Netflix prize

$\min \|y - UV^T\|_F^2$

$\min \|_x (y - X)\|_F$ — X cont Nones.

$\begin{bmatrix} 0 & 1 & 0 \\ & 0 & 1 \end{bmatrix}$ — binary matrix.

$\min \|X\|_*$ nuclear norm.

$\text{s.t.} \ \|W \times (X - y)\|_F^2 \leq \delta$

$+ \frac{\lambda}{2}\|X - Z\|_F^2$

$y(X) + \langle \lambda, X - Z \rangle$

$\min y(X)$

$\text{s.t.} \ X = Z$

$\min_{X, Z} \|X\|_*$

$\text{s.t.} \ \|W \otimes (Y - Z)\|_F^2 \leq \delta^2$ — constr is independent of X — that make problem simpler.

$\text{s.t.} \ X = Z$

$\mathcal{L}(X, Z, T) = \|X\|_* + i_\delta(Z) + \frac{\lambda}{2}\left[\|X - Z - T\|_F^2 - \|T\|_F^2\right)$

$c(Z) = \|W \otimes (y - Z)\|^2 \leq \delta.$

Indicator $i_\delta(Z) = \begin{cases} 0 & \text{if } c(Z) \leq \delta \\ \infty & \text{otherwise} \end{cases}$

$\min \mathcal{L}(X, Z, T),$

$X^{k+1} = \operatorname{argmin} \mathcal{L}(X, Z^k, T^k)$

$= \operatorname{argmin} \|X\|_* + \frac{\lambda}{2}\|X - \underbrace{Z^k - T^k}_{A}\|_F^2 =$

$= \operatorname{argmin} \|X\|_* + \frac{\lambda}{2}\|X - A\|_F^2$

singular value thresholding

$A = U \Sigma V^T \Rightarrow X^{k+1} =$ in slide
look

$Z^{k+1} = \text{argmin } i_\delta(z) + \frac{\rho}{2} \| X^{k+1} - z - \Gamma^k \|_F^2 =$

$= \text{argmin } \frac{\rho}{2} \| X^{k+1} - \Gamma - z \|_F^2 =$

s.t $\| W * (y - z) \|_F^2 < \delta$

can rem $\frac{\rho}{2}$

$= \text{argmin } \| B - z \|_F^2$

s.t $\| W * (y - z) \|_F^2 < \delta$

$B = X^{k+1} - \Gamma^k$ vectorize W

$b = \text{vec}(B)$

$z = \text{vec}(z)$

$y = \text{vec}(y)$

$y = \begin{bmatrix} y_i \\ y_i \\ y_{ik} \end{bmatrix}$ $w \begin{bmatrix} 0 \\ 1 \\ y_w \end{bmatrix}$

$y_w = \begin{bmatrix} y_{i1} \\ y_{i2} \end{bmatrix}$

$\min \| b - z \|^2$

s.t $\| y_w - z_w \| < \delta$

vec of available elements

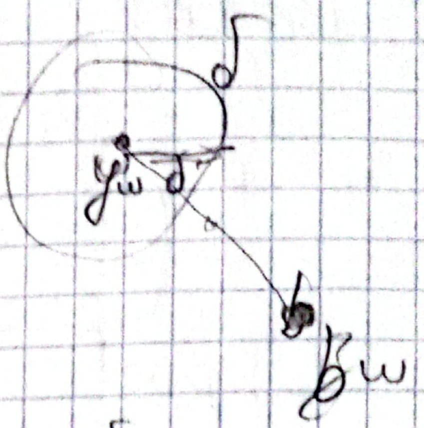
$\begin{bmatrix} b_w \\ b_{\bar{w}} \end{bmatrix} - \begin{bmatrix} z_w \\ z_{\bar{w}} \end{bmatrix} =$

$z_w = \begin{bmatrix} z_{i1} \\ z_{ik} \end{bmatrix}$

$= [b_w - z_w]^2 + \| b_{\bar{w}} - z_{\bar{w}} \|^2$

$\Rightarrow b_{\bar{w}} = z_{\bar{w}}$

$$\min \|b_\omega - z_\omega\|^2$$
$$\text{s.t. } \|y_\omega - z_\omega\|^2 \le \delta.$$



$$\begin{cases} z_\omega^* = y_\omega + \delta \dfrac{b_\omega - y_\omega}{\|b_\omega - y_\omega\|_2}; \\ z_\omega^* = b_\omega \quad \text{if } |b_\omega - y_\omega| \le \delta \end{cases}$$

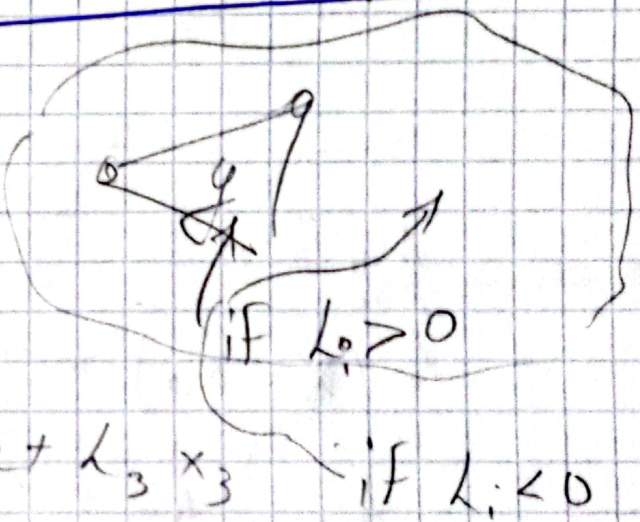
$$z_\omega^* = y_\omega + \min(\delta, \|b_\omega - y_\omega\|) \cdot \frac{b_\omega - y_\omega}{|b_\omega - y_\omega|}$$

$$z = (1-\omega) * B + \omega\left(y + \min(\dots)\frac{B-y}{\dots}\right)$$

но на слайде получилось лучше.

---

lection 2



$x_1 \quad x_2 \quad x_3$

$$y = \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3$$
$$\lambda_1 + \lambda_2 + \lambda_3 = 1$$

---

def subspace V

$0 \in V$

$x \in V \Rightarrow \alpha x \in V$

$x, y \in V \Rightarrow x + y \in V$

$\lambda y \in V$

$\lambda y = \lambda(x - x_0)$

$= (\lambda x + (1-\lambda)x_0) - x_0$

$y \in V: \exists x \in C$

$: y = x - x_0$

$\lambda y \in V$

proof

proof $y \in V, t \in V$, $y + t \in V$

$\exists x \in C\ \ y = x - x_0$

$\exists b \in C: t = b - x_0$

$y + t = x - x_0 + b - x_0 =$

$= (x + b - x_0) - x_0$

$\in C?$

$\alpha_1 = 1\ \alpha_2 = 1\ \alpha_3 = -1 \Rightarrow$

$\Rightarrow \sum \alpha_i = 1 \Rightarrow$

$\Rightarrow (x + b - x_0) \in C$

$Ay = 0$

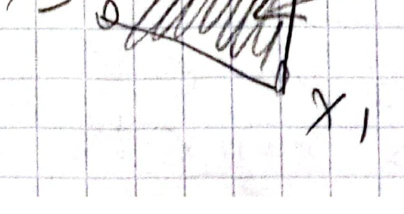
$A(x - x_0) = 0$

$Ax = Ax_0 \rightarrow b$

---

if $0 \le \alpha_1, \alpha_2, \alpha_3 \le 1 \Rightarrow$

$\Rightarrow$



convex set

Affine combination

$$y = L_1 x_1 + L_2 x_2$$

$$L_1 + L_2 = 1$$

Convex combination

$$L_1 + L_2 = 1$$

$$L_1, L_2 \geq 0$$

$C = \{ x \in R_+^2 : x_1 x_2 \geq 0 \}$ is convex

$y \in C : \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \rightarrow y_1 y_2 > 0.$

$\xi \in C = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \rightarrow \xi_1 \xi_2 \geq 0.$

$$x = Ly + (1-L)\xi =$$

$$= \begin{bmatrix} Ly_1 + (1-L)\xi_1 \\ Ly_2 + (1-L)\xi_2 \end{bmatrix} \in R_+ \quad \Big[\text{все полож.}$$

$$x_1 \cdot x_2 = (Ly_1 + (1-L)\xi_1)(Ly_2 + (1-L)\xi_2) =$$

$$= \cancel{L^2 y_1 \xi_1 + (1-L)\xi_1 \xi_2}$$

$$L^2 y_1 y_2 + (1-L)^2 \xi_1 \xi_2 + L(1-L)(y_1 \xi_2 + y_2 \xi_1)$$

$$\geq L^2 0 + (1-L) 0 + \ldots$$

минковский - коши

$x^2 + y^2 \geq 2xy$

$x + y \geq 2\sqrt{xy}$

$\geq \lambda^2 a + (1-\lambda)^2 a + 2\lambda(1-\lambda) a$ $=$

$s^t (y_1 \xi_2 + y_2 \xi_1) \geq 2\sqrt{y_1 y_2 \xi_1 \xi_2} =$

$\geq 2a$

$= (\lambda^2 + (1-\lambda)^2 + 2\lambda(1-\lambda))\, a =$

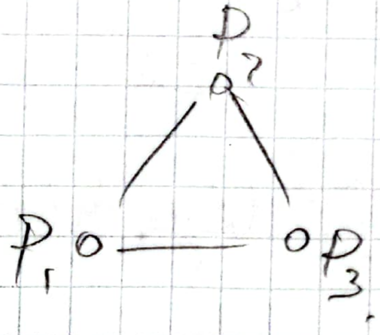
$= (\cancel{\lambda^2} + (1 - 2\lambda + \cancel{\lambda^2}) + 2\lambda(1-\lambda))\, a =$

$= (\lambda + (1-\lambda))^2 a = a$

$x_1 x_2 \geq a$

---

convex combination and convex hull



не понял

кажется

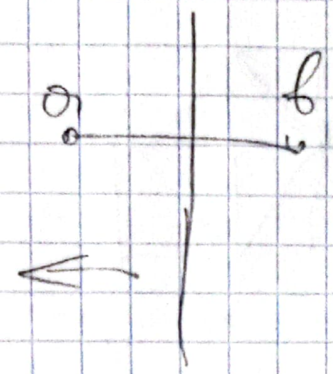
$y = \lambda_1 x_1 + \lambda_2 x_2$,

где y - твой у. сем,

x_1, x_2 - это источники?

convex.

$\|x-a\| \le \|x-b\|$



Euclidean ball.

Norm balls and norm cones

$S = \{x \mid \|x-a\|_2 \le \theta \|x-b\|_2\}$

$\|x-a\|^2 \le \theta \|x-b\|$

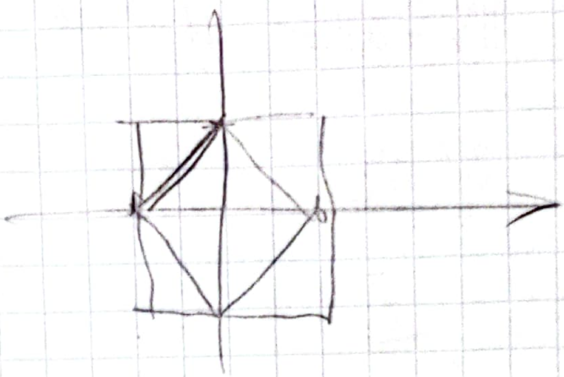
$\|x\|^2 + \|a\|^2 - 2x^Ta \le \theta^2(\|x\|^2 + \|b\|^2 - 2x^Tb)$

$(1-\theta^2)\|x\|^2 - 2(a-b\theta^2)^Tx \le \theta^2\|b\|^2 - \|a\|^2$

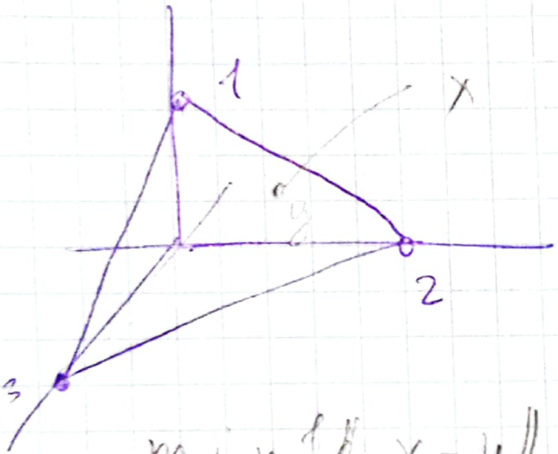
$(1-\theta^2)\left\|x - \frac{(a-b\theta^2)^Tx}{(1-\theta^2)}\right\| \le \theta$

$$\left\|x - \frac{a-b\theta}{1-\theta^2}\right\| \le \frac{\theta^2\left(\|b\|^2 - \|a\|^2 + \frac{\|a-b\theta^2\|}{1-\theta^2}\right)}{1-\theta^2}$$

Polyhedra



simplex



$$\min \frac{1}{2}\|x-y\|_2^2$$

$$x \in \Delta_3 \left| \begin{array}{l} 1^T x = 1 \\ x \geq 0 \end{array} \right.$$

$$\Rightarrow \quad \min L(x, \mu, \lambda) = \frac{1}{2}\|x-y\|_2^2 + \mu(1_n^T x - 1) - \lambda^T x$$

$$\lambda_i x_i = 0.$$

$$\frac{\partial L}{\partial x} = x - y + \mu \cdot 1_n - \lambda = 0$$

$$x^* = y - \mu + \lambda$$

$$x^* = \left[x_I^* = y_I - 1_I \mu \right.$$; понимаю, что

$\underline{\lambda = 0}$ не понял.

$x \quad 1 = 1_I^T x_I = 1_{y_I}^T - I\mu$

$\mu^* = \frac{1_I^T y_I - 1}{I}$

ничего не понял

$x^*_{i \in I} = y_{i \in I} - \mu^* > 0$

$y_{i \in I} > \mu = \frac{1_I^T y_I - 1}{I}$

$y = \begin{bmatrix} y_1 \\ y_i > 0 \\ \hline y_n \end{bmatrix}$ $\} I$

$\mu = \frac{y_1 + y_2 - 1}{2}$

$\mu = \frac{y_2 + y_3 + y - 1}{}$

$\mu < y_3$

$\min \| y - X \|_F^2 \qquad \| y - X \| = \| \frac{y + y^T}{2}$

на слайде $A = A^T$

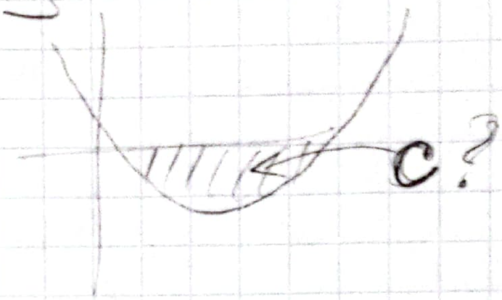
$B = -B^T$

$\| A - B \|_F^2 = \| A \|_F^2 + \| B \|_F^2 - 2\, \cancel{tr(AB^T)}$
$= 0$

на слайд прув.

Lecture № 3.

$C = \{x \in \mathbb{R}^n \mid f(x) = x^T A x + b^T x + c \le 0\}$



Let $x_1, x_2 \in C$, and $z =$
$= \theta x_1 + (1-\theta) x_2$
for $0 \le \theta \le 1$

на слайде

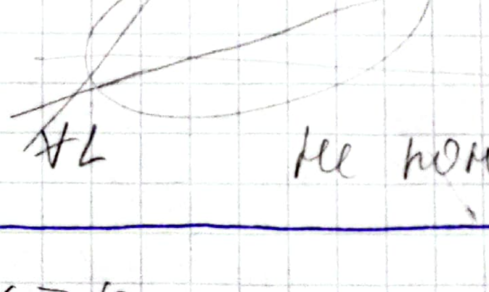
$vec(X) = vec\,[x_1 \mid x_2 \mid x_3] = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

вопрос.

$vec(z^T X z) = vec(z z^T X) = vec(z z^T)^T vec(X) =$

$= (z \otimes z)^T \cdot vec(X) =$

кроон



не понял

$x > y$

$x - y > 0 \Rightarrow x - y \in \mathbb{R}_+$

$x \succeq_{S_+^n} y$ $\qquad S_+^n = \{X : n\times n \ \ \xi^T X \xi \geq 0\}$

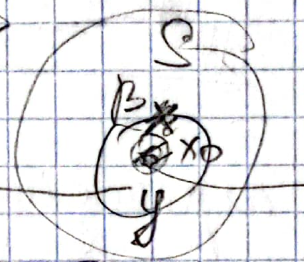
Вопрос про размер-ть X

$x - y \in S_+^n$

$U\Lambda U^T = X \in S_+^n \Rightarrow \lambda \geq 0$

$U(-\Lambda)U^T = -X \in S_+^n \qquad -\lambda \geq 0$

int



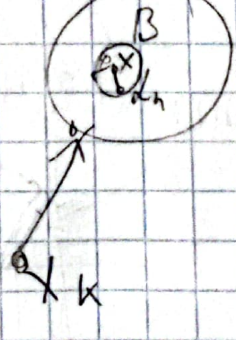
$z \in B(x,\delta) \subset S$

$\lambda_1(x) > 0$ of X

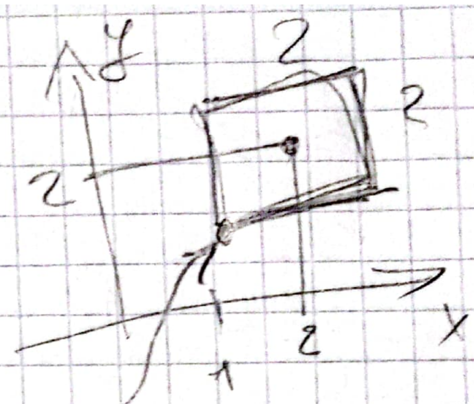
smallest ev.

$u^T Y u > 0$

step 2



$K = \mathbb{R}^2_+$



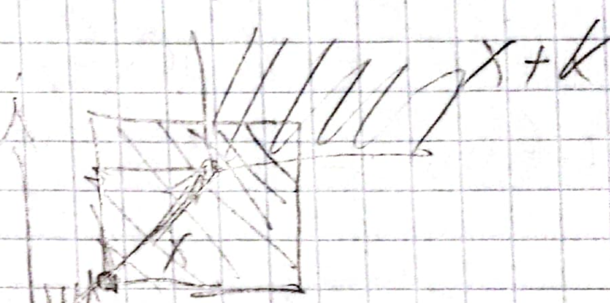
$x \leq_K y$

$y - x \in K = \mathbb{R}^2_+$

$S \subseteq x + K$

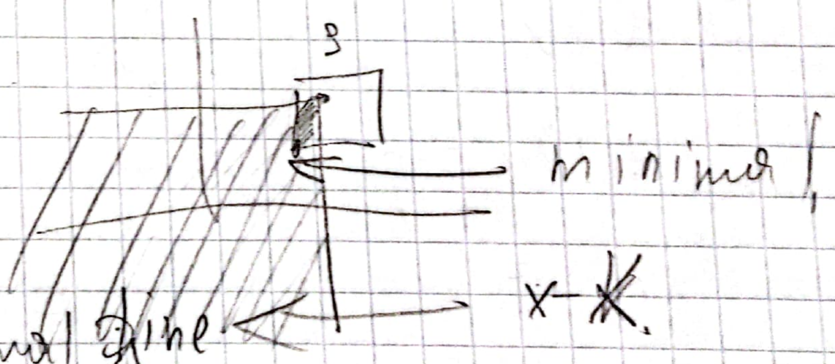
$X + K$

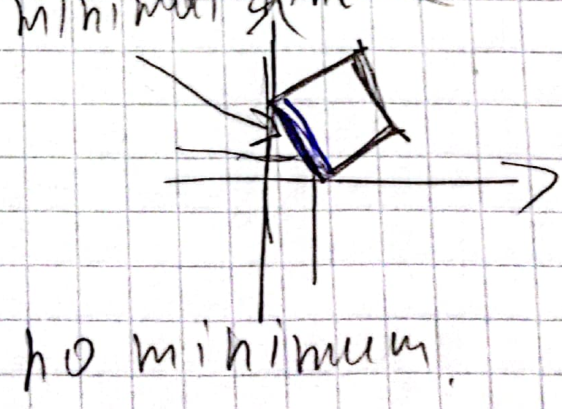
$K = \mathbb{R}^2_+$

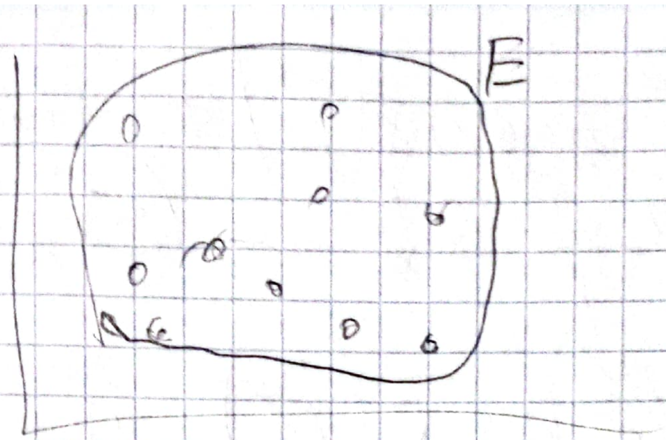


x − min if пересекаются обе фигуры!

$S \cap (x - K) = x$







$[x_1, x_2 \ldots x_k]$

$x_1 = u_1 v.$

$$E = \{(x - x_c)^T Q (x - x_c) \le 1\}$$

$x = 0$

$Q > 0$

$\Downarrow$

$Q \in S^2_{++}$

we need

min $\mathrm{vol}(E) \sim \det Q.$

min $-\log\det Q$

v_k s.t. $x_k^T Q x_k \le 1$

$k = 1 \ldots k$

$$\mathcal{L}(Q, v) = -\log\det Q + \sum v_k (x_k^T Q x_k - 1)$$

$v_k \ge 0$

$$\frac{\partial x_k^T Q x_k}{\partial Q} = \frac{\partial \operatorname{tr}(x_k x_k^T Q)}{\partial Q} = x_k x_k^T$$

$$\frac{\partial(\log\det Q)}{\partial Q} = Q^{-1}$$

$$\frac{\partial \mathcal{L}}{\partial Q} = -Q^{-1} + \sum u_k x_k x_k^T = 0$$

$$Q^* = (X \operatorname{diag}(v) X^T)^{-1}$$

$$\mathcal{L} = -\log\det\left((X \operatorname{diag}(v) X^T)^{-1}\right) + \sum u_k x_k^T Q^{*}_{*} - \sum v_k$$?

$$\operatorname{tr}\left(\sum u_k x_k x_k^T\, Q^{*}\right) \quad tr(I) = n$$

$$X^{-1}$$

$$\mathcal{L} = +\log\det\left((X \operatorname{diag}(v) X^T)^{-1}\right) * - 1^T v +$$

$$v = \lambda \bar{u} \quad \sum u_k = 1$$

$$\mathcal{L}(\bar{u}, \lambda) = \log\det\left(\lambda X \operatorname{diag}(\bar{u}) X^T\right) =$$

$$\log\det\left((X \operatorname{diag}(v) X^T)^{-1}\right) =$$

$$= \log\det A^{-1} = \log \frac{1}{\det A} =$$

$$= -1 \log\det A$$

$$\log\det(\lambda X \operatorname{diag}(u) X^T) - \lambda 1^T \bar{u} + n$$

$$1$$

$$\det(\lambda \underset{n \times n}{A}) = \log \lambda^n \det A =$$

$$=$$

$$\mathcal{L} = n \log \lambda - \lambda + \log\det(X \operatorname{diag}(u) X^T)$$

$$+ n$$

$$\frac{\partial L}{\partial \alpha} = \frac{n}{2} - 1 = 0 \quad \alpha^* = n.$$

$$\max \ \log\det\left(X \operatorname{diag}(u) X^T\right).$$
$$\text{s.t. } 1^T u = 1$$

Prime

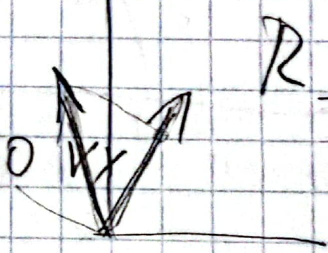
$$\min \ c^T x$$
$$\text{s.t. } Ax - b \succeq 0$$
$$\Downarrow$$
$$Ax - b \in K = \mathbb{R}^n_+$$

Dual

$$\max \ b^T y$$
$$\text{s.t. } A^T y = c$$
$$y \in K^*$$

dual cone

$\mathbb{R}^n_+$



step 1 $A = x x^T$

$$\operatorname{tr}(A^T B) = x^T B x \geq 0 \ \forall x$$
$$B \in S_+.$$

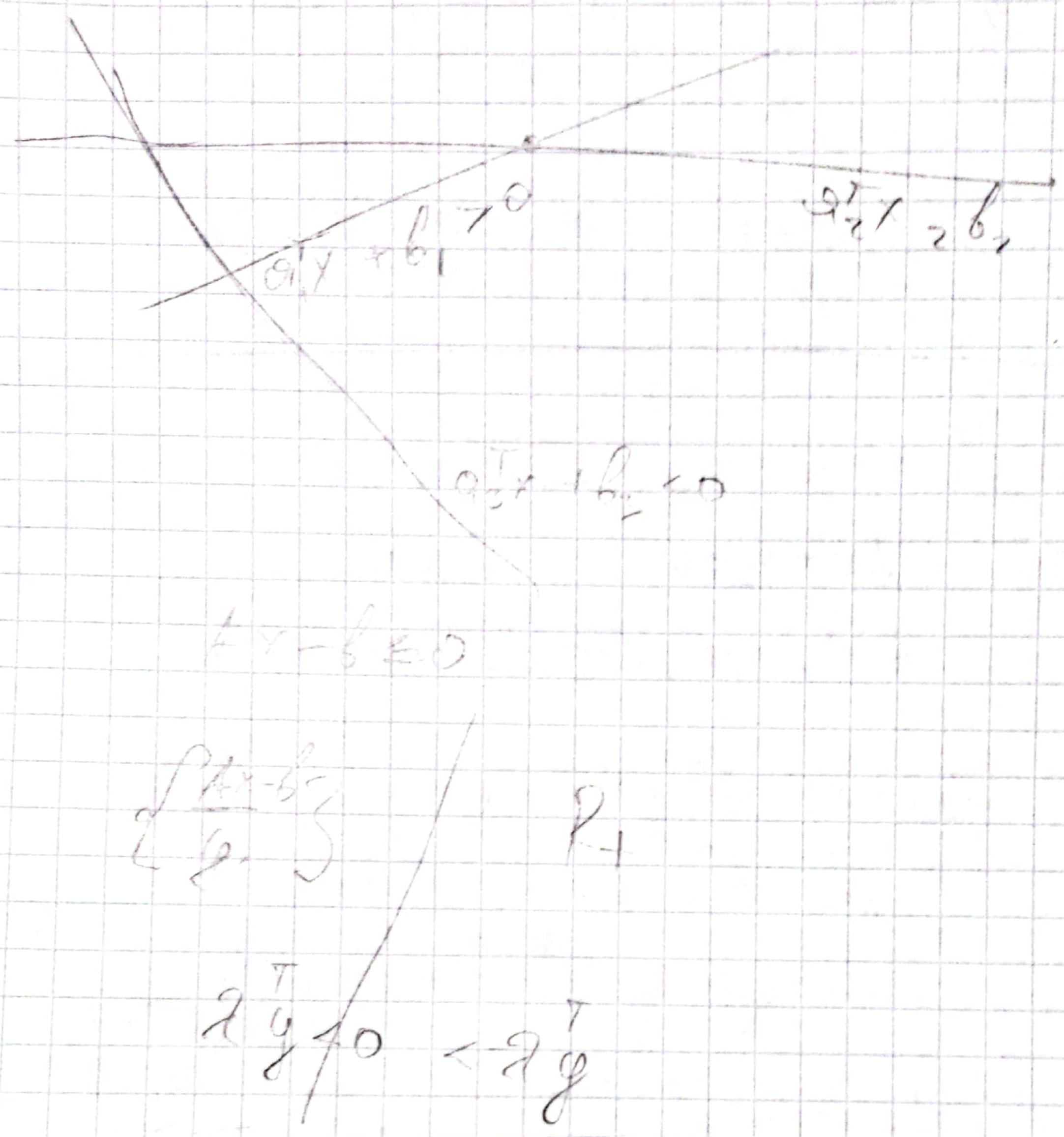
step2 $S_+ \subset K^* \rightarrow K^* \subset S_+$

$$A = U \operatorname{diag}(\lambda) U^T$$

$B \in S^n_+$ prove: $\operatorname{tr}(A^T B) \geq 0 \ \forall A \in S_+$

$$\operatorname{tr}(A^T B) = \sum \underbrace{\lambda_i}_{\geq 0} \underbrace{u_i^T B u_i}_{\geq 0}$$

$B \in K^* \geq 0$



convex function.

$f(x) = e^x$

$x, y \in \mathrm{dom}\, f(\mathbb{R})$

$$e^{\theta x + (1-\theta)y} \leq \theta e^x + (1-\theta)e^y.$$

~~so to~~ hard to prove

$x_1, x_2 \dots x_n$

$$\frac{x_1 + x_2 + \dots + x_n}{n} \geq \sqrt[n]{x_1 x_2 \dots x_n}$$

$$\omega_1 x_1 + \omega_2 x_2 + \dots \omega_n x_n \geq x_1^{\omega_1} x_2^{\omega_2} \dots x_n^{\omega_n}$$

$$i_C(x) = \begin{cases} 0 & x \in C \\ \infty & x \notin C \end{cases}$$

C : convex set

$i_C(x)$: convex func. $H \Rightarrow$?

$x, y \in C$

$z = \lambda x + (1-\lambda) y \in C$

$i_C(x) = i_C(y) = 0 \Rightarrow i_C(z) = 0$

$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$

$0 \leq 0$

if $x, y \notin C$ if $(x) \in C$, $y \notin C$

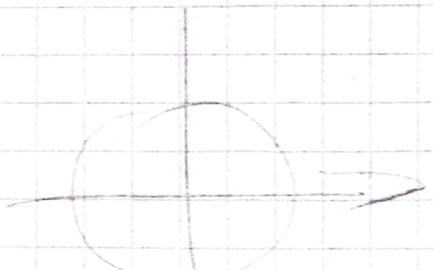
if $x \notin C$, $y \in C$

$\Downarrow$

$$\underbrace{f(\lambda x + (1-\lambda) y)}_{0 \,||\, \infty} \leq \underbrace{\lambda f(x) + (1-\lambda) f(y)}_{\infty}$$

■

$h(x) = \sup x^T z \quad \forall z \in A$



$B = (\|x\|_1 \leq 1)$.

$\log \det (X + tV) \Leftrightarrow$

$\det(AB) = \det(A)\det(B)$

$\Leftrightarrow \log \det (X^{\frac{1}{2}} X^{\frac{1}{2}} + tV) =$

$= \log \det (X^{\frac{1}{2}} (I + t X^{-\frac{1}{2}} V X^{-\frac{1}{2}}) X^{\frac{1}{2}}) =$

$= \log \det (X^{\frac{1}{2}}) \cdot \det (\quad) \det (X^{\frac{1}{2}}) =$

$= \log \det (X) \det (I + t X^{-\frac{1}{2}} V X^{-\frac{1}{2}})$

$\det (I + t V \operatorname{diag}(\lambda) V^T) =$

$= \det (V V^T + t V \operatorname{diag}(\lambda) V^T) =$

$= \cancel{\det} (1 + t\lambda_1)(1 + t\lambda_2) \ldots =$

$\Rightarrow$

concave?

$= \log \det X + \sum \log (1 + t\lambda_k) =$

$h(t) = \log(1 + t z_k)$

$h'(t) = \dfrac{z_k}{1 + t z_k}$

$h''(t) =$

1st order condition.

$t f(x) + (1-t) f(y) \geq f(tx + (1-t)y)$

$t(f(x) - f(y)) \geq \dfrac{f(tx + (1-t)y) - f(y)}{}$

$f(x) \geq f(y) + \lim \dfrac{f(tx + (1-t)y) - f(y)}{t}$

$f(x) \geq f(y) + \nabla f(y)(x - y)$

Lecture 4.

remind!

~~f(x)~~ strongly convex fun - $f(x)$

$\tilde{f}(x) = f(x) - \dfrac{\sigma}{2} \|x\|_2^2$: convex $\Rightarrow$

$\Rightarrow f(x)$ - strongly convex

$\tilde{f}(x)$ - convex

$x, y \in dom(f(x))$

$\theta \in [0, 1]$

$\tilde{f}(\theta x + (1-\theta) y) \leq \theta \tilde{f}(x) + (1-\theta)\tilde{f}(x$

$$- f(\quad) - \frac{\sigma}{2}\|\quad\| \le \theta\left(f(x) - \frac{\sigma}{2}\|x\|_2^2\right) +$$

$$+ (1-\theta)(\quad)$$

$$- \frac{\sigma}{2}\left(\theta^2\|x\|^2 + (1-\theta)^2\|y\|^2 + 2\theta(1-\theta)x^T y\right.$$

$$\left.- \theta\|x\|^2 - (1-\theta)\|y\|^2\right) =$$

$$= \frac{\sigma}{2}\left(\theta(\theta-1)\|x\|^2 + (1-\theta)(-\theta)\|y\|^2 +\right.$$

$$\left.+ 2\theta(1-\theta)x^T y\right) =$$

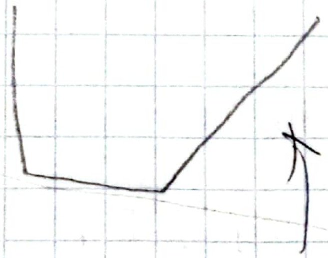
$$= -\frac{\sigma}{2}\theta(1-\theta)\left(\|x\|^2 + \|y\|^2 - 2x^T y\right)$$

---

first ord cond. ||

---

strict conv. fun



not strict conv.

f-a for strong conv-ty.

$$f(y) \geq f(x) + \nabla f^T(x)(y-x) + \frac{\sigma}{2}\|x-y\|_2^2$$

if $f(x)$ - σ strongconv $\Longleftrightarrow$ $\tilde{f}(x)$: conv

$$\tilde{f}(y) \geq \tilde{f}(x) + \nabla \tilde{f}(x)^T(y-x)$$

$$f(y) - \frac{\sigma}{2}\|y\|^2 \geq f(x) - \frac{\sigma}{2}\|x\|^2 + (\nabla f(x) - \sigma x)^T(y-x)$$

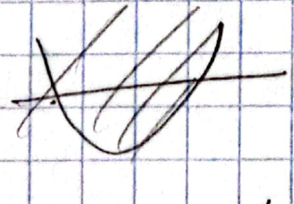
$$f(y) \geq f(x) + \nabla f(x)^T(y-x) + \frac{\sigma}{2}\|y\|^2 - \frac{\sigma}{2}\|x\|^2 \underbrace{- \sigma x^T(y-x)}_{\sigma(-x^Ty + \|x\|^2)} =$$

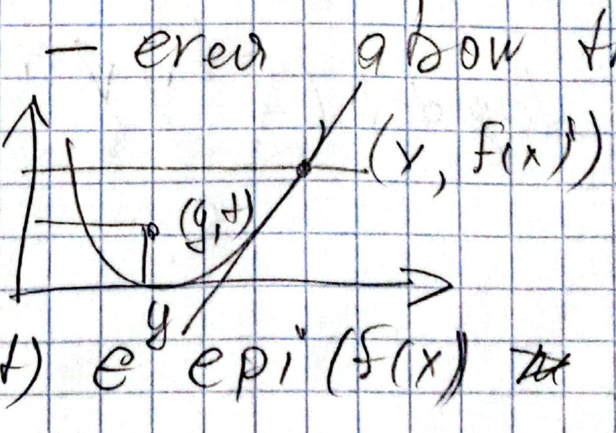
$$= \frac{\sigma}{2}\left(\|x\|^2 + \|y\|^2 - 2x^Ty\right) =$$

$$= \frac{\sigma}{2}\|x-y\|^2$$

---

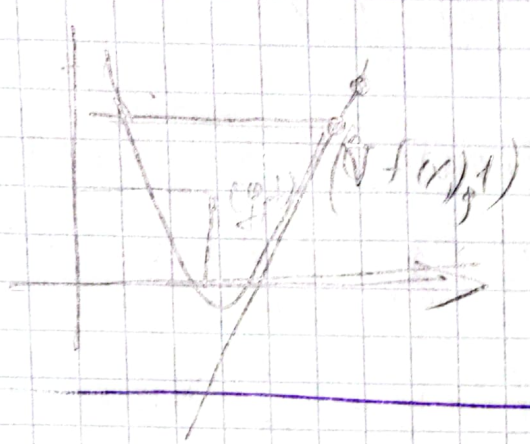
epigraph - erea abow the gr.





$(y,t) \in epi(f(x))$ $t \geq f(y)$

$(\nabla f(x), -1)^T (y, t) \le -f(x) + \nabla f(x)^T x.$



$\nabla_X \log\det(X) = X^{-1}$

* $f(X) = \log\det(X)$

$X, Y \in S^n_+$

$\log\det(Y) - \log\det(X) - \text{trace}(X^{-1}(Y - X))$

$\log(\det(Y)\det(X)^{-1}) - \langle X^{-1}, Y - X \rangle$

$= \log \det(Y)\det(X^{-1}) - tr(X^{-1}Y) + tr(X^{-1}X)$

$$= \log\det(\underbrace{X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}}_{Q}) - tr(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}) + n =$$

$=$

$Q = X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} = U \operatorname{diag} \lambda U^T$

$\log\det(Q) - \operatorname{tr}(Q) + n \le 0$

$\log(\lambda_1 \cdot \lambda_2 \cdots \lambda_n) - \sum \lambda_i + n \le 0$?

$\sum \log \lambda_i - \lambda_i + 1 \le 0$?

gotaqur

$$\begin{cases} \log t - t + 1 \le 0 \\ \text{for } t > 0 \end{cases}$$ — наш. д

Second order condition

$f(x) = \log \sum \exp(x_i)$

$\frac{\partial f(x)}{\partial x_i} = \frac{\exp(x_i)}{\sum \exp x_i}$

$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \begin{cases} i = j \\ \end{cases}$ same

$$H = \left[\operatorname{diag}\left(\frac{e^{x_i}}{\sum e^{x}}\right) - \frac{e^{x_i} e^{x_j}}{(\sum e^{x})^2} \right] =$$

$$H = \left[\operatorname{diag}\left(\underbrace{\frac{e^{x_j}}{\sum_j e^{x_j}}}_{\xi}\right)\right] + \frac{1}{(\sum e^{x_k})^2} e^x (e^x)^T$$

$$H = \operatorname{diag} \xi - \xi \xi^T$$

$$v^T H v = \sum \xi_k v_k^2 - \left(\sum v_k \xi_k\right)^2 =$$

= по сладге по-чему > 0

$$\left(\sum a_i b_i\right)^2 \le \left(\sum a_i^2\right)\left(\sum b_i^2\right)^2$$

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1} u u^T A^{-1}}{1 + u^T A^{-1} u}$$

$$(A - u u^T)^{-1} = A^{-1} + \frac{A^{-1} u u^T A^{-1}}{1 - u^T A^{-1} u}$$

$$H^{-1} = \operatorname{diag} \frac{1}{z} + \frac{\operatorname{diag}(z)\, z z^T \operatorname{diag}(z)}{1 - z^T \operatorname{diag}(\frac{1}{z}) \cdot z} =$$

$$= \operatorname{diag} \frac{1}{z} + \frac{1 \cdot 1^T}{}$$

$$(A+uv^T)^{-1} = A^{-1} - A^{-1}u(I + v^T A^{-1} u)^{-1} v^T A^{-1} =$$

$$(A-uv)^{-1} = A^{-1} - A^{-1}(-u)(I + v^T A^{-1}(-u))^{-1} u^T A^{-1}$$

$$= A^{-1} + \frac{A^{-1} u u^T A^{-1}}{1 - u^T A^{-1} u}$$

что то пошло не так

у проф-ря знаменатель 0.

$$f(x_1, x_2) = x_1^2 + 2x_1x_2 + 3x_2^2 + 2x_1 - 3x_2$$

$$Q = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}$$

$u^T Q u > 0$ может это не док-ть

минор 1й

$$\begin{bmatrix} \boxed{1} & \\ & \boxed{2} \end{bmatrix}$$

$Q > 0$ if all minors ≥ 0 проверить

$$f(x, y, z) = x^2 + y^2 + 3z^2 - xy + 2xz + yz$$

$$Q = \begin{bmatrix} 1 & -\frac{1}{2} & 1 \\ -\frac{1}{2} & 1 & \frac{1}{2} \\ 1 & \frac{1}{2} & 3 \end{bmatrix}$$

Minors

$$\begin{bmatrix} \boxed{1.1} & & \\ & 1 \times \left(\frac{1}{2}\right)^2 & \\ & & \boxed{?} \end{bmatrix}$$

Примеры

$$\begin{cases} y_i = w x_i^k \\ \hat{y}_i = c x_i^k \end{cases}$$

a) for $x_i = [0, 1, 2, 3, 4]$

$k = 2$

$$\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n} = \sum \frac{\left((w-c) x_i^2\right)^2}{n} =$$

$$= \frac{(c-w)^2 \sum x_i^4}{n}.$$

$$\theta_t = \theta_{t-1} + \Delta\theta$$

$$\Delta\theta = -\eta \nabla J_t$$

$$V(t) = \gamma V(t-1) + (1-\gamma) g(t)$$

$$\theta_t = \theta_{t-1} - \left(\gamma V_t + (1-\gamma)\, \eta g(t)\right)$$

Monotonicity

$$(\nabla f(y) - \nabla f(x))^T (y - x) \geq 0.$$

$$(\nabla \tilde{f}(y) - \nabla \tilde{f}(x))^T (y-x)$$

$$\nabla \tilde{f}(x) = \nabla f(x) - \sigma x$$

$$(\nabla f(y) - \sigma y) - \nabla f(x) + \sigma x)^T (y-x) \geq 0$$

$$(\nabla f(y) - \nabla f(x))^T (y-x \geq \sigma (y-x)^T (y-x).$$

Optimality cond.

$\min f(x)$ s.t $x \in \Omega$

минимизация м б вне Ω.

$$g'(\theta) = \nabla f(\theta x - (1-\theta) y)^T (x-y)$$

$$g'(0) = \nabla f(y)^T (x-y) < 0$$

θ	0	1
$g'(\theta)$	<0	

$$g'(1) = \nabla f(x) (x-y)$$

$$g(1) = g(0) + \int_0^1 g'(\theta) \, d\theta \geq g(0) +$$

$$+ g'(0).$$

$\nabla f(x)(x-y) \geq \nabla$

Opt. prob with linear constr.

$\min f(x)$

s.t. $Ay = b$

$(\nabla f(y) - \nabla f(x))^T (y-x) \geq 0$

$y = x^* + v$

$(\nabla f(y) - \nabla f(x^*)) v \geq 0$

$\nabla f(y) v \geq \nabla f(x^*) v$

на слайде.

$\nabla f(x^*)^T \underbrace{(y-x)}$ $= 0$

$\begin{cases} A(x^*) + Av = b \\ A(x^*) = b \end{cases} \Rightarrow Av = 0$

$\nabla f(x^*)^T (\pm v) \geq 0 \Rightarrow$

$\Rightarrow \nabla f(x^*) v = 0 \Rightarrow$

$\Rightarrow \nabla f(x^*) = A^T \lambda$, т.к v в нуль-пространстве A.

$$\mathrm{vec}(a b^T) = b \otimes a$$

$$\mathrm{vec}(ABC^T) = \sum_{ij} b_{ij} a_i c_j^T = \sum b_{ij} \mathrm{ve}$$

$$= \sum b_{ij} \mathrm{vec}(a_i c_j^T) =$$

$$= \sum_{ij} b_{ij} (c_j \otimes a_i) =$$

$$= (C \otimes A) \mathrm{vec}(B)$$

тейлор от разл-я
инверс от
суммы матриц

~~стр 28~~
стр 7

- explain RL. Q-function?
- typo.
- сказать что OPE больше подходит.
- caple remarks.

Lecture 5

non neg constr.

$\nabla f(x^*)^T (y - x^*) \geq 0$

$y, x \geq 0 \quad x \in \mathbb{R}^n_+$

$g = \begin{bmatrix} g_1 \\ \vdots \\ g_n \end{bmatrix}$

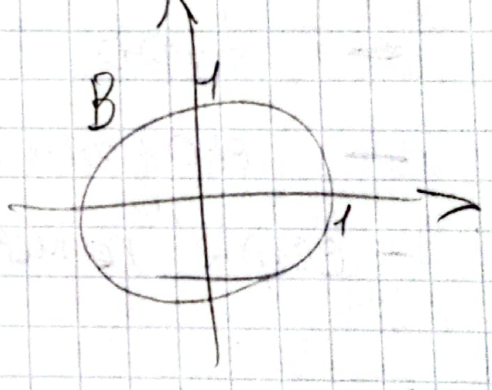
if ~~$g_i < 0 \Rightarrow y_i > x_i$~~

~~$g_i > 0 \quad y_i < x_i$~~

на слайде.

---

over unit ball



$\min f(x)$

s.t. $x \in B(0,1)$

$\nabla f(x^*)^T (y - x^*) \geq 0$

$\min_y \nabla f(x^*)^T y \geq \nabla f(x^*)^T x \quad \forall y \in B$

$\min_y (lhs) \geq rhs$

$$y^* = -\frac{\nabla f(x)}{\|\nabla f(x^*)\|_2}$$

$-\nabla f(x^*)y^* \geq \nabla f(x^*) x^*$,

$-\|\nabla f(x^*)\|_2 = \nabla f(x^*)^T x^*$.

$$x^* = -\frac{\nabla f(x^*)}{\|\nabla f(x^*)\|}$$

$\nabla f(x^*) = \lambda x^*$ for $\lambda < 0$

Example

$$\min \log \left(\sum \exp(x_k)\right) - c^T x$$

s.t. $1x = 1$

$x \geq 0$

$$g = \nabla f(x) = \begin{bmatrix} e^{x_1} \\ \\ e^{x_n} \end{bmatrix} \frac{1}{\sum e^{x_k}} - c = 1_n \cdot u$$

$$\sum g = \frac{\sum e^{x_k}}{\sum e^{x_k}} - 1^T c = \frac{1^T 1}{n} u. =$$

из слайда №47

$\Rightarrow \quad 1 - \sum c_k = n\, u$

так мы находим c

Epigraphs.

слайды 50, 49 надо доказать t внутри эпиграфа

Jensen's inequality

$$\begin{cases} f(t_1 x + t_2 x_2) \le t_1 f(x_1) + t_2 f(x_2) \\ t_1 + t_2 = 1, \quad t_k \ge 0 \end{cases}$$

$$\Downarrow$$

$$f\left(\sum_{k=1}^{n} t_k x_k\right) \le \sum_{k=1}^{n} t_k f(x_k)$$

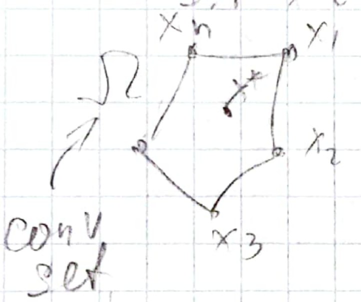
слайды

---

example

$\max f(x)$

s.t. $x \in \Omega$



$x^* =$ ~~point~~ $\operatorname{argmax} f(x_k)$

От обратного, скажем есть $v_k \notin \{x_k\}$

$x^* = \sum x_k \cdot \lambda_k$

$\lambda_k \ge 0$

$\sum \lambda_k = 1$

$f(x^*) = f(\sum x_k \lambda_k) \le \sum \lambda_k f(x_k) \le$

$$\leq \left(\sum \lambda_k\right) f(x^*) = f(x^*)$$

Power of a nonneg. functions

x^2 $\quad x^4$ — strictly convex

$$\max f(x)$$
$$\text{s.t. } x \in B(x_c, r)$$

x^* ~~on~~ on the boundary of B,

$x^* \neq x_c$

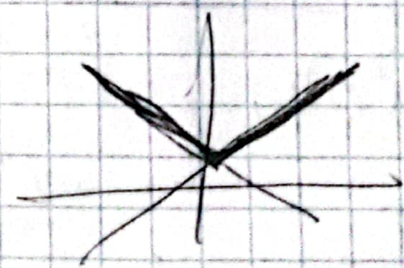
на слайде 58

$$\nabla^2 f \succeq 0$$

$\nabla f(x) \nabla f^T(x) \succeq 0$, т.к. $u^T \nabla f(x) \nabla f^T(x) u \geq 0$

с 60

$$f(x) = \max_{i=1,\dots m} (a_i^T x + b_i)$$



$$\min t$$
$$\text{s.t. } a_1^T x + b_1 - t \leq 0$$
$$a_2^T x + b_2 - t \leq 0$$

C 62.

$h'' > 0$

$g'^{2}(x) > 0$

C 64

$f(x,y) = x^T A x + 2x^T B y + y^T C y$

$g(x) = \min_y f(x,y)$

$\nabla_y f(x,y) = 2B^T x + 2Cy = 0$

$y^* = -C^{-1} B x$

$g(x) = f(x, y^*) =$

$= x^T A x - 2x^T B C^{-1} B x +$

$+ x^T B C^{-1} \cancel{C C^{-1}} B^T x =$

$= x^T (A - B C^{-1} B^T) x.$

Conjugate C 66.

Example

$f(x) = -\log(x)$

$\sup(yx - f(x)) = \sup_D \underbrace{yx + \log(x)}_{g(x)}$

$g' = y + \frac{1}{x} = 0$

$x^* = -\frac{1}{y} \geq 0$

$y \geq 0 \to f^*(y) = \infty$ ← underfined.

$y < 0 \to f^*(y) = y \frac{-1}{y} + \log(-\frac{1}{y})$

$f(x) = \frac{1}{2} x^T Q x$

$g^*(y) = \sup_x y^T x - f(x) = \sup y^T x - \frac{1}{2} x^T Q x$

$\frac{\partial f}{\partial y} = y - Qx = 0$

$x^* = Q^{-1} y$

$f^*(y) = y^T Q^{-1} y - \frac{1}{2} y^T Q^{-1} Q Q^{-1} y =$

$= \frac{1}{2} y^T Q^{-1} y$

---

c 68

what is support function??

не понял.

---

c 71

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

$f^*(y) = \sup \; y^T x - \frac{1}{2} x^T A x - b^T x - c$

$\frac{\partial y}{\partial x} = y - Ax - b = 0$

$Ax = y - b$

?

$x^* = A^{-1}(y - b)$ when $y \in \text{range}(A) + b$

$f^*(y) = y^T A^{-1}(y-b) - \frac{1}{2}(y-b)^T A^{-1} A A^{-1}(y-b) -$

$- b^T A^{-1}(y-b) - c =$

$= (y-b)^T A^{-1}(y-b) - \frac{1}{2}(y-b)^T A^{-1}(y-b) - c =$

= на слайде.

с 72

$f(x) = \sum x_i \ln x_i \qquad 1^T x = 1$

$g = y^T x - \sum x_i \ln x_i \qquad x \geq 0$

$\frac{\partial g}{\partial x} = y - \ln x - 1 = 1_n \mu$

$y_i - \ln x_i - 1 = \mu$

$\ln x_i = y_i - \mu - 1$

$x_i = \exp y_i \cdot \exp^{-(\mu+1)}$

α

сус 44

$$\sum x_i = 1 \Rightarrow \left(\sum \exp(y_i)\right) \lambda = 1$$

$$\lambda = \frac{1}{\sum \exp(y_i)}$$

$$x_i^* = \frac{e^{y_i}}{\sum \exp(y_i)}$$

$$f^* = \sum y_i x_i^* - \sum \quad =$$

$$= \sum y_i \underbrace{\frac{e^{y_i}}{\sum \exp(y_i)}}_{\lambda} - \sum \cancel{\frac{e^{y_i}}{\lambda}} (y_i \cancel{\ln\lambda}) =$$

$$= \ln \sum e^{y_i}$$

Fenchen's Inequality

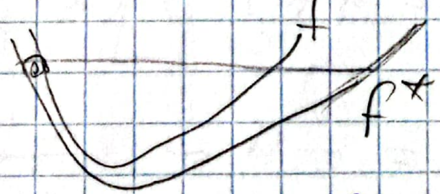
$$f^*(y) = \sup\ y^T x - f(x)$$

$$f^*(y) \geq y^T x - f(x)$$

$$f(x) + f^*(y) \geq x^T y$$

---

c 75



$$a^T(z - x) + b(s - f^{**}(x)) \geq 0$$

$$\cancel{\frac{a^T z}{b}} \leq s.$$

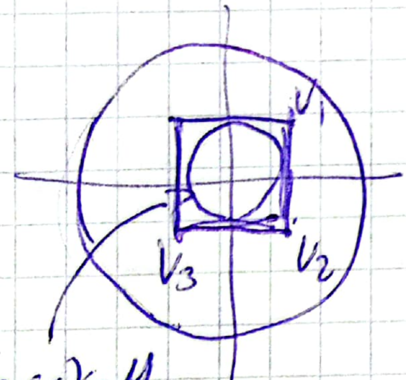
$$-\frac{a^T}{b}(z-x) - s + f^{**}(x) \le -\frac{c}{b}$$

$$y^T z - s + y^T x + f^{**}(x) \le -\frac{c}{b}$$

$$\frac{y^T z - s}{y^T - f(z)}$$

– слайды 15

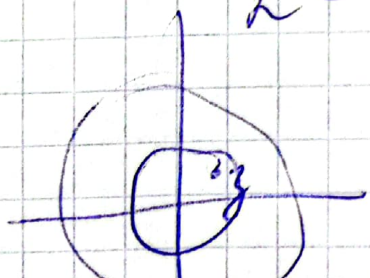
с 17. software license ..



$f(x) \le M$

$$f(x \in B) = f\left(\sum \lambda_k v_k\right) \le$$
$$\le \sum \lambda_k f(v_k)$$

$$\lambda = \frac{\|x - x_0\|}{\varepsilon} \le 1 \; ?$$



$$z = x_0 + \frac{1}{\lambda}(x - x_0) =$$

$$= \frac{1}{\lambda} x + \frac{\lambda - 1}{\lambda} x_0 =$$

$$x = \lambda z + (1 - \lambda) x_0$$

Приколы

$$y = \mathrm{Relu}(x) = \max(x, 0)$$

$$= \min_y \|x - y\|_2^2$$
$$\text{s.t. } y \geq 0.$$

$$\min \|x - y\|_2^2$$

$$\alpha y \geq d \qquad \alpha \sim N(1, \sigma^2)$$

$$\alpha y - d \geq 0 \qquad d \sim N(\sigma, \gamma^2)$$

$$\underbrace{[\alpha, -d]}_{w^T} \begin{bmatrix} y \\ 1 \end{bmatrix} \geq 0$$

$$w = \begin{bmatrix} \alpha \\ -d \end{bmatrix} \sim \left(\begin{bmatrix} 1 \\ -d \end{bmatrix}, \begin{bmatrix} \sigma^2 & \\ & \gamma^2 \end{bmatrix} \right)$$

$$t = w^T \begin{bmatrix} y \\ 1 \end{bmatrix} \sim N\left(\begin{bmatrix} y - d \end{bmatrix}, \begin{bmatrix} & \end{bmatrix} \right)$$

$$E(t) = E(w^T y) = y - d$$

$$\mathrm{var}(t) = E(t^2) - E(t)^2 =$$

$$= E\left(w^T \begin{bmatrix} y \\ 1 \end{bmatrix}\right)^2 - (y - d)^2 =$$

$$=$$

$$\mathrm{var}(t) = [y \;\; 1]^T \begin{bmatrix} \sigma^2 & \\ & \gamma^2 \end{bmatrix} \begin{bmatrix} y \\ 1 \end{bmatrix} = \sigma^2 y^2 + \gamma^2.$$

$$Pr\ (\alpha y \geq t) = Pr\left(\omega^T \begin{bmatrix} y \\ 1 \end{bmatrix} \geq 0\right) =$$

$$= Pr\ (t \geq 0)$$

$$\Rightarrow Pr\left(\frac{t - \bar{t}}{\sigma_t} \geq \frac{-\bar{t}}{\sigma_t}\right) \Rightarrow$$

$$\Rightarrow \frac{\bar{t}}{\sigma_t} \geq \Phi^{-1}(1-\varepsilon)$$

$$1 - \varepsilon = Pr\ (\alpha y \geq d) = Pr\left(\omega^T \begin{bmatrix} y \\ 1 \end{bmatrix}\right) \geq 0$$

$$\frac{y - \bar{d}}{\sqrt{\sigma^2 y^2 + \gamma^2}} = \frac{\bar{t}}{\sigma_t} \geq \Phi^{-1}(1-\varepsilon)$$

$$\bar{d} = 0 \quad \gamma = \gamma.$$

$$\min \|y - x\|_2^2$$

$$\text{s.t.} \quad \frac{y}{\sqrt{y^2+1}} \geq \sigma \Phi^{-1}(1-\varepsilon)$$

$$y = \tan(u)$$

$$\frac{y}{\sqrt{y^2+1}} = \sin u \geq \sigma \Phi^{-1}(1-\varepsilon)$$

$$y \geq \operatorname{atan}\left(\sigma \Phi^{-1}(1-\varepsilon)\right)$$

Lecture 6

CVXPY

Relu : $y = \max(x, 0)$

$y = \operatorname{argmin} \| x - y \|^2$

s.t $y \geq 0$

$y = \operatorname{argmin} \| x - y \|^2$

s.t. $\text{Prob}(y \geq 0) = 1 - \varepsilon$

0.70 0.5

$y = \_\_\_\_\_\_$

$\text{Prob}(Ly \geq d) = 1 - \varepsilon$

$L \sim N(1, \sigma^2)$

$d \sim N(\bar{d}, \gamma^2)$

$y = \operatorname{argmin} \| y - x \|^2$

$\text{Prob}(Ly - d \geq 0)$

$w \sim N\left(\begin{bmatrix} 1 \\ -\bar{d} \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 \\ 0 & \gamma^2 \end{bmatrix}\right)$

$w \boxed{\begin{matrix} y \\ 1 \end{matrix}} = [L; -d] \begin{bmatrix} y \\ 1 \end{bmatrix}$

$$t = \omega \begin{bmatrix} y \\ 1 \end{bmatrix} \sim N\left[\underbrace{y - d}_{\mu_t}\right] \quad \underbrace{\sigma_y^2 y^2 + \gamma^2}_{\sigma_t^2}$$

$$\mathrm{Prob}(t \geq 0) = 1 - \varepsilon \longrightarrow \mathrm{Prob}\left(\frac{\mu - t}{\sigma_t} \leq \frac{\mu}{\sigma_t}\right)$$

$$\mathrm{Prob}\Big(\underbrace{\frac{t - \mu_t}{\sigma_t}}_{\sim N(0,1)} \geq \frac{-\mu}{\sigma_t}\Big) = 1 - \varepsilon$$

$$\Phi\left(\frac{\mu_t}{\sigma_t}\right) = 1 - \varepsilon$$

$$\frac{\mu}{\sigma_t} \geq \Phi^{-1}(1 - \varepsilon)$$

$$\mathrm{Prob}(\alpha \leq x) = 1 - \varepsilon \quad \text{normal d. } N(0,1)$$

$$x > \Phi$$

$$\operatorname*{argmin}_{y} \|x - y\|^2$$

s.t.
$$\frac{y - d}{\sqrt{\sigma_y^2 y^2 + \gamma^2}} \geq \Phi^{-1}(1 - \varepsilon)$$

if $\sigma = \gamma$

$$\frac{y - d}{\sqrt{\sigma_y^2 y^2 + \sigma^2}} = \frac{y - d}{\sigma\sqrt{y + 1}} \geq \Phi^{-1}(1 - \varepsilon)$$

$$\frac{y - d}{\sqrt{y + 1}} \geq \sigma\,\Phi^{-1}(1 - \varepsilon)$$

$$\frac{y}{\sqrt{y^2+1}} - \frac{d}{\sqrt{y^2+1}} \geq \sin(u-\theta)$$

if $y = \tan u$

$\sin u$

$$\cos(\theta)\sin(u) - \cos(u)\sin(\theta)$$

$$\cos\theta = \frac{1}{d^2+1}$$

$$\min \|x - \tan(u)\|^2$$

$$\text{s.t. } \sin(u-\theta) \geq \frac{\sigma\Phi^{-1}(1-\epsilon)}{\sqrt{d^2 - ?}}$$

$y \sim N(\bar{y}, \sigma_y)$

$$\text{Prob}(\underline{ay - d} > 0) = 1 - \epsilon$$

$t \sim N(\mu_t, \sigma_t^2)$

$a \sim N(1, \sigma^2)$

$d \sim N(\bar{d}, \delta^2)$

$E(ay) = \bar{y} \qquad \text{var}(ay) = E((ay)^2) - E(ay)^2 =$

$= E(a^2) E(y^2) = \bar{y}^2 =$

$\cancel{= \sigma^2 \sigma_y^2 - \bar{y}^2} =$

$t = ay - d = \qquad = (\text{var}(a) + \mu_a^2)(\text{var}(y) +$

$+\mu^2) - \bar{y}^2 = (1+\sigma^2)\sigma_y^2 + \sigma^2 \bar{y}^2$

$t = ay - d$

$E(t) = \bar{y} - d$

$var(t) = (1+\sigma^2)\sigma_y^2 + \sigma^2 \bar{y}^2 + \sigma^2$

$g(t) = f(y) - f(x) - \nabla f(x)^T (y-x)$

$g'(t) = (\nabla f(x+th) - \nabla f(x))^T h \leq$

$a^T b \leq \|a\| \|b\|$

на слайде

$g(x) = f(y) - f(x) - \nabla f(x)^T (y-x) \leq \frac{L}{2}$

с 6

$$f(y) \leq f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\|x-y\|^2$$

$y = x_k - \frac{1}{L} \nabla f(x_k)$

слайд

$$f(x^*) \leq f(x_{k+1}) \leq f(x_k) - \frac{1}{2L} \|\nabla f(x_k)\|^2$$

c 8

$$f(y) \geq f(x) + \nabla f(x)^T (y-x)$$

$$\nabla f(x)^T (y-x) \leq f(y) - f(x)$$

$$\|x_{k+1} - x^*\| = \|x_k - x^*\|^2 +$$

$+ \quad t^2 \|\nabla$ ...

$$\sum_{k=0} f(x_{k+1}) - f(x^*) \leq \sum \frac{\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2}{2t} =$$

$f(x_0)$

$$N(f(x_N) - f(x^*)) \leq \frac{\|x_0 - x^*\|^2 - \|x_N - x^*\|^2}{2t} \leq \frac{\|x_0 - x^*\|^2}{2t}$$

симуляция

strongly convex if $\lambda_{min}(H) \geq m > 0$

L-smooth if $\lambda_{max}(H) \leq L$

$$f(x) = \frac{1}{2} x^T Q x$$

$$Q = \begin{bmatrix} 1 & \\ & \gamma \end{bmatrix} \quad \gamma > 1$$

$$g(x) = \dot{f} = Qx$$

$$x_{k+1} = x_k - t\, g(x_k)$$

$= x_k + Q x_k = (I + Q) x_k =$

$= (I + Q)^{k+1} x_0$

$\min f(x)$
$x \in C$

$\Downarrow$

$\min f(x) + i_C(z)$

$x = z$

$z \in Z$

$\mathcal{L}(x, z, \lambda) = f(x) + i_C(z) + \lambda^T(x - z) + \frac{\gamma}{2} \|x - z\|^2$

$x^{k+1} = \arg\min \mathcal{L}(x, z^k, \lambda^k) =$

$= \arg\min f(x) + \lambda^{k^T}(x - z_k) + \frac{\gamma}{2} \|x - z_k\|^2$

$z^{k+1} = \arg\min \mathcal{L}(x^{k+1}, z, \lambda^k) =$

$= \arg\min i_C(z) + \frac{\gamma}{2} \|x^k - \lambda^k - z\|_2^2$

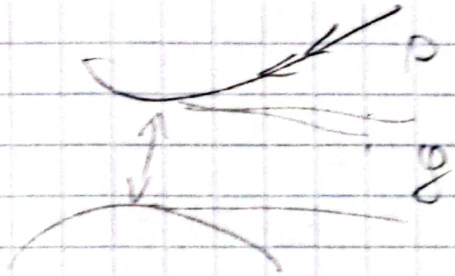
было на перв. занятии,
я нихуя не понимаю.

c 5 primal dual,

$$\min_{x} \max_{\lambda} \mathcal{L}(x,\lambda) \geq \max_{\lambda} \min_{x} \mathcal{L}(x,\lambda)$$

$$\max_{\lambda} \mathcal{L}(x,\lambda) \geq \min \mathcal{L}(x,\lambda)$$

min max



gap dual-primal

слайды с 9

керотив ?

$$\mathcal{L}(x,\lambda) = \frac{1}{2}(x-2)^2 + \lambda(x^2-1)$$

≤ 0

$$g(\lambda) = \min \mathcal{L}(x,\lambda)$$

$$\frac{\partial \mathcal{L}}{\partial x} = x - 2 + 2\lambda x = 0$$

$$x^* = \frac{2}{1+2\lambda}$$

$g(\lambda) = \frac{1}{2}$ — подставим x^*

на слайде 13

$\max g(\lambda)$

с 16 Sat

$\underset{m \times n}{A} \, x = b$ → plenty solutions

$$\mathcal{L}(\underset{n \times 1}{x}, \underset{m \times 1}{\mu}) = \frac{1}{2} \|x\|_2^2 + \mu^T (Ax - b)$$

$$g(\mu) = \min_x \mathcal{L}(x, \mu)$$

$$\frac{\partial \mathcal{L}}{\partial x} = x + A^T \mu = 0$$

$$x^* = -A^T \mu$$

$$g(\mu) = \frac{1}{2} \|A^T \mu\|_2^2 + \mu^T (-AA^T \mu - b) =$$

$$= \frac{1}{2} \mu^T A A^T \mu - \mu^T A A^T \mu - \mu^T b$$

$$= -\frac{1}{2} \mu^T A A^T \mu - \mu^T b$$

$$\frac{\partial \varphi}{\partial \mu} = -AA^T\mu - b = 0$$

$$\mu = \cancel{\quad} -(AA^T)^{-1} b$$

$$x^* = -A^T\mu = +\underset{n\times m}{A^T}(\underset{m\times n\ n\times m}{AA^T})^{-1}\underset{m\times 1}{b}$$

For L_1 norm.

$$\min \|x\|_1$$
$$\text{s.t. } Ax = b$$

Lecture 7

$Ax = b$ ← fat matrix

e 17

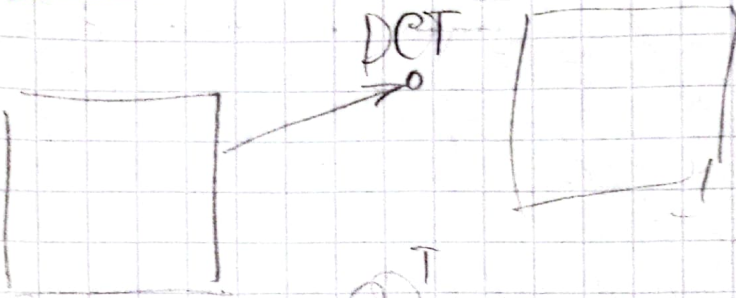
$$\min \|x\|_2 \quad\rightarrow\quad \min \sum w_i x_i^2$$
$$\text{s.t. } Ax = b \qquad\qquad \text{s.t. } Ax = b$$

$$\min \sum z_i^2 \qquad z_i = \sqrt{w_i}\, x_i$$

$$\text{s.t. } \underbrace{A \cdot \operatorname{diag}\left(\frac{1}{\sqrt{w}}\right)}_{\tilde{A}} \cdot z = b$$

$\Downarrow$

$$\min \|z\|_2^2$$
$$\text{s.t. } \tilde{A} \cdot z = b$$



$\Phi^T \to$

$\Phi = y$

$\|x\| = \sum_{i=1}^{m} |x_i|$

$f(y) \geq f(x) + \nabla f(x)^T (y - x)$

g

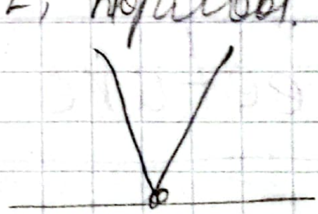
проблема в т-ой L_1 нормы

sub gradient

$x < 0: \quad g = -1$

$x > 0: \quad g = 1$

$x = 0: \quad -1 \leq g \leq 1$



$\min \|x\|_1$

$\text{s.t. } Ax = b$

$\mathcal{L}(x, \mu) = \|x\|_1 + \mu^T (Ax - b)$

$g(\mu) = \inf_x \|x\|_1 + \mu^T (Ax - b)$

$0 \in \frac{\partial \mathcal{L}}{\partial x} = g + A^T \cdot \mu$

subgradient of $\|x\|$ with respect to x

$$\zeta : m \times 1 = \begin{bmatrix} 1 \\ -1 \\ \vdots \\ t \in [-1; 1] \\ \vdots \end{bmatrix} \quad \begin{matrix} \to x_+ : \{x_i > 0\} \\ \to x_- : \{x_i < 0\} \end{matrix}$$

$$\zeta + A^T \mu = 0.$$

$$\begin{matrix} A_+ : \{x_i > 0\} \\ A_- : \{x_i < 0\}. \end{matrix}$$

$$\begin{bmatrix} 1 \\ -1 \\ \zeta_0 \end{bmatrix} + \begin{bmatrix} A_+ \\ A_- \\ A_0 \end{bmatrix}^T \cdot \mu = 0$$

$$\begin{bmatrix} 1 \\ -1 \\ \zeta_0 \end{bmatrix} + \begin{bmatrix} A_+ & A_- & A_0 \end{bmatrix} \begin{bmatrix} \mu \end{bmatrix} = 0.$$

$$\begin{bmatrix} A_+ \circ \mu_+ \\ A_- \ \mu_- \\ A_0 \ \mu_0 \end{bmatrix} \begin{matrix} = -1 \\ = 1 \\ \le 1 \end{matrix} \qquad -1 \le A_0 \mu_0$$

$$\mathcal{L}(x, \mu) = \sum |x_i| + \underbrace{\mu^T A x}_{-x_i} - \mu^T b =$$

$$= \sum_{I_+} x_i - x_i + \sum_{i \in I_-} -x_i + x_i - \mu^T b.$$

Dual problem: $g = -\mu^T b$

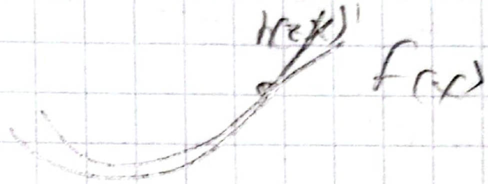
max $-\mu^T b$

s.t. $-1 \le A^T \mu \le 1$

решаем то же реформулируя

$$f(x) = \sum |x_i|$$

$$h(x) = \sum z_i + \frac{x_i^2}{z_i}$$



$$z_i + \frac{x_i^2}{z_i} \ge 2\sqrt{z_i \frac{|x_i|^2}{z_i}} = 2|x_i|$$

хорошая function ??

step 1:

$$z^{k+1} = \arg\min_{z} h(z, x^k) \quad \boxed{\begin{aligned} &\min \sum z_k + \frac{x_k^2}{z_k} \\ &\text{s.t. } Ax = b \end{aligned}}$$

$$x^{k+1} = \arg\min h(z^{k+1}, x) \Rightarrow x^* = |z|$$

$$\frac{1}{2}\left(\sum_k z_k + \frac{x_k^2}{z_k}\right) \ge \sum_k \sqrt{z_k \frac{|x_k|^2}{z_k}} = \sum |x_k|$$

equal only if $z_k^* = |x_k|$

$$z_k^* = |x_k|$$

$$z^{k+1} = \arg\min_z h(z, x^k) \rightarrow \min \sum \frac{x_k^2}{z_k}$$

$$s.t.\ Ax = b$$

$$\Downarrow$$

$$\sum w_i x_i$$

$$x^* = (\mathrm{diag}(z) A^T A \cdot \mathrm{diag}(z) A^T)^{-1} b \quad s.t.\ Ax = b$$

C 2.2

$$g(v) = \frac{1}{2}(x^* A^T A x + b^T b - 2 x^{*T} A^T b) + v^T (Bx - d) =$$

$$= \frac{1}{2} x^{*T} A^T A x^* - x^* (A^T b - B^T v) + \frac{1}{2} b^T b - v^T d$$

$$g(v) = -\frac{1}{2} (A^T b - B^T v)^T (A^T A)^{-1}$$

$$\frac{\partial g}{\partial v} = B (A^T A)^{-1} (A^T b - B v) - d = 0$$

$$v^* = (B (A^T A)^{-1} B^T)^{-1} (B (A^T A)^{-1} A^T b - d)$$

e 23. $A = V \operatorname{diag}(\sigma) U^T$

$A^T A = U \operatorname{diag}(\sigma^2) U^T$,

$(A^T A + \lambda I)^{-1} = (U \operatorname{diag}(\sigma^2 + \lambda) U^T)^{-1}$

$= U \operatorname{diag} \frac{1}{\sigma^2+\lambda} U^T$.

$A(A^T A + \lambda I)^{-1} A^T =$ поменять

$U \to V$

$V^T \to U^T$

$= V \operatorname{diag}(\sigma) \cancel{U^T} \cdot (\cancel{U} \operatorname{diag} \frac{1}{\sigma^2+\lambda} \cancel{U^T}) \cancel{U} \operatorname{diag}(\sigma) V^T =$

$= \frac{1}{2} \| V \operatorname{diag}\left(\frac{\sigma^2}{\sigma^2+\lambda}\right) V^T b - b \| +$

$+ \frac{\lambda}{2} \| U \operatorname{diag}\left(\frac{\sigma}{\sigma^2+\lambda}\right) V^T b \|^2 - \lambda c =$

$= \frac{1}{2} \| \operatorname{diag} \frac{\sigma^2}{\sigma^2+\lambda} \underbrace{V^T b}_{=t} - \underbrace{V^T b}_{t} \|^2 +$

$+ \frac{\lambda}{2} \| \operatorname{diag} \frac{\sigma}{\sigma^2+\lambda} V^T b \|^2 - \lambda c =$

$= \| \frac{\sigma^2}{\sigma^2+\lambda} \cdot t - t \| =$

$= \frac{1}{2} \| \operatorname{diag} \frac{-\lambda}{\sigma^2+\lambda} \cdot t \| + \ldots =$

$$= -\frac{1}{2}\left\| \frac{\lambda t}{\sigma^2+\lambda} \right\|^2 + \frac{\lambda}{2}\left\| \frac{\sigma \lambda t}{\sigma^2+\lambda} \right\|^2 - \lambda c =$$

$$= -\frac{1}{2}\sum \frac{\lambda^2 t_i^2}{(\sigma_i^2+\lambda)} + \frac{\lambda}{2}\frac{\sigma_i^2 t_i}{\sigma_i^2+\lambda^2} - \lambda c =$$

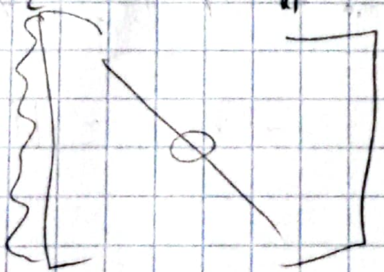
$$= \frac{\lambda}{2}\sum \frac{t_i^2(\lambda+\sigma^2)}{(\sigma_i^2+\lambda)^2} - \lambda c.$$

$$= \frac{\lambda}{2}\sum \frac{t_i^2(}{\sigma^2+\lambda} - \lambda c.$$

---

c 27.

$x_1, x_2 \;--\; x_n$

$W =$  $\omega_{ij} = \|x_i - x_j\|^2$

$x_1, x_2, x_3 \mid x_4 \;..\; x_n$

labels : $1 \quad 1 \quad 1 \mid -1 , \; , -1$

$x_i = 1$ or -1

на слайдах с 29

dual of dual

dual - dual

$\mathcal{L}(\nu, Z) = -1^T\nu + \langle Z, -W_? - \mathrm{diag}(\nu)\rangle$

$Z \in S^n_+$

trace( )

$g(Z) = \inf_{\nu} \; 1^T\nu - tr(Z\, \mathrm{diag}(\nu)) - tr(ZW)$

$\mathrm{diag}(Z)^T\nu$

~~diag Z~~

$= \inf (1^T - \mathrm{diag}(Z)^T\nu - tr(ZW)) =$

$=$ ~~inf Z~~ $-tr(ZW)$ when $\mathrm{diag}\, Z = 1$

Primal

$x^T W x = tr(xx^T W)$, $xx^T = Z$

s.t. $x^2 = 1$, $x^2 = \mathrm{diag}\, Z$

$tr\, Z$

$z^* = x^* x^T$

$\inf f(x) = -\sup -f(x)$

$\inf (f(x) + (A^T\lambda + C^T u)^T x - b^T\lambda - d^T\nu)$

$= -\sup \underbrace{\underbrace{(-A^T\lambda - C^T u)}^T x - f(x)}_{y} + b^T\lambda + d^T\nu$

$f_0^*(-A^T\lambda - C^T\nu)$

$\frac{1}{2} x^T x \to \frac{1}{2} y^T y$

$\|x\|_1 \to I_\infty(y)$

индикатор.

$\min x_i \log x_i$

$A = -I \quad b = 0$

$C = 1^T \quad d = 1$

$-\nu - \sum e^{\lambda_i - \nu - 1} \to \min$ dll

$g'(\nu) = -1 + n \exp(-(1+\nu)) =$

$= 0 \qquad \frac{1}{n} = e^{-(1+v)}$

$1 + v = \ln(n)$

$\min \max \{a_i x + b_i\}$

$\max\{a_i x + b_i\} = t$

$\min t$

s.t. $A^T x + b \le t$

c 38

---

c 40

$g_i(x) \le 0 \qquad g = \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix}$

---

c 43.

$\mathcal{L}(x, \lambda) = x^T A x + 2 b^T x + \lambda(x^T x - 1)$

$\lambda \geq 0$

$$= x^T(A + \lambda I)x + 2b^T x - \lambda$$

Q должен $\in S^n_+$.

$$\cancel{x^* = (A+\lambda I)^{-1}}$$

$$\frac{\partial f}{\partial x} = 2Qx + 2b = 0$$

$$x = -Q^{-1}b$$

$$x^* = -Q^{-1}b.$$

$b \in R(Q)$

$b \in$ column space Q.

$$\max -b^T(A+\lambda I)^{-1}b - \lambda.$$

$$\text{s.t. } A + \lambda I \succeq 0$$

$$b \in R(A + \lambda I)$$

$$A = U \operatorname{diag}(\sigma) U^T.$$

$$b^T(A+\lambda I)^{-1}b = b^T(U \operatorname{diag}(\sigma) U^T + \lambda I$$

$$= c^T\left(\operatorname{diag}\frac{1}{\sigma+\lambda}\right)c = \sum \frac{c_k^2}{\sigma_k+\lambda}$$

$$\min \sum \frac{\sigma_k}{\sigma_k + a} + a \quad \text{s.t. } a \geq -\sigma_n.$$

~~st~~ SVM e45

$$\min \|w\|_2^2$$

s.t. $y_k (w^T x + b) \geq 1.$

$$-y_k w^T x_k - y_k b \leq -1$$

$$\begin{bmatrix} y_k x_k \\ y_k \end{bmatrix}^T \begin{bmatrix} w \\ b \end{bmatrix} \leq -1$$

scalar

$$A = -\begin{bmatrix} y_1 x_1 & y_2 x_2 & \cdots & y_n x_n \\ y_1 & y_2 & \cdots & y_n \end{bmatrix}$$

$$A^T \begin{bmatrix} w \\ b \end{bmatrix} \leq -1.$$

$$\frac{\partial f}{\partial x} = x - A^T \lambda = 0$$

$$x^* = A^T \lambda$$

$$g(\ ) = -\frac{1}{2} \lambda^T A A^T \lambda + c\,1^T z + \lambda(1-\varepsilon) - v$$

$$\cancel{-v^T z_1}$$

$$\underbrace{(c\,1 - \lambda - v^T)}_{=0} z$$

иначе нет минимума?

Lecture 8

№3 HW

$U = \{u_1, u_2, u_3\}$

$\min \|Y - UX\|_F^2$

$\min \|U^T Y - X\|_F^2 \rightarrow \min \sum \|y_k - x_k\|^2$

$x_k \geq 0$

$1^T x_k = 1$

$\min \|Y - U X\|_F^2$

s.t. $U^T U = I$

$\mathcal{L}(U, T) = \frac{1}{2} \|Y - UX\|_F^2 + \langle T, U^T U - I \rangle$

$\frac{\partial \mathcal{L}}{\partial U} = -(Y - UX) X^T + UT =$

$U^* = Y$

$\boxed{U^* = AB^T}$

$\min f(X)$

$\left.\text{s.t. } X^T X = I \right\} \rightarrow f(X) - \frac{1}{2} tr(\Lambda(X^T X - I))$

$\frac{\partial f}{\partial X} = \overbrace{\frac{\partial f}{\partial X}}^{G} - \Lambda X = G - \Lambda X$

$$\nabla f = G - X\Lambda = 0$$

$$X^T G - X^T X \Lambda = 0$$

$$\Lambda = X^T G$$ < close потому что симм.

$$\nabla f = G - \underbrace{X G^T X}_{} =$$

$$= \underbrace{(G X^T - X G^T)}_{F} X.$$

$$F = [G \quad X] \begin{bmatrix} X^T \\ -G^T \end{bmatrix} =$$

$$=$$

$$Y^{k+1} = X^k - \eta F \underbrace{X^k}_{\frac{X^{k+1} + X^k}{2}}$$

$$X^{k+1} = X^k - \frac{\eta}{2} F X^{k+1} - \frac{\eta}{2} F X^k$$

$$\left(I + \frac{\eta}{2} F\right) X^{k+1} = X^k - \frac{\eta}{2} F X^k = \left(I - \frac{\eta}{2} F\right) X^k$$

$$X^{k+1} = \left(I + \frac{\eta}{2} F\right)^{-1} \left(I - \frac{\eta}{2} F\right) X^k$$

$$F = G X^T - X G^T$$

$$\left(I + \eta \underbrace{[G \mid X]}_{U} \underbrace{\begin{bmatrix} X^T \\ -G^T \end{bmatrix}}_{V^T}\right) =$$

$$\Rightarrow \quad U = [G \mid X]$$

$$V = [X \mid -G]$$

$= I - y\,V(I + \quad)^{-1}$

$$(G + VV^T)^{-1} = G^{-1} - G^{-1}V(I + V^T G^{-1} U)^{-1} U G^{-1}$$

Woodbury identity

Orthogonal Procrustes problem

Lasso.

$$\min \|x\|_1$$
$$\text{s.t. } Ax = b.$$

$$\min \tfrac{1}{2}\|Ax - b\|_2^2 + \lambda \|x\|_1$$

$$\min \tfrac{1}{2}|x - b|^2 + \lambda |x|$$

$$h(x, z) = \tfrac{1}{2}(x - b)^2 + \lambda\left(z + \frac{x^2}{z}\right)$$

$$\geq \sqrt{z \cdot \frac{|x|^2}{z}}$$

$$z \geq 0$$

$$z = |x|$$

$$\frac{\partial h}{\partial x} = x - b + \lambda\left(\frac{2}{z}x\right) = 0$$

$$x\left(1 + \frac{2\lambda}{z}\right) = b.$$

$x > 0$:

$$x + 2\lambda = b \qquad x^* = b - 2\lambda$$

$$b > 2\lambda$$

$x < 0$

$$x^* = b + 2\lambda \qquad b < -2\lambda$$

$$\Rightarrow \quad x = \max(|b-\lambda|, 0)\cdot \mathrm{sign}(b-\lambda)$$

soft thresholding

$$\min \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1 \quad \text{e47}$$

$$g(v) = \left(\frac{1}{2}(-v)^T(-v) + y^T(-v)\right) - I_\infty(v)$$

$$\max g(v)$$

$$\Downarrow$$

$$\min -g(v) = \frac{1}{2}v^Tv - y^Tv + I_\infty(v)$$

$$\Downarrow$$

$$\min \frac{1}{2}v^Tv - y^Tv$$

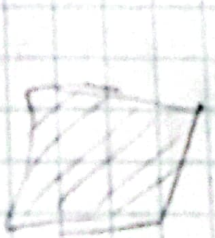
$$\text{s.t.} \quad v \in B_\infty(\gamma)$$

$$-\gamma \le v \le \gamma.$$

$$\min \frac{1}{2}v^Tv - y^Ta$$

$$\text{s.t.} \quad -\gamma \le v \le \gamma.$$

пример. DCT

$$\boxed{\Phi} \; \boxed{Y} \; \boxed{\Phi^T} =$$ 

original

$$\|y - Ax\|^2 + \gamma^T\|x_1\|$$

$\rightarrow$ sparse

$\boxed{Y}$

transform 2D-DCT

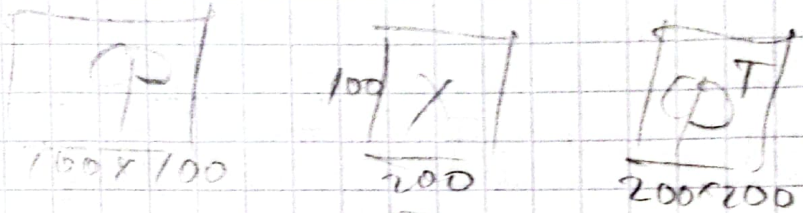
γ - sparsity cond-n

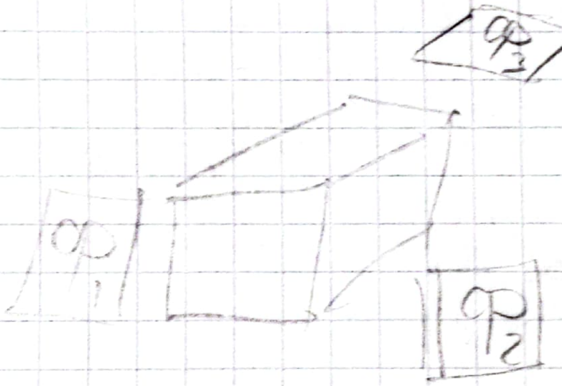
$X = \Phi^T Y \Phi$

A

$y = \mathrm{vec}(Y),$

$x = \mathrm{vec}(X) = \underbrace{(\Phi^T \otimes \Phi^T)}_{\not\equiv} \mathrm{vec}\, Y$





$$\min \ \frac{1}{2} v^T v - y^T v$$
$$\text{s.t.} \ -\gamma \le A^T v \le \gamma.$$

$x = \mathrm{vec}(X) = (\Phi_3 \otimes \Phi_2 \otimes \Phi_1)^T \mathrm{vec}(Y)$

[illegible]

$$\mathcal{L}(v, t_1, t_2) = \frac{1}{2} v^T v - y^T v + t_1^T (A^T v - 1\gamma) + t_2^T(-A^T v - 1\gamma)$$

$$\frac{\partial \mathcal{L}}{\partial v} = v - y + A t_1 - A t_2 = 0$$

$$v^* = A(t_2 - t_1) + y.$$

$$g(t_1, t_2) = (y - A(t_1 - t_2))^T (y - A(t_1 - t_2)) -$$

$$\cancel{A t_1 - A} - y^T (y - A(t_1 - t_2)) +$$

$$+ v^T A(t_1 - t_2) - \gamma (1^T t_1 + 1^T t_2).$$

$\max_{\gamma} g(t_1, t_2)$

s.t. $t_1 \geq 0, \ t_2 \geq 0$

$\Downarrow$

$$\min \frac{1}{2} \| y - A(t_1 - t_2) \|^2 + \gamma (1^T t_1 + 1^T t_2)$$

$\Rightarrow t_1, t_2 \geq 0$

$H = A^T A$

$$\Rightarrow \left\| y - [A \mid -A] \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \right\|^2 + \gamma 1^T \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$$

$$t = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \geq 0$$ потому что DCT

$$H = \begin{bmatrix} A^T A & -A^T A \\ -A^T A & A^T A \end{bmatrix} = \begin{bmatrix} I & -I \\ -I & I \end{bmatrix}$$

$$\min \frac{1}{2} t^T H t - y^T [A \mid -A] t + \gamma 1^T t.$$

$$\frac{1}{2} t^T \begin{bmatrix} I & -I \\ -I & I \end{bmatrix} t - [\, DCT(y) \mid -DCT(y) \,] t$$

$$+ \gamma 1^T t =$$

$$\min \frac{1}{2} t^T \begin{bmatrix} I & -I \\ -I & I \end{bmatrix} t + \left(-\begin{bmatrix} b \\ -b \end{bmatrix} - \gamma\right)^T t$$

$$s.t. \; t \geq 0$$

$z = y - v$

$x^* = A(t_1 - t_2)$

KKT

$$\nabla \mathcal{L} = A^T A x - A^T b + \lambda x = 0$$

$$x = (A^T A + \lambda I)^{-1} A^T b.$$

comp. slac $\longrightarrow$ $\lambda \cdot \left(\frac{1}{2}\|x\|^2 - c\right) = 0$

if x^* inside $B(0, 2c) \Rightarrow \lambda = 0$

$x^* = (A^T A)^{-1} A^T b.$

$\Downarrow$

$\|x^*\|^2 < 2c.$

otherwise: $\|x^*\|^2 = 2c.$

Solman.

$\|x\|_x = \sum \sigma_i$? или $\sum \|\sigma_i\|$ можно ли $\alpha\|\sigma_i\| > 0$? привести что-то.

посчитать Lasso сколько t.

ADMM

пример алгоритма. | доказать двол.

Lecture 9

Solman

$\sum c_i |x_i| \leq \sum c_i t_i$

$\min \sum c_i t_i$

s.t $|x_i| \leq t_i \rightarrow x_i \leq t_i$

$x_i \geq -t_i$

$x_i = y_1^i - y_2^i$

$y_1^i \geq y_2^i \geq 0$

$\min c^T x$ $\Rightarrow$ $\min\max_c c^T x$

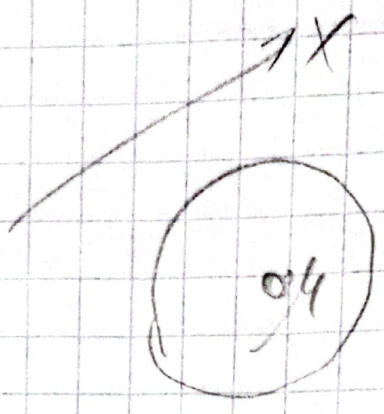
s.t $Ax = b$. $\qquad$ c.

$Ax = b$

$c \in B(\mu, \sigma)$

$\max c^T x$

$c \in B(\mu, \delta)$



$c^* = \mu + \delta \frac{x}{|x|}$

$c = \mu + \delta u$

$\|u\|_2 \le 1$

$\max c^T x = \max \mu^T x + \delta \underset{\Downarrow\, x}{u^T x}$

$\min \left(\mu + \delta \frac{x}{\|x\|}\right)^T x$

s.t. $Ax = b$

$\Downarrow$

$\min \mu^T x + \delta \|x\|_2$

s.t. $Ax = b$

$\langle w, x \rangle + b = \hat{y}_k$

$\min \|W\|_F^2 + \gamma \|W\|$

s.t. $\hat{y}_k y_k \ge 1$

$$\hat{y} = (x_k, w) + b$$

$$w = v$$

$$\min \|W\|_F^2 + \gamma |V|_*$$

s.t. $W = V \quad | : u$

$$W^{k+1} = \operatorname{argmin} \frac{1}{2}\|W\|_F^2 + \|W - V + u\|_F^2 \frac{\rho}{2}$$

$$\text{s.t. } y_k (\operatorname{tr}(W^T x_k) + b) \geq 1$$

$$V^{k+1} = \operatorname{argmin} \gamma \|V\|_* + \frac{\rho}{2} \|W^k - u - u^k$$

$$u^{k+1} = u^k + V^{k+1} - W^{k+1}$$

$$\mathrm{SVD}(W^{k+1} - u^k) = A \Sigma B^T$$

$$V^{k+1} = A \operatorname{diag}(\max(\sigma - \frac{\rho}{2\gamma}, 0)) B^T$$

$$W^{k+1} =$$

closed form?

$$\mathcal{L} = \frac{1}{2} \|y - Ax\| + \lambda \|z\|_1 + \mu^T (x - z)$$

$$g = \inf(\mu^T x + \frac{1}{2}\|y - Ax\|^2) - \inf$$

$$-\inf_z (\mu^T z + \lambda \|z\|_1)$$

делаем как на слайде.

вопросы

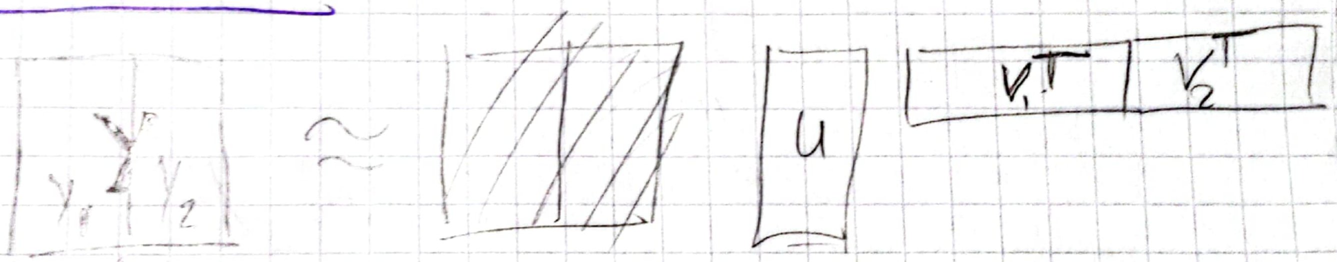
proper func

closed func.

g^* - of dual

matlab and license. exponent.ru.

Lecture 10



realy large.

$$\min \frac{1}{2}\| Y - UV^T \|_F^2$$

$$\min \frac{1}{2}\| Y_1 - U_1V_1^T \|_F^2 + \frac{1}{2}\| Y_2 - U_2V_2^T \|_F^2$$

$$U_1 = U_2.$$

$$L(U_1, U_2, V_1, V_2, T) = \frac{1}{2}\| Y_1 - U_1 V_1^T \|_F^2 +$$
$$+ \frac{1}{2}\| Y_2 - U_2 V_2^T \|_F^2 + \langle T, U_1 - U_2 \rangle +$$
$$+ \frac{\lambda}{2}\| U_1 - U_2 \|_F^2$$

$$\Downarrow$$

$$\frac{\lambda}{2}\| U_1 - U_2 - T \|_F^2 - \ldots ?$$

for $U_1 V_1$ вытащим все и берем
arg min.

$$(V_2 + T)$$

$$\min \frac{1}{2}\| Y - U V^T \|_F^2 + \frac{\lambda}{2}\| V - D \|_F^2 -$$

$$- \cancel{V(Y - UV^T) \ldots}$$

$$\begin{cases} V^* = ((U^T U)^{-1} U^T Y)^T \\ U^* = (Y V + \lambda D)(V^T V + \lambda I)^{-1} \end{cases}$$

$$\frac{1}{2}\| Y - U (U^T U)^{-1} U^T Y \|_F^2$$

$$(I - U (U^T U)^{-1} U^T)$$ — orthogonal projection?
householder projection

$$U = \underset{n \times R}{A} \cancel{\Sigma} \underset{R \times R}{B^T}$$

$$n > R$$

$$Y - A \Sigma B^T (B \Sigma A^T A \Sigma B^T)^{-1} B \Sigma A^T Y =$$

$$= \|Y - A\cancel{\Sigma} A^T \cancel{Y}\| = \|Y\|_F^2 + \|AA^T Y\|_F^2 -$$

$$- 2\,tr(Y^T A A^T Y) =$$

$$\|A^T Y\|_F^2 \qquad \|A^T Y\|_F^2$$

потому что A орт.

$$= \frac{1}{2}\left(\|Y\|_F^2 - \|A^T Y\|_F^2 + \frac{2}{2}\|A\Sigma B^T + D\|_F^2\right.$$

$$\min \|\Sigma - A^T D B\|_F^2$$

$$\Sigma = A^T D B$$

$$A$$

$$\frac{1}{2}\left(\|Y\|^2 - \|A^T Y\|_F^2 + \frac{2}{2}\|D\|_F^2 - 2\|A^T D\|_F^2\right) =$$

$$= \min \frac{2}{2}\|A^T Y\|_F^2 - \frac{2}{2}\|A^T D\|_F^2 =$$

$$= \max\ tr\left(A^T (yy^T + 2DD^T) A\right)$$

A: R - leading eigin vectors of

$yy^T + 2DD^T$

$$\min \frac{1}{2}\|Y_1 - U_1 V_1^T\|_F^2 + \|Y_2 - U_2 V_2^T\|_F^2 -$$

$$U_1 = U_2$$

$$\{U_1, V_1\} = \arg\min \frac{1}{2}\|Y_1 - U_1 V_1^T\|_F^2 + \frac{\rho}{2}\|U_1 - \underbrace{(U_2 + T)}_{D}\|_F^2$$

$$\{U_2, V_2\} = \arg\min \frac{1}{2}\|Y_2 - U_2 V_2^T\|_F^2 + \frac{\rho}{2}\|U_2 - U_1 - T\|_F^2$$

$$T = T + U_2 - U_1$$

---

recomended sys

$$\min \frac{1}{2}\|Y_1 - U_1 V_1^T\|_F^2 + \frac{1}{2}\|Y_2 - U_2 V_2^T\|_F^2 + \dots + \frac{1}{2}\|Y_k - U_k V_k^T\|_F^2 =$$

$$\begin{array}{|c|c|c|c|}\hline Y_1 & Y_2 & \cdots & Y_k \\ \hline\end{array}$$

$$U_k - \underbrace{\frac{\sum U_k}{k}}_{\tilde{U}} = 0$$

$$\min \sum \frac{1}{2}\|Y_k - U_k V_k^T\|_F^2 + \frac{\rho}{2}\sum \|U_k - \tilde{U} - T_k\|_F^2 +$$

$$+\frac{\gamma}{2}\left\|\bar{U}-\frac{\sum U_k}{k}\right\|_F^2$$

$$\{U_k V_k\} = \operatorname{argmin} \frac{1}{2}\|Y_k - U_k V_k^T\|_F^2 +$$
$$+\frac{\alpha}{2}\|U_k-(\bar{U}+T_k)\|_F^2+$$
$$+\frac{\gamma}{2}\left\|\bar{U}-\frac{\sum U_l}{k}-\frac{U_k}{k}\right\|_F^2$$

$$\frac{\alpha}{2}\|U_k-D\|_F^2+\frac{\gamma}{2k^2}\|U_k-G\|_F^2 =$$

$$G = k\bar{U}-\sum_{l\neq k}U_l$$

$=$

для простоты:

$$\alpha\|U_k-D\|_F^2+\gamma\|U_k-G\|_F^2=$$
$$=\left\|U_k-\frac{\alpha D+\gamma G}{\alpha+\gamma}\right\|_F^2$$

выводим формулу

$$\alpha\|U-D\|_F^2+\gamma\|U-G\|_F^2=$$
$$=(\alpha+\gamma)\|U\|_F^2-2\operatorname{tr}(\alpha U^T D+\gamma U^T G)+\ldots$$
$$=(\alpha+\gamma)\left(\|U\|_F^2-2\operatorname{tr}\left(U^T\left(\frac{\alpha D+\gamma G}{\alpha+\gamma}\right)\right)+\ldots\right.$$

$$\Rightarrow \lambda \| U_k - D \|_F^2 + \gamma \| U_k - G \|_F^2 = \| U_k - \frac{\lambda D + \gamma G}{\lambda + \gamma} \|_F^2$$

---

step 1

$\{ U_k, V_k \} =$

$T_k = T_k + \bar{U} - U_k$

step 2 update $\bar{U}$

$$\bar{U} = \operatorname{argmin} \frac{\lambda}{2} \sum \| U_k - T_k - \bar{U} \|_F^2 + \frac{\gamma}{2} \| \bar{U} - \frac{\sum U_k}{k} \|_F^2$$

$$\lambda (U_k - T_k - \bar{U}) + \gamma (\bar{U} - \frac{\sum U_k}{k})$$

$$\bar{U} = \frac{\lambda (\sum U_k - \sum T_k) + \frac{\gamma}{k} (\sum U_k)}{\gamma + \lambda k}$$

$$\bar{U} = \frac{\sum U_k}{k} - \frac{\lambda \sum T_k}{\gamma + \lambda k}$$

$$\min \| y_k - A x_k \|$$

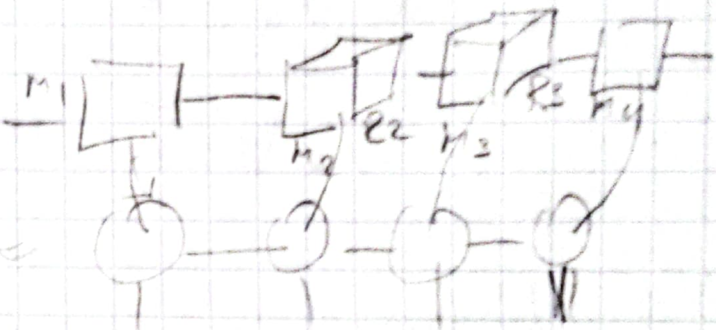
$m \times 1 \quad n \times 1$

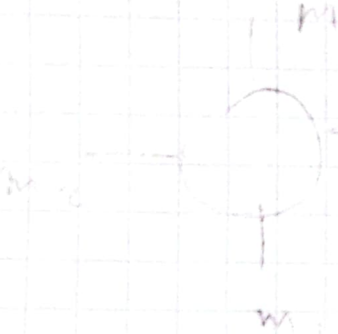
$m \times n$

$A: m_1 \times m_2 \times m_3 \times m_4.$

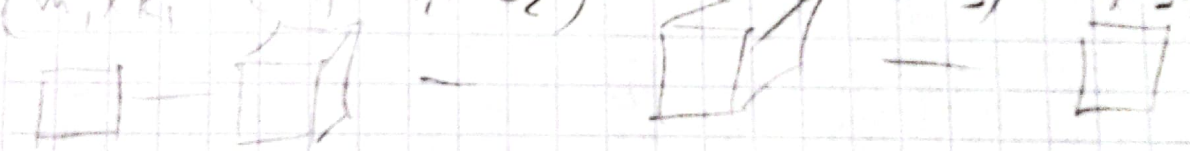
Tensor train.

$A(i_1, i_2, i_3, i_4) =$





$(m_1 \times R_1) \quad (R_1 \times m_2 \times R_2) \quad (R_2 \times m_3 \times R_3) \quad (R_3 \times m_4)$

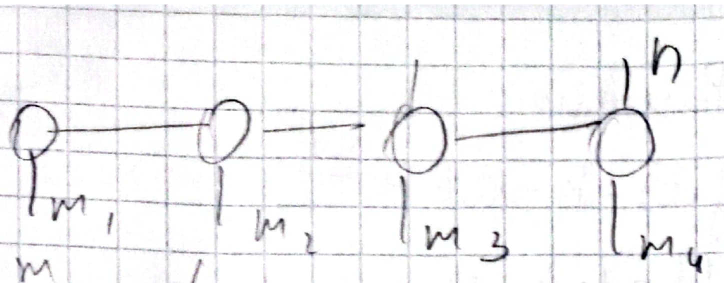


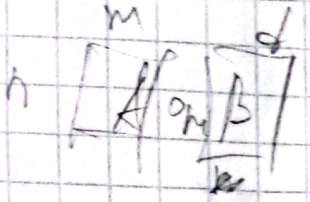
$$= X_1(i_1, :) \cdot X_2(:, i_2, :) \cdot X_3(:, i_3, :) \, X_4(:, i_4)$$

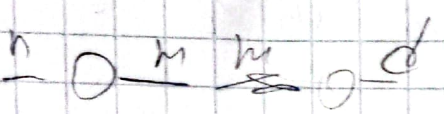
$A: m \times n$

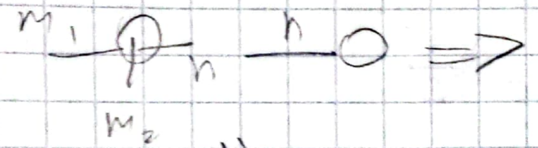
$m = m_1 \, m_2 \, m_3 \, m_4$

$A: m_1 \times m_2 \times m_3 \times m_4 \times n.$

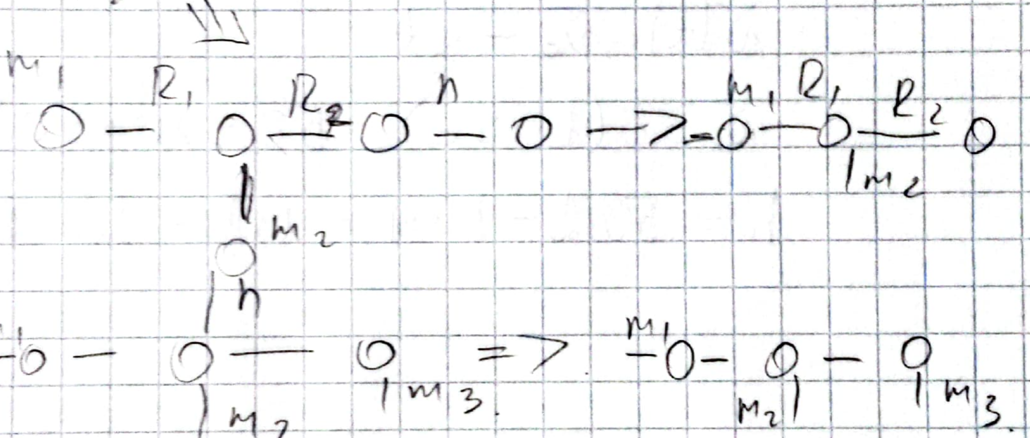








~~p~~result $m_1 \times m_2$



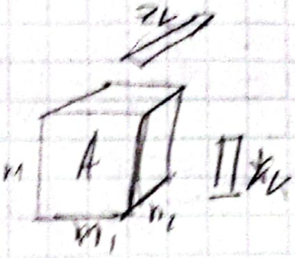
$\min \| y_k - A x_k \|$

kth : $\{x_k, z_k\}, \{y_k\}$

input input output

скорость коэфф-е

$y_k =$ 

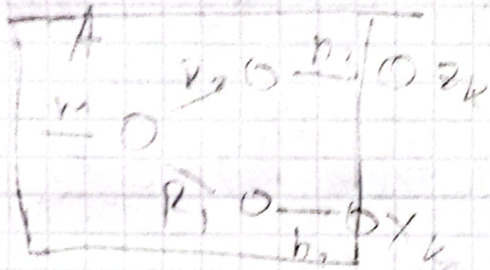


$y_k = A \times_2 x_k \times_3 z_k$

$y_k = A \cdot (z_k \otimes x_k)$

A = reshape

$\cdots \times (n, n_1 n_2)$



$\min \|y_k - A x_k\|$

$A = [U \otimes V]$

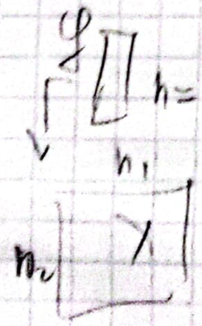
$$A = \begin{bmatrix} u_{11} V & \cdots & u_{1 d_1} V \\ \vdots & & \\ u_{n_1 1} V & & u_{n_1 d_1} V \end{bmatrix}$$

$\min \sum_k \|y_k - A x_k\|^2$

s.t. $A = U \otimes V$

$$(U \otimes V) x_k = \mathrm{vec}\left(V \underbrace{[\,X_k\,]}_{d_2 \times d_1} U^T\right)$$

$n = n_1 \times n_2$



$$\|y_k - A x_k\|_2^2 = \|Y - V X_k U\|_F^2$$

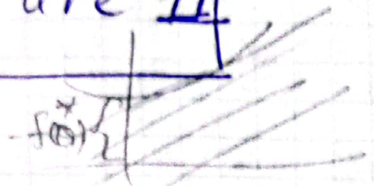
Cayley transform

$B^{k+1} = Q B^k$; orthogonal

A

$\|D\|$

Lecture 11



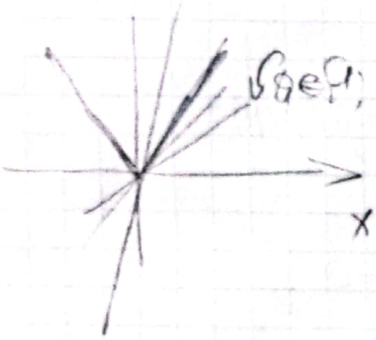
$$\langle a, x \rangle + \lambda \le f(x)$$

$$\lambda \le -[\langle a, x\rangle - f(x)]$$

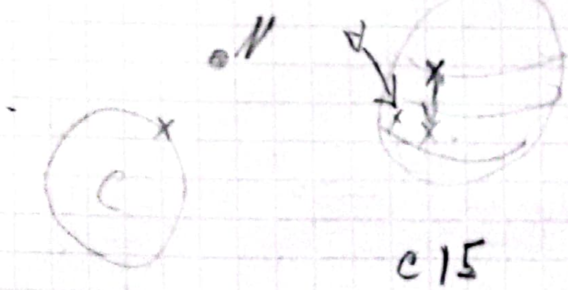
$$\lambda \le \inf(\ldots)$$

$$\lambda = \underbrace{-\sup \langle a,x\rangle - f(x)}_{f^*(a)}.$$

$a \notin [-1,1]$



$$f^*(a) = \begin{cases} 0 & \text{if } a \in [-1;1] \\ \infty & \text{if } a \notin [-1,1] \end{cases}$$



$\operatorname{dom} \Phi(x)$

c 15

$$\min_{x \in X} f(x) + g(x)$$

CP: $x^{k+1} := x^k + \lambda_k \nabla f(x^k)$

$$:= \operatorname*{argmin}_{x \in X} \left\{ f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2\lambda_k} \|x - x_k\|_2^2 \right\}$$

primal dual method

P-D: $\min_{x \in X} f(x) + g(Kx)$

x use the convex conjugate of $g(Kx)$.

$g^*(\mu) = \sup_{x} \{ \langle \mu, Kx \rangle - g(Kx) \}$

$\Rightarrow g^*(\mu) \geq \langle \mu, Kx \rangle - g(Kx)$:

$\Rightarrow g(Kx) \geq \langle \mu, Kx \rangle - g^*(\mu)$

$g(Kx) \geq \sup \langle \mu, Kx \rangle - g^*(\mu)$

$\min f(x) + g(Kx) \geq \min \max f(x) + \langle \mu, Kx \rangle - g^*$

$\geq \max \min f(x) - \langle -x, K^* \mu \rangle - g^*(\mu)$

$\langle \mu, Kx \rangle = \langle K^* \mu ; x \rangle = -\langle K^* \mu, -x \rangle$

$\min_{x \in X} \max_{u \in U} f(x) + \langle u, Kx \rangle - g^*(u)$

$\min f(x) + g(z)$ s.t. $Kx - z = 0$

PD $\to$ SGD - ?

---

input image Y noisy $\to$ $y := \begin{bmatrix} Y_{11} \\ Y_{21} \\ \vdots \end{bmatrix}$

Я уныым надо пересмотреть лекцию

$\frac{\partial}{\partial x} \|Cx - b\|^2 = C^T(Cx - b) = C^T C x - C^T b$

$x = \cancel{b}$ $\quad I$?

$P_{\frac{1}{C}} = (I - CC^T)$

$$\Sigma (x_i - d)^T (I - CC^T)(x_i - d)^T =$$

$$= \Sigma (x_i - d)(x_i - d) - \Sigma (x - d) CC^T (x, d)$$

$$\Downarrow$$

$$\max \operatorname{Trace}(U^T R_c (U))$$

def of Volume $\dfrac{|\det(\Delta(\beta_1 \dots \beta_N))|}{(N-1)!}$

$$\Delta(\beta_1 \dots \beta_N) = \begin{bmatrix} \beta_1 & \dots & \beta_N \\ 1 & & 1 \end{bmatrix}$$

HW discusions

y_k - labels

x_k - data

$\neq 6\%$ → 60% для 1й проблемы!

Lecture 12

describe HW 2

$$\max_{e_k} (y_k - w^T(x_k + e_k))^2$$
$$\text{s.t. } \|e_k\| \le \delta.$$

$$\mathcal{L}(e_k, \lambda) = -\frac{1}{2}(y - w^T(x + e))^2 + \frac{\lambda}{2}(\|e\|^2 - \delta^2)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = -\frac{1}{2}(\cancel{y}\ b - w^T e)^2 + \frac{\lambda}{2}(\|e\|^2 - \delta^2)$$

→ число ⇒ $e \parallel w$

$$\frac{\partial \mathcal{L}}{\partial e} = w(b - w^T e) + \lambda e = 0.$$

$$wb - ww^T e + \lambda e = 0.$$

$$(\lambda I - ww^T) e = -wb$$

$$e = \alpha w.$$

~~e~~ max of convex (func? line) over a conv set on the boundary.

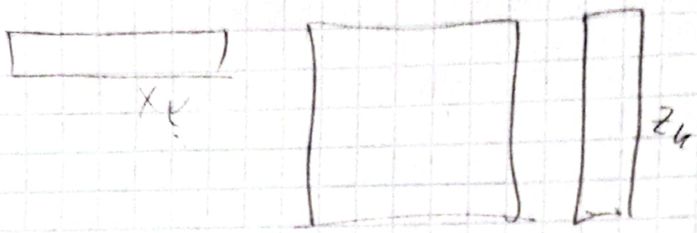
$e^* = \pm \frac{w}{\|w\|} \delta^2$ $\qquad t = 2^{-1}, \dots$

$= \frac{1}{2} (y - w^T (x + e^*))^2 + \frac{\lambda}{2} (\|e\|^2 - \delta^2)$

$\min \mathcal{L}(e^*, \lambda) = \frac{1}{2}$

$(|b| + \delta \|w\|_2)^2$

$\sum \|b\|^{\|} + \delta \|w\|_2 = \|b\|_1 + \delta K \|w\|_2$



~~vect~~

$\mathrm{trac}(W(z_k x_k^T)$

$\mathrm{tr}(W(z + v_k)(x_k + u_k)^T)$

$\mathrm{tr}(W(z_k x_k^T + E_k)$

$\|E_k\| \leq \delta^2$

$u_1 V_1^T = u_1 Q Q^T V_1^T$

$u_1 = [\Sigma_1 b \ A_1]$

$u_2 = [\Sigma_2 b \ A_2]$

$\varphi_1 = u_1 V_1^T = u_1 \operatorname{diag}(\beta_1) \operatorname{diag}(\frac{1}{\beta}) V_1^T$

$\varphi_2 = u_2 V_2^T = u_2 \operatorname{diag}$

pok.

$u_1 : R \to 1$

$u_2 : R \to 1$

control the rank u_1, u_2

$Y := I \times J$

Y_1 1000 × 500. matrix

$U_1 = 1000 \times R$

$\searrow$ 10 − 20.

$V_1 = 1000 \times R$

Rank should be R.

$\|\alpha\|_* \geq \delta$?

$U = QR$. $\longrightarrow QR(V_1) = U_1 F_1$

$Y - U_1 V_1^T$

$V_1^* = Y_1^T U_1$

$\min \; \|Y_1 - U_1 U_1^T Y_1\|_2^2 + \|$ makoo ma $\|$

s.t. $U_1 = U_2$ orthogonal

$U_1^T U_1 = I$

$U_2^T U_2 = I$

$d(U_1, U_2) =$

$\searrow$ sum eigenvals

$U_1^T U_2 = I$

$\min \; \|Y_1 - U_1 U_1^T Y_1\|_F^2 + \|Y_2 - U_2 U_2^T Y_2\|^2$

$U_1^T U_1 = I_R$

$U_2^T U_2 = I_R$

$U_1^T U_2 = I$

$\max \; \|U_1^T Y_1\|_F^2 + \|U_2^T Y_2\|_F^2$

$$K = \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} = \begin{bmatrix} \phi & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \leftarrow \text{nonsingular},$$

$$x = \lambda u + \beta v = \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} \lambda \\ \beta \end{bmatrix}$$

$$x^T Q x > 0$$

$$\| $$

$$U^T Q U = \begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}^T \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix} =$$

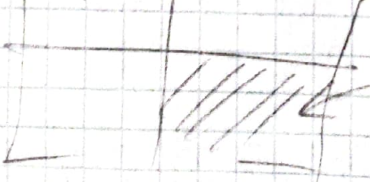
$$= \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & \phi \end{bmatrix}$$

$$\frac{1}{2} (Fz + \hat{x})^T Q (Fz + \hat{x}) =$$

$$= \frac{1}{2} z^T F^T Q F z + (FQ\hat{x} + Fp)^T z$$

$$x^* = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ + \\ + \end{bmatrix} \quad p = \begin{bmatrix} + \\ + \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{matrix} A = \{x_i = 0\} \\ I = \{x_i > 0\} \end{matrix}$$

$$x_I = Q_{II}^{-1} p_I$$

$Q =$  $\leftarrow Q_{II}$

x - feasible

$$x = \begin{bmatrix} x_I \\ x_A = 0 \end{bmatrix} + d \begin{bmatrix} d_I \\ 0 \end{bmatrix}$$

$$\min \tfrac{1}{2}(x_I - d_I)^T Q_{II} (x_I + d_I) + p_I^T d_I$$

$$d_I = -Q_{II}^{-1}(Q_{II} x_I + p_I)$$

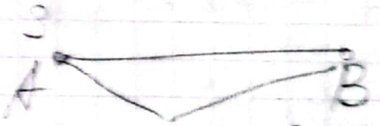
$$d_I = -x_I - Q_{II}^{-1} p_I$$

$$x^* = \begin{bmatrix} -Q_{II}^{-1} p_I \\ 0 \end{bmatrix}$$

c 28

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} < \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Lecture 15



$y = \alpha_1 s + \alpha_2 s(t - \delta_1) + \alpha_3 s + s(t - \delta_2)$

$y = s \circledast h$ convolution

$\to$ FFT $\quad y_f = S * H$ hadamard prod.

y - given

Toeplitz

$y = x \circledast h = \mathrm{Toepliz}(h) \begin{bmatrix} x \end{bmatrix} =$

$= \boxed{A} \, (h \otimes x)$ Kronecker

Z - rank $\quad x \cdot h^T$

$y = A \cdot \mathrm{vec}(Z)$

$$\min \|Z\|_*$$
$$\text{s.t } y = A\,\mathrm{vec}(Z)$$

$$\mathcal{L} = \|Z\|_* + \gamma\|X\|_{TV} + i(F) +$$
$$+ \frac{\rho}{2}\left(\|X - F - T_X\|_F^2 + \|Z - F - T_Z\|_F^2\right)$$

$$Z^{(k+1)} = \underset{Z}{\mathrm{argmin}}\ \|Z\|_* + \frac{\rho}{2}\|Z - F^k - T_Z^k\|_F^2$$
$$= \mathrm{prox}_{\|\,\|_*, \frac{1}{\rho}}(F^k + T^k) = \longrightarrow U \Sigma V^T$$

singular value tresholding.

$$= U\,\mathrm{diag}\left(\max\left(0, \sigma - \frac{1}{\rho}\right)\right) V^T$$

$$X^{k+1} = \underset{X}{\mathrm{argmin}}\ \gamma\|X\|_{TV} + \frac{\rho}{2}\|X - F^k - T_X\|_F^2$$
$$= \mathrm{prox}_{\|\,\|_{TV}, \frac{\gamma}{\rho}}(F^k + T_X^k)$$

$$F^{k+1} = \underset{F}{\mathrm{argmin}}\ i(F) + \frac{\rho}{2}\left(\|F - (X^{k+1} - T_X)\| +\right.$$
$$\left. + \|F - (Z^{k+1} - T_Z^k)\|\right) =$$
$$\text{s.t. } y = A\,\mathrm{vec}(F) \Leftarrow$$

$$= \underset{F}{\mathrm{argmin}}\ \|F - \underbrace{\tfrac{1}{2}(X^{k+1} - T_X + Z^{k+1} - T_Z)}_{\mathrm{vec}(\ \ )}\|$$
$$\text{s.t } y = A\,\mathrm{vec}(F)$$

$$F^{k+1} = \underset{F}{\mathrm{argmin}}\ \|f - D\|_F^2$$
$$\text{s.t } y = Af.$$

$$\min \|f - D\|_F^2$$
$$\text{s.t. } y = Af.$$

$$\mathcal{L} = \frac{1}{2}\|f - D\| + \mu^T(y - Af)$$

$$\frac{\partial f}{\partial f} = f - d - A^T\mu \Rightarrow f^* = A^T\mu + d.$$

$$g(\mu) = \frac{1}{2}\frac{1}{2}\|A^T\mu\|^2 + \mu^T(y - AA^T\mu - A^T d)$$

$$= -\frac{1}{2}\mu^T A A^T \mu + \mu^T(y - Ad)$$

$$\mu^* = (AA^T)^{-1}(y - Ad)$$

$$f^* = A^T(AA^T)^{-1}(y - Ad) + d$$

скажем что AA^T имеет простую структуру и легко посчитать inv. самому.