

Решение систем линейных алгебраических уравнений

Мальковский Н. В.

Санкт-Петербургский Академический Университет



Матричная декомпозиция

Определение

Декомпозицией или разложением матрицы A называется представление A в виде некоторого произведения

$$A = A_1 A_2 \dots A_n$$

Матричная декомпозиция

Определение

Декомпозицией или разложением матрицы A называется представление A в виде некоторого произведения

$$A = A_1 A_2 \dots A_n$$

Замечание. Обычно $n = 2, 3$, в интересующих нас случаях A и A_i – квадратные матрицы.

Матричная декомпозиция

Определение

Декомпозицией или разложением матрицы A называется представление A в виде некоторого произведения

$$A = A_1 A_2 \dots A_n$$

Замечание. Обычно $n = 2, 3$, в интересующих нас случаях A и A_i – квадратные матрицы.

Зачем нужно: Пусть $A = A_1 \dots A_n$, дана система

$$Ax = b$$

Матричная декомпозиция

Определение

Декомпозицией или разложением матрицы A называется представление A в виде некоторого произведения

$$A = A_1 A_2 \dots A_n$$

Замечание. Обычно $n = 2, 3$, в интересующих нас случаях A и A_i – квадратные матрицы.

Зачем нужно: Пусть $A = A_1 \dots A_n$, дана система

$$Ax = b$$

Если $A_1 y = b$, систему $Ax = b$ можно редуцировать до системы $A_2 \dots A_n x = y$.

Матричная декомпозиция

Определение

Декомпозицией или разложением матрицы A называется представление A в виде некоторого произведения

$$A = A_1 A_2 \dots A_n$$

Замечание. Обычно $n = 2, 3$, в интересующих нас случаях A и A_i – квадратные матрицы.

Зачем нужно: Пусть $A = A_1 \dots A_n$, дана система

$$Ax = b$$

Если $A_1 y = b$, систему $Ax = b$ можно редуцировать до системы $A_2 \dots A_n x = y$.

Таким образом, интересны разложения с небольшим n и простой структурой A_i .

Ортогональные матрицы

Определение

Матрица $P \in \mathbb{R}^{n \times n}$ называется ортогональной, если $P^T P = P P^T = I$.

Ортогональные матрицы

Определение

Матрица $P \in \mathbb{R}^{n \times n}$ называется ортогональной, если $P^T P = P P^T = I$.

Замечание 1. Непосредственное следствие, если P ортогональна, то $P^{-1} = P^T$.

Ортогональные матрицы

Определение

Матрица $P \in \mathbb{R}^{n \times n}$ называется ортогональной, если $P^T P = P P^T = I$.

Замечание 1. Непосредственное следствие, если P ортогональна, то $P^{-1} = P^T$. Замечание 2. Пусть $p_1, \dots, p_n \in K^n$ – столбцы матрицы P , т. е.

$P = [p_1 \dots p_n]$, тогда

$$P^T P = \begin{bmatrix} p_1^T \\ \dots \\ p_n^T \end{bmatrix} [p_1 \dots p_n] = \begin{pmatrix} p_1^T p_1 & p_1^T p_2 & \dots & p_1^T p_n \\ p_2^T p_1 & p_2^T p_2 & \dots & p_2^T p_n \\ \vdots & \vdots & \ddots & \vdots \\ p_n^T p_1 & p_n^T p_2 & \dots & p_n^T p_n \end{pmatrix} = I$$

Ортогональные матрицы

Определение

Матрица $P \in \mathbb{R}^{n \times n}$ называется ортогональной, если $P^T P = P P^T = I$.

Замечание 1. Непосредственное следствие, если P ортогональна, то $P^{-1} = P^T$. Замечание 2. Пусть $p_1, \dots, p_n \in K^n$ – столбцы матрицы P , т. е.

$P = [p_1 \dots p_n]$, тогда

$$P^T P = \begin{bmatrix} p_1^T \\ \dots \\ p_n^T \end{bmatrix} [p_1 \dots p_n] = \begin{pmatrix} p_1^T p_1 & p_1^T p_2 & \dots & p_1^T p_n \\ p_2^T p_1 & p_2^T p_2 & \dots & p_2^T p_n \\ \vdots & \vdots & \ddots & \vdots \\ p_n^T p_1 & p_n^T p_2 & \dots & p_n^T p_n \end{pmatrix} = I$$

Таким образом

$$p_i^T p_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

иначе говоря, столбцы P – ортонормированы. Это свойство выполняется и для строк.

Треугольные матрицы

Определение

Матрица $A = (a_{ij}) \in K^{n \times n}$ называется верхней треугольной матрицей, если $a_{ij} = 0$ при $i > j$. Аналогично A называется нижней треугольной, если $a_{ij} = 0$ при $i < j$.

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix} \quad \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

Решение СЛАУ для ортогональных матриц

Пусть P – ортогональная матрица, рассмотрим систему

$$Px = b$$

Решение СЛАУ для ортогональных матриц

Пусть P – ортогональная матрица, рассмотрим систему

$$Px = b$$

Из свойств P получаем

$$x = P^T Px = P^T b$$

Решение СЛАУ для ортогональных матриц

Пусть P – ортогональная матрица, рассмотрим систему

$$Px = b$$

Из свойств P получаем

$$x = P^T Px = P^T b$$

Фактически, свойство ортогональности можно переписать в виде $P^{-1} = P^T$.

Треугольные матрицы

Свойства:

Треугольные матрицы

Свойства:

1. Произведение верхних треугольных матриц есть верхняя треугольная матрица, при $i > j$

$$\sum_{k=1}^n a_{ik} b_{kj} = \sum_{k=1}^j \underbrace{a_{ik}}_{=0} b_{kj} + \sum_{k=j+1}^n a_{ik} \underbrace{b_{kj}}_{=0} = 0$$

Треугольные матрицы

Свойства:

1. Произведение верхних треугольных матриц есть верхняя треугольная матрица, при $i > j$

$$\sum_{k=1}^n a_{ik} b_{kj} = \sum_{k=1}^j \underbrace{a_{ik}}_{=0} b_{kj} + \sum_{k=j+1}^n a_{ik} \underbrace{b_{kj}}_{=0} = 0$$

2. Пусть $A = (a_{ij})$ – верхняя треугольная матрица с ненулевыми диагональными элементами, тогда A^{-1} существует и также является верхней треугольной матрицей. Пусть $i = n$, тогда для $j < n$ имеем

$$\sum_{k=1}^n a_{ik} b_{kj} = \sum_{k=1}^{n-1} \underbrace{a_{ik}}_{=0} b_{kj} + a_{nn} b_{nj} = 0 \Rightarrow b_{nj} = 0$$

Повторяя для $i = n - 1 \dots j + 1$ получаем

$$\sum_{k=1}^n a_{ik} b_{kj} = \sum_{k=1}^{i-1} \underbrace{a_{ik}}_{=0} b_{kj} + a_{ii} b_{ij} + \sum_{k=i+1}^n a_{ik} \underbrace{b_{kj}}_{=0, \text{ индукция}} = 0 \Rightarrow b_{ij} = 0$$

Треугольные матрицы

3. Если A – верхняя треугольная, то $A - \lambda I$ тоже верхняя треугольная матрица, а её определитель легко вычисляется

$$\det(A - \lambda I) = \prod_{i=1}^n (a_{ii} - \lambda)$$

Таким образом, диагональные элементы A являются её собственными числами с учетом кратности.

Треугольные матрицы

3. Если A – верхняя треугольная, то $A - \lambda I$ тоже верхняя треугольная матрица, а её определитель легко вычисляется

$$\det(A - \lambda I) = \prod_{i=1}^n (a_{ii} - \lambda)$$

Таким образом, диагональные элементы A являются её собственными числами с учетом кратности.

4. Атомарная треугольная матрица – нижняя треугольная матрица следующего вида

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & a_{(i+1)i} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_{ni} & 0 & \dots & 1 \end{pmatrix}$$

Домножение слева на такую матрицу добавляет к строке i линейную комбинацию строк $i + 1, \dots, n$.

5. Обратная матрица к атомарной вычисляется довольно просто

$$A = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & a_{(i+1)i} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_{ni} & 0 & \dots & 1 \end{pmatrix} \quad A^{-1} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & -a_{(i+1)i} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & -a_{ni} & 0 & \dots & 1 \end{pmatrix}$$

Решение СЛАУ для треугольных матриц

Пусть L – нижняя треугольная матрица, тогда система $Lx = b$ имеет вид

$$\begin{array}{rcll} l_{11}x_1 & & & = b_1 \\ l_{21}x_1 + l_{22}x_2 & & & = b_2 \\ \vdots & \vdots & \ddots & = \vdots \\ l_{n1}x_1 + l_{n2}x_2 + \dots + l_{nn}x_n & & & = b_n \end{array}$$

Решение СЛАУ для треугольных матриц

Пусть L – нижняя треугольная матрица, тогда система $Lx = b$ имеет вид

$$\begin{array}{rcll} \ell_{11}x_1 & & & = b_1 \\ \ell_{21}x_1 + \ell_{22}x_2 & & & = b_2 \\ \vdots & \vdots & \ddots & = \vdots \\ \ell_{n1}x_1 + \ell_{n2}x_2 + \dots + \ell_{nn}x_n & & & = b_n \end{array}$$

Из-за особого вида системы решение легче всего получить последовательным решением уравнений в заданном порядке, что дает рекуррентные формулы

$$x_i = \frac{b_i - \sum_{k=1}^{i-1} \ell_{ik}x_k}{\ell_{ii}}$$

Замечание 1. Если все диагональные элементы ненулевые, то формулы корректны. Если же $\ell_{ii} = 0$, но при этом $b_i - \sum_{k=1}^{i-1} \ell_{ik}x_k \neq 0$, то решения не существует, если же $b_i - \sum_{k=1}^{i-1} \ell_{ik}x_k = 0$, то x_i можно выбрать любое. *Замечание 2.* Нахождение решения по этим формулам принято называть прямой подстановкой.

Решение СЛАУ для треугольных матриц

Пусть L – нижняя треугольная матрица, тогда система $Lx = b$ имеет вид

$$\begin{array}{rcccccc} u_{11}x_1 & + & u_{12}x_2 & & \dots & & + & u_{1n}x_n & = & b_1 \\ & & \ddots & & \vdots & & & \vdots & = & \vdots \\ & & & & u_{(n-1)(n-1)}x_{n-1} & + & u_{(n-1)n}x_n & = & b_{n-1} \\ & & & & & & u_{nn}x_n & = & b_n \end{array}$$

Решение СЛАУ для треугольных матриц

Пусть L – нижняя треугольная матрица, тогда система $Lx = b$ имеет вид

$$\begin{array}{rcccccc} u_{11}x_1 & + & u_{12}x_2 & & \dots & & + & u_{1n}x_n & = & b_1 \\ & & \ddots & & \vdots & & & \vdots & = & \vdots \\ & & & & u_{(n-1)(n-1)}x_{n-1} & + & u_{(n-1)n}x_n & = & b_{n-1} \\ & & & & & & u_{nn}x_n & = & b_n \end{array}$$

Эту систему решать проще в обратном порядке, по аналогии с верхней треугольной матрицей получаем следующие формулы

$$x_i = \frac{b_i - \sum_{k=i+1}^n u_{ik}x_k}{u_{ii}}$$

Решение СЛАУ для треугольных матриц

Пусть L – нижняя треугольная матрица, тогда система $Lx = b$ имеет вид

$$\begin{array}{rcccccc} u_{11}x_1 & + & u_{12}x_2 & & \dots & & + & u_{1n}x_n & = & b_1 \\ & & \ddots & & \vdots & & & \vdots & = & \vdots \\ & & & & u_{(n-1)(n-1)}x_{n-1} & + & u_{(n-1)n}x_n & = & b_{n-1} \\ & & & & & & u_{nn}x_n & = & b_n \end{array}$$

Эту систему решать проще в обратном порядке, по аналогии с верхней треугольной матрицей получаем следующие формулы

$$x_i = \frac{b_i - \sum_{k=i+1}^n u_{ik}x_k}{u_{ii}}$$

Замечание. Такой процесс принято называть обратной подстановкой.

Определение

LU-декомпозицией матрицы $A \in K^{n \times n}$ называется представление $A = LU$, где $L, U \in K^{n \times n}$, L – нижняя треугольная матрица, а U – верхняя треугольная матрица.

Определение

LU-декомпозицией матрицы $A \in K^{n \times n}$ называется представление $A = LU$, где $L, U \in K^{n \times n}$, L – нижняя треугольная матрица, а U – верхняя треугольная матрица.

Замечание. Уравнение $Ax = b$ переписывается в виде $L^{-1}Ax = Ux = L^{-1}b$. Метод Гаусса заключается в последовательном домножении левой и правой части исходного уравнения на атомарные матрицы, что в итоге дает приведенное уравнение $Ux = L^{-1}b$.

Метод Гаусса

Метод Гаусса заключается в последовательном вычитании одного уравнения из других, таким образом множество решений системы не меняется.

Метод Гаусса

Метод Гаусса заключается в последовательном вычитании одного уравнения из других, таким образом множество решений системы не меняется. Метод делится на n шагов, на шаге k система имеет вид

$$\begin{pmatrix} u_{11}^k & u_{12}^k & \dots & u_{1k}^k & \dots & u_{1n}^k \\ 0 & u_{22}^k & \dots & u_{2k}^k & \dots & u_{2n}^k \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & u_{kk}^k & \dots & u_{kn}^k \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & u_{nk}^k & \dots & u_{nn}^k \end{pmatrix} x = \begin{pmatrix} b_1^k \\ b_2^k \\ \vdots \\ b_k^k \\ \vdots \\ b_n^k \end{pmatrix}$$

Здесь под u_{ij}^k подразумевается значение соответствующего коэффициента на шаге k . $u_{ij}^k = 0$ при выполнении двух условий:

- $i > j$
- $j \leq k$

Метод Гаусса

Итерация k : вычесть промасштабированную строку k из строк $k + 1, \dots, n$ так, чтобы обнулить столбец k .

Метод Гаусса

Итерация k : вычесть промасштабированную строку k из строк $k + 1, \dots, n$ так, чтобы обнулить столбец k .

Непосредственным вычислением получаем для:

$$u_{ij}^k = \begin{cases} u_{ij}^{k-1} - \frac{u_{kk}^{k-1}}{u_{ik}^{k-1}} u_{kj}, & i > k \\ u_{ij}^{k-1}, & i \leq k \end{cases} \quad b_i^k = \begin{cases} b_i^{k-1} - \frac{u_{kk}^{k-1}}{u_{ik}^{k-1}} b_k^{k-1}, & i > k \\ b_i^{k-1}, & i \leq k \end{cases}$$

Метод Гаусса

Итерация k : вычесть промасштабированную строку k из строк $k + 1, \dots, n$ так, чтобы обнулить столбец k .

Непосредственным вычислением получаем для:

$$u_{ij}^k = \begin{cases} u_{ij}^{k-1} - \frac{u_{kk}^{k-1}}{u_{ik}^{k-1}} u_{kj}, & i > k \\ u_{ij}^{k-1}, & i \leq k \end{cases} \quad b_i^k = \begin{cases} b_i^{k-1} - \frac{u_{kk}^{k-1}}{u_{ik}^{k-1}} b_k^{k-1}, & i > k \\ b_i^{k-1}, & i \leq k \end{cases}$$

Итерация k требует

- $n - k$ делений.

Метод Гаусса

Итерация k : вычесть промасштабированную строку k из строк $k + 1, \dots, n$ так, чтобы обнулить столбец k .

Непосредственным вычислением получаем для:

$$u_{ij}^k = \begin{cases} u_{ij}^{k-1} - \frac{u_{kk}^{k-1}}{u_{ik}^{k-1}} u_{kj}, & i > k \\ u_{ij}^{k-1} & i \leq k \end{cases} \quad b_i^k = \begin{cases} b_i^{k-1} - \frac{u_{kk}^{k-1}}{u_{ik}^{k-1}} b_k^{k-1}, & i > k \\ b_i^{k-1}, & i \leq k \end{cases}$$

Итерация k требует

- $n - k$ делений.
- $(n - k)^2 - 1$ сложений и умножений.

Метод Гаусса

Итерация k : вычесть промасштабированную строку k из строк $k + 1, \dots, n$ так, чтобы обнулить столбец k .

Непосредственным вычислением получаем для:

$$u_{ij}^k = \begin{cases} u_{ij}^{k-1} - \frac{u_{kk}^{k-1}}{u_{ik}^{k-1}} u_{kj}, & i > k \\ u_{ij}^{k-1} & i \leq k \end{cases} \quad b_i^k = \begin{cases} b_i^{k-1} - \frac{u_{kk}^{k-1}}{u_{ik}^{k-1}} b_k^{k-1}, & i > k \\ b_i^{k-1}, & i \leq k \end{cases}$$

Итерация k требует

- $n - k$ делений.
- $(n - k)^2 - 1$ сложений и умножений.

Метод Гаусса

Итерация k : вычесть промасштабированную строку k из строк $k + 1, \dots, n$ так, чтобы обнулить столбец k .

Непосредственным вычислением получаем для:

$$u_{ij}^k = \begin{cases} u_{ij}^{k-1} - \frac{u_{kk}^{k-1}}{u_{ik}^{k-1}} u_{kj}, & i > k \\ u_{ij}^{k-1} & i \leq k \end{cases} \quad b_i^k = \begin{cases} b_i^{k-1} - \frac{u_{kk}^{k-1}}{u_{ik}^{k-1}} b_k^{k-1}, & i > k \\ b_i^{k-1}, & i \leq k \end{cases}$$

Итерация k требует

- $n - k$ делений.
- $(n - k)^2 - 1$ сложений и умножений.

Замечание. Если A обратима, то все формулы корректны и u_{kk}^{k-1} отлично от нуля.

Пример

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad U = \begin{pmatrix} 2 & 3 & -2 & -1 & 0 \\ -2 & -2 & 4 & 4 & 1 \\ -4 & -7 & 0 & -1 & -1 \\ 2 & 5 & 0 & 6 & 1 \\ -2 & -5 & -4 & -4 & 0 \end{pmatrix}$$

Пример

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ -2 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 & -2 & -1 & 0 \\ 0 & 1 & 2 & 3 & 1 \\ 0 & -1 & -4 & -3 & -1 \\ 0 & 2 & 2 & 7 & 1 \\ 0 & -2 & -6 & -5 & 0 \end{pmatrix}$$
$$A = \begin{pmatrix} 2 & 3 & -2 & -1 & 0 \\ -2 & -2 & 4 & 4 & 1 \\ -4 & -7 & 0 & -1 & -1 \\ 2 & 5 & 0 & 6 & 1 \\ -2 & -5 & -4 & -4 & 0 \end{pmatrix} L_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ -2 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Пример

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ -2 & -1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 \\ -1 & -2 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 & -2 & -1 & 0 \\ 0 & 1 & 2 & 3 & 1 \\ 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & -2 & 1 & -1 \\ 0 & 0 & -2 & 1 & 2 \end{pmatrix}$$
$$A = \begin{pmatrix} 2 & 3 & -2 & -1 & 0 \\ -2 & -2 & 4 & 4 & 1 \\ -4 & -7 & 0 & -1 & -1 \\ 2 & 5 & 0 & 6 & 1 \\ -2 & -5 & -4 & -4 & 0 \end{pmatrix} L_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 2 & 0 & 1 & 0 \\ 0 & -2 & 0 & 0 & 1 \end{pmatrix}$$

Пример

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ -2 & -1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 1 & 0 \\ -1 & -2 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 & -2 & -1 & 0 \\ 0 & 1 & 2 & 3 & 1 \\ 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & 2 \end{pmatrix}$$
$$A = \begin{pmatrix} 2 & 3 & -2 & -1 & 0 \\ -2 & -2 & 4 & 4 & 1 \\ -4 & -7 & 0 & -1 & -1 \\ 2 & 5 & 0 & 6 & 1 \\ -2 & -5 & -4 & -4 & 0 \end{pmatrix} L_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Пример

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ -2 & -1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 1 & 0 \\ -1 & -2 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 & -2 & -1 & 0 \\ 0 & 1 & 2 & 3 & 1 \\ 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$
$$A = \begin{pmatrix} 2 & 3 & -2 & -1 & 0 \\ -2 & -2 & 4 & 4 & 1 \\ -4 & -7 & 0 & -1 & -1 \\ 2 & 5 & 0 & 6 & 1 \\ -2 & -5 & -4 & -4 & 0 \end{pmatrix} L_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

Рекурсивная форма метода Гаусса

Пусть $A \in K^{n \times n}$, $\ell \in K^{n-1}$, $u \in K^{n-1}$, $\tilde{A} \in K^{(n-1) \times (n-1)}$ и при этом

$$A = \begin{bmatrix} a_{11} & u^T \\ \ell & \tilde{A} \end{bmatrix}$$

Рекурсивная форма метода Гаусса

Пусть $A \in K^{n \times n}$, $\ell \in K^{n-1}$, $u \in K^{n-1}$, $\tilde{A} \in K^{(n-1) \times (n-1)}$ и при этом

$$A = \begin{bmatrix} a_{11} & u^T \\ \ell & \tilde{A} \end{bmatrix}$$

Обозначим за $GU(A)$, $GL(A)$ – верхнюю и нижнюю треугольную матрицу, получающуюся в результате применения метода Гаусса к матрице A , тогда имеют место соотношения

$$GU(A) = \begin{bmatrix} a_{11} & u^T \\ 0 & GU(\tilde{A} - \frac{1}{a_{11}}\ell u^T) \end{bmatrix}$$

и

$$GL(A) = \begin{bmatrix} 1 & 0_{n-1}^T \\ -\frac{1}{a_{11}}\ell & I_{n-1} \end{bmatrix} \begin{bmatrix} 1 & 0_{n-1}^T \\ 0_{n-1} & GL(\tilde{A} - \frac{1}{a_{11}}\ell u^T) \end{bmatrix}$$

Разложение Холецкого

Определение

Разложением Холецкого называется представление квадратной матрицы в виде $A = LL^T$, где L – нижняя треугольная матрица.

Разложение Холецкого

Определение

Разложением Холецкого называется представление квадратной матрицы в виде $A = LL^T$, где L – нижняя треугольная матрица.

Теорема

Для квадратной матрицы A разложение Холецкого существует, если она симметрична и положительно определена, и не существует, если не симметрична или не положительно полуопределена.

Разложение Холецкого

Определение

Разложением Холецкого называется представление квадратной матрицы в виде $A = LL^T$, где L – нижняя треугольная матрица.

Теорема

Для квадратной матрицы A разложение Холецкого существует, если она симметрична и положительно определена, и не существует, если не симметрична или не положительно полуопределена.

Док-во. Необходимость: легко увидеть, что матрица LL^T всегда симметрична и положительно определена

$$x^T LL^T x = (L^T x)^T L^T x = \|L^T x\|^2 \geq 0$$

Разложение Холецкого

Достаточность: пусть A – симметричная положительно определенная матрица и пусть, $A = LU$, где L – обратимая нижняя треугольная матрица, а U – верхняя треугольная матрица (существует в следствии метода Гаусса), тогда

$$A = A^T = LU = U^T L^T$$

Разложение Холецкого

Достаточность: пусть A – симметричная положительно определенная матрица и пусть, $A = LU$, где L – обратимая нижняя треугольная матрица, а U – верхняя треугольная матрица (существует в следствии метода Гаусса), тогда

$$A = A^T = LU = U^T L^T$$

$$U(L^T)^{-1} = L^{-1}U^T.$$

Разложение Холецкого

Достаточность: пусть A – симметричная положительно определенная матрица и пусть, $A = LU$, где L – обратимая нижняя треугольная матрица, а U – верхняя треугольная матрица (существует в следствии метода Гаусса), тогда

$$A = A^T = LU = U^T L^T$$

$$U(L^T)^{-1} = L^{-1}U^T.$$

В левой части последнего равенства находится произведение верхних треугольных матриц, в правой – произведение нижних треугольных матриц. Равенство возможно только в том случае, когда в обеих частях диагональные матрицы.

Разложение Холецкого

Достаточность: пусть A – симметричная положительно определенная матрица и пусть, $A = LU$, где L – обратимая нижняя треугольная матрица, а U – верхняя треугольная матрица (существует в следствии метода Гаусса), тогда

$$A = A^T = LU = U^T L^T$$

$$U(L^T)^{-1} = L^{-1}U^T.$$

В левой части последнего равенства находится произведение верхних треугольных матриц, в правой – произведение нижних треугольных матриц. Равенство возможно только в том случае, когда в обеих частях диагональные матрицы.

Обозначим $D = \text{diag}\{d_1, \dots, d_n\} = U(L^T)^{-1}$, тогда

$$A = LDL^T$$

Разложение Холецкого

Достаточность: пусть A – симметричная положительно определенная матрица и пусть, $A = LU$, где L – обратимая нижняя треугольная матрица, а U – верхняя треугольная матрица (существует в следствии метода Гаусса), тогда

$$A = A^T = LU = U^T L^T$$

$$U(L^T)^{-1} = L^{-1}U^T.$$

В левой части последнего равенства находится произведение верхних треугольных матриц, в правой – произведение нижних треугольных матриц. Равенство возможно только в том случае, когда в обоих частях диагональные матрицы.

Обозначим $D = \text{diag}\{d_1, \dots, d_n\} = U(L^T)^{-1}$, тогда

$$A = LDL^T$$

В силу обратимости L и положительной определенности A получаем положительную определенность D , т. е. $d_i \geq 0$, отсюда $A = L\sqrt{D}(L\sqrt{D})^T$, где $\sqrt{D} = \text{diag}\{\sqrt{d_1}, \dots, \sqrt{d_n}\}$.

Разложение Холецкого

Замечание 1. Если матрица A строго положительно определена, то разложение Холецкого единственно.

Разложение Холецкого

Замечание 1. Если матрица A строго положительно определена, то разложение Холецкого единственно.

Расписывая равенство

$$LL^T = A$$

покоординатно получаем получаем для $i \geq j$

$$a_{ij} = \sum_{k=1}^n l_{ik}l_{jk} = \sum_{k=1}^j l_{ik}l_{jk}$$

Разложение Холецкого

Замечание 1. Если матрица A строго положительно определена, то разложение Холецкого единственно.

Расписывая равенство

$$LL^T = A$$

покоординатно получаем получаем для $i \geq j$

$$a_{ij} = \sum_{k=1}^n \ell_{ik} \ell_{jk} = \sum_{k=1}^j \ell_{ik} \ell_{jk}$$

Из этих равенств можно получаем следующие рекуррентные формулы

$$\begin{aligned} \ell_{ii} &= \pm \sqrt{a_{ii} - \sum_{j=1}^{i-1} \ell_{ij}^2} \\ \ell_{ij} &= \frac{1}{\ell_{jj}} \left(a_{ij} - \sum_{k=1}^{j-1} \ell_{ik} \ell_{jk} \right), \quad j < i \end{aligned} \quad (1)$$

Разложение Холецкого

Если A положительно определена, то выражение под корнем всегда положительно, так как система должна иметь решение. В случае, если $\ell_{jj} = 0$, то ℓ_{ij} может быть выбрано произвольно.

Разложение Холецкого

Если A положительно определена, то выражение под корнем всегда положительно, так как система должна иметь решение. В случае, если $\ell_{jj} = 0$, то ℓ_{ij} может быть выбрано произвольно.

Предположим, что 0 является собственным числом L , тогда $0 = L^T v = LL^T v = Av$. Пусть теперь 0 является собственным числом A , $Av = 0$, тогда

$$0 = v^T Av = v^T LL^T v = \|L^T v\|^2 \Rightarrow L^T v = 0$$

Разложение Холецкого

Если A положительно определена, то выражение под корнем всегда положительно, так как система должна иметь решение. В случае, если $\ell_{jj} = 0$, то ℓ_{ij} может быть выбрано произвольно.

Предположим, что 0 является собственным числом L , тогда $0 = L^T v = LL^T v = Av$. Пусть теперь 0 является собственным числом A , $Av = 0$, тогда

$$0 = v^T Av = v^T LL^T v = \|L^T v\|^2 \Rightarrow L^T v = 0$$

Таким образом L содержит собственное число 0 тогда и только тогда когда A содержит с. ч. 0 . Следовательно для строго положительно определенной A формулы (1) задают единственную матрицу L с точностью до знака перед корнем. Итого имеем следующую теорему единственности

Разложение Холецкого

Если A положительно определена, то выражение под корнем всегда положительно, так как система должна иметь решение. В случае, если $\ell_{jj} = 0$, то ℓ_{ij} может быть выбрано произвольно.

Предположим, что 0 является собственным числом L , тогда $0 = L^T v = LL^T v = Av$. Пусть теперь 0 является собственным числом A , $Av = 0$, тогда

$$0 = v^T Av = v^T LL^T v = \|L^T v\|^2 \Rightarrow L^T v = 0$$

Таким образом L содержит собственное число 0 тогда и только тогда когда A содержит с. ч. 0 . Следовательно для строго положительно определенной A формулы (1) задают единственную матрицу L с точностью до знака перед корнем. Итого имеем следующую теорему единственности

Теорема

Для положительно определенной матрицы A существует единственная положительно определенная нижняя треугольная матрица L такая, что $A = LL^T$.

Рекурсивная форма разложения Холецкого

Пусть $A \in K^{n \times n}$ – симметричная строго положительно определенная матрица, $\ell \in K^{n-1}$, $\tilde{A} \in K^{n-1 \times n-1}$ и при этом

$$A = \begin{bmatrix} a_{11} & \ell^T \\ \ell & \tilde{A} \end{bmatrix}$$

Рекурсивная форма разложения Холецкого

Пусть $A \in K^{n \times n}$ – симметричная строго положительно определенная матрица, $\ell \in K^{n-1}$, $\tilde{A} \in K^{(n-1) \times (n-1)}$ и при этом

$$A = \begin{bmatrix} a_{11} & \ell^T \\ \ell & \tilde{A} \end{bmatrix}$$

Отметим, что имеет место равенство

$$\begin{bmatrix} a_{11} & \ell^T \\ \ell & \tilde{A} \end{bmatrix} = \begin{bmatrix} \sqrt{a_{11}} & 0_{n-1}^T \\ \frac{1}{\sqrt{a_{11}}}\ell & I_{n-1} \end{bmatrix} \begin{bmatrix} 1 & 0_{n-1}^T \\ 0_{n-1} & \tilde{A} - \frac{1}{a_{11}}\ell\ell^T \end{bmatrix} \begin{bmatrix} \sqrt{a_{11}} & \frac{1}{\sqrt{a_{11}}}\ell^T \\ 0_{n-1} & I_{n-1} \end{bmatrix}$$

Рекурсивная форма разложения Холецкого

Пусть $A \in K^{n \times n}$ – симметричная строго положительно определенная матрица, $\ell \in K^{n-1}$, $\tilde{A} \in K^{(n-1) \times (n-1)}$ и при этом

$$A = \begin{bmatrix} a_{11} & \ell^T \\ \ell & \tilde{A} \end{bmatrix}$$

Отметим, что имеет место равенство

$$\begin{bmatrix} a_{11} & \ell^T \\ \ell & \tilde{A} \end{bmatrix} = \begin{bmatrix} \sqrt{a_{11}} & 0_{n-1}^T \\ \frac{1}{\sqrt{a_{11}}}\ell & I_{n-1} \end{bmatrix} \begin{bmatrix} 1 & 0_{n-1}^T \\ 0_{n-1} & \tilde{A} - \frac{1}{a_{11}}\ell\ell^T \end{bmatrix} \begin{bmatrix} \sqrt{a_{11}} & \frac{1}{\sqrt{a_{11}}}\ell^T \\ 0_{n-1} & I_{n-1} \end{bmatrix}$$

Таким образом, обозначив за $HL(A)$ – единственную матрицу L , что $A = LL^T$ и используя факт, что при $B = RR^T$ выполняется

$$\begin{bmatrix} I & 0 \\ 0 & B \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & R^T \end{bmatrix}$$

Рекурсивная форма разложения Холецкого

Пусть $A \in K^{n \times n}$ – симметричная строго положительно определенная матрица, $\ell \in K^{n-1}$, $\tilde{A} \in K^{(n-1) \times (n-1)}$ и при этом

$$A = \begin{bmatrix} a_{11} & \ell^T \\ \ell & \tilde{A} \end{bmatrix}$$

Отметим, что имеет место равенство

$$\begin{bmatrix} a_{11} & \ell^T \\ \ell & \tilde{A} \end{bmatrix} = \begin{bmatrix} \sqrt{a_{11}} & 0_{n-1}^T \\ \frac{1}{\sqrt{a_{11}}}\ell & I_{n-1} \end{bmatrix} \begin{bmatrix} 1 & 0_{n-1}^T \\ 0_{n-1} & \tilde{A} - \frac{1}{a_{11}}\ell\ell^T \end{bmatrix} \begin{bmatrix} \sqrt{a_{11}} & \frac{1}{\sqrt{a_{11}}}\ell^T \\ 0_{n-1} & I_{n-1} \end{bmatrix}$$

Таким образом, обозначив за $HL(A)$ – единственную матрицу L , что $A = LL^T$ и используя факт, что при $B = RR^T$ выполняется

$$\begin{bmatrix} I & 0 \\ 0 & B \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & R^T \end{bmatrix}$$

получаем

$$HL(A) = \begin{bmatrix} \sqrt{a_{11}} & 0_{n-1}^T \\ \frac{1}{\sqrt{a_{11}}}\ell & HL(\tilde{A} - \frac{1}{a_{11}}\ell\ell^T) \end{bmatrix}$$

Итеративные схемы

Рассмотрим снова СЛАУ

$$Ax = b.$$

Итеративные схемы

Рассмотрим снова СЛАУ

$$Ax = b.$$

Если D – обратимая матрица, то система может быть эквивалентна переписана в виде

$$x = x - D(Ax - b) = (I - DA)x + Db$$

Итеративные схемы

Рассмотрим снова СЛАУ

$$Ax = b.$$

Если D – обратимая матрица, то система может быть эквивалентна переписана в виде

$$x = x - D(Ax - b) = (I - DA)x + Db$$

Общая итеративная схема решения СЛАУ заключается в построении последовательности

$$x_{k+1} = (I - DA)x_k + Db \quad (2)$$

Итеративные схемы

Рассмотрим снова СЛАУ

$$Ax = b.$$

Если D – обратимая матрица, то система может быть эквивалентна переписана в виде

$$x = x - D(Ax - b) = (I - DA)x + Db$$

Общая итеративная схема решения СЛАУ заключается в построении последовательности

$$x_{k+1} = (I - DA)x_k + Db \quad (2)$$

Теорема

Если матрица $I - DA$ положительно определена, $\sigma(I - DA) < 1$ и D обратима, то (2) сходится к решению системы $Ax = b$.

Док-во. Сходимость и скорость сходимости была показана раньше (сходимость линейных итеративных процессов). Пусть $x_k \rightarrow x^*$, тогда $x^* = x^* - D(Ax^* - b)$, т. е. $D(Ax^* - b) = 0$, из обратимости D следует $Ax^* = b$. ■

Два способа построения итеративных схем

Основная проблема обычно заключается в том, чтобы подобрать D так, что $\sigma(I - DA) < 1$. Два наиболее распространенных случая:

1. $0 \preceq A \preceq \gamma I$, тогда годится выбор $D = \gamma^{-1}I$. Для симметричной матрицы A этот метод идентичен применению градиентного спуска к функции

$$f(x) = \frac{1}{2}x^T Ax - b^T x + c$$

Два способа построения итеративных схем

Основная проблема обычно заключается в том, чтобы подобрать D так, что $\sigma(I - DA) < 1$. Два наиболее распространенных случая:

1. $0 \preceq A \preceq \gamma I$, тогда годится выбор $D = \gamma^{-1}I$. Для симметричной матрицы A этот метод идентичен применению градиентного спуска к функции

$$f(x) = \frac{1}{2}x^T Ax - b^T x + c$$

2. A имеет диагональное преобладание, т. е.

$$|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|.$$

В этом случае годится выбор $D = \text{diag}\{a_{11}^{-1}, \dots, a_{nn}^{-1}\}$. Этот метод принято называть *методом Якоби*.

Теорема Гершгорина

Теорема (Гершгорина о кругах)

Пусть $A = (a_{ij}) \in \mathbb{C}^{n \times n}$. Для любого собственного числа λ матрицы A найдется $1 \leq i \leq n$ такое, что

$$|\lambda - a_{ii}| \leq \sum_{j \neq i} |a_{ij}|$$

Теорема Гершгорина

Теорема (Гершгорина о кругах)

Пусть $A = (a_{ij}) \in \mathbb{C}^{n \times n}$. Для любого собственного числа λ матрицы A найдется $1 \leq i \leq n$ такое, что

$$|\lambda - a_{ii}| \leq \sum_{j \neq i} |a_{ij}|$$

Док-во. Пусть $Ax = \lambda x$ и $i = \operatorname{argmax}_j |x_j|$. Таким образом

$$\sum_j a_{ij} x_j = \lambda x_i$$

Поделив обе части на x_i получаем

$$\sum_{j \neq i} a_{ij} \frac{x_j}{x_i} + a_{ii} = \lambda$$

Теорема Гершгорина

Теорема (Гершгорина о кругах)

Пусть $A = (a_{ij}) \in \mathbb{C}^{n \times n}$. Для любого собственного числа λ матрицы A найдется $1 \leq i \leq n$ такое, что

$$|\lambda - a_{ii}| \leq \sum_{j \neq i} |a_{ij}|$$

Док-во. Пусть $Ax = \lambda x$ и $i = \operatorname{argmax}_j |x_j|$. Таким образом

$$\sum_j a_{ij} x_j = \lambda x_i$$

Поделив обе части на x_i получаем

$$\sum_{j \neq i} a_{ij} \frac{x_j}{x_i} + a_{ii} = \lambda$$

Таким образом

$$|\lambda - a_{ii}| = \left| \sum_{j \neq i} a_{ij} \frac{x_j}{x_i} \right| \leq \sum_{j \neq i} |a_{ij}| \quad \blacksquare$$

Теорема Гершгорина

Замечание 1. Вещественная часть любого собственного числа вещественной матрицы с диагональным преобладанием и строго положительными диагональными элементами неотрицательна.

Теорема Гершгорина

Замечание 1. Вещественная часть любого собственного числа вещественной матрицы с диагональным преобладанием и строго положительными диагональными элементами неотрицательна.

Замечание 2. В методе Якоби выполняется $0 \preceq (I - DA) \preceq I$.

Определение

Пусть A – симметричная матрица. Вектора u, v называются A -ортогональными или сопряженными, если

$$u^T A v = 0$$

Определение

Пусть A – симметричная матрица. Вектора u, v называются A -ортогональными или сопряженными, если

$$u^T A v = 0$$

Рассмотрим задачу минимизации

$$f(x) = \frac{1}{2} x^T A x - b^T x + c,$$

где A – симметричная обратимая матрица, таким образом $\nabla f(x) = Ax - b$, а значит нахождение точки минимума f равносильно решению системы $Ax = b$. Обозначим за x^* единственную точку минимума f .

Метод сопряженных направлений

Предположим, что нам известны n попарно сопряженных направлений d_0, \dots, d_{n-1} . Выберем произвольную точку x_0 и сделаем по очереди n шагов градиентного спуска по каждому из направлений, выбирая размер шага как минимум по направлению.

$$x_{k+1} = x_k - \alpha_k d_k$$

Метод сопряженных направлений

Предположим, что нам известны n попарно сопряженных направлений d_0, \dots, d_{n-1} . Выберем произвольную точку x_0 и сделаем по очереди n шагов градиентного спуска по каждому из направлений, выбирая размер шага как минимум по направлению.

$$x_{k+1} = x_k - \alpha_k d_k$$

Получаем α_k из уравнения $\frac{d}{d\alpha} f(x_k - \alpha d_k) = 0$:

$$0 = \frac{d}{d\alpha} f(x_k - \alpha d_k) = -d_k^T (A(x_k - \alpha d_k) - b) = \alpha d_k^T A d_k^T - d_k^T (A x_k - b)$$

Метод сопряженных направлений

Предположим, что нам известны n попарно сопряженных направлений d_0, \dots, d_{n-1} . Выберем произвольную точку x_0 и сделаем по очереди n шагов градиентного спуска по каждому из направлений, выбирая размер шага как минимум по направлению.

$$x_{k+1} = x_k - \alpha_k d_k$$

Получаем α_k из уравнения $\frac{d}{d\alpha} f(x_k - \alpha d_k) = 0$:

$$0 = \frac{d}{d\alpha} f(x_k - \alpha d_k) = -d_k^T (A(x_k - \alpha d_k) - b) = \alpha d_k^T A d_k^T - d_k^T (A x_k - b)$$

$$\alpha_k = \frac{d_k^T (A x_k - b)}{d_k^T A d_k}$$

Метод сопряженных направлений

Теперь предположим, что (d_0, \dots, d_{n-1}) – базис в \mathbb{R}^n , тогда

$$x_0 - x^* = \sum_{i=0}^{n-1} \delta_i d_i$$

Метод сопряженных направлений

Теперь предположим, что (d_0, \dots, d_{n-1}) – базис в \mathbb{R}^n , тогда

$$x_0 - x^* = \sum_{i=0}^{n-1} \delta_i d_i$$

Умножая это равенство на $d_k^T A$ получаем

$$d_k^T A(x_0 - x^*) = \sum_{i=0}^{n-1} \delta_i d_k^T A d_i = \delta_k d_k^T A d_k$$

и получаем следующие равенства для δ

$$\begin{aligned} \delta_k &= \frac{d_k^T A(x_0 - x^*)}{d_k^T A d_k} \\ &= \frac{d_k^T A(x_0 - x^* - \sum_{i=0}^{k-1} \alpha_i d_i)}{d_k^T A d_k} = \frac{d_k^T A(x_k - x^*)}{d_k^T A d_k} = \alpha_k \end{aligned}$$

Метод сопряженных направлений

Таким образом получаем

$$x_k - x^* = x_0 - \sum_{i=0}^{k-1} \alpha_i d_i - x^* = \sum_{i=0}^{n-1} \alpha_i d_i - \sum_{i=0}^{k-1} \alpha_i d_i = \sum_{i=k}^{n-1} \alpha_i d_i,$$

что гарантирует сходимость этой процедуры за n шагов.

Метод сопряженных направлений

Таким образом получаем

$$x_k - x^* = x_0 - \sum_{i=0}^{k-1} \alpha_i d_i - x^* = \sum_{i=0}^{n-1} \alpha_i d_i - \sum_{i=0}^{k-1} \alpha_i d_i = \sum_{i=k}^{n-1} \alpha_i d_i,$$

что гарантирует сходимость этой процедуры за n шагов. Более того,

$$\|x_k - x^*\|_A^2 = (x_k - x^*)^T A (x_k - x^*) = \sum_{i=k}^{n-1} \sum_{j=k}^{n-1} \alpha_i \alpha_j d_i^T A d_j = \sum_{i=k}^{n-1} \alpha_i^2 d_i^T A d_i$$

Пусть $x \in x_0 + \text{Span}\{d_0, \dots, d_{k-1}\}$, $x = x_0 + \sum_{i=0}^{k-1} \beta_i d_i$, тогда

$$\begin{aligned} \|x - x^*\|_A^2 &= \left(\sum_{i=0}^{k-1} \beta_i d_i + \sum_{i=0}^{n-1} \alpha_i d_i \right)^T A \left(\sum_{i=0}^{k-1} \beta_i d_i + \sum_{i=0}^{n-1} \alpha_i d_i \right) \\ &= \sum_{i=0}^{k-1} (\beta_i + \alpha_i)^2 d_i^T A d_i + \sum_{i=k}^{n-1} \alpha_i^2 d_i^T A d_i \geq \|x_k - x^*\|_A^2 \end{aligned}$$

Метод сопряженных направлений

Важное свойство метода сопряженных направлений:

$$x_k = \operatorname{argmin}_{x \in x_0 + \operatorname{Span}\{d_0, \dots, d_{k-1}\}} \|x - x^*\|_A$$

Метод сопряженных направлений

Важное свойство метода сопряженных направлений:

$$x_k = \operatorname{argmin}_{x \in x_0 + \operatorname{Span}\{d_0, \dots, d_{k-1}\}} \|x - x^*\|_A$$

Остается вопрос: как найти d_0, \dots, d_{n-1} ?

Сопряжение Грама-Шмидта

Пусть $v_1, \dots, v_n \in \mathbb{R}^n$, рассмотрим u_1, \dots, u_n :

$$u_k = v_k - \sum_{i=1}^{k-1} \frac{u_i^T A v_k}{u_i^T A u_i} u_i$$

Сопряжение Грама-Шмидта

Пусть $v_1, \dots, v_n \in \mathbb{R}^n$, рассмотрим u_1, \dots, u_n :

$$u_k = v_k - \sum_{i=1}^{k-1} \frac{u_i^T Av_k}{u_i^T Au_i} u_i$$

Аналогично обычной процедуре Грама-Шмидта докажем по индукции, что

$u_j^T Au_k = 0$ при $j \neq k$. Пусть это утверждение верно вплоть до $k - 1$, тогда для $j < k$

$$u_j^T Au_k = u_j^T Av_k - \sum_{i=1}^{k-1} \frac{u_i^T Av_k}{u_i^T Au_i} u_j^T Au_i = u_j^T Av_k - u_j^T Av_k = 0.$$

Сопряжение Грама-Шмидта

Пусть $v_1, \dots, v_n \in \mathbb{R}^n$, рассмотрим u_1, \dots, u_n :

$$u_k = v_k - \sum_{i=1}^{k-1} \frac{u_i^T A v_k}{u_i^T A u_i} u_i$$

Аналогично обычной процедуре Грама-Шмидта докажем по индукции, что

$u_j^T A u_k = 0$ при $j \neq k$. Пусть это утверждение верно вплоть до $k - 1$, тогда для $j < k$

$$u_j^T A u_k = u_j^T A v_k - \sum_{i=1}^{k-1} \frac{u_i^T A v_k}{u_i^T A u_i} u_j^T A u_i = u_j^T A v_k - u_j^T A v_k = 0.$$

Также по индукции доказывается равенство линейных оболочек $\text{Span}\{v_1, \dots, v_k\} = \text{Span}\{u_1, \dots, u_k\}$.

Общая схема метода сопряженных направлений

Итого, на данный момент имеем следующую общую схему:

Общая схема метода сопряженных направлений

Итого, на данный момент имеем следующую общую схему:

1. Выбрать n линейно независимых векторов v_0, \dots, v_{n-1} .

Общая схема метода сопряженных направлений

Итого, на данный момент имеем следующую общую схему:

1. Выбрать n линейно независимых векторов v_0, \dots, v_{n-1} .
2. Построить n сопряженных относительно матрицы A направлений d_0, \dots, d_{n-1} по формулам

$$d_k = v_k - \sum_{i=0}^{k-1} \frac{d_i^T A v_k}{d_i^T A d_i} d_i$$

Общая схема метода сопряженных направлений

Итого, на данный момент имеем следующую общую схему:

1. Выбрать n линейно независимых векторов v_0, \dots, v_{n-1} .
2. Построить n сопряженных относительно матрицы A направлений d_0, \dots, d_{n-1} по формулам

$$d_k = v_k - \sum_{i=0}^{k-1} \frac{d_i^T A v_k}{d_i^T A d_i} d_i$$

3. Выбрать произвольную точку x_0 и построить последовательность

$$x_{k+1} = x_k - \frac{d_k^T (A x_k - b)}{d_k^T A d_k} d_k$$

Общая схема метода сопряженных направлений

Итого, на данный момент имеем следующую общую схему:

1. Выбрать n линейно независимых векторов v_0, \dots, v_{n-1} .
2. Построить n сопряженных относительно матрицы A направлений d_0, \dots, d_{n-1} по формулам

$$d_k = v_k - \sum_{i=0}^{k-1} \frac{d_i^T A v_k}{d_i^T A d_i} d_i$$

3. Выбрать произвольную точку x_0 и построить последовательность

$$x_{k+1} = x_k - \frac{d_k^T (A x_k - b)}{d_k^T A d_k} d_k$$

Общая схема метода сопряженных направлений

Итого, на данный момент имеем следующую общую схему:

1. Выбрать n линейно независимых векторов v_0, \dots, v_{n-1} .
2. Построить n сопряженных относительно матрицы A направлений d_0, \dots, d_{n-1} по формулам

$$d_k = v_k - \sum_{i=0}^{k-1} \frac{d_i^T A v_k}{d_i^T A d_i} d_i$$

3. Выбрать произвольную точку x_0 и построить последовательность

$$x_{k+1} = x_k - \frac{d_k^T (A x_k - b)}{d_k^T A d_k} d_k$$

Из показанного ранее вытекает, что $A x_n = b$. Если m – количество ненулевых элементов A , то умножение A на вектор требует $\mathcal{O}(m)$ сложений и умножений. Таким образом, шаг 2 требует суммарно $\mathcal{O}(n^2 m)$ действий, шаг 3 – $\mathcal{O}(nm)$.

Подпространства Крылова

Определение

Подпространством Крылова порядка k матрицы A и вектора b называется

$$\mathcal{K}_k(A, b) = \text{Span}\{b, Ab, \dots, A^{k-1}b\}, \quad \mathcal{K}_0(A, b) = \{0\}.$$

Подпространства Крылова

Определение

Подпространством Крылова порядка k матрицы A и вектора b называется

$$\mathcal{K}_k(A, b) = \text{Span}\{b, Ab, \dots, A^{k-1}b\}, \quad \mathcal{K}_0(A, b) = \{0\}.$$

Пусть $\chi_A(t) = \det(A - tI) = \sum_{i=0}^{n-1} \alpha_i t^i - t^n$ – характеристический полином A .
Из теоремы Гамильтона-Кэли

$$\chi(A) = \sum_{i=0}^{n-1} \alpha_i A^i - A^n = 0$$

Подпространства Крылова

Определение

Подпространством Крылова порядка k матрицы A и вектора b называется

$$\mathcal{K}_k(A, b) = \text{Span}\{b, Ab, \dots, A^{k-1}b\}, \quad \mathcal{K}_0(A, b) = \{0\}.$$

Пусть $\chi_A(t) = \det(A - tI) = \sum_{i=0}^{n-1} \alpha_i t^i - t^n$ – характеристический полином A .
Из теоремы Гамильтона-Кэли

$$\chi(A) = \sum_{i=0}^{n-1} \alpha_i A^i - A^n = 0$$

Умножая на $A^{-1}b$ и учитывая $\alpha_0 \neq 0$ для положительно определенной матрицы A получаем

$$A^{-1}b = \frac{1}{\alpha_0} \left(A^{n-1}b - \sum_{i=1}^{n-1} \alpha_i A^{i-1}b \right) \in \mathcal{K}_n(A, b)$$

Последовательность Крылова

Определение

Последовательностью Крылова функции $f(x) = \frac{1}{2}x^T Ax - b^T x$ и начальной точкой x_0 называется последовательность

$$x_k = \operatorname{argmin}_{x \in x_0 + \mathcal{K}_k(A, Ax_0 - b)} f(x)$$

Последовательность Крылова

Определение

Последовательностью Крылова функции $f(x) = \frac{1}{2}x^T Ax - b^T x$ и начальной точкой x_0 называется последовательность

$$x_k = \operatorname{argmin}_{x \in x_0 + \mathcal{K}_k(A, Ax_0 - b)} f(x)$$

Из условий оптимальности для x_k

$$\nabla f(x_k) = Ax_k - b \perp \mathcal{K}_k(A, Ax_0 - b)$$

С другой стороны, раз $x_k \in x_0 + \mathcal{K}_k(A, Ax_0 - b)$, то очевидным образом $Ax_k - b \in Ax_0 - b + A\mathcal{K}_k(A, Ax_0 - b) \subset \mathcal{K}_{k+1}(A, Ax_0 - b)$.

Последовательность Крылова

Определение

Последовательностью Крылова функции $f(x) = \frac{1}{2}x^T Ax - b^T x$ и начальной точкой x_0 называется последовательность

$$x_k = \operatorname{argmin}_{x \in x_0 + \mathcal{K}_k(A, Ax_0 - b)} f(x)$$

Из условий оптимальности для x_k

$$\nabla f(x_k) = Ax_k - b \perp \mathcal{K}_k(A, Ax_0 - b)$$

С другой стороны, раз $x_k \in x_0 + \mathcal{K}_k(A, Ax_0 - b)$, то очевидным образом $Ax_k - b \in Ax_0 - b + A\mathcal{K}_k(A, Ax_0 - b) \subset \mathcal{K}_{k+1}(A, Ax_0 - b)$.

Таким образом $\nabla f(x_k) \notin \mathcal{K}_k(A, Ax_0 - b)$, но $\nabla f(x_k) \in \mathcal{K}_{k+1}(A, Ax_0 - b) \Rightarrow$

$$\mathcal{K}_k(A, Ax_0 - b) = \operatorname{Span}\{\nabla f(x_0), \dots, \nabla f(x_{k-1})\},$$

в силу $\nabla f(x_k) \in \mathcal{K}_k(A, Ax_0 - b)^\perp$ при $i \neq j$

$$\nabla f(x_i)^T \nabla f(x_j) = 0$$

Метод сопряженных градиентов

Основная идея метода сопряженных градиентов: использование $v_k = \nabla f(x_k)$ в методе сопряженных направлений.

Метод сопряженных градиентов

Основная идея метода сопряженных градиентов: использование $v_k = \nabla f(x_k)$ в методе сопряженных направлений.

1. Используя индукцию покажем, что при $v_k = \nabla f(x_k)$ метод сопряженных направлений генерирует последовательность Крылова для f , x_0 :

База. $k = 0$ – тривиально.

Индукционный переход. Пусть метод сгенерировал последовательность Крылова вплоть до $k - 1$, тогда так как метод сопряженных направлений выбирает

$$x_k = \operatorname{argmin}_{x \in x_0 + \operatorname{Span}\{d_0, \dots, d_{k-1}\}} \|x - x^*\|_A^2,$$

учитывая

$$\|x - x^*\|_A^2 = (x - x^*)^T A(x - x^*) = x^{*T} A x^* - 2x^{*T} A x + x^T A x = \|x^*\|_A^2 + 2f(x)$$

получаем, что минимизация $\|x - x^*\|_A$ равносильна минимизации $f(x)$. По построению

$$\operatorname{Span}\{d_0, \dots, d_{k-1}\} = \operatorname{Span}\{\nabla f(x_0), \dots, \nabla f(x_{k-1})\},$$

по индукционному предположению

$$\operatorname{Span}\{\nabla f(x_0), \dots, \nabla f(x_{k-1})\} = \mathcal{K}_k(A, Ax_0 - b). \blacksquare$$

Метод сопряженных градиентов

2. Вычисление d_k сильно упрощается:

$$v_i^T (Ax_k - b) = v_i^T (Ax_{k-1} - b) - \alpha_{k-1} v_i^T Ad_{k-1}$$

$$v_i^T Ad_{k-1} = \frac{1}{\alpha_{k-1}} (v_i^T v_{k-1} - v_i^T v_k^T)$$

Метод сопряженных градиентов

2. Вычисление d_k сильно упрощается:

$$v_i^T (Ax_k - b) = v_i^T (Ax_{k-1} - b) - \alpha_{k-1} v_i^T A d_{k-1}$$

$$v_i^T A d_{k-1} = \frac{1}{\alpha_{k-1}} (v_i^T v_{k-1} - v_i^T v_k^T)$$

Так как $v_i^T v_j = 0$ при $i \neq j$, то $d_i^T A v_k$ отлично от нуля только при $i = k$ или $i = k - 1$. Итого шаг 2 имеет вид

$$\begin{aligned} d_k &= v_k - \sum_{i=0}^{k-1} \frac{d_i^T A v_k}{d_i^T A d_i} d_i = v_k - \frac{v_k^T v_k}{\alpha_{k-1} d_{k-1}^T A d_{k-1}} d_{k-1} \\ &= v_k - \frac{v_k^T v_k}{d_{k-1}^T v_{k-1}} d_{k-1} = v_k - \frac{v_k^T v_k}{v_{k-1}^T v_{k-1}} d_{k-1} \end{aligned}$$

Метод сопряженных градиентов

2. Вычисление d_k сильно упрощается:

$$v_i^T (Ax_k - b) = v_i^T (Ax_{k-1} - b) - \alpha_{k-1} v_i^T A d_{k-1}$$

$$v_i^T A d_{k-1} = \frac{1}{\alpha_{k-1}} (v_i^T v_{k-1} - v_i^T v_k^T)$$

Так как $v_i^T v_j = 0$ при $i \neq j$, то $d_i^T A v_k$ отлично от нуля только при $i = k$ или $i = k - 1$. Итого шаг 2 имеет вид

$$\begin{aligned} d_k &= v_k - \sum_{i=0}^{k-1} \frac{d_i^T A v_k}{d_i^T A d_i} d_i = v_k - \frac{v_k^T v_k}{\alpha_{k-1} d_{k-1}^T A d_{k-1}} d_{k-1} \\ &= v_k - \frac{v_k^T v_k}{d_{k-1}^T v_{k-1}} d_{k-1} = v_k - \frac{v_k^T v_k}{v_{k-1}^T v_{k-1}} d_{k-1} \end{aligned}$$

Это соотношение уменьшает сложность шага 2 с $\mathcal{O}(km)$ до $\mathcal{O}(m)$ на итерации k .

Последовательность Крылова

Замечание.

$$x \in x_0 + \mathcal{K}_k(A, Ax_0 - b) \Leftrightarrow x = x_0 + \sum_{i=0}^{k-1} \phi_i A^i (Ax_0 - b)$$

Если $Ax^* = b$, то

$$x - x^* = x_0 - x^* + \sum_{i=0}^{k-1} \phi_i A^i A(x_0 - x^*) = \left(I + \sum_{i=1}^k \phi_{i-1} A^i \right) (x_0 - x^*)$$

Последовательность Крылова

Замечание.

$$x \in x_0 + \mathcal{K}_k(A, Ax_0 - b) \Leftrightarrow x = x_0 + \sum_{i=0}^{k-1} \phi_i A^i (Ax_0 - b)$$

Если $Ax^* = b$, то

$$x - x^* = x_0 - x^* + \sum_{i=0}^{k-1} \phi_i A^i A(x_0 - x^*) = \left(I + \sum_{i=1}^k \phi_{i-1} A^i \right) (x_0 - x^*)$$

Таким образом выбор x_k как минимум $\|x - x^*\|_A$ на множестве $x_0 + \mathcal{K}_k(A, Ax_0 - b)$ можно описать следующим образом

$$x_k = x^* + \operatorname{argmin}_{P(0)=1, \deg P \leq k} P(A)(x_0 - x^*)$$

Асимптотический анализ метода СГ

Отметим для начала, что для любого многочлена P и собственного вектора ν : $A\nu = \lambda\nu$ выполняется

$$P(A)\nu = \sum_{i=0}^k \alpha_i A^i \nu = \sum_{i=0}^k \alpha_i \lambda^i \nu = P(\lambda)\nu$$

Асимптотический анализ метода СГ

Отметим для начала, что для любого многочлена P и собственного вектора ν : $A\nu = \lambda\nu$ выполняется

$$P(A)\nu = \sum_{i=0}^k \alpha_i A^i \nu = \sum_{i=0}^k \alpha_i \lambda^i \nu = P(\lambda)\nu$$

Пусть $\lambda_{min} = \lambda_1 \leq \dots, \leq \lambda_n = \lambda_{max}$ – собственные значения A (с учетом кратности), ν_1, \dots, ν_n – собственные векторы.

Асимптотический анализ метода СГ

Так как A – вещественная симметричная матрица, то ν образуют ортогональный базис,

$$x_0 - x^* = \sum_{j=1}^n \xi_j \nu_j.$$

Асимптотический анализ метода СГ

Так как A – вещественная симметричная матрица, то ν образуют ортогональный базис,

$$x_0 - x^* = \sum_{j=1}^n \xi_j \nu_j.$$

Тогда

$$x_k - x^* = \sum_{j=1}^n \xi_j P_k(\lambda_j) \nu_j$$

$$A(x_k - x^*) = \sum_{j=1}^n \xi_j P_k(\lambda_j) \lambda_j \nu_j$$

Асимптотический анализ метода СГ

Так как A – вещественная симметричная матрица, то ν образуют ортогональный базис,

$$x_0 - x^* = \sum_{j=1}^n \xi_j \nu_j.$$

Тогда

$$\begin{aligned} x_k - x^* &= \sum_{j=1}^n \xi_j P_k(\lambda_j) \nu_j & \|x_k - x^*\|_A^2 &= \sum_{j=1}^n P_k(\lambda_j)^2 \lambda_j \xi_j^2 \|\nu_j\|^2 \\ A(x_k - x^*) &= \sum_{j=1}^n \xi_j P_k(\lambda_j) \lambda_j \nu_j & &= \sum_{j=1}^n P_k(\lambda_j)^2 \|x_0 - x^*\|_A^2 \\ & & &\leq \max_{\lambda \in \Lambda(A)} P_k(\lambda)^2 \|x_0 - x^*\|_A^2, \end{aligned}$$

где $\Lambda(A)$ – спектр матрицы A (множество её собственных значений).

Многочлен градиентного спуска

Выбирая многочлен $G_k(x) = \left(1 - \frac{2x}{\lambda_{\min} + \lambda_{\max}}\right)^k$ получаем

$$\begin{aligned}\|x_k - x^*\|_A^2 &\leq \max_{\lambda \in \Lambda(A)} G_k(\lambda)^2 \|x_0 - x^*\|_A^2 \\ &= \max_{\lambda \in \Lambda(A)} \left(\frac{\lambda_{\min} + \lambda_{\max} - 2\lambda}{\lambda_{\min} + \lambda_{\max}} \right)^{2k} \|x_0 - x^*\|_A^2 \\ &= \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\min} + \lambda_{\max}} \right)^{2k} \|x_0 - x^*\|_A^2,\end{aligned}$$

что соответствует сходимости градиентного спуска.

Многочлен Чебышева

Выбирая $P_k(\lambda) = \frac{T_k\left(\frac{\lambda_{\max} + \lambda_{\min} - 2\lambda}{\lambda_{\max} - \lambda_{\min}}\right)}{T_k\left(\frac{\lambda_{\max} + \lambda_{\min}}{\lambda_{\max} - \lambda_{\min}}\right)}$, где T_k – многочлен Чебышева

$$T_k(x) = \frac{1}{2} \left[(x + \sqrt{x^2 - 1})^k + (x - \sqrt{x^2 - 1})^k \right]$$

учитывая $T_k(x) \leq 1$ при $x \in [-1, 1]$ получаем $P_k(\lambda) \leq T_k\left(\frac{\lambda_{\max} + \lambda_{\min}}{\lambda_{\max} - \lambda_{\min}}\right)^{-1}$ при $\lambda \in [\lambda_{\min}, \lambda_{\max}]$

Многочлен Чебышева

Выбирая $P_k(\lambda) = \frac{T_k\left(\frac{\lambda_{\max} + \lambda_{\min} - 2\lambda}{\lambda_{\max} - \lambda_{\min}}\right)}{T_k\left(\frac{\lambda_{\max} + \lambda_{\min}}{\lambda_{\max} - \lambda_{\min}}\right)}$, где T_k – многочлен Чебышева

$$T_k(x) = \frac{1}{2} \left[(x + \sqrt{x^2 - 1})^k + (x - \sqrt{x^2 - 1})^k \right]$$

учитывая $T_k(x) \leq 1$ при $x \in [-1, 1]$ получаем $P_k(\lambda) \leq T_k\left(\frac{\lambda_{\max} + \lambda_{\min}}{\lambda_{\max} - \lambda_{\min}}\right)^{-1}$ при $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ и следовательно

$$\begin{aligned} \|x_k - x^*\|_A^2 &\leq \max_{\lambda \in \Lambda(A)} P_k(\lambda)^2 \|x_0 - x^*\|_A^2 \\ &\leq T_k\left(\frac{\lambda_{\max} + \lambda_{\min}}{\lambda_{\max} - \lambda_{\min}}\right)^{-1} \|x_0 - x^*\|_A^2 \\ &\leq 2 \left[\left(\frac{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}{\sqrt{\lambda_{\max}} - \sqrt{\lambda_{\min}}}\right)^k + \left(\frac{\sqrt{\lambda_{\max}} - \sqrt{\lambda_{\min}}}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}\right)^k \right]^{-1} \|x_0 - x^*\|_A^2 \\ &\leq 2 \left(\frac{\sqrt{\lambda_{\max}} - \sqrt{\lambda_{\min}}}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}\right)^k \|x_0 - x^*\|_A^2 \end{aligned}$$

Ссылки на литературу

Шарый С. П. Курс вычислительных методов // параграф 3.6

Vandenberghe L. Лекция по разложению Холецкого

Shewchuk J. R. An Introduction to the Conjugate Gradient Method Without the Agonizing Pain