\ \vdots & & & & & \ddots \end{bmatrix}$$

(7)

(niewypełnione elementy są zerami) będzie macierzą *pełną*.

Minimum Degree Algorithm*

Niech $\mathbf{A} \in \mathbb{R}^{N \times N}$ będzie macierzą posiadającą faktoryzację Cholesky'ego, dla której zachodzi niebezpieczeństwo pojawienia się wypełnienia. Wypełnienie zależy od struktury macierzy, nie od wartości jej poszczególnych elementów. Zamiast równania

$$\mathbf{Ax} = \mathbf{b} \quad (8a)$$

możemy rozwiązywać równanie

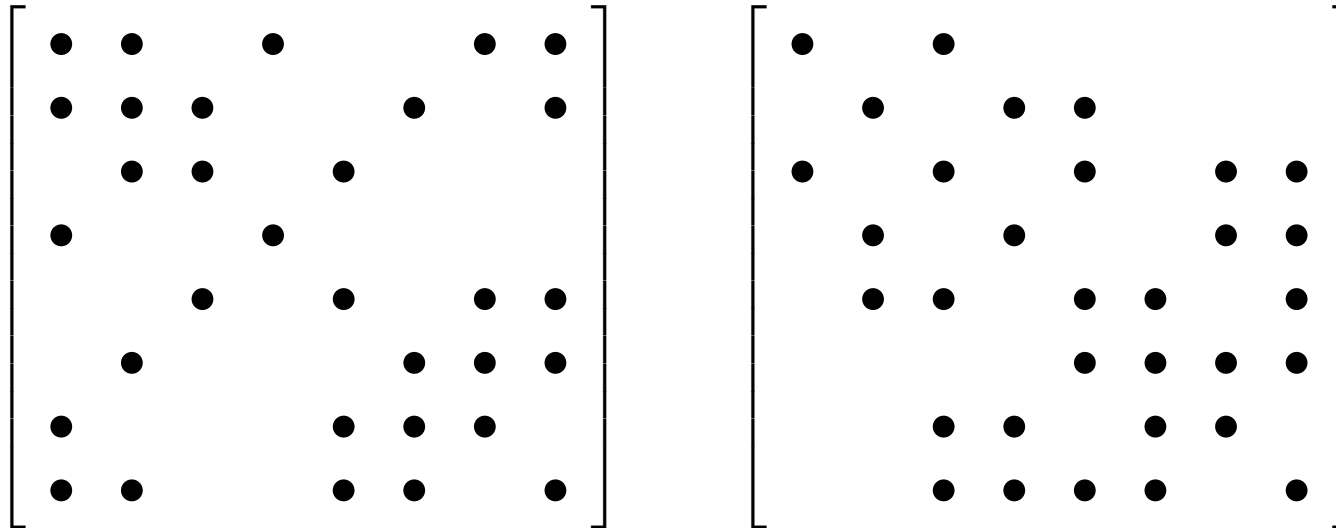
$$(\mathbf{PAP}^T) (\mathbf{Px}) = \mathbf{Pb} \quad (8b)$$

gdzie \mathbf{P} jest macierzą permutacji. (Jest to macierz ortogonalna.) Jeśli \mathbf{A} jest symetryczna i dodatnio określona, także macierz \mathbf{PAP}^T jest symetryczna i dodatnio określona, a więc posiada ona swoją faktoryzację Cholesky'ego. Macierz \mathbf{P} staramy się dobrać tak, aby wypełnienie w czynniku Cholesky'ego spemutowanej macierzy było możliwie małe. Problem

znalezienia permutacji takiej, aby wypełnienie było najmniejsze z możliwych jest NP-zupełny, w praktyce do poszukiwania P posługujemy się algorytmami heurystycznymi. Z historycznych powodów, z uwagi na związek pomiędzy macierzami symetrycznymi a grafami (struktura zerowych/niezerowych elementów pozadiagonalnych macierzy symetrycznej odpowiada macierzy sąsiedztwa pewnej klasy grafów nieskierowanych), algorytmy te nazywa się *minimum degree algorithms*. Przegląd tych algorytmów wykracza, niestety, poza zakres wykładu ze *wstępu* do metod numerycznych.

Jeżeli znalezienie efektywnej permutacji nie wydaje się tanie i wygodne, można rozważyć użycie zupełnie innej klasy algorytmów, na przykład algorytmów iteracyjnych.

Przykład



Macierz po prawej stronie stanowi optymalną permutację macierzy po lewej. W obu macierzach 32 elementy (dokładnie połowa) jest pusta. Gdyby zastosować faktoryzację Cholesky'ego do macierzy po lewej, czynnik trójkątny miałby tylko 4 (zamiast 16) elementów zerowych. Czynnik Cholesky'ego macierzy po prawej będzie miał 13 elementów zerowych (trzy elementy się wypełnią).

Faktoryzacja QR

Innym rodzajem faktoryzacji jest faktoryzacja QR : daną macierz $\mathbf{A} \in \mathbb{R}^{N \times N}$ przedstawiam w postaci iloczynu

$$\mathbf{A} = \mathbf{Q} \mathbf{R}, \quad (9)$$

gdzie \mathbf{Q} jest macierzą ortogonalną, a \mathbf{R} jest macierzą trójkątną górną. Złożoność obliczeniowa faktoryzacji QR wynosi dla macierzy pełnych $O(N^3)$, czyli tyle samo, co faktoryzacji LU , jednak współczynnik przy wyrazie wiodącym jest gorszy niż dla LU . QR nie jest więc metodą “z wyboru” rozwiązywania układów równań liniowych. Jeśli jednak z jakichś względów

faktoryzację QR możemy łatwo (lub musimy) obliczyć, układ równań liniowych rozwiązujemy jak następuje:

$$\mathbf{Ax} = \mathbf{b} \quad (10a)$$

$$\mathbf{QRx} = \mathbf{b} \quad (10b)$$

$$\mathbf{Rx} = \mathbf{Q}^T \mathbf{b} \quad (10c)$$

Koszt obliczeniowy przejścia od (10b) do (10c) wynosi $O(N^2)$. Równanie (10c) rozwiązujemy metodą *backsubstitution*, co także kosztuje $O(N^2)$. Jest to więc koszt mały w porównaniu z dokonaniem samej faktoryzacji.

Pozostaje pytanie: Jak dokonać tej faktoryzacji?

Transformacja Householdera

Niech $\mathbf{u} \in \mathbb{R}^N$, $\mathbf{u} \neq 0$. Tworzymy macierz

$$\mathbf{P} = \mathbb{I} - 2 \frac{\mathbf{u}\mathbf{u}^T}{\|\mathbf{u}\|^2}. \quad (11)$$

W sposób oczywisty $\mathbf{P}^T = \mathbf{P}$. Obliczmy

$$\begin{aligned} \mathbf{P}^2 &= \left(\mathbb{I} - 2 \frac{\mathbf{u}\mathbf{u}^T}{\|\mathbf{u}\|^2} \right) \left(\mathbb{I} - 2 \frac{\mathbf{u}\mathbf{u}^T}{\|\mathbf{u}\|^2} \right) \\ &= \mathbb{I} - 2 \frac{\mathbf{u}\mathbf{u}^T}{\|\mathbf{u}\|^2} - 2 \frac{\mathbf{u}\mathbf{u}^T}{\|\mathbf{u}\|^2} + 4 \frac{\overbrace{\mathbf{u}\mathbf{u}^T\mathbf{u}\mathbf{u}^T}^{\|\mathbf{u}\|^2}}{\|\mathbf{u}\|^4} \\ &= \mathbb{I} - 4 \frac{\mathbf{u}\mathbf{u}^T}{\|\mathbf{u}\|^2} + 4 \frac{\mathbf{u}\mathbf{u}^T}{\|\mathbf{u}\|^2} \\ &= \mathbb{I} \end{aligned} \quad (12)$$

Skoro $\mathbf{P}^T = \mathbf{P}$ oraz $\mathbf{P}^2 = \mathbb{I}$, $\mathbf{P}^T = \mathbf{P}^{-1}$: macierz (11) jest macierzą symetryczną, rzeczywistą oraz ortogonalną. Macierz taką nazywamy ortogonalną macierzą rzutową.

Niech teraz w (11)

$$\mathbf{u} = \mathbf{x} \mp \|\mathbf{x}\| \hat{\mathbf{e}}_1, \quad (13)$$

gdzie $\hat{\mathbf{e}}_1$ jest pierwszym wektorem jednostkowym. Macierz (11) wraz z (13) nazywam *macierzą Householdera*. Obliczam

$$\mathbf{P}\mathbf{x} = \mathbf{x} - \frac{2\mathbf{u}\mathbf{u}^T\mathbf{x}}{\|\mathbf{u}\|^2}. \quad (14a)$$

Zauważmy, że $\mathbf{u}^T\mathbf{x} = \mathbf{x}^T\mathbf{x} \mp \|\mathbf{x}\| \hat{\mathbf{e}}_1^T\mathbf{x} = \|\mathbf{x}\|^2 \mp \|\mathbf{x}\| x_1$, gdzie x_1 jest pierwszą składową wektora \mathbf{x} .

Analogicznie $\|\mathbf{u}\|^2 = (\mathbf{x} \mp \|\mathbf{x}\| \hat{\mathbf{e}}_1)^T (\mathbf{x} \mp \|\mathbf{x}\| \hat{\mathbf{e}}_1) = \mathbf{x}^T \mathbf{x} \mp \|\mathbf{x}\| x_1 \mp \|\mathbf{x}\| x_1 + \|\mathbf{x}\|^2 = 2 (\|\mathbf{x}\|^2 \mp \|\mathbf{x}\| x_1)$. Wobec tego

$$\mathbf{P}_\mathbf{x} = \mathbf{x} - \frac{2\mathbf{u} (\|\mathbf{x}\|^2 \mp \|\mathbf{x}\| x_1)}{2 (\|\mathbf{x}\|^2 \mp \|\mathbf{x}\| x_1)} = \mathbf{x} - \mathbf{u} = \pm \|\mathbf{x}\| \hat{\mathbf{e}}_1. \quad (14b)$$

Efektem działania macierzy Householdera na wskazany wektor jest wyzerowanie wszystkich jego składowych, poza pierwszą, i “przelanie” całej jego długości na pierwszą składową. Złożoność obliczeniowa tej procedury wynosi $O(N)$.

Faktoryzacja QR

Niech $\mathbf{A} \in \mathbb{R}^{N \times N}$. Niech \mathbf{P}_1 oznacza transformację Householdera zbudowaną na pierwszej kolumnie macierzy \mathbf{A} . Otrzymujemy

$$\mathbf{P}_1 \mathbf{A} = \mathbf{A}_1 \quad (15)$$
$$\mathbf{P}_1 \begin{bmatrix} \bullet & \bullet & \bullet & \bullet & \cdots \\ \bullet & \bullet & \bullet & \bullet & \cdots \\ \bullet & \bullet & \bullet & \bullet & \cdots \\ \bullet & \bullet & \bullet & \bullet & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \begin{bmatrix} \bullet & \bullet & \bullet & \bullet & \cdots \\ & \bullet & \bullet & \bullet & \cdots \\ & \bullet & \bullet & \bullet & \cdots \\ & \bullet & \bullet & \bullet & \cdots \\ & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Transformacja Householdera \mathbf{P}_1 wyzerowała pierwszą kolumnę macierzy \mathbf{A} , za wyjątkiem elementu diagonalnego. Złożoność obliczeniowa tego kroku wynosi $O(N^2)$ (transformacja Householdera działa na N kolumn).

Niech teraz

$$\mathbf{P}_2 = \left[\begin{array}{c|cccc} 1 & 0 & 0 & \dots & 0 \\ \hline 0 & & & & \\ 0 & & & & \\ 0 & & & & \\ 0 & & & & \end{array} \right] \quad (16)$$

$(N-1)\mathbf{P}_2$

gdzie $(N-1)\mathbf{P}_2 \in \mathbb{R}^{(N-1) \times (N-1)}$ jest transformacją Householdera, zbudowaną na drugiej kolumnie macierzy \mathbf{A}_1 , poczynając od elementu diagonalnego w dół. Otrzymujemy

$$\mathbf{P}_2\mathbf{P}_1\mathbf{A} = \mathbf{P}_2\mathbf{A}_1 = \mathbf{A}_2 = \begin{bmatrix} \bullet & \bullet & \bullet & \bullet & \dots \\ & \bullet & \bullet & \bullet & \dots \\ & & \bullet & \bullet & \dots \\ & & \bullet & \bullet & \dots \\ & & \vdots & \vdots & \ddots \end{bmatrix} \quad (17)$$

Następnie definiuję

$$\mathbf{P}_3 = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ \hline 0 & 0 & & & & \\ 0 & 0 & & & & \\ 0 & 0 & & & & \\ 0 & 0 & & & & \end{array} \right] \quad (18)$$

gdzie $(N-2)\mathbf{P}_3 \in \mathbb{R}^{(N-2) \times (N-2)}$ jest transformacją Householdera, zbudowaną na trzeciej kolumnie macierzy \mathbf{A}_2 , poczynając od elementu diagonalnego w dół. Stojąca w lewym górnym rogu macierz jednostkowa służy do tego, żeby nie zepsuć struktury, którą osiągnęliśmy w poprzednich kro-

kach. Otrzymujemy

$$\mathbf{P}_3\mathbf{P}_2\mathbf{P}_1\mathbf{A} = \mathbf{P}_3\mathbf{A}_2 = \mathbf{A}_3 = \begin{bmatrix} \bullet & \bullet & \bullet & \bullet & \cdots \\ & \bullet & \bullet & \bullet & \cdots \\ & & \bullet & \bullet & \cdots \\ & & & \bullet & \cdots \\ & & & \vdots & \ddots \end{bmatrix} \quad (19)$$

Widać, że po $N-1$ krokach osiągnę

$$\mathbf{R} = \begin{bmatrix} \bullet & \bullet & \bullet & \cdots & \bullet \\ & \bullet & \bullet & \cdots & \bullet \\ & & \bullet & \cdots & \bullet \\ & & & \ddots & \vdots \\ & & & & \bullet \end{bmatrix} = \underbrace{\mathbf{P}_{N-1}\mathbf{P}_{N-2}\cdots\mathbf{P}_1}_{\mathbf{Q}^T} \mathbf{A} \quad (20)$$

\mathbf{R} jest macierzą trójkątną górną. Ponieważ macierze \mathbf{P}_i są ortogonalne, ich iloczyn, oznaczony przez \mathbf{Q}^T , także jest macierzą ortogonalną. Nie

musimy zapamiętywać poszczególnych macierzy P_i , wystarczy zapamiętać ich iloczyn.

Otrzymaliśmy zatem dla dowolnej macierzy kwadratowej *faktoryzację na macierz ortogonalną i trójkątną górną*:

$$A = QR, \quad (21)$$

czyli poszukiwaną faktoryzację *QR*.

Nie jest to jedyny algorytm pozwalający uzyskać faktoryzację *QR*, pokazuje jednak, że jest ona w ogólnym (czyli dowolnym) wypadku możliwa.

Obroty Givensa

Transformacja Householdera służy do zerowania wielu składowych jakiegoś wektora. Jeżeli chcemy selektywnie wyzerować jakieś składowe — lub jeśli interesujący nas wektor ma jakąś szczególną postać — bardziej efektywne od transformacji Householdera będą *obroty Givensa*.

Macierz Givensa ma postać (niezaznaczone elementy są zerami)

$$\mathbf{G}(i, j) = \begin{bmatrix} 1 & & & & & & & & & \\ & \dots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & c & & & & & & \\ & & & & \dots & & & & & \\ & & & & & 1 & & & & \\ & & & & & & \dots & & & \\ & & & -s & & & & c & & \\ & & & & & & & & 1 & \\ & & & & & & & & & \dots \\ & & & & & & & & & & 1 \end{bmatrix} \quad (22)$$

gdzie wyróżnione elementy znajdują się na pozycjach, odpowiednio, (i, i) , (i, j) , (j, i) , (j, j) . Przyjmujemy, że $c = \cos \theta$, $s = \sin \theta$. Macierz (22) jest macierzą obrotu w płaszczyźnie (x_i, x_j) o kąt θ przeciwnie do ruchu wskazówek zegara. Jest to macierz ortogonalna.

Niech x będzie pewnym wektorem i niech $y = G(i, j)x$. Składowe wektora y wynoszą

$$y_k = \begin{cases} cx_i + sx_j & k = i \\ -sx_i + cx_j & k = j \\ x_k & \text{poza tym} \end{cases} \quad (23)$$

Zażądajmy, aby $y_j = 0$. Widać, że musi zachodzić

$$c = \frac{x_i}{\sqrt{x_i^2 + x_j^2}}, \quad s = \frac{x_j}{\sqrt{x_i^2 + x_j^2}}. \quad (24)$$

Obrót Givensa (22) wraz z warunkami (24) zeruje j -tą składową wybranego wektora. Składowa i -ta przybiera wartość $\sqrt{x_i^2 + x_j^2}$.

Faktoryzacja QR macierzy trójdzielnej symetrycznej

Rozpatrzmy macierz $\mathbf{A} \in \mathbb{R}^{N \times N}$, trójdzielną symetryczną

$$\mathbf{A} = \begin{bmatrix} a & b & & & & \\ b & d & e & & & \\ & e & f & g & & \\ & & g & h & l & \\ & & & \ddots & \ddots & \ddots \end{bmatrix} \quad (25)$$

Zadziałajmy na nią macierzą Givensa taką, aby zerowała drugi element pierwszej kolumny

$$\mathbf{G}_1 = \begin{bmatrix} c_1 & s_1 & & & & \\ -s_1 & c_1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & \ddots & \end{bmatrix} \quad (26)$$

$$\mathbf{A}_1 = \mathbf{G}_1 \mathbf{A} = \begin{bmatrix} \sqrt{a^2 + b^2} & \frac{ab+bd}{\sqrt{a^2+b^2}} & \frac{be}{\sqrt{a^2+b^2}} & & & & & & & \\ & \frac{ad-b^2}{\sqrt{a^2+b^2}} & \frac{ae}{\sqrt{a^2+b^2}} & & & & & & & \\ & & e & & & & & & & \\ & & & f & & & & & & \\ & & & & g & & & & & \\ & & & & & g & & & & \\ & & & & & & h & & & \\ & & & & & & & l & & \\ & & & & & & & & \dots & \\ & & & & & & & & & \dots & \\ & & & & & & & & & & \dots \end{bmatrix} \quad (27)$$

Wiersze macierzy \mathbf{A}_1 począwszy od trzeciego w dół zgadzają się z wierszami macierzy \mathbf{A} . Obliczenie macierzy \mathbf{A}_1 wymaga wykonania stałej, *niezależnej od rozmiaru macierzy*, liczby operacji. W wyniku otrzymaliśmy macierz, w której poddiagonalne elementy pierwszej kolumny są zerami.

Macierz A_1 mnożymy przez macierz Givensa

$$G_2 = \begin{bmatrix} 1 & & & & & \\ & c_2 & s_2 & & & \\ & -s_2 & c_2 & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & \ddots \end{bmatrix} \quad (28)$$

dobraną tak, aby zerowała trzeci element drugiej kolumny macierzy A_1 . Pierwszy wiersz i pierwsza kolumna nie zmieniają się, podobnie jak wiersze począwszy od czwartego. W rezultacie macierz $A_2 = G_2 A_1 = G_2 G_1 A$ ma zera w poddiagonalnych miejscach dwu pierwszych kolumn. Ten krok także wymaga stałej, niezależnej od rozmiaru macierzy, liczby operacji.

W kolejnym kroku macierz A_2 mnożymy przez taką macierz Givensa, która wyzeruje czwarty element trzeciej kolumny. I tak dalej.

W ten sposób, po $N-1$ krokach, ponosząc koszt numeryczny $O(N)$ (stały koszt na krok, $\sim N$ kroków), otrzymujemy

$$\mathbf{G}_{N-1} \cdots \mathbf{G}_2 \mathbf{G}_1 \mathbf{A} = \mathbf{R} = \begin{bmatrix} \bullet & \bullet & \bullet & & & & \\ & \bullet & \bullet & \bullet & & & \\ & & \bullet & \bullet & \bullet & & \\ & & & \cdots & \cdots & \cdots & \\ & & & & \bullet & \bullet & \\ & & & & & \bullet & \bullet \\ & & & & & & \bullet \end{bmatrix}, \quad (29a)$$

czyli

$$\mathbf{A} = \underbrace{\mathbf{G}_1^T \mathbf{G}_2^T \cdots \mathbf{G}_{N-1}^T}_{\mathbf{Q}} \mathbf{R} \quad (29b)$$

Macierz \mathbf{Q} jest ortogonalna. Macierz \mathbf{R} jest trójkątna górna (tak naprawdę ma ona tylko dwie niezerowe diagonale nad diagonalą główną). Widzimy, że (29b) jest faktoryzacją QR macierzy trójdzielnej symetrycznej.

Zastosowanie do rozwiązywania układu równan liniowych

Jeżeli chcemy użyć obrotów Givensa do rozwiązania układu równań liniowych

$$\mathbf{Ax} = \mathbf{b}, \quad (30)$$

gdzie \mathbf{A} jest trójdagonalną macierzą symetryczną, postępując jak poprzednio otrzymujemy kolejno

$$\mathbf{G}_1 \mathbf{Ax} = \mathbf{G}_1 \mathbf{b} \quad (31a)$$

$$\mathbf{G}_2 \mathbf{G}_1 \mathbf{Ax} = \mathbf{G}_2 \mathbf{G}_1 \mathbf{b} \quad (31b)$$

...

$$\mathbf{G}_{N-1} \cdots \mathbf{G}_2 \mathbf{G}_1 \mathbf{Ax} \equiv \mathbf{Rx} = \mathbf{G}_{N-1} \cdots \mathbf{G}_2 \mathbf{G}_1 \mathbf{b}. \quad (31c)$$

Oczywiście istotne jest tylko równanie (31c) — lewych stron poprzednich równań nie musimy wyliczać. Każde kolejne mnożenie po stronie prawej

wykonujemy w stałym czasie, a więc do postaci $\mathbf{R}\mathbf{x} = \tilde{\mathbf{b}}$ dochodzimy w czasie $O(N)$. To równanie rozwiązujemy metodą *backsubstitution*, co, z uwagi na szczególną postać macierzy \mathbf{R} , także da się wykonać w czasie liniowym.

Przykład ten pokazuje, że możemy odnieść duży zysk na złożoności obliczeniowej, jeśli tylko dobierzemy odpowiedni algorytm odpowiadający strukturze — w tym wypadku rzadkości i symetryczności — macierzy.

Uwaga: Skumulowanej macierzy Givensa \mathbf{Q} nie musimy wyliczać w sposób jawny — gdybyśmy to chcieli zrobić, wymagałoby to $O(N^2)$ operacji.

Wzór Shermana-Morrisona

Twierdzenie: Niech $\mathbf{A} \in \mathbb{R}^{N \times N}$, $\det \mathbf{A} \neq 0$ oraz $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$. Niech $\mathbf{A}_1 = \mathbf{A} + \mathbf{u}\mathbf{v}^T$. Wówczas

$$\mathbf{A}_1^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}. \quad (32)$$

Zauważmy, że ponieważ $\det \mathbf{A} \neq 0$, macierz \mathbf{A}^{-1} istnieje. Ponadto wyrażenie $\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}$ jest *liczbą* (skalarem).

Wzór ten jest przydatny w przypadku, gdy chcemy wyliczyć \mathbf{A}_1^{-1} , gdzie \mathbf{A}_1 jest pewną szczególną modyfikacją macierzy \mathbf{A} , a odwrotność \mathbf{A}^{-1} znamy.

Przykład

Niech $\mathbf{u} = \mathbf{v} = [1, 0, 0, 0, 1]^T$. Wówczas

$$\mathbf{uv}^T = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} [1 \ 0 \ 0 \ 0 \ 1] = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (33)$$

Niech teraz

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 0 & 0 & 0 \\ 1 & 4 & 1 & 0 & 0 \\ 0 & 1 & 4 & 1 & 0 \\ 0 & 0 & 1 & 4 & 1 \\ 0 & 0 & 0 & 1 & 3 \end{bmatrix}, \quad \mathbf{A}_1 = \mathbf{A} + \mathbf{uv}^T = \begin{bmatrix} 4 & 1 & 0 & 0 & 1 \\ 1 & 4 & 1 & 0 & 0 \\ 0 & 1 & 4 & 1 & 0 \\ 0 & 0 & 1 & 4 & 1 \\ 1 & 0 & 0 & 1 & 4 \end{bmatrix}. \quad (34)$$

Dowód.

$$\begin{aligned} & (\mathbf{A} + \mathbf{u}\mathbf{v}^T) \left(\mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} \right) \\ &= \mathbf{A}\mathbf{A}^{-1} - \frac{1}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} \mathbf{A}\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1} + \mathbf{u}\mathbf{v}^T\mathbf{A}^{-1} \\ & \quad - \frac{1}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} \mathbf{u} \underbrace{\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}_{\text{to jest liczba!}} \mathbf{v}^T\mathbf{A}^{-1} \\ &= \mathbb{I} - \frac{1}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} \mathbf{u}\mathbf{v}^T\mathbf{A}^{-1} + \mathbf{u}\mathbf{v}^T\mathbf{A}^{-1} - \frac{\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} \mathbf{u}\mathbf{v}^T\mathbf{A}^{-1} \\ &= \mathbb{I} + \left(1 - \frac{1}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} - \frac{\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} \right) \mathbf{u}\mathbf{v}^T\mathbf{A}^{-1} = \mathbb{I}. \quad (35) \end{aligned}$$

□

Algorytm Shermana-Morrisona

Wzór Shermana-Morrisona (32) pozwala skonstruować odwrotność macierzy A_1 jeśli znamy odwrotność A . Jednak w praktyce prawie **nigdy nie konstruujemy jawnej odwrotności macierzy!** Jak więc zastosować ten wzór?

Zauważmy, że **zapewne** chcemy obliczyć jakieś $A_1^{-1}b$, gdzie b jest znanym wektorem, przy założeniu, że **łatwo** potrafimy obliczyć $A^{-1}b$. Interesuje nas znalezienie

$$w = A_1^{-1}b = \left(A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1}u} \right) b \quad (36)$$

Algorytm wygląda następująco:

(a) Rozwiąż równanie

$$\mathbf{Az} = \mathbf{b} \quad (37a)$$

(b) Rozwiąż równanie

$$\mathbf{Aq} = \mathbf{u} \quad (37b)$$

(c) Oblicz

$$\mathbf{w} = \mathbf{z} - \frac{\mathbf{v}^T \mathbf{z}}{1 + \mathbf{v}^T \mathbf{q}} \mathbf{q}. \quad (37c)$$

Problem sprowadza się więc do rozwiązania dwu równań (37a),(37b) z taką samą macierzą, które umiemy szybko rozwiązać, gdyż — na przykład — znamy faktoryzację macierzy \mathbf{A} . Zauważmy, że $\mathbf{v}^T \mathbf{A}^{-1} \mathbf{b} = \mathbf{v}^T \mathbf{z}$ jest *liczbą*.

Przykład (c.d.)

Pierwsza z macierzy (34) jest macierzą symetryczną, dodatnio określoną i trójdziagonalną, a więc jej czynnik Cholesky'ego ma tylko dwie niezerowe diagonale, a koszt jego wyliczenia jest rzędu $O(N)$. Czynnik Cholesky'ego drugiej z tych macierzy jest pełną macierzą trójkątną (nastąpi *wypełnienie*) i koszt jego wyliczenia jest rzędu $O(N^3)$ (wyobraźmy sobie, że zamiast o macierzach 5×5 , mówimy o macierzach 1000×1000). Zastosowanie algorytmu (37) redukuje problem do znalezienia i dwukrotnego zastosowania rzadkiego czynnika Cholesky'ego pierwszej z macierzy (34). Da się to zrobić w czasie liniowym.