Домашнее задание I

Кобылянский А.В.
Группа 5381
Вариант 2

Задача 1

Запишите в эквивалентной форме каждое из следующих выражений, используя указанные в скобках матрицы/векторы и три операции: матричное произведение, транспонирование и конструирование диагональной матрицы diag().

2. $(x \in \mathbb{R}^m, y \in \mathbb{R}^n, A \in \mathbb{R}^{m\times n}), \sum_{i=1}^m \sum_{j=1}^n a_{ij}x_iy_j$

$$\sum_{i=1}^m \sum_{j=1}^n a_{ij}x_iy_j = x^TAy$$

4. $(V = [v_1, ..., v_k] \in \mathbb{R}^{n\times k}, U = [u_1, ..., u_k] \in \mathbb{R}^{m\times k}) \sum_{i=1}^k u_iv_i^T$

$$\sum_{i=1}^k u_iv_i^T = UV^T$$

6. $(y \in \mathbb{R}^n, A \in \mathbb{R}^{m\times n}), B := [a_{ij}y_i]_{ij}$

$$B = A\,\mathrm{diag}(y)$$

8. $(\Sigma \in \mathbb{R}^{k\times s}, V = [v_1, ..., v_s] \in \mathbb{R}^{n\times s}, U = [u_1, ..., u_k] \in \mathbb{R}^{m\times k}), \sum_{i=1}^k \sum_{j=1}^s \Sigma_{ij}u_iv_j^T$

$$\sum_{i=1}^k \sum_{j=1}^s \Sigma_{ij}u_iv_j^T = U\Sigma V^T$$

Задача 2

Пусть $A = A^T \in \mathbb{R}^{n\times n}, D = D^T \in \mathbb{R}^{m\times m}$ и $B \in \mathbb{R}^{n\times m}$

2. Показать, что если $A \succ 0$ (положительно определенная матрица), то

$$\min_{x\in\mathbb{R}^n} \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} A & B \\ B^T & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = y^T(D - B^TA^{-1}B)y$$

$A \succ 0 \Rightarrow det(A) \neq 0$, из пункта 1 имеем:

$$\begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} A & B \\ B^T & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = (x + A^{-1}By)^TA(x + A^{-1}By) + y^T(D - B^TA^{-1}B)y$$

Правое слагаемое не зависит от x, необходимо минимизировать левое слагаемое. Выражение z^TAz достигает минимума при $z = 0$, т.к. A - положительно определена. При $x = -A^{-1}By$:

$$\begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} A & B \\ B^T & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0^TA0 + y^T(D - B^TA^{-1}B)y = y^T(D - B^TA^{-1}B)y$$

Задача 3

Пусть $H = H^T \in \mathbb{R}^{n\times n}$, а также $H \succ 0$. Найти

$$\arg\max_{x\in\mathbb{R}^n} \frac{x^T H x}{x^T x}$$

Т.к. H - симметрична, то $H = U\Lambda U^T$, где Λ - диагональная матрица собственных значений, $U^{-1} = U^T$. Все элементы Λ действительны и положительны, т.к. $H \succ 0$. Пусть $y = U^T x$

$$\frac{x^T H x}{x^T x} = \frac{y^T U^T U \Lambda U^T U y}{y^T U^T U y} = \frac{y^T \Lambda y}{y^T y} = \frac{\sum_i \lambda_i y_i^2}{\sum_i y_i^2}$$

$$\frac{\partial f}{\partial y_i} = \frac{2y_i\lambda_i \sum_i y_i^2 - 2y_i \sum_i \lambda_i y_i^2}{(\sum_i y_i^2)^2} = 0 \Rightarrow \left[\begin{array}{l} \quad y_i = 0 \\ \lambda_i = \dfrac{\sum_i \lambda_i y_i^2}{\sum_i y_i^2} \end{array}\right.$$

Пусть U выбрана так, чтобы собственные числа в $\Lambda = \mathrm{diag}([\lambda_1, ..., \lambda_n])$ шли по невозрастанию $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$.

y будет стационарной точкой, если $f(y) = \lambda_i$ и $\forall j : \lambda_j \neq \lambda_i \Rightarrow y_j = 0$, следовательно $\max f(y) = \lambda_1$. Пусть первые k собственных чисел одинаковы, т.е. $\lambda_1 = ... = \lambda_k \neq \lambda_{k+1}$, тогда максимум достигается для $y = [y_1, ..., y_k, 0, ..., 0]$, где $y_1, ..., y_k$ - принимают любые значения, кроме всех нулей, т.е. $y \neq 0$. Uy - линейная комбинация собственных векторов, соответствующих собственному числу λ_1

Ответ: функция достигает максимума на x из собственного подпространства максимальнго собственного числа матрицы H, $x \neq 0$.

Задача 4

Для каждой из следующих функций $f(x)$ выписать первый дифференциал. Для скалярных функций векторного аргумента найти вектор-градиент $\nabla f(x)$ и матрицу Гессе $\nabla^2 f(x)$. Для скалярных функций матричного аргумента найти матрицу-градиент $\nabla f(X)$. Для векторных функций векторного аргумента найти матрицу Якоби $J(x)$.

Для каждого пункта сделать численную проверку градиента.

Функция для численной проверки градиента.

```
function error_mean = grad_check(fun, dfun, x, varargin)
  # fun(x) - function for which gradient is checked
  # dfun(x, dx) - derivative of fun(x)
  # x - the point at which gradient is checked

  eps = 1e-4;
  tests_count = 100;

  error_sum = 0;
  for k = 1:tests_count
    dx = eps*rand(size(x));
    difference = fun(x + dx, varargin{:}) - fun(x,
        varargin{:});

    error = norm(difference - dfun(x, dx, varargin{:}), "
        fro") / norm(difference, "fro");
    error_sum += error;
  endfor

  error_mean = error_sum / tests_count;
endfunction
```

2. $f(x) = \exp(A\exp(B\exp(Cx)))$, где $\exp(y) = [e^{y_1}, ..., e^{y_n}]^T, x \in \mathbb{R}^n$

Посчитав частные производные $\exp(x)$ легко видеть, что матрица Якоби для этой функции равна $\mathrm{diag}(\exp(x))$, следовательно $d\exp(x) = \mathrm{diag}(\exp(x))dx$.

$$
\begin{aligned}
&d\exp(A\exp(B\exp(Cx))) = \mathrm{diag}(\exp(A\exp(B\exp(Cx))))Ad\exp(B\exp(Cx))) = \\
&= \mathrm{diag}(\exp(A\exp(B\exp(Cx))))A\,\mathrm{diag}(\exp(B\exp(Cx))))Bd\exp(Cx)) = \\
&= \mathrm{diag}(\exp(A\exp(B\exp(Cx))))A\,\mathrm{diag}(\exp(B\exp(Cx)))B\,\mathrm{diag}(\exp(Cx))Cdx
\end{aligned}
$$

Матрица Якоби $J(x)$

$$J(x) = \mathrm{diag}(\exp(A\exp(B\exp(Cx))))A\,\mathrm{diag}(\exp(B\exp(Cx)))B\,\mathrm{diag}(\exp(Cx))C$$

Численная проверка

```
function res = f(x, A, B, C)
  res = exp(A*exp(B*exp(C*x)));
endfunction

function res = df(x, dx, A, B, C)
  res = diag(exp(A*exp(B*exp(C*x))))*...
    A*diag(exp(B*exp(C*x)))*B*diag(exp(C*x))*C*dx;
endfunction

x = 0.01*rand(n, 1);
A = rand(n);
B = rand(n);
C = rand(n);
grad_check(@f, @df, x, A, B, C)  # ans = 5.2212e-04
```

4. $f(X) = \operatorname{tr}(AX^{-1}B), X \in \mathbb{R}^{n\times n}$

$$\begin{aligned}
&d\operatorname{tr}(AX^{-1}B) = \operatorname{tr}(d(AX^{-1}B)) = \operatorname{tr}(Ad(X^{-1})B) = \\
&= \operatorname{tr}(-AX^{-1}dXX^{-1}B) = \operatorname{tr}(-X^{-1}BAX^{-1}dx) = \\
&= \operatorname{tr}(-(X^{-T}A^TB^TX^{-T})^Tdx)
\end{aligned}$$

Матрица градиент $\nabla f(x)$

$$\nabla f(x) = -X^{-T}A^TB^TX^{-T}$$

Численная проверка

```
function res = f(x, A, B)
  res = trace(A*inv(x)*B);
endfunction

function res = df(x, dx, A, B)
  res = trace(-inv(x)*B*A*inv(x)*dx);
endfunction

x = rand(n);
A = rand(n);
B = rand(n);
grad_check(@f, @df, x, A, B)  # ans = 3.1501e-04
```

6. $f(X) = \operatorname{tr}\left(\begin{bmatrix} X & B \\ B^T & A \end{bmatrix}\right), X \in \mathbb{R}^{n\times n}$

$$d\operatorname{tr}\left(\begin{bmatrix} X & B \\ B^T & A \end{bmatrix}\right) = d(\operatorname{tr}(X) + \operatorname{tr}(A)) = \operatorname{tr}(dX)$$

Матрица градиент $\nabla f(x)$

$$\nabla f(x) = I$$

Численная проверка

```
function res = f(x, A, B)
  res = trace([x, B; B', A]);
endfunction

function res = df(x, dx)
  res = trace(dx);
endfunction

x = rand(n);
A = rand(n);
B = rand(n);
grad_check(@f, @df, x, A, B) # ans = 3.2477e-12
```

8. $f(X) = \det(X^T AX), X \in \mathbb{R}^{n\times n}$

$$\begin{aligned}
&d\det(X^TAX) = \det(X^TAX)\operatorname{tr}((X^TAX)^{-1}d(X^TAX)) = \\
&= \det(X)^2\det(A)\operatorname{tr}((X^TAX)^{-1}(dX^TAX + X^TAdX)) = \\
&= \det(X)^2\det(A)(\operatorname{tr}(X^{-1}A^{-1}X^{-T}dX^TAX) + \operatorname{tr}(X^{-1}A^{-1}X^{-T}X^TAdX)) \\
&= \det(X)^2\det(A)tr(2X^{-1}dX)
\end{aligned}$$

Матрица градиент $\nabla f(x)$

$$\nabla f(x) = 2\det(X)^2\det(A)X^{-T}$$

Численная проверка

```
function res = f(x, A)
  res = det(x'*A*x);
endfunction

function res = df(x, dx, A)
  res = 2*det(x)^2*det(A)*trace(inv(x)*dx);
endfunction

x = rand(n);
A = rand(n);
grad_check(@f, @df, x, A) # ans = 2.2538e-04
```

10. $f(X) = \log\det(X^p), X \in \mathbb{R}^{n\times n}$

$$d\log\det(X^p) = d(\log\det(X)^p) = d(p\log\det(X)) =$$
$$= p\frac{d\det(X)}{\det(X)} = p\frac{\det(X)\operatorname{tr}(X^{-1}dX)}{\det(X)} = p\operatorname{tr}(X^{-1}dX)$$

Матрица градиент $\nabla f(x)$

$$\nabla f(x) = pX^{-T}$$

Численная проверка

```
function res = f(x, p)
  res = log(det(x^p));
endfunction

function res = df(x, dx, p)
  res = p*trace(inv(x)*dx);
endfunction

x = rand(n);
p = floor(20*rand());
grad_check(@f, @df, x, p) # ans = 4.2008e-04
```

12. $f(X) = \frac{1}{2}\|AX - B\|_F^2, x \in \mathbb{R}^{m\times n}$

$$d\frac{1}{2}\|X\|_F^2 = \frac{1}{2}d((\operatorname{tr}(X^TX)^{0.5})^2) = \frac{1}{2}\operatorname{tr}(d(X^TX)) = \frac{1}{2}\operatorname{tr}(dX^TX + X^TdX) = \operatorname{tr}(X^TdX)$$

$$d\frac{1}{2}\|AX - B\|_F^2 = \operatorname{tr}((AX - B)^TAdX)$$

Матрица градиент $\nabla f(x)$

$$\nabla f(x) = A^T(AX - B)$$

Численная проверка

```
function res = f(x, A, B)
  res = 0.5*norm(A*x - B, "fro")^2;
endfunction

function res = df(x, dx, A, B)
  res = trace((A*x - B)'*A*dx);
endfunction

x = rand(n);
A = rand(n);
B = rand(n);
grad_check(@f, @df, x, A, B) # ans = 1.7692e-04
```

14. $f(X) = \log(\sum_{i=1}^n \exp(a_i^T x)), x \in \mathbb{R}^n$

Пусть $A = [a_1, ..., a_n]$ - матрица из столбцов a_i, $1 = [1, ..., 1]^T \in \mathbb{R}^n$, тогда

$$\log(\sum_{i=1}^n \exp(a_i^T x)) = \log(1^T \exp(A^T x))$$

$$d\log(1^T \exp(A^T x)) = \frac{1^T d\exp(A^T x)}{1^T \exp(A^T x)} = \frac{1^T \operatorname{diag}(\exp(A^T x))A^T dx}{1^T \exp(A^T x)} = \frac{\exp(A^T x)^T A^T dx}{1^T \exp(A^T x)}$$

Вектор градиент $\nabla f(x)$

$$\nabla f(x) = \frac{A\exp(A^T x)}{1^T \exp(A^T x)}$$

Найдем вторую производную

$$d\left(\frac{\exp(A^T x)^T A^T dx_1}{1^T \exp(A^T x)}\right) = \frac{d(\exp(A^T x)^T A^T dx_1) 1^T \exp(A^T x) - \exp(A^T x)^T A^T dx_1 d(1^T \exp(A^T x))}{(1^T \exp(A^T x))^2} =$$

$$= \frac{dx_2^T A \operatorname{diag}(\exp(A^T x)) A^T dx_1 1^T exp(A^T x) - \exp(A^T x)^T A^T dx_1 1^T \operatorname{diag}(\exp(A^T x)) A^T dx_2}{(1^T \exp(A^T x))^2)} =$$

$$= |v := \exp(A^T x)| = \frac{((dx_1)^T A \operatorname{diag}(v) A^T dx_2) 1^T v - (dx_1)^T A v v^T A^T dx_2}{(1^T v)^2} =$$

$$= \frac{(dx_1)^T A(1^T v \operatorname{diag}(v) - vv^T) A^T dx_2}{(1^T v)^2}$$

Матрица Гессе $\nabla^2 f(x)$

$$\nabla^2 f(x) = \frac{A(1^T v \operatorname{diag}(v) - vv^T)A^T}{(1^T v)^2}, \text{где } v = \exp(A^T x)$$

Численная проверка

```
function res = f(x, A, B)
  o = ones(1, rows(x)); # 1'
  res = log(o*exp(A'*x));
endfunction

function res = df(x, dx, A, B)
  o = ones(1, rows(x)); # 1'
  grad = A*exp(A'*x) / (o*exp(A'*x));
  res = grad'*dx;
endfunction

x = rand(n, 1);
A = rand(n);
B = rand(n);
grad_check(@f, @df, x, A, B) # ans = 2.8365e-06
```

16. $f(X) = \operatorname{tr}((X^TCX)^{-1}A), X \in \mathbb{R}^{n\times n}, A, C$ - симметричные матрицы

$$
\begin{aligned}
&d\operatorname{tr}((X^TCX)^{-1}A) = \operatorname{tr}(d(X^{-1}C^{-1}X^{-T}A)) = \\
=&\operatorname{tr}(dX^{-1}C^{-1}X^{-T}A + X^{-1}C^{-1}dX^{-T}A) = \\
=&\operatorname{tr}(C^{-1}X^{-T}AdX^{-1} + AdX^{-1}C^{-1}X^{-T}) = \\
=&\operatorname{tr}(2C^{-1}X^{-T}AdX^{-1}) = \\
=&-2\operatorname{tr}(C^{-1}X^{-T}AX^{-1}dXX^{-1}) = \\
=&-2\operatorname{tr}(X^{-1}C^{-1}X^{-T}AX^{-1}dX) = \\
=&-2\operatorname{tr}((X^TCX)^{-1}AX^{-1}dX)
\end{aligned}
$$

Матрица градиент $\nabla f(x)$

$$\nabla f(x) = X^{-T}A^TX^{-1}C^{-T}X^{-T}$$

Численная проверка

```
function res = f(x, A, C)
  res = trace(inv(x'*C*x)*A);
endfunction

function res = df(x, dx, A, C)
  res = -2*trace(inv(x'*C*x)*A*inv(x)*dx);
endfunction

x = rand(n);
A = rand(n); A = A*A';   # making symmetric
C = rand(n); C = C*C';   #
grad_check(@f, @df, x, A, C) # ans = 1.7692e-04
```

Задача 5

Для каждой из следующих функций найти все стационарные точки и определить их тип (локальный максимум/минимум, седловая точка).

2. $f(x) = (1 - x_1)^2 + 100(x_2 - x_1)^2$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 202x_1 - 200x_2 - 2 \\ -200_1 + 200x_2 \end{bmatrix}$$
$$\nabla f(x) = 0 \Rightarrow x_1 = 1, x_2 = 1$$

$f(x)$ имеет единственную стационарную точку $x = [1, 1]$. Очевидно, что это точка глобального минимума. Убедимся в этом, рассмотря $\nabla^2 f(x)$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1} \end{bmatrix} = \begin{bmatrix} 202 & -200 \\ -200 & 200 \end{bmatrix}$$

Собственные числа $\nabla^2 f(x)$ равны $\lambda_{1,2} = 201 \pm \sqrt{40001} > 0$. Оба числа положительны, значит $\nabla^2 f(x) \succ 0$ и $x = [1, 1]$ - точка строгого глобального минимума.

4. $f(x) = \operatorname{vec}(xa^T)^T \operatorname{vec}(xb^T)$

$$\begin{aligned} \operatorname{vec}(xa^T)^T \operatorname{vec}(xb^T) &= \operatorname{tr}((xa^T)^T xb^T) = \\ &= \operatorname{tr}(ax^T xb^T) = \operatorname{tr}(a^T bx^T x) = a^T bx^T x \\ d(a^T bx^T x) &= 2a^T bx^T dx \\ \nabla f(x) &= 2a^T bx = 0 \Rightarrow x = 0 \lor a^T b = 0 \\ \nabla^2 f(x) &= 2a^T b I_n \end{aligned}$$

При $a^T b \neq 0$ $f(x)$ имеет единственную стационарную точку $x = 0$. При $a^T b > 0$ это точка глобального минимума, при $a^T b < 0$ это точка глобального максимума.

При $a^T b = 0$ $f(x) \equiv 0$ и все точки $\mathbb{R}^n$ будут точками точками минимума и максимума одновременно.

Задача 6

Для каждой задачи оптимизации найти множество решений и оптимальное значение целевой функции.

2. $\min\limits_{x\in\mathbb{R}^n} x^T Ax - x^T b, A \succ 0$

$$\begin{aligned} d(x^T Ax - x^T b) &= 2x^T Adx - dx^T b = (2x^T A - b^T)dx = \\ \nabla f(x) &= 2Ax - b = 0 \Rightarrow x^* = 0.5A^{-1}b \\ d((2x^T A - b^T)dx_1) &= (dx_1)^T 2Adx_2 \\ \nabla^2 f(x) &= 2A \succ 0 \end{aligned}$$

Так как $\nabla^2 f(x) \succ 0$, то функция выпукла и любая стационарная точка будет глобальным минимумом, значи x^* - точка глобального минимума.

$$\begin{aligned} f(x^*) &= (\frac{1}{2}A^{-1}b)^T A(\frac{1}{2}A^{-1}b) - (\frac{1}{2}A^{-1}b)^T b = \\ &= \frac{1}{4}b^T A^{-1}b - \frac{1}{2}b^T A^{-1}b = -\frac{1}{2}b^T A^{-1}b \end{aligned}$$

Ответ: $x^* = \frac{1}{2}A^{-1}b, f(x^*) = -\frac{1}{2}b^T A^{-1}b$

4. $\min\limits_{x\in\mathbb{R}^n} c^T x + \frac{\alpha}{3}\left\|x\right\|^3, \alpha > 0$

$$d\frac{1}{3}\left\|x\right\|^3 = \frac{1}{3}d(x^Tx)^{\frac{3}{2}} = \frac{1}{3}\frac{3}{2}(x^Tx)^{\frac{1}{2}}d(x^Tx) = \frac{1}{2}\left\|x\right\|(2x^Tdx) = \left\|x\right\|x^Tdx$$

$$d(c^Tx + \frac{\alpha}{3}\left\|x\right\|^3) = (c^T + \alpha\left\|x\right\|x^T)dx$$

$$\nabla f(x) = c + \alpha\left\|x\right\|x = 0 \Rightarrow \left\|(\alpha\left\|x\right\|x)\right\| = \left\|-c\right\| \Rightarrow \left\|x\right\| = \sqrt{\frac{\left\|c\right\|}{a}} \Rightarrow x^* = -\frac{c}{\sqrt{\alpha\left\|c\right\|}}$$

Найдем $\nabla^2 f(x)$

$$\begin{aligned} d((c^T + \alpha\left\|x\right\|x^T)dx_1) &= \alpha(d(\left\|x\right\|)x^Tdx_1 + \left\|x\right\|d(x^Tdx_1)) = \\ &= \alpha((\frac{1}{2}(x^Tx)^{-\frac{1}{2}}(2x^Tdx_2))(x^Tdx_1)\left\|x\right\|((dx_2)^Tdx_1)) = \\ &= \alpha(\left\|x\right\|^{-1}(x^Tdx_2)(x^Tdx_1) + \left\|x\right\|(dx_2^Tdx_1)) = \\ &= (dx_1)^T\alpha(\left\|x\right\|^{-1}xx^T + \left\|x\right\|I_n)dx_2 \\ \nabla^2 f(x) &= \alpha(\left\|x\right\|^{-1}xx^T + \left\|x\right\|I_n) = \alpha\left\|x\right\|\left(\frac{xx^T}{\left\|x\right\|^2} + I_n\right) \end{aligned}$$

Пусть $x \neq 0$. $\frac{xx^T}{\left\|x\right\|^2}$ - одноранговая матрица у которой все собственные числа равны 0, кроме одного, соответствующего собственному вектору x, равного 1.

$$\left(\frac{xx^T}{\|x\|^2}\right)x = \frac{x(x^Tx)}{\|x\|^2} = \frac{x\,\|x\|^2}{\|x\|^2} = x$$

Собственные числа $xx^T/\|x\|^2 + I_n$ равны соответственно 1 и 2. Т.к. все собственные числа положительны, матрица положительно определена, $\alpha\,\|x\| > 0$, значит $\nabla^2 f(x) \succ 0$.

Пусть $x = 0$. $\nabla^2 f(0) = 0$, значит

$$\forall x \in \mathbb{R}^n\ \nabla^2 f(x) \succeq 0 \Rightarrow f(x) - \text{выпукла } \Rightarrow x^* - \text{глобальный минимум}$$

$$f(x^*) = -\frac{c^Tc}{\sqrt{\alpha\,\|c\|}} + \frac{\alpha}{3}\frac{\|c\|^3}{(\alpha\,\|c\|)^{\frac{3}{2}}} = -\frac{\|c\|^{\frac{3}{2}}}{\sqrt{\alpha}} + \frac{1}{3}\frac{\|c\|^{\frac{3}{2}}}{\sqrt{\alpha}} = -\frac{2}{3}\frac{\|c\|^{\frac{3}{2}}}{\sqrt{\alpha}}$$

Ответ: $x^* = -\frac{c}{\sqrt{\alpha\|c\|}}, f(x^*) = -\frac{2}{3}\frac{\|c\|^{\frac{3}{2}}}{\sqrt{\alpha}}$

6. $\min\limits_{x \in \mathbb{R}^n_+, 1^Tx=1} a^Tx, a \in \mathbb{R}^n_+$

Сумма компонент x_i должна быть равна 1. Чтобы значение функции было минимально, необходимо найти минимальную из компонент a_i и уствновить $x_i = 1, x_j = 0\ j \neq i$. Если минимальны a_i несколько, то обозначим

$$I_{min} = \{i \mid \forall a_j \in a : a_i \leq a_j\}$$

индексы минимальных компонент a, тогда

$$\operatorname*{arg\,min}_{x \in \mathbb{R}^n_+, 1^Tx=1} a^Tx = \left\{x \in \mathbb{R}^n_+ \mid V \subset I_{min} \wedge \sum_{i \in V} x_i = 1 \wedge \forall i \notin V : x_i = 0\right\}$$

Минимальное значение функции равно минимальной компоненте a

$$\min_{x \in \mathbb{R}^n_+, 1^Tx=1} a^Tx = \min a$$

Задача 7

Для данного набора переменных $u \in \mathbb{R}^n, v \in \mathbb{R}^m, z \in \mathbb{R}$ и набора констант $B \in \mathbb{R}^{k\times n}, C \in \mathbb{R}^{k\times m}, a \in \mathbb{R}^k, c \in \mathbb{R}^m$ записать систему уравнений

$$Bu + Cv = a, v_i + z = c_i, \sum_i u_i = z$$

в виде $Ax = b$, если обозначить $x = [u; v; z] \in \mathbb{R}^{n+m+1}$. То есть выразить A и b через B, C, a, c.

Пусть $0_v = [0, ..., 0]^T, 1_v = [1, ..., 1]^T, I_n$ - единичная матриц, 0_m - матрица полностью из нулей, тогда

$$A = \begin{bmatrix} B & C & 0_v \\ 0_m & I_n & 1_v \\ 1_v^T & 0_v^T & -1 \end{bmatrix}, \; b = \begin{bmatrix} a \\ c \\ 0 \end{bmatrix}$$

Задача 8

Рассмотрим задачу предсказания оценок пользователей для некоторых объектов, например, фильмов. В этом случае данные представлены матрицей размера $N \times M$, где N – количество фильмов в базе данных, M – количество зарегистрированных пользователей. Часть элементов этой матрицы известны, а значения остальных требуется предсказать.

Целевая функция для настройки параметров имеет вид

$$f(P, Q) = \frac{1}{2} \sum_{n,m:r_{n,m}\neq 0} (r_{n,m} - p_n^T q_m)^2 + \frac{\lambda_p}{2} \|P\|_F^2 + \frac{\lambda_q}{2} \|Q\|_F^2$$

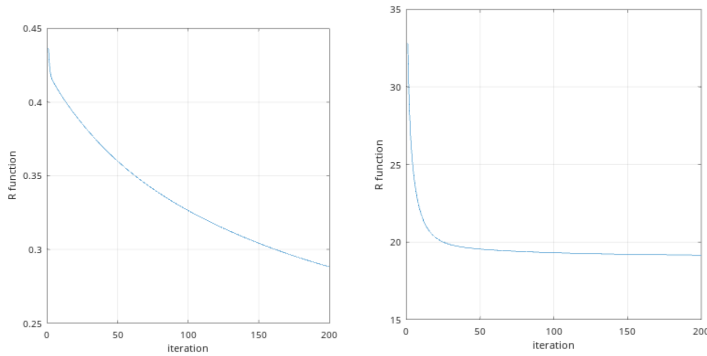
Использовать часть оценок (R) для нахождения оптимальных параметров $(P^*, Q^*) = \arg\min_{P,Q} f(P, Q)$ с помощью ALS. Реализовать ALS. Использовать $K = 10 - 15$ и $\lambda_p = \lambda_q = \lambda$, где $\lambda = \{0.001, 0.01, 0.1\}$. Построить графики зависимости целевой функции от номера итерации для нескольких значений λ. Построить графики зависимости среднеквадратичной ошибки (root-meansquare error, RMSE) для тестовой (R test) и обучающей (R) выборок от номера итерации для нескольких значений λ.

Функция, реализующая ALS

```
function [Ps, Qs] = als(R, k, lambda, iteration_count)
  n = rows(R);
  m = columns(R);

  P = rand(k, n);
  Q = rand(k, m);

  filter = R;
  filter(filter != 0) = 1;

  for _ = 1:iteration_count
    V = Q*R';
    for u = 1:n
      M = Q*diag(filter(u, :))*Q' + lambda*eye(k);
      P(:, u) = M \ V(:, u);
    endfor

    V = P*R;
    for i = 1:m
      M = P*diag(filter(:, i))*P' + lambda*eye(k);
      Q(:, i) = M \ V(:, i);
    endfor

    Ps(end + 1, :, :) = P;
    Qs(end + 1, :, :) = Q;
  endfor
endfunction
```

Функция, расчитывающая $f(P, Q)$ и RMSE

```
function [fun, RMSE] = tests(R, P, Q, lambda=0.01)
  pq = P'*Q;
  pq(R == 0) = 0;
  res = norm(R - pq, "fro")^2;

  fun = 0.5*(res + lambda*(norm(P, "fro")^2 + norm(Q, "fro")^2));
  RMSE = (res / nnz(R))^0.5;
endfunction
```



(a) $\lambda = 0.001$ (b) $\lambda = 0.1$

Рис. 1: Графики зависимости целевой функции от итерации

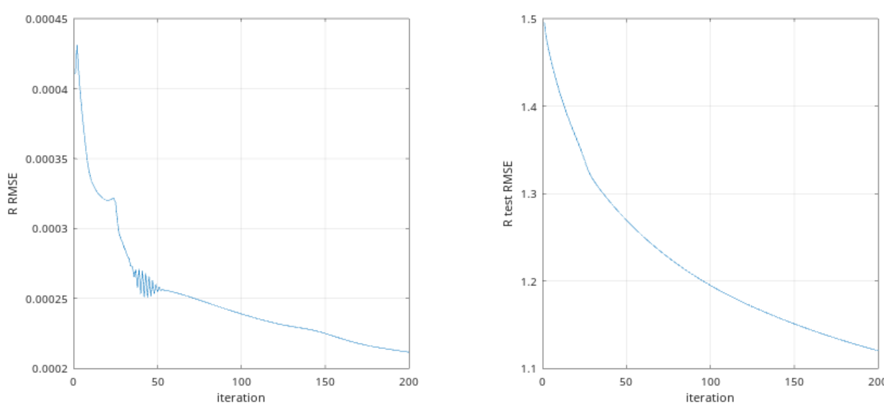


Рис. 2: Графики зависимости RMSE для R и R_test от итерации при $\lambda = 0.001$

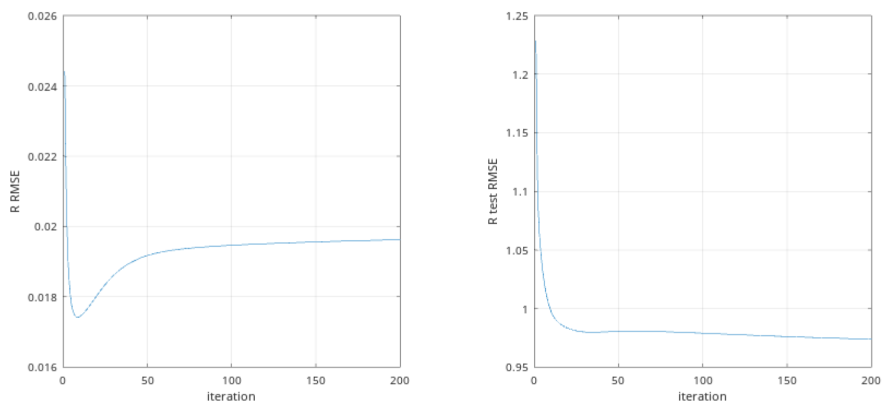


Рис. 3: Графики зависимости RMSE для R и R_test от итерации при $\lambda = 0.1$