

On the feasibility for the system of quadratic equations.

Minimizing  $z(c)$  when  $|c_{\text{bad}}| = \infty$

12 ноября 2016 г.

## The framework

Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a quadratic map:  $f_i(x) = x^T A_i x + 2b_i^T x$ ,  $A_i^T = A_i$ . We want to explore the image  $F = f(\mathbb{R}^n)$  and find convex parts of  $F$ .

1. Consider a problem

$$g(c) = \min_{x \in \mathbb{R}^n} (c, f(x)) \quad (1)$$

It has a finite solution iff

$$\begin{cases} c \cdot A \geq 0 \\ (c \cdot b)^T \text{Ker}(c \cdot A) = 0 \end{cases}$$

The  $x$ , minimizing 1 is ( $Q^g$  is a pseudoinverse of  $Q$ )

$$x = \text{Ker}(c \cdot A) - (c \cdot A)^g (c \cdot b)$$

Define the boundary points of  $F$  on the supporting hyperplane with a normal vector  $c$ :

$$X(c) = \arg \min (c, f(x)) = \text{Ker}(c \cdot A) - (c \cdot A)^g (c \cdot b)$$

2. The boundary points  $f(X(c))$  can be nonconvex.
3. Let  $c_+$  be a vector from  $\mathbb{R}^m$  s.t.  $c_+ A > 0$ . Then the supporting hyperplane with the normal vector  $c_+$  is touching  $F$  at a single point  $y_0 = f(x_0)$ , where

$$x_0 = -(c_+ \cdot A)^{-1} (c_+ \cdot b)$$

We want to find the maximum value  $z_{\max}$  such that a convex cut  $Z \subseteq F$  is convex:

$$Z = \{y \mid (c_+, y - y_0) \in [0, z_{\max}]\}$$

To do that, we discover all of the nonvexities  $f(X(c))$  and find

$$z_{\max} = \inf_c \tilde{z}(c), \quad \tilde{z}(c) = \inf (c_+, f(X(c)) - f(x_0))$$

4. To find  $z(c) = \inf_{x \in \text{Ker}(c \cdot A) - (c \cdot A)^g (c \cdot b)} (c_+, f(x))$ , we introduce the plus norm

$$\|q\|_+^2 = q^T (c_+ \cdot A) q$$

Note that  $(c_+, f(x) - f(x_0(c))) = \|x - x_0(c)\|^2$ . Therefore,  $\tilde{z}(c) = \rho_+^2(\text{Ker}(c \cdot A) - (c \cdot A)^g (c \cdot b), x_0(c))$

5. Next, we change variables to find  $\tilde{z}(c)$  explicitly. We use the following change  $\begin{cases} x = S\hat{x} + x^0 \\ y = \hat{y} + y^0 \end{cases}$

$$y = f(x), \hat{y} = \hat{f}(\hat{x}).$$

and demand  $\begin{cases} c_+ \cdot A = I \\ c_+ \cdot b = 0 \end{cases}$ . It leads to

$$x^0 = x_0, y^0 = y_0, S = S_1 S_2, S_1^T (c_+ \cdot A) S_1 = \Lambda, S_1^T S_1 = I, S_2 = \Lambda^{-1/2}$$

,

$$\begin{cases} \hat{A}_i = S^T A_i S \\ \hat{b}_i = S^T (b_i + A_i x^0) \end{cases}$$

In new variables  $\|\hat{q}\|_+^2 = \hat{q}^T \hat{q}$ , i.e. the new basis is orthonormal w.r.t.  $\|\cdot\|_+$ , because  $c_+ \cdot A = I$ .

Notice that  $\tilde{z}(c) = \rho_+^2(x_0, \text{Ker}(c \cdot A) - (c \cdot A)^g(c \cdot b)) = \text{diam}(\text{Ker}(c \cdot \hat{A}) - (c \cdot \hat{A})^g(c \cdot \hat{b})) = \|(c \cdot A)^g(c \cdot b)\|^2$  since  $\text{Ker}(c \cdot \hat{A})^T((c \cdot A)^g(c \cdot b)) = 0$

6. We associate the conditions  $\text{Rg}(c \cdot A) < n$ ,  $c \cdot A \geq 0$  with nonconvexity of  $f(X(c))$ . The real sufficient condition for nonconvexity is given by the Theorem 3.4, which requires some additional equalities:

$$\begin{cases} \text{Rg}(c \cdot A) = n - 1 \\ f^1 \nparallel f^2 \end{cases}$$

But we assume that other cases, in particular,  $\text{Rg}(c \cdot A) < n - 1$  or  $f^1 \parallel f^2$  are rare.

7. Let us note that if  $f(X(c))$  is nonconvex, then  $\lambda_{\min}(c \cdot A) = 0$ . Consider  $\tilde{z}(c + \alpha c_+) = \|(c \cdot A + \alpha I)^g(c \cdot b)\|^2$ , since  $c_+ \cdot \hat{A} = I$ ,  $c_+ \cdot \hat{b} = 0$ . Therefore, adding  $\alpha c_+$  to  $c$  adjusts the spectrum of  $c \cdot A$ . If and only if we choose  $\alpha = -\lambda_{\min}(c \cdot A)$ ,  $\tilde{z}$  will be associated with a  $\tilde{c} = c + \alpha c_+$  for which  $\lambda_{\min}(\tilde{c} \cdot A) = 0$ . This way, for every given  $c \in \mathbb{R}^m$  we can find one and only one  $\tilde{c} = c + \alpha c_+$  such that  $\lambda_{\min}(\tilde{c} \cdot A) = 0$

We define  $z(c) = \|((c \cdot \hat{A}) - \lambda_{\min}(c \cdot \hat{A}))(c \cdot \hat{b})\|^2$ , which is the same as  $\tilde{z}(c - \lambda_{\min} c_+)$ .

Notice that  $z(c + \mu c_+) = z(c)$ .

8. Next, we define  $Q(c) = c \cdot \hat{A} - \lambda_{\min}(c \cdot \hat{A})$ ,  $\hat{b}_c = c \cdot \hat{b}$ ,  $c_{\text{bad}} = \{c \mid \text{Ker } Q(c) \perp b_c\}$ ,  $v(c) = Q(c)^{-1} b_c$ ,  $z(c) = \|v(c)\|^2$ .

## Minimization

We want to solve a problem

$$z_{\max} = \inf_{c \in c_{\text{bad}}} z(c) \quad (2)$$

We use gradient projection method.

Nonconvexity certificate gives a single point  $c \in c_{\text{bad}} = \{c \mid \text{Ker } Q \perp b_c\}$ , which is used as a start point. We assume  $\text{Rg } Q(c) = n - 1$ , define  $k(c) \in \mathbb{R}^n$  s.t.  $k(c) \in \text{Ker } Q(c)$ ,  $\|k(c)\| = 1$ . We assume that for any  $c'$  the dot product  $k(c')^T k(c) > 0$ . Each iteration has two key elements: gradient step and projection.

### Gradient step

Define  $R_i = \hat{A}_i - k^T \hat{A}_i k$ . Then

$$\nabla_i z \equiv \frac{\partial z}{\partial c_i} = 2v^T Q^{-1}(\hat{b}_i - R_i v)$$

Define  $n_i = (\hat{b}_i^T - v^T R_i) k$ . Then  $n$  is a normal vector to  $c_{\text{bad}}$  at point  $c$ . We eliminate normal part of the gradient to decrease distance to  $c_{\text{bad}}$  (see picture):

$$c' = c - \beta \nabla z \left( \hat{1} - n \frac{(\cdot, n)}{|n|^2} \right)$$

## Projection



We next assume that for some  $\lambda$  vector  $\tilde{c}(\lambda) = c' + \lambda n \in c_{\text{bad}}$ . To find  $\lambda$ , we write the condition  $\tilde{c} \in c_{\text{bad}} \Leftrightarrow (\tilde{c} \cdot \tilde{b})^T k(\tilde{c}(\lambda)) = 0$ . We define a function

$$m(\lambda) = (\tilde{c}(\lambda) \cdot \tilde{b})^T k(\tilde{c}(\lambda))$$

This function is continuous if  $\text{Rg } Q(\lambda) = n - 1$  in the neighbourhood of  $\lambda$ . Next, we find its root using bisection method on  $[-\lambda_0, \lambda_0]$ ,  $\lambda_0 = \|c - c'\|$ .

If it does not converge, or other problems occur, we shrink the gradient step  $\beta$  and try the projection again.

## Explanations

### 1. $F$ and $\text{conv } F$

Consider  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , s.t.  $f_i(x) = x^T A_i x + 2b_i^T x$ ,  $A_i = A_i^T$ . Define  $F = f(\mathbb{R}^n)$ .

Then why  $A = \inf_{y \in F} (c, y) = \inf_{y \in \text{conv } F} (c, y) = B$ ?

1. First,  $F \subseteq \text{conv } F$ , therefore,  $B \leq A$ .

2. Secondly, let  $y_k \in \text{conv } F$  be a sequence s.t.  $g_k = (c, y_k) \xrightarrow{k \rightarrow \infty} B$ .  $y_k = \sum_{i=1}^{n_k} \alpha_i^k y_i^k$ .

$$g_k(c, y_k) = \sum_{i=1}^{n_k} \alpha_i^k (c, y_i^k) = \sum_{i=1}^{n_k} \alpha_i^k g_i^k. \text{ Define } g_0^k = \min_{i \in \{1, \dots, n_k\}} g_i^k. \text{ Then } B \leq g_0^k \leq g^k. \text{ Therefore, } g_0^k \rightarrow B$$

also. This way, we have constructed a sequence  $y_0^k \in F$  s.t.  $(c, y_0^k) \rightarrow B$ , therefore,  $A \leq B$ .

### 2. Minimum of $f(x)$

Consider  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ .  $f_i(x) = x^T A_i x + 2b_i^T x$ .  $A_i^T = A_i$ . Let  $c \in \mathbb{R}^m$

We want to find  $g(c) = \inf_{x \in \mathbb{R}^n} (c, f(x))$ .

$$\text{Define } A_c \equiv c \cdot A = \sum_{i=1}^m c_i A_i, b_c = c \cdot b = \sum_{i=1}^m c_i b_i.$$

$$(c, f(x)) = \sum_{i=1}^m c_i f_i(x) = \sum_{i=1}^m c_i (x^T A_i x + 2b_i^T x) = x^T A_c x + 2b_c^T x.$$

If  $\exists v: -\alpha = v^T A_c v < 0$  then  $g(c) = -\infty$ :  $g(\beta v) = -\beta^2 \alpha + 2\beta b_c^T v \rightarrow -\infty, \beta \rightarrow +\infty$ .

From this point on, we assume  $A_c \geq 0$ . Let  $R_0$  be a zero eigenspace (=kernel) of  $A_c$ :  $R_0 = \{v: A_c v = 0\} = \ker A_c$

If  $\exists v \in R_0: v^T b_c \neq 0$  then  $g(c) = -\infty$ : Consider  $f(\beta v) = \beta^2 v^T \overset{0}{A_c v} + 2\beta \underbrace{b_c^T v}_{\neq 0} \rightarrow -\infty, \beta \rightarrow \infty$

Then  $R_0 \subseteq \{b_c\}^\perp$

Consider  $A = \sum_{i=1}^n \lambda_i s_i s_i^T = S \Lambda S^T$ ,  $S = \|s_1 \dots s_n\|$ ,  $S^T S = E$ ,  $s_i^T s_j = \delta_{ij}$ .

$f$  is differentiable, then for finding  $g(c)$  the gradiend  $\nabla(c, f(x)) = 2A_c x + 2b_c = 0$ .

$$S\Lambda S^T x = -b_c \Leftrightarrow \Lambda S^T x = -S^T b_c (*)$$

Let  $x$  be  $x = x^\parallel + x^\perp$ ,  $x^\parallel \in R_0$ ,  $x^\perp \perp R_0$ .

Then neither  $f(x)$  nor  $\Lambda S^T x$  depend on  $x^\parallel$ . This means that the  $x$  minimizing  $g(c)$  is defined in terms of  $x^\perp$  and  $x^\parallel$  is arbitrary.

Define  $\lambda_i^g = \begin{cases} 0, & \lambda_i = 0 \\ 1/\lambda_i, & \lambda_i \neq 0 \end{cases}$ . Define  $\Lambda^g = \text{diag}(\lambda_1^g, \dots, \lambda_n^g)$ . Then  $\Lambda \Lambda^g = \delta_{ij} [\lambda_i \neq 0]$ . Then  $S \Lambda^g \Lambda S^T$  is a projector on  $R_0^\perp$ .

Consider  $(*) \Leftrightarrow \Lambda^g \Lambda S^T (x^\parallel + x^\perp) = -\Lambda^g S^T b_c$ . But  $\Lambda S^T x^\parallel = 0$ , therefore,  $\underbrace{S \Lambda^g \Lambda S^T}_{\text{projector}} x^\perp = -S \Lambda^g S^T b_c$ .

But  $x^\perp$  is already in  $R_0^\perp$ , therefore,  $x^\perp = -\underbrace{S \Lambda^g S^T}_{A_c^g} b_c$ . Here  $A_c^g$  is a pseudoinverse of  $A_c$ .

Therefore,  $\boxed{x = -A_c^g b_c + x^\parallel}$ , where  $x^\parallel \in R_0$ .

Consider  $(c, f(x)) = (c, f(x^\perp)) = b_c^T A_c^g A_c A_c^g b_c - 2b_c^T A_c^g b_c$ . Consider  $A_c A_c^g b_c = S \Lambda^g S^T S^T \Lambda^g S^T b_c$ . Because  $R_0 \subseteq \{b_c\}^\perp$ ,  $\Lambda \Lambda^g S^T b_c = S^T b_c$ . Therefore,  $A_c A_c^g b_c = b_c$ . Then  $(c, f(x)) = b_c^T A_c^g b_c - 2b_c^T A_c^g b_c = \boxed{-b_c^T A_c^g b_c}$

### 3. Finding $c$ provided $d$

Let  $H: \mathbb{R}^{n+1, n+1} \rightarrow \mathbb{R}^n$  be a map s.t.  $H_i(X) = \text{Tr}(H_i X)$ ,

$$H_i = \left\| \begin{array}{cc} A_i & b_i \\ b_i^T & 0 \end{array} \right\|^\square$$

Consider a boundary point  $X$ , which is a solution of (main article, (4)):

$$\boxed{\begin{array}{c} \sup \quad t \\ \left\{ \begin{array}{l} H(X) = y^0 + td \\ X \geq 0 \\ X_{n+1, n+1} = 1 \end{array} \right. \end{array}}$$

Define  $f(t, X) = t$ ,  $D_0 = \{(t, X) | X \geq 0, X_{n+1, n+1} = 1\}$ ,  $D_1 = \{(t, X) | H(X) = y^0 + td\}$ .

Then supremum is equivalent to

$$\sup_{(t, X) \in D_0 \cap D_1} f(t, X)$$

Define a Lagrange function  $L(c, t, X) = \underbrace{t}_{f(t, X)} + \sum_{i=1}^m c_i (y_i^0 + td_i - H_i(X))$ .

Here we divided the constraints into two parts:  $D_1$  goes to the Lagrange function,  $D_0$  goes to the inner supremum. *Stephen Boyd, Lieven Vandenberghe. Convex Optimization. Page ????. Cambridge University Press*

Then the dual function is  $g(c) = \sup_{(t, X) \in D_0} L(c, t, X)$ .

Because  $L = t(1 + \sum_{i=1}^m c_i d_i) + \sum_{i=1}^n c_i (y_i^0 - H_i(X))$ ,  $g = +\infty$  when  $(c, d) \neq -1$ . From this point we assume that  $\boxed{(c, d) = -1}$ .

Now,  $g(c) = \sup_{X_{n+1, n+1}=1, X \geq 0} (c, y^0 - H(X)) = (c, y^0) + \sup_{y \in \text{conv } F} -(c, y) = (c, y^0) - \inf_{y \in \text{conv } F} (c, y)$ .

Then the dual problem is

$$g(c) \rightarrow \inf_{(c, d) = -1}$$

Let us prove that  $\inf_{y \in \text{conv } F} (c, y) = \inf_{\substack{H = \left\| \begin{array}{cc} A_c & b_c \\ b_c^T & \gamma \end{array} \right\| \geq 0}} (-\gamma)$

$$\text{Via Schur complement } H \geq 0 \Leftrightarrow \begin{cases} A_c \geq 0 \\ \gamma - b_c^T A_c^g b_c \geq 0 \\ (E - A_c A_c^g) b_c = 0 \end{cases}.$$

$A_c \geq 0$  is a necessary condition for  $\exists g(c) \in \mathbb{R}$  (see part 2).

$(E - A_c A_c^g) b_c = 0$  is another necessary condition for  $\exists g(c) \in \mathbb{R}$ .

Statement  $\gamma \geq b_c^T A_c b_c$  means  $-\gamma \leq -b_c^T A_c b_c = \inf_{y \in \text{conv } F} (c, y)$ , which means that  $-\gamma$  is a lower bound

for  $\inf_{y \in \text{conv } F} (c, y)$ .

Then  $H \geq 0 \Leftrightarrow -\gamma \leq \inf_{y \in \text{conv } F} (c, y)$ .

Then  $g(c) = (c, y^0) - \inf_{H \geq 0} -\gamma$ .

Then the dual problem is:

$$\inf_{(c,d)=-1} g(c) \Leftrightarrow \inf_{(c,d)=-1} \left[ (c, y^0) - \inf_{H \geq 0} (-\gamma) \right] \Leftrightarrow \inf_{(c,d)=-1} \inf_{H \geq 0} (c, y^0) + \gamma = \boxed{\begin{cases} \inf & \gamma + (c, y^0) \\ H \geq 0 \\ (c, d) = -1 \end{cases}}.$$

This problem is exactly (5) from main article ■.

#### 4. What is $z_{\max}$ ?

Consider  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $f_i(x) = x^T A_i x + 2b_i^T x$ ,  $A_i = A_i^T$ .

Let  $c_+ \in \mathbb{R}^m$  s.t.  $A_+ = \sum c_i A_i > 0$ . Then the minimum  $\inf_x (c, f(x))$  is obtained at a single point

$x_0 = -A_+^{-1} b_+$ ,  $b_+ = \sum c_i b_i$ .

Consider  $S_\varepsilon^+ = \{x \in \mathbb{R}^n \mid (x - x_0)^T A_+ (x - x_0) = \varepsilon^2\}$ . Then  $f(S_\varepsilon^+) = \{y \in \mathbb{R}^m \mid (c_+, y) = (c_+, f(x_0)) + \varepsilon^2\}$ .

Indeed, if  $\begin{cases} P = (c_+, f(x) - f(x_0)) \\ Q = (x - x_0)^T A_+ (x - x_0) \end{cases}$  then  $P - Q = \underline{x^T A_+ x} - x_0^T A_+ x_0 + 2b_+^T x - 2b_+^T x_0 - \underline{x^T A_+ x} - x_0^T A_+ x_0 + 2x^T A_+ x_0 = /x_0 = -A_+^{-1} b_+ / = 2x_0^T b_+ + 2b_+^T x - 2b_+^T x_0 - 2x^T b_+ = 0$ .

**Therefore, the image of  $B_\varepsilon^+$  is a convex cut  $\{y \mid (c_+, y) \in (c_+, f(x_0)) + [0, z_{\max}]\}$**

#### 5. Variables s.t. $c \cdot A = I$ , $c \cdot b = 0$

Given: the map  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $f_i(x) = x^T A_i x + 2b_i^T x$ ,  $A_i = A_i^T$ , vector  $c \in \mathbb{R}^m$ :  $c \cdot A > 0$

We need to find a new pair of bases s.t.  $\begin{cases} \tilde{c} \cdot \hat{A} = I \\ \tilde{c} \cdot \hat{b} = 0 \end{cases}$  (1) (2)

We choose  $x = S\hat{x} + x^0$ ,  $y = \hat{y} + y^0$ ,  $f(x) = y$ ,  $\hat{f}(\hat{x}) = \hat{y}$ . Consider  $f_i(x) = x^{0T} A_i x^0 + \hat{x}^T S^T A_i S \hat{x} + 2x^{0T} A_i S \hat{x} + 2b_i^T S \hat{x} + 2b_i^T x^0 = \hat{x}^T \underbrace{S^T A_i S}_{\hat{A}_i} \hat{x} + 2\hat{x}^T \underbrace{S^T (b_i + A_i x^0)}_{\hat{b}_i} + \underbrace{x^{0T} A_i x^0 + 2b_i^T x^0}_{y_i^0}$

$c \cdot \hat{A} = S^T c \cdot A S = I$ , then  $S = S_1 S_2$ ,  $S_1^T A_i S_1 = \Lambda = \text{diag}$ ,  $S_1^T S_1 = I$ ,  $S_2 = \Lambda^{-1/2}$ .

$c \cdot \hat{b} = S^T (c \cdot b + (c \cdot A) x^0)$ . Then  $x^0 = -(c \cdot A)^{-1} (c \cdot b)$ .

The algorithm:

1. Compute  $S_1$  via the eigenbasis of  $c \cdot A$ ,  $S_1^T (c \cdot A) S_1 = \Lambda$
2. Compute  $S_2 = \Lambda^{-1/2}$ ,  $S = S_1 S_2$ .
3. Compute  $x^0 = -(c \cdot A)^{-1} (c \cdot b)$
4. Compute  $\tilde{A}_i = S^T A_i S$ ,  $\tilde{b}_i = S^T (b_i + A_i x^0)$

#### 6. $z(c) = ?$

Given: the map  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $f_i(x) = x^T A_i x + 2b_i^T x$ ,  $A_i = A_i^T$ , two vectors  $c, c_+ \in \mathbb{R}^m$ .  $c_+ A = I$ ,  $c_+ b = 0$ . Find:  $z(c) = \inf_{y \in Y} (c, y)$  where  $Y$  is an intersection of  $f(\mathbb{R}^n)$  with a tangent hyperplane defined by its normal vector  $c$ .

Define  $\sigma(Q) = \{\lambda \mid \dim \text{Ker}(Q - \lambda E) > 0\}$ . Define  $\lambda_{\min}(Q) = \min \sigma(Q)$  — minimal eigenvalue of  $Q$ .

1. If  $\lambda_{\min}(c \cdot A) < 0$  then the tangent hyperplane does not exist, and  $z(c) = +\inf$
2. Then  $\lambda_{\min}(c \cdot A) > 0$ , then there is no nonconvexity, and  $z(c) = \inf$
3. For  $\lambda_{\min} = 0$  in part 2  $Y$  was found explicitly:  $Y = \{f(x) | x = x^{\parallel} - A_c^g b_c, x^{\parallel} \in \text{Ker } A_c\}$ . Then for  $y \in Y$   $(c_+, y) = x^T (c_+ \xrightarrow{A} I) x + 2(c_+ \xrightarrow{b} 0)^T x = \boxed{x^T x}$
4.  $x^T x = \|x\|^2 = \|x^{\parallel}\|^2 + \|A_c^g b_c\|^2$ . We want to minimize  $(c_+, y)$ , therefore, we choose  $x^{\parallel} = 0$ . Then  $z(c) = \|A_c^g b_c\|^2 = \boxed{|(c \cdot A)^g (c \cdot b)|^2}$
5. Consider  $\inf_{x \in \mathbb{R}^n} (c_+, f(x)) = -(c_+ \xrightarrow{b} 0)(A_{c_+} \xrightarrow{I})^{-1} (c_+ \cdot b) = 0$ . Therefore,  $z(c) = \inf_{y \in Y} (c_+, y) - \inf_{y \in F} (c_+, y)$

$$\text{Now consider } z(c) = \begin{cases} |(c \cdot A)^g (c \cdot b)|^2, & \lambda_{\min}(c \cdot A) = 0 \\ +\infty, & \text{otherwise} \end{cases}$$

Consider  $z(c + \alpha c_+) = |(c \cdot A + \alpha c_+ \xrightarrow{A} I)^g b_c|^2 = |(c \cdot A + \alpha I)^g b_c|^2$ . We need  $\alpha$  s.t.  $\lambda_{\min}(c \cdot A + \alpha I) = 0$ , therefore,  $\alpha = -\lambda_{\min}(c \cdot A)$ .

Define  $\hat{z}(c) = |(c \cdot A - \lambda_{\min}(A))^g (c \cdot b)|^2$ . Consider  $\mathbb{R}^m \ni c = c^{\parallel} + c^{\perp}$ ,  $c^{\parallel} \parallel c_+$ ,  $(c_+, c^{\perp}) = 0$ .

Then  $\hat{z}(c) = \hat{z}(c^{\perp})$ , i.e.  $\hat{z}$  does not depend on  $c^{\parallel}$ . It depends only on  $c^{\perp}$ , and  $c^{\parallel}$  is chosen in a way that  $\lambda_{\min}(c \cdot A) = 0$ .

## 7. Theorem 3.4 (Nonconvexity certificate)

Given.

1. The map  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $f_i(x) = x^T A_i x + 2b_i^T x$ ,  $A_i = A_i^T$ ,  $m, n \geq 3$ . The vector  $c \in \mathbb{R}^m$ .
2.  $A_c \geq 0$ ,  $\dim \text{Ker } A_c = 1$  (=simple zero eigenvalue),  $\text{Ker } A_c = \text{Lin}\{e\}$
3.  $b_c \perp \text{Ker } A_c$
4.  $b \perp \text{Ker } A_c$ ,  $e^0 = -A_c^g b_c$
5.  $f(\alpha e + e^0) = f^0 + 2\alpha f^1 + \alpha^2 f^2$ ,  $f^1 \not\parallel f^2$

Then  $F = \text{Im } f$  is nonconvex.

Consider  $\inf_{y \in F} (c, y)$  (part 2).  $x = \underbrace{x^{\parallel}}_{\alpha e} + \underbrace{(-A_c^g b_c)}_{e^0}$ . Then

$$f_i(\alpha e + e^0) = \alpha^2 \underbrace{e^T A_i e}_{f_i^2} + 2\alpha \underbrace{(b_i^T e + e^T A_i e^0)}_{f_i^1} + \underbrace{2b_i^T e^0 + e^{0T} A_i e^0}_{f_i^0}$$

If  $f^1 \not\parallel f^2$ , then  $\{f(\alpha e + e^0) | \alpha \in \mathbb{R}\} = F \cap \{y | (c, y) = \inf_{y \in F} (c, y)\}$  is nonconvex. Then  $F$  is nonconvex.

## 8. Equations (0.18)-(0.21)

Consider  $A(t): n \times n$ ,  $\exists \dot{A}$ ,  $A^T = A$ ,  $A \geq 0$ ,  $A$  has a simple zero eigenvalue:  $\forall t A(t)x_0(t) = 0$ ,  $x_0^T x_0 = 0$ .

Then  $A = S \Lambda S^T$ ,  $S^T S = E$ ,  $A^g = S \Lambda^g S^T$ . Define  $\lambda_i = \Lambda_{ii}$ .  $\Lambda_{ii}^g = \begin{cases} \frac{1}{\lambda_i}, & \lambda_i \neq 0 \\ 0, & \lambda_i = 0 \end{cases}$ .

Then  $AA^g = A^g A = S \Lambda S^T S \Lambda^g S^T = S \Lambda \Lambda^g S^T = \sum_{\lambda_i \neq 0} s_i s_i^T = 1 - x_0 x_0^T$  (0.20)

Consider  $\frac{d}{dt} A x_0 = \dot{A} x_0 + A \dot{x}_0$ . Multiplying by  $A^g$  from the left:  $A^g \dot{A} x_0 + A^g A \dot{x}_0 = 0$ . Consider  $A^g A \dot{x}_0 = (1 - x_0 x_0^T) \dot{x}_0 = \dot{x}_0 - x_0 x_0^T \dot{x}_0$ . Since  $\|x_0\|^2 = x_0^T x_0 = 1$ ,  $x_0^T \dot{x}_0 = 0$ . Then  $-A^g \dot{A} x_0 = \dot{x}_0$  (0.19).

Consider  $\dot{A} x_0 + A \dot{x}_0 = 0$ . Multiplying by  $x_0^T$  from the left:  $x_0^T \dot{A} x_0 + x_0^T A \dot{x}_0 = 0$ . Then  $x_0^T \dot{A} x_0 = 0$  (0.18)

Consider  $AA^g = 1 - x_0 x_0^T$ . Then  $\dot{A} A^g + A \dot{A}^g = -\dot{x}_0 x_0^T - x_0 \dot{x}_0^T = A^g \dot{A} x_0 x_0^T + x_0 x_0^T \dot{A} A^g$  (a)

Consider  $A^g x_0 = S \Lambda^g S^T x_0 = 0$ .

Multiplying (a) by  $x_0$  from the right:  $A \dot{A}^g x_0 = A^g \dot{A} x_0$ . Multiplying by  $A^g$  from the left:  $A A^g \dot{A}^g x_0 = A^g A^g \dot{A} x_0$ . Then  $(1 - x_0 x_0^T) \dot{A}^g x_0 = A^g A^g \dot{A} x_0$ . Then  $\dot{A}^g x_0 = A^{-2} \dot{A} x_0 + x_0 x_0^T \dot{A}^g x_0$ .

Let's multiply (a) by  $A^g$  from the left:  $A^g \dot{A} A^g + A^g A \dot{A}^g = A^{-2} \dot{A} x_0 x_0^T + \cancel{A^g x_0^T}^0 x_0^T \dot{A} A^g$ .

Consider  $A^g A \dot{A}^g = \dot{A}^g - x_0 x_0^T \dot{A}^g = \dot{A}^g - x_0 x_0^T \dot{A} A^{-2} - x_0 x_0^T \dot{A}^g x_0 x_0^T$ .

Consider  $x_0^T \dot{A}^g x_0 = x_0^T (\dot{S} \Lambda^g S^T + S \dot{\Lambda}^g S^T + S \Lambda^g \dot{S}^T) x_0 = x_0^T \dot{S} \Lambda^g S^T x_0 = 0$ .

Then  $\frac{d}{dt} A^{-1} = -A^{-1} \dot{A} A^{-1} + x_0 x_0^T \dot{A} A^{-2} + A^{-2} \dot{A} x_0 x_0^T$  (0.21).

Case  $\text{Rg} A < n - 1$  is not considered since probability of such event is small.

## 9. Gradient descent

1. (0.5):  $\underbrace{(c \cdot A - \lambda_{\min}(c \cdot A))}_{Q} x_0 = 0$ , then  $\dot{x}_0 = -Q^{-1} \dot{Q} x_0$ .  $\frac{d}{dt} \lambda_{\min} = \frac{d}{dt} x_0^T (c \cdot A) x_0 = 2 \dot{x}_0^T (c \cdot A) x_0 + x_0^T (\dot{c} \cdot A) x_0 = 2 \lambda_{\min} \dot{x}_0^T x_0 + x_0^T (\dot{c} \cdot A) x_0$ .  
Then  $\dot{x}_0 = -(A_c - \lambda_{\min}(A_c))^{-1} (\dot{c} \cdot A - x_0^T (\dot{c} \cdot A) x_0) x_0$  (correct).
2. (0.6).  $x_0^T (c \cdot b) = 0$ , use (0.5) (correct).
3. (0.7).  $\frac{\partial}{\partial t} \|v(c)\|^2 = \frac{\partial}{\partial t} \sum_j v_j^2(c) = 2 \sum_j v_j \frac{\partial}{\partial t} v_j = 2 v^T(c) \frac{d}{dt} v(c(t))$  (correct).  $v(c) = \underbrace{(c \cdot A - \lambda_{\min}(c \cdot A))^{-1}}_Q (c \cdot b) = Q^{-1}(c \cdot b)$ .  
4. (0.8) Define  $Q = c \cdot A - \lambda_{\min}(c \cdot A)$ . Define  $v = Q^{-1}(c \cdot b)$ . Then  $z(c) = \|v\|^2$  and  $\dot{z} = 2 v^T \dot{v}$ .  
 $\dot{v} = \dot{Q}^{-1}(c \cdot b) + Q^{-1}(\dot{c} \cdot b)$ .  
Consider (0.21)  $\dot{Q}^{-1} = -Q^{-1} \dot{Q} Q^{-1} + x_0 x_0^T \dot{Q} Q^{-2} + Q^{-2} \dot{Q} x_0 x_0^T$ .  
Then  $\dot{z} = 2 v^T \left( Q^{-1}(\dot{c} \cdot b) + (-Q^{-1} \dot{Q} Q^{-1} + x_0 x_0^T \dot{Q} Q^{-2} + Q^{-2} \dot{Q} x_0 x_0^T)(c \cdot b) \right) \equiv$   
 $\equiv 2 v^T Q^{-1}(\dot{c} \cdot b) - 2 v^T Q^{-1} \dot{Q} Q^{-1}(c \cdot b) + 2 v^T x_0^T x_0^T \dot{Q} Q^{-2}(c \cdot b) + 2 v^T Q^{-2} \dot{Q} x_0 x_0^T (c \cdot b) \equiv$   
Since  $x_0 \in \text{Ker } Q \perp (c \cdot b)$ , we have  $x_0^T (c \cdot b) = 0$ .  
Since  $Q x_0 = 0$ ,  $Q^{-1} x_0 = 0$ :  $Q^{-1} x_0 = S \Lambda^{-1} S^T x_0 = S * 0 = 0$ . Since  $v^T = (c \cdot b)^T Q^{-1}$ . Then  $v^T x_0 = 0$   
 $\equiv 2 v^T Q^{-1}(\dot{c} \cdot b) - 2 v^T Q^{-1} \dot{Q} \underbrace{Q^{-1}(c \cdot b)}_v = 2 v^T Q^{-1}(\dot{c} \cdot b) - 2 v^T Q^{-1} \dot{Q} v = \boxed{\dot{z} = 2 v^T Q^{-1}(\dot{c} \cdot b - \dot{Q} v)}$ ,  
 $\dot{Q} = \dot{c} \cdot A - x_0^T (\dot{c} \cdot A) x_0$   
Since  $\dot{z} = \sum \frac{\partial z}{\partial c_i} \frac{\partial c_i}{\partial t}$ ,  $\frac{\partial z}{\partial c_i}$  can be found as a coefficient at  $\dot{c}_i$  in  $\dot{z}$   
 $\dot{z} = 2 v^T Q^{-1} \sum_i (\dot{c}_i b_i - \dot{c}_i A_i v + x_0^T \dot{c}_i A_i x_0 v) = \sum_i \dot{c}_i [2 v^T Q^{-1} (b_i - (A_i - x_0^T A_i x_0) v)]$   
Thus,  $\boxed{\frac{\partial z}{\partial c_i} = 2 v^T Q^{-1} (b_i - (A_i - x_0^T A_i x_0) v)}$ ,  $Q = c \cdot A - \lambda_{\min}(c \cdot A)$ ,  $v = Q^{-1}(c \cdot b)$ ,  $x_0 \in \text{Ker } Q$ ,  $\|x_0\| = 1$   
**not the same as (0.8) in draft.pdf (but numerically same):**  
 $\dot{z}_{(0.8)} = 2 \underbrace{(c \cdot b)^T Q^{-2}(\dot{c} \cdot b)}_{v^T Q^{-1}} - v^T (Q^{-1} \dot{Q} + \dot{Q} Q^{-1}) v = 2 v^T Q^{-1}(\dot{c} \cdot b) - v^T Q^{-1} \dot{Q} v - v^T \dot{Q} Q^{-1} v$   
5. (0.10). If  $\dot{c} = \beta c_+$ , then  $\dot{z} = 2 v^T Q^{-1}(\dot{c} \cdot b - \dot{Q} v) = \equiv$ . Since  $c_+ \cdot b = 0$ ,  $c_+ \cdot A = I$ ,  $\dot{Q} = c_+ \cdot A - x_0^T (c_+ \cdot A) x_0 = I - x_0^T x_0 = I - 1 = 0$ . And  $\equiv 0$  (correct with new  $\dot{z}$ ).  
6. (0.14)  $n_i = (b_i^T - v^T (A_i - x_0^T A_i x_0)) x_0$   
7. (0.16)  $P(\lambda) = Q^{-1} Q = S \Lambda^g S^T S \Lambda S^T = S \Lambda \Lambda^g S^T = 1 - x_0 x_0^T$  — projector on  $(\text{Ker } Q)^\perp$  (correct)  
8. (0.15)  $P(\lambda)(c(\lambda) \cdot b) = c(\lambda) \cdot b \Leftrightarrow \underbrace{(c(\lambda) \cdot b) \perp \text{Ker } Q}_{\Leftrightarrow c(\lambda) \in c_{\text{bad}}} \Leftrightarrow c(\lambda) \cdot b \in \text{Im } Q \Leftrightarrow \exists \hat{x}: Q \hat{x} = c(\lambda) \cdot b$  (0.17)  
(correct)

## 10. Gradient descent. Projection

We have  $c \in \mathbb{R}^m$ ,  $c \in c_{\text{bad}} = \{c \mid \|c\| = 1, \text{Ker } Q(c) \perp (c \cdot b)\}$ .  $Q = c \cdot A - \lambda_{\min}(c \cdot A)$ ,  $v = Q^{-1}(c \cdot b)$ ,  $x_0 \in \text{Ker } Q$ ,  $\|x_0\| = 1$

1. Calculate  $\frac{\partial z}{\partial c_i} = 2v^T Q^{-1}(b_i - (A_i - x_0^T A_i x_0)v)$ . Define  $\Delta c = -\nabla z(c)$
2. Calculate  $\hat{n}_i = (b_i^T - v^T(A_i - x_0^T A_i x_0))x_0$ , define  $n_i = \frac{\hat{n}_i}{|\hat{n}|}$
3. Define  $c' = c + \Delta c - n(\Delta c, n)$
4. Define  $c(\lambda) = c' + \lambda n$ . Define  $x_0(\lambda)$  s.t.  $x_0(\lambda) \in \text{Ker } Q(\lambda)$ ,  $x_0(\lambda)^T x_0 > 0$ ,  $\|x_0(\lambda)\| = 1$ . Define  $m(\lambda) = (c(\lambda) \cdot b)^T x_0(\lambda)$ . Beware of  $\text{Rg } Q < n - 1$ .
5. Find root of  $m(\lambda)$  using binary search on  $[-\lambda^0, \lambda^0]$ ,  $\lambda^0 = \|c - c'\|$ .
6. Next  $c$ :  $c(\lambda)$

$$m(\lambda) = 0 \Leftrightarrow (c(\lambda) \cdot b) \perp \{x_0\} = \text{Ker } Q \text{ if } \text{Rg } Q = n - 1.$$

## 11. Equation 0.19

Consider  $A = S\Lambda S^T$ ,  $S^T S = E$ ,  $x_0$  is a simple zero eigenvector of  $A$ :  $Ax_0 = 0$ ,  $\|x_0\| = 1$ .  
 $A^{-1} = S\Lambda_1 S^T$ ,  $\Lambda = (\lambda_1, \dots, \lambda_{n-1}, 0)$ ,  $\Lambda_1 = (\lambda_1^{-1}, \dots, \lambda_{n-1}^{-1}, 0)$ .

Equation (0.19):

$$x_0 = -A^{-1}\dot{A}x_0$$

Consider  $A^{-1}\dot{A}x_0 = S\Lambda_1 S^T(\dot{S}\Lambda S^T + S\dot{\Lambda}S^T + S\Lambda\dot{S}^T)x_0 \boxed{=}$ .

Consider  $x_0 = Sy_0$ , where  $y_0 = (0, 0, \dots, 1)$ . Therefore,  $0 = \dot{y}_0 = \dot{S}^T x_0 + S^T \dot{x}_0$

Going back to 0.19, the part  $\Lambda S^T x_0 = \Lambda y_0 = 0$ , another part  $\dot{\Lambda} S^T x_0 = 0$ . Consequently,

$$\boxed{=} S\Lambda_1 S^T S\Lambda\dot{S}^T x_0 = S\Lambda_1 \Lambda\dot{S}^T x_0 = -S\Lambda_1 \Lambda S^T \dot{x}_0 = -\sum_{i=1}^{n-1} s_i s_i^T \dot{x}_0 = -(E - x_0 x_0^T) \dot{x}_0 = -\dot{x}_0 + x_0 x_0^T \dot{x}_0 \boxed{=}$$

Taking a derivative  $\|x_0\| = 1$ , we get  $x_0^T \dot{x}_0 = 0$ , therefore,

$$\boxed{=} -\dot{x}_0$$

## 12. Nonconvexity certificate in $b_i = 0$ case

Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a quadratic map:  $f_i(x) = x^T A_i x$ ,  $A_i = A_i^T$ .

Consider  $c \in \mathbb{R}^m$  and boundary points  $\partial F_c$ :

$$c \cdot f(x) \rightarrow \min_c$$

Where  $A = \sum c_i A_i$ . Assuming  $A \geq 0$ .

Minimization leads to  $Ax = 0$ . The following cases hold:

1.  $\text{Rg } A = n$ .  $x = 0$  is a unique solution
2.  $\text{Rg } A = n - 1$ .  $x = \alpha e$ ,  $f(x) = \alpha^2 f(e)$
3.  $\text{Rg } A = n - 2$ . In this case  $x = \alpha_1 e^1 + \alpha_2 e^2$ . Consider  $f(x) = \alpha_1^2 f_{11} + 2\alpha_1 \alpha_2 f_{12} + \alpha_2^2 f_{22}$ .
  - (a)  $f_{11}, f_{22}, f_{12}$  are linearly independent. In this case  $\partial F_c$  is nonconvex
  - (b)  $\text{Rg } |f_{11} f_{12} f_{22}| = 1$ .  $\partial F_c$  convex.

**Result** If exist  $c \in \mathbb{R}^m$ :

1.  $\text{Rg } A = n - 2$
2.  $A \geq 0$
3.  $f_{ij}$  are linearly independent

Then  $F = \text{Im } f$  is nonconvex.



### 13. conv F

Пусть  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $f_i(x) = x^T A_i x + 2b_i^T x$ ,  $A_i = A_i^T$ .

Обозначим  $F = f(\mathbb{R}^n)$ ,  $G = \text{conv} F$

Обозначим  $H_i = \begin{bmatrix} A_i & b_i \\ b_i^T & 0 \end{bmatrix}$

Обозначим  $X = \begin{bmatrix} x \\ 1 \end{bmatrix} \begin{bmatrix} x^T & 1 \end{bmatrix} = \begin{bmatrix} xx^T & x \\ x^T & 1 \end{bmatrix}$

Тогда  $f_i(x) = \text{tr} H_i X$ ,  $f(x) = H(X)$

Обозначим  $V = \{X \in \mathbb{R}^{(n+1) \times (n+1)} | X = X^T, X \geq 0, X_{n+1, n+1} = 1\}$

Обозначим  $G_1 = H(V)$ .

Доказать:  $G_1 = G$  (On the feasibility for the system of quadratic equations, Theorem 3.1. (Convex hull))

1.  $G \subseteq G_1$ . Пусть  $y \in G$ . Тогда  $\exists \{y_i\}_{i=1}^l \subset F$ ,  $\{\lambda_i\}_{i=1}^l: y = \sum_{i=1}^l \lambda_i y_i$ , где  $\lambda_i \geq 0$ ,  $\sum \lambda_i = 1$ .

Поскольку  $y_i \in F$ ,  $\exists \{X_i\}: y_i = H(X_i)$ , причем  $X_i = \begin{bmatrix} x_i x_i^T & x_i \\ x_i^T & 1 \end{bmatrix} \in V$ . Рассмотрим  $j$ -ю компоненту  $y^j = \sum \lambda_i y_i^j = \sum \lambda_i \text{tr} H_j X_i = \text{tr} H_j \underbrace{\sum \lambda_i X_i}_X$ . То есть, найден  $X \in V: y = H(X)$ .

Значит,  $y \in G_1$

2.  $G_1 \subseteq G$ . Пусть  $y \in G_1$ . Тогда  $y = H(X)$ ,  $X \in V$ . Доказать:  $y \in G = \text{conv} F$ . Представим  $X$  в виде выпуклой комбинации  $X = \sum \lambda_i X_i$ , где  $X_i \in V$ , причем  $X_i = \begin{bmatrix} x_i x_i^T & x_i \\ x_i^T & 1 \end{bmatrix} \in V$  для некоторого  $x_i$ . Это докажет  $y \in G$ .

Рассмотрим  $X = \sum_{k=1}^n \lambda_k s_k s_k^T$  — спектральное разложение. Поскольку  $X \in V$ ,  $X \geq 0$ , значит,  $\lambda_k \geq 0$ . Обозначим  $\Lambda = \sum \lambda_k$ . Пусть  $s_{k, n+1} \neq \frac{1}{\Lambda}$ . Тогда переопределим  $s_{k, n+1} = \frac{1}{\Lambda}$ . Это можно сделать, т.к.  $H_i(s_k s_k^T)$  не изменится (прямая проверка, использовать  $H_{i, n+1, n+1} = 0$ ).

Рассмотрим  $X = \sum \underbrace{\frac{\lambda_k}{\Lambda}}_{\alpha_k} \underbrace{s_k s_k^T}_{X_k}$ . Получили представление  $X = \sum \alpha_k X_k$ ,  $X_k$  имеет нужный вид,

$\alpha_k$  — выпуклая комбинация. Получаем  $y \in G$

### 14. Small ball

Пусть  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $f_i(x) = x^T A_i x - 2b_i^T x$ ,  $A_i = A_i^T$ . Пусть  $c \in \mathbb{R}^m$ .

Обозначим  $c \cdot A = \sum_{i=1}^n c_i A_i$ ,  $c \cdot b = \sum_{i=1}^n c_i b_i$ ,  $F_c(x) = c^T f(x)$

**Хотим найти:**

$$\min_{\|x\|^2=1} F_c(x)$$

Функция Лагранжа:  $L(x, \lambda) = x^T (c \cdot A) x - 2(c \cdot b)^T x - \lambda(\|x\|^2 - 1)$ .

Находим  $L_x = 2(c \cdot A)x - 2c \cdot b - 2\lambda x = 0$ ,  $L_\lambda = \|x\|^2 - 1 = 0$ ,

Получаем систему  $\begin{cases} \|x\| = 1 \\ (c \cdot A - \lambda)x = c \cdot b \end{cases}$  Это совпадает с (2.3).

Далее переходим в базис из собственных векторов  $\{x_i\}$  симметричной матрицы  $c \cdot A$

$$S = \|x_1 \dots x_n\|, S^T S = E$$

$$x = Sy, \Lambda = S^T (c \cdot A) S, c \cdot b = S\alpha$$

Получаем

$$\begin{cases} \|y\| = 1 & (1) \\ (\Lambda - \lambda)y = \alpha & (2) \end{cases}$$

Выражаем из (2)

$$y_k = \frac{\alpha_k}{\lambda_k - \lambda}$$

Подставляем в (1) — это совпадает с (2.7)

$$\sum_{i=1}^n \left( \frac{\alpha_k}{\lambda_k - \lambda} \right)^2 = 1$$

Это соотношение определяет  $\lambda$ , по  $\lambda$  находим  $y_k$ , затем находим  $x$  в старом базисе.