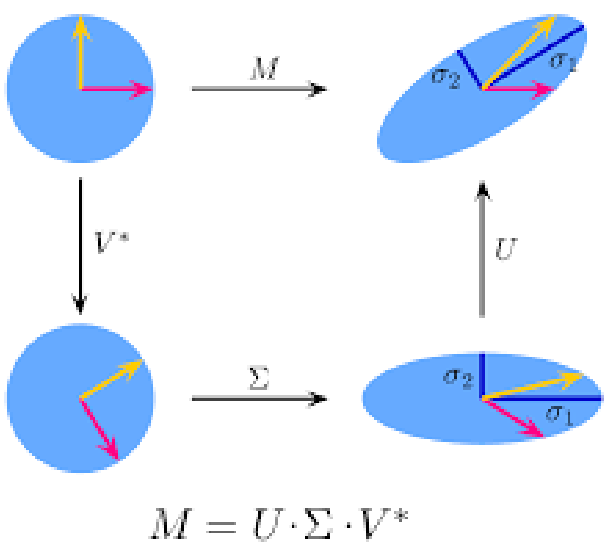
09.04.03 Прикладная информатика

Профиль «Машинное обучение и анализ данных»

Дисциплина «Математические основы анализа данных»

Лекция 5

Разложения матриц и их приложения



План лекции

- LU-разложение матриц
- Решение СЛУ методом LU-разложения
- Метод Холецкого решения СЛУ
- Сингулярное разложение матриц
- Приложения сингулярного разложения

LU-разложение матриц

- **LU-разложение (LU-декомпозиция, LU-факторизация)** — представление матрицы A в виде произведения двух матриц, $A = LU$, где L — нижняя (Lower) треугольная матрица, а U — верхняя (Upper) треугольная матрица.
- LU-разложение используется для решения систем линейных уравнений, обращения матриц и вычисления определителя.
- LU-разложение существует только в том случае, когда матрица A обратима, а все главные миноры матрицы A невырождены.

Алгоритм LU-разложения

- Пусть A – прямоугольная матрица порядка m×n любого ранга.
- С правой стороны матрицы A приписываем единичную матрицу E порядка $m \times m$.
- Применяем к матрице $A|E$ метод исключения Гаусса.
- Если на каком-то этапе ведущий элемент равен нулю, и существует ненулевой элемент, расположенный ниже ведущего элемента, то LU-разложение данной матрицы невозможно.
- Если же элементы ниже ведущего элемента нулевые, то выбираем новый ведущий элемент той же строки и следующего столбца.

- Приводим матрицу $A\,|\,E$ к треугольному или ступенчатому виду.
- Получим матрицу $U\,|\,L_0$, где $U-$ верхняя треугольная или ступенчатая матрица, а L_0- нижняя треугольная матрица.
- Заметим, что полученная матрица L_0 приводит A к треугольному или ступенчатому виду:

$$L_0A = U.$$

- Так как L_0- квадратная невырожденная матрица, следовательно имеет обратную матрицу. Тогда $A = L_0^{-1}U$ или $A = LU$.

Пример LU – разложения:

$$A=\left\|\begin{matrix} 1 & 2 & 1 \\ 2 & 1 & 1 \\ 1 & -1 & 2 \end{matrix}\right\| \qquad \begin{array}{c} \begin{array}{cc} A \qquad\qquad & \qquad E \end{array} \\ \left(\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 2 & 1 & 1 & 0 & 1 & 0 \\ 1 & -1 & 2 & 0 & 0 & 1 \end{array}\right) \end{array} \to \left(\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & -3 & -1 & -2 & 1 & 0 \\ 0 & -3 & 1 & -1 & 0 & 1 \end{array}\right) \to$$

$$\begin{array}{c} \begin{array}{cc} U \qquad\qquad & \qquad L_0 \end{array} \\ \left(\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & -3 & -1 & -2 & 1 & 0 \\ 0 & 0 & 2 & 1 & -1 & 1 \end{array}\right) \end{array} \qquad L=L_0^{-1}=\left\|\begin{matrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 1 & 1 \end{matrix}\right\| \quad U=\left\|\begin{matrix} 1 & 2 & 1 \\ 0 & -3 & -1 \\ 0 & 0 & 2 \end{matrix}\right\|$$

Решение СЛАУ методом LU-разложения

- Если матрица A исходной СЛУ разложена в произведение треугольных матриц L и U, то, значит, можно записать эквивалентное уравнение:

$$Ax = b \to LUx = b$$

- Введём вектор вспомогательных переменных y, тогда последнее выражение можно переписать в виде системы:

$$\begin{cases} Ly = b \\ Ux = y \end{cases}$$

- Таким образом, решение данной системы с квадратной матрицей коэффициентов сводится к последовательному решению двух систем с треугольными матрицами коэффициентов.

- Пример: Используя метод LU-разложения, решить систему
$$\begin{cases} 2x_1 + 6x_2 + 4x_3 = 8 \\ x_1 + 5x_2 + 4x_3 = 8 \\ x_1 + 5x_2 + 7x_3 = 17 \end{cases}$$

- Матрица системы:
$$A = \begin{pmatrix} 2 & 6 & 4 \\ 1 & 5 & 4 \\ 1 & 5 & 7 \end{pmatrix}$$

- Находим её LU-разложение:
$$L = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 3 \end{pmatrix} \quad U = \begin{pmatrix} 1 & 3 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

- Решение исходной системы сводится к последовательному решению систем

$$\begin{cases} 2y_1 = 8 \\ y_1 + 2y_2 = 8 \\ y_1 + 2y_2 + 3y_3 = 17 \end{cases} \qquad \begin{cases} x_1 + 3x_2 + 2x_3 = y_1 \\ x_2 + x_3 = y_2 \\ x_3 = y_3 \end{cases}$$

- Решаем первую: y1 = 8 / 2 = 4

  2y2 = 8 – y1 = 8 – 4 = 4 => y2 = 4 / 2 = 2

  3y3 = 17 – y1 – 2y2 = 17 – 4 – 2 × 2 = 9 => y3 = 3.

- Находим столбец неизвестных x:

  x3 = y3 = 3;

  x2 = y2 – x3 = 2 – 3 = -1;

  x1 = y1 – 3x2 – 2x3 = 4 – 3 (-1) – 2 × 3 = 1.

Метод Холецкого (метод квадратного корня)

- Метод использует **разложение Холецкого**, которое впервые было предложено французским офицером и математиком Андре-Луи Холецким (André-Louis Cholesky, 15 октября 1875 – 31 августа 1918).
- Сам Холецкий погиб в бою в Первой Мировой войне, а идея разложения была опубликована в 1924 году его сослуживцем.
- Название метода связано с характерными операциями, отсутствующими в родственном разложении Гаусса.

- Если матрица системы является <u>симметричной и положительно определённой</u>, то для решения системы применяют метод Холецкого (метод квадратного корня).
- В основе метода лежит алгоритм LU-разложения матрицы A, в результате чего она приводится к виду

$$A = LL^T,$$

где L — нижняя треугольная матрица со строго положительными элементами на диагонали.

$$L = \begin{bmatrix} l_{11} & 0 & 0 & \cdots & \cdots & 0 \\ l_{21} & l_{22} & 0 & \cdots & \cdots & 0 \\ l_{31} & l_{32} & l_{33} & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ l_{n-11} & \cdots & \cdots & l_{n-1\,n-2} & l_{n-1\,n-1} & 0 \\ l_{n1} & \cdots & \cdots & l_{n\,n-2} & l_{n\,n-1} & l_{n\,n} \end{bmatrix}$$

Теорема. Для любой симметричной положительно определённой матрицы разложение Холецкого существует и единственно.

Система 1

$$\begin{cases} 3x - y + 2z = 1 \\ -x - 3z = 1 \\ 2x - 3y + z = 1 \end{cases}$$

Метод Холецкого применять нельзя!

Система 2

$$\begin{cases} 5x + 2y - 4z = 1 \\ 2x + y - 2z = 2 \\ -4x - 2y + 5z = 3 \end{cases}$$

Метод Холецкого применять можно

Алгоритм метода Холецкого

1. Строится разложение матрицы A в виде $A = LL^T$
2. Решение системы $Ax = b$ сводится к последовательному решению двух систем с треугольными матрицами:

$$Ly = b \text{ и } L^T x = y$$

- Элементы матрицы L можно вычислить, начиная с верхнего левого угла матрицы A, по следующим формулам:

$$l_{11} = \sqrt{a_{11}},$$
$$l_{j1} = \frac{a_{j1}}{l_{11}}, \quad j \in [2, n],$$
$$l_{ii} = \sqrt{a_{ii} - \sum_{p=1}^{i-1} l_{ip}^2}, \quad i \in [2, n],$$
$$l_{ji} = \left(a_{ji} - \sum_{p=1}^{i-1} l_{ip} l_{jp}\right) / l_{ii}, \quad i \in [2, n-1], j \in [i+1, n].$$

Пример. Решить систему $Ax = b$ методом Холецкого,

если $A = \begin{pmatrix} 81 & -45 & 45 \\ -45 & 50 & -15 \\ 45 & -15 & 38 \end{pmatrix}, b = \begin{pmatrix} 531 \\ -460 \\ 193 \end{pmatrix}$

- Находим элементы матрицы L:

$$l_{11} = \sqrt{81} = 9 \; l_{22} = \sqrt{50 - (-5)^2} = 5 \; l_{33} = \sqrt{38 - 5^2 - 2^2} = 3$$

$$l_{21} = \frac{-45}{9} = -5 \; l_{32} = \frac{-15 - 5 \cdot (-5)}{5} = 2$$

$$l_{31} = \frac{45}{9} = 5$$

- Разложение матрицы A:

$$A = LL^T = \begin{pmatrix} 9 & 0 & 0 \\ -5 & 5 & 0 \\ 5 & 2 & 3 \end{pmatrix} \cdot \begin{pmatrix} 9 & -5 & 5 \\ 0 & 5 & 2 \\ 0 & 0 & 3 \end{pmatrix}$$

- Последовательно решаем системы:

1) $Ly = b$

$$\begin{pmatrix} 9 & 0 & 0 \\ -5 & 5 & 0 \\ 5 & 2 & 3 \end{pmatrix} \quad b = \begin{pmatrix} 531 \\ -460 \\ 193 \end{pmatrix} \qquad y = \begin{pmatrix} 59 \\ -33 \\ -12 \end{pmatrix}$$

2) $L^T x = y$

$$\begin{pmatrix} 9 & -5 & 5 \\ 0 & 5 & 2 \\ 0 & 0 & 3 \end{pmatrix} \quad y = \begin{pmatrix} 59 \\ -33 \\ -12 \end{pmatrix} \qquad x = \begin{pmatrix} 6 \\ -5 \\ -4 \end{pmatrix}$$

Ответ: (6; -5; -4).

Другие виды разложений матриц

- Ранговая факторизация: $A = CF$, где C — матрица $m \times r$ и F — матрица $r \times n$.
- QR-разложение: $A = QR$, где Q — ортогональная матрица размера $m \times m$, и R— верхняя треугольная размера $m \times n$.
- Интерполяционное разложение
- Спектральное разложение: $A = VDV^{-1}$, где D — диагональная матрица, образованная из собственных значений A, а столбцы V — соответствующие собственные вектора.
- Жорданова нормальная форма: $A = CJC^{-1}$, где J — матрица специального вида, называемая жордановой, а C — матрица перехода к новому базису.
- **Сингулярное разложение**

Сингулярное разложение

- **Сингулярное разложение** (Singular Value Decomposition, SVD) — декомпозиция вещественной матрицы с целью её приведения к каноническому виду.
- Сингулярное разложение является удобным методом при работе с матрицами.
- Оно показывает геометрическую структуру матрицы и позволяет <u>наглядно представить имеющиеся данные</u>.

Некоторые предварительные сведения о матрице A

- Обе матрицы AA^T и A^TA — симметричны (хотя не обязательно одинаковых порядков).
- Если $\mathrm{rank}\, A = r$, то $\mathrm{rank}\, AA^T = \mathrm{rank}\, A^TA = r$.
- Характеристические полиномы матриц AA^T и A^TA могут различаться только лишь на степень переменной.
- Собственные числа обеих матриц вещественны и неотрицательны. Количество ненулевых совпадает с $\mathrm{rank}\, A = r$.
- Обозначим положительные собственные числа $\lambda_1^2, \ldots, \lambda_r^2$
- Таким образом спектры матриц:

$A \cdot A^\top$	$A^\top A$
$\{\lambda_1^2, \ldots, \lambda_r^2, \underbrace{0, \ldots, 0}_{m-r}\}$	$\{\lambda_1^2, \ldots, \lambda_r^2, \underbrace{0, \ldots, 0}_{n-r}\}$.

Теорема (о сингулярном разложении)

- Для вещественной матрицы A размерности $m \times n$, ранга r существует представление её в виде произведения

$$A = V\Sigma W^T,$$

- где матрицы V (размерности $m \times m$) и W (размерности $n \times n$) — ортогональные;
- матрица Σ размерности $m \times n$ имеет ненулевыми только диагональные элементы $\sigma_{jj} = \sigma_j$, которые расположены в порядке неубывания и равны арифметическим квадратным корням из ненулевых собственных чисел матрицы AA^T.
- Столбцы матрицы V — собственные векторы AA^T
- Столбцы матрицы W — собственные векторы $A^T A$

- Диагональные элементы матрицы Σ называются **сингулярными числами** матрицы A.
- Столбцы матрицы V называют левыми сингулярными векторами, а столбцы матрицы W — правыми сингулярными векторами матрицы A.
- Левый и правый сингулярные векторы, соответствующие одному и тому же сингулярному числу σ_j, связаны соотношениями:

$$AW_{[j]} = \sigma_j V_{[j]},$$
$$A^T V_{[j]} = \sigma_j W_{[j]}.$$

Схема разложения при m<n:

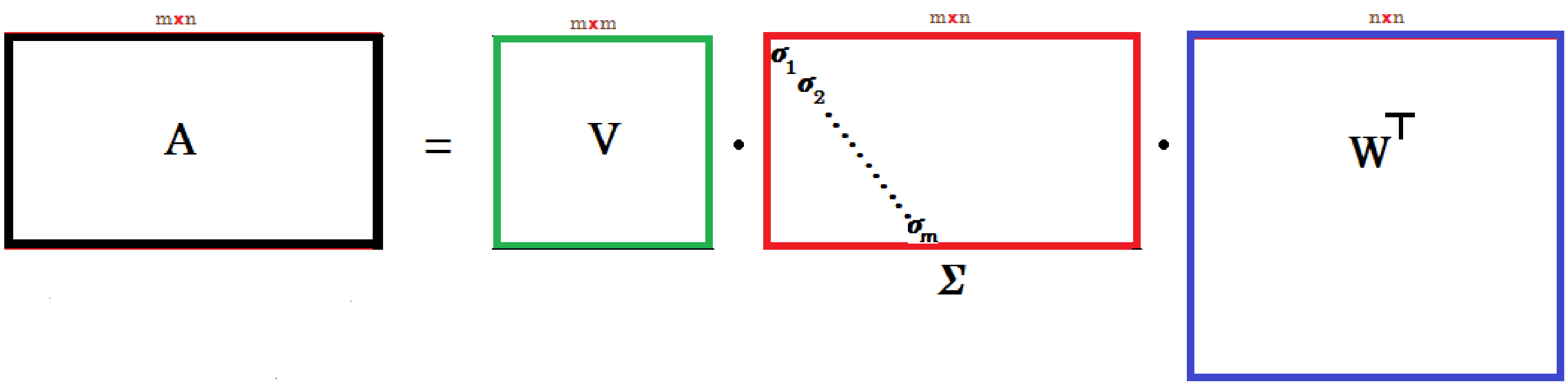
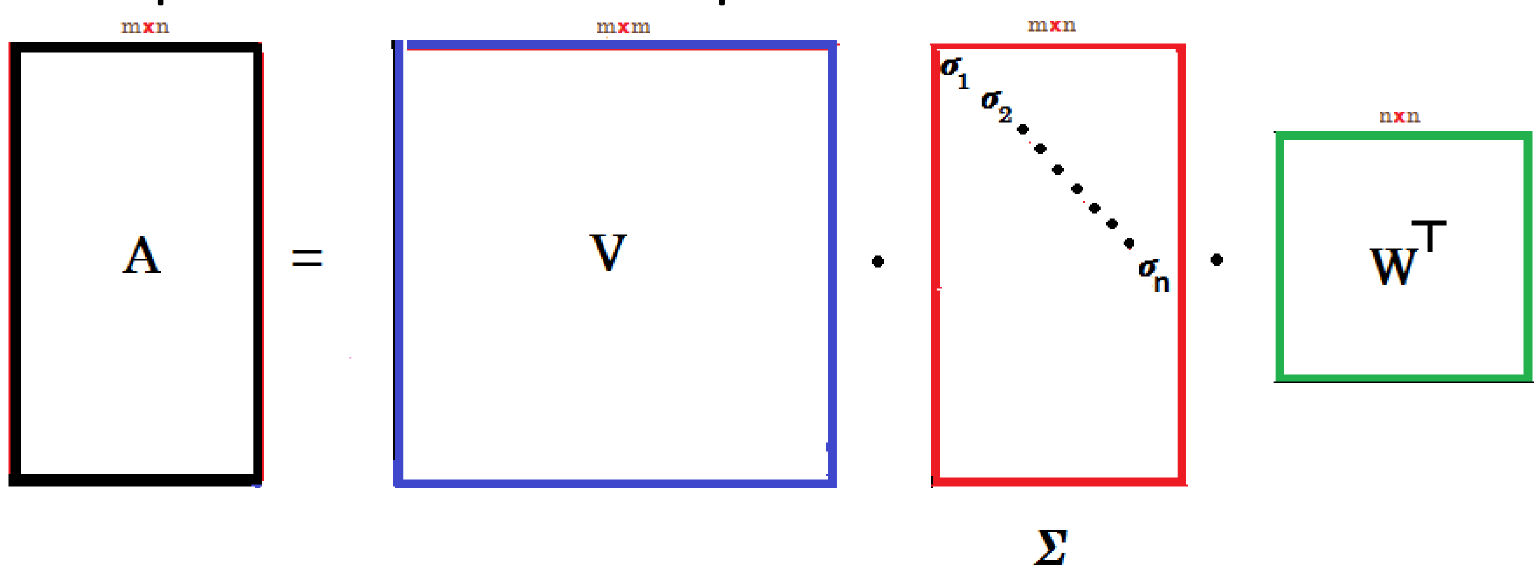


Схема разложения при m>n:



Пример:

- Для матрицы $M = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \end{bmatrix}$
- Одним из сингулярных разложений является разложение

$$V = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \ \Sigma = \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \ W = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \sqrt{0.2} & 0 & 0 & 0 & \sqrt{0.8} \\ 0 & 0 & 0 & 1 & 0 \\ -\sqrt{0.8} & 0 & 0 & 0 & \sqrt{0.2} \end{bmatrix}$$

- Имеет место соотношение $VW = WV = E$.

Приложения сингулярного разложения

1. Псевдообратная матрица

- Если матрица A является вырожденной или прямоугольной, то обратной матрицы A^{-1} для неё не существует.
- Однако для A может быть найдена псевдообратная матрица A^{+} — такая матрица, для которой выполняются условия:

$$A^{+}A = E_n$$

$$AA^{+} = E_m$$

$$A^{+}AA^{+} = A^{+}$$

$$AA^{+}A = A$$

- Матрица A^{+}, удовлетворяющая этим условиям существует и единственна.

- Если $M = V\Sigma W$, то псевдообратная к ней матрица находится по формуле:

$$M^{+} = V\Sigma^{+} W$$

- где Σ^{+} — псевдообратная к матрице Σ, получающаяся из неё заменой каждого ненулевого элемента σ_j на диагонали на обратный к нему: σ_j^{-1}, с последующим транспонированием самой матрицы.

2. Приближение матрицей меньшего ранга

- В некоторых практических задачах требуется приближать заданную матрицу M некоторой другой матрицей M_k с заранее заданным рангом k.
- Известна следующая теорема (Эккарта–Янга):
- Если потребовать, чтобы такое приближение было наилучшим в том смысле, что Фробениусова норма разности матриц M и M_k минимальна, при ограничении $rank\, M_k = k$, то оказывается, что наилучшая такая матрица M_k получается из сингулярного разложения матрицы M по формуле:

$$M_k = V\Sigma_k W$$

- где Σ_k — матрица Σ, в которой заменили нулями все диагональные элементы, кроме k наибольших элементов.

3. Метод наименьших квадратов и число обусловленности

- Задача наименьших квадратов ставится следующим образом: Даны действительная $m \times n$ матрица A и действительный m-мерный вектор Y.
- Требуется найти действительный n-мерный вектор w, минимизирующий Евклидову длину вектора невязки
$$|Y - Aw|_E \to min$$
- Решение задачи наименьших квадратов — вектор
$$w = (A^T A)^{-1} (A^T Y)$$
- Для квадратных матриц A число обусловленности определено отношением $cond(A) = \|A\| \cdot \|A^{-1}\|$.
- Число обусловленности матрицы есть отношение её первого сингулярного числа к последнему: $cond(A) = \sigma_1 / \sigma_n$

4. Сокращённое представление

- Для матрицы M порядка $m \times n$ при необходимости приближения матрицей ранга r меньшего чем n, часто используют компактное представление разложения: $M = V_r \Sigma_r W_r$
- Вычисляются только r столбцов V и r строк W. Остальные столбцы V и строки W не вычисляются.
- Это экономит большое количество памяти при $r << n$ (r много больше n).
- Если мы хотим сохранить только 90% информации, то мы должны взять r, таким образом, чтобы сумма квадратов первых элементов Σ была 90% от общей суммы всех квадратов диагональных элементов Σ.

Пример. Для передачи матрицы

$$A = \begin{pmatrix} 72 & -24 & 0 & 569 & 356 & -63 & 278 & -404 & 464 & -148 & 497 & 425 \\ 301 & 337 & -335 & -41 & 78 & -308 & 414 & 321 & -362 & 394 & 337 & 325 \\ 215 & -191 & -243 & 563 & 327 & 264 & 426 & -65 & 264 & -240 & 738 & 425 \\ -168 & -79 & 136 & -389 & -263 & 111 & -323 & 149 & -197 & -6 & -483 & -398 \\ 74 & 96 & -83 & -214 & -216 & -10 & -207 & 196 & -184 & 42 & -141 & -174 \\ -70 & -290 & 104 & 26 & -302 & 453 & -664 & -111 & 307 & -542 & -207 & -439 \\ 434 & 443 & -248 & 670 & 184 & -433 & 11 & -420 & 536 & -73 & 714 & 540 \\ -61 & 105 & 172 & 118 & -39 & -172 & -308 & -349 & 325 & -108 & -129 & -48 \\ -255 & 78 & 251 & -260 & 71 & -273 & 133 & -73 & -160 & 311 & -386 & -34 \\ -297 & -275 & 225 & -423 & -233 & 292 & -289 & 155 & -203 & -84 & -573 & -480 \end{pmatrix}_{10\times 12}$$

по каналу связи требуется переслать 120 её элементов.

- Попробуем уменьшить это количество за счёт предварительного сингулярного разложения матрицы.

- Имеем:

$$A \approx 2700.156919 \begin{pmatrix} -0.390369 \\ -0.193164 \\ -0.412283 \\ 0.337413 \\ 0.140855 \\ 0.222348 \\ -0.524734 \\ 0.001110 \\ 0.131021 \\ 0.405810 \end{pmatrix} \begin{pmatrix} -0.229044 \\ -0.143812 \\ 0.176558 \\ -0.429270 \\ -0.243324 \\ 0.156198 \\ -0.279937 \\ 0.166302 \\ -0.232733 \\ 0.003274 \\ -0.537057 \\ -0.423297 \end{pmatrix}^{\top} + 1599.933523 \begin{pmatrix} 0.240669 \\ -0.572289 \\ 0.171253 \\ 0.011588 \\ -0.095070 \\ 0.637038 \\ 0.160932 \\ 0.269379 \\ -0.250291 \\ 0.095578 \end{pmatrix} \begin{pmatrix} -0.051775 \\ -0.232734 \\ 0.119333 \\ 0.283435 \\ -0.061748 \\ 0.298356 \\ -0.403913 \\ -0.317640 \\ 0.480821 \\ -0.486426 \\ 0.031925 \\ -0.151288 \end{pmatrix}^{\top}$$

$$+999.456577 V_{[3]} W_{[3]}^{\top} + 799.790045 V_{[4]} W_{[4]}^{\top} +$$

$$+\color{red}{2.945378} V_{[5]} W_{[5]}^{\top} + \color{red}{2.305654} V_{[6]} W_{[6]}^{\top} + \color{red}{2.087323} V_{[7]} W_{[7]}^{\top} + \color{red}{1.898662} V_{[8]} W_{[8]}^{\top} + \color{red}{1.235953} V_{[9]} W_{[9]}^{\top} + \color{red}{0.539705} V_{[10]} W_{[10]}^{\top}$$

- Если отбросить в этом разложении все слагаемые, соответствующие 6 минимальным сингулярным значениям (обозначены красным), то получим приближение матрицы A в виде **матрицы ранга 4**

$$A_4 \approx \begin{pmatrix} 71.26 & -23.70 & -0.53 & 568.70 & 356.77 & -62.66 & 277.37 & -403.65 & 464.44 & -147.54 & 497.19 & 425.11 \\ \dots & & & & & & & & & & & \dots \end{pmatrix}$$

- Эта матрица отличается от A. Однако, последняя «узнаваема» из A_4.
- Если в ходе пересылки предполагается, что отправляемая матрица имеет только целочисленные элементы, то при получении матрицы A_4 мы вправе округлить ее элементы до ближайшего целого.
- Сделав это, увидим, что в 88 элементах из 120 ошибок не будет, а оставшиеся имеют ошибку в последней цифре в пределах 1 по абсолютной величине.
- Если такая погрешность допустима для нашей цели, имеет смысл пересылать именно матрицу A_4, потому что это позволит сэкономить на количестве пересылаемых элементов.

$$A_4 = \sigma_1 V_{[1]} W_{[1]}^\top + \sigma_2 V_{[2]} W_{[2]}^\top + \sigma_3 V_{[3]} W_{[3]}^\top + \sigma_4 V_{[4]} W_{[4]}^\top$$

Программные реализации сингулярного разложения матриц

- SVD входит в список основных методов многих математических библиотек, в том числе свободно распространяемых.
- В GNU Scientific library (GSL): https://www.gnu.org/software/gsl/manual/html_node/Singular-Value-Decomposition.html
- библиотеке Intel® Math Kernel Library (Intel® MKL). https://software.intel.com/en-us/intel-mkl
- В библиотеке numpy для линейной алгебры в Python: https://docs.scipy.org/doc/numpy-1.12.0/reference/generated/numpy.linalg.svd.html
- В библиотеке для машинного обучения tensorflow:

  https://www.tensorflow.org/api_docs/python/tf/svd

Литература и источники

- Хорн Р., Джонсон Ч. Матричный анализ. М.: Мир. 1989.
- Жданов А. И. Введение в вычислительную линейную алгебру: электрон. учеб. пособие. М-во образования и науки РФ, Самарский гос. аэрокосм. ун-т им. С. П. Королева (нац. исслед. ун-т). Самара, 2011.
- http://pmpu.ru/vf4/algebra2/svd
- http://www.machinelearning.ru/wiki/index.php?title=Сингулярное_разложение