

**Exercise 2.37 [Boyd & Vandenberghe, 2004]**

*Nonnegative polynomials and Hankel LMIs.* Let  $K_{pol}$  be the set of (coefficients of) nonnegative polynomials of degree  $2k$  on  $\mathbb{R}$ :

$$K_{pol} = \{x \in \mathbb{R}^{2k+1} \mid x_1 + x_2 t + x_3 t^2 + \cdots + x_{2k+1} t^{2k} \geq 0 \forall t \in \mathbb{R}\}$$

(a) Show that  $K_{pol}$  is a proper cone.

(b) A basic result states that a polynomial of degree  $2k$  is nonnegative on  $\mathbb{R}$  if and only if it can be expressed as the sum of squares of two polynomials of degree  $k$  or less. In other words,  $x \in K_{pol}$  if and only if the polynomial

$$p(t) = x_1 + x_2 t + x_3 t^2 + \cdots + x_{2k+1} t^{2k}$$

can be expressed as

$$p(t) = r(t)^2 + s(t)^2,$$

where  $r$  and  $s$  are polynomials of degree  $k$ .

Use this result to show that

$$K_{pol} = \left\{ x \in \mathbb{R}^{2k+1} \mid x_i = \sum_{m+n=i+1} Y_{mn}, Y \in \mathbf{S}_+^{k+1} \right\}$$

In other words,  $p(t) = x_1 + x_2 t + x_3 t^2 + \cdots + x_{2k+1} t^{2k}$  is nonnegative if and only if there exists a matrix  $Y \in \mathbf{S}_+^{k+1}$  such that

$$\begin{aligned} x_1 &= Y_{11} \\ x_2 &= Y_{12} + Y_{21} \\ x_3 &= Y_{13} + Y_{22} + Y_{31} \\ &\vdots \\ x_{2k+1} &= Y_{k+1,k+1}. \end{aligned}$$

(c) Show that  $K_{pol}^* = K_{han}$  where

$$K_{han} = \{z \in \mathbb{R}^{2k+1} \mid H(z) \succeq 0\}$$

and

$$H(z) = \begin{bmatrix} z_1 & z_2 & z_3 & \cdots & z_k & z_{k+1} \\ z_2 & z_3 & z_4 & \cdots & z_{k+1} & z_{k+2} \\ z_3 & z_4 & z_5 & \cdots & z_{k+2} & z_{k+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ z_k & z_{k+1} & z_{k+2} & \cdots & z_{2k-1} & z_{2k} \\ z_{k+1} & z_{k+2} & z_{k+3} & \cdots & z_{2k} & z_{2k+1} \end{bmatrix}$$

(This is the *Hankel matrix* with coefficients  $z_1, \dots, z_{2k+1}$ .)

(d) Let  $K_{mom}$  be the conic hull of the set of all vectors of the form  $(1, t, t^2, \dots, t^{2k})$ , where  $t \in \mathbb{R}$ . Show that  $y \in K_{mom}$  if and only if  $y_1 \geq 0$  and

$$y = y_1(1, \mathbf{E}u, \mathbf{E}u^2, \dots, \mathbf{E}u^{2k})$$

for some random variable  $u$ . In other words, the elements of  $K_{mom}$  are nonnegative multiples of the moment vectors of all possible distributions on  $\mathbb{R}$ . Show that  $K_{pol} = K_{mom}^*$ .

(e) Combining the results of (c) and (d), conclude that  $K_{han} = \mathbf{cl}K_{mom}$ .

AS an example illustrating the relation between  $K_{mom}$  and  $K_{han}$ , take  $k = 2$  and  $z = (1, 0, 0, 0, 1)$ . Show that  $z \in K_{han}$ ,  $z \notin K_{mom}$ . Find an explicit sequence of points in  $K_{mom}$  which converge to  $z$ .

*Ans:* Here only presents solutions to (b) and (c).

(b) Write  $r(t) = r_1 + r_2 t + r_3 t^2 + \dots + r_{k+1} t^k$ ,  $s(t) = s_1 + s_2 t + s_3 t^2 + \dots + s_{k+1} t^k$  and the coefficient vectors  $r = [r_1, r_2, r_3, \dots, r_{k+1}]$ ,  $s = [s_1, s_2, s_3, \dots, s_{k+1}]$ , then

$$r(t) = r^T \begin{bmatrix} 1 \\ t \\ t^2 \\ \vdots \\ t^k \end{bmatrix}, s(t) = s^T \begin{bmatrix} 1 \\ t \\ t^2 \\ \vdots \\ t^k \end{bmatrix}$$

Denote  $e(t) = \begin{bmatrix} 1 \\ t \\ t^2 \\ \vdots \\ t^k \end{bmatrix}$  to avoid using the exhausting package for arrays; we see

$$r(t)^2 = e(t)^T r r^T e(t), s(t) = e(t)^T s s^T e(t)$$

Let  $R = r r^T, S = s s^T$ , due to the structure these two matrices are positive semidefinite. (e.g. for any given  $z \in \mathbb{R}^{k+1}$ ,  $z^T R z = (r^T z)^2 \geq 0$ .) Let  $Y = R + S \in \mathbf{S}_+^{k+1}$ . (sum of two positive semidefinite matrices is still positive semidefinite.) It's now evident that

$$\begin{aligned} p(t) &= e(t)^T Y e(t) \\ &= \sum_{m,n=1}^{k+1} t^{m-1} Y_{mn} t^{n-1} \\ &= \sum_{m,n=1}^{k+1} Y_{mn} t^{m+n-2} \\ &= \sum_{j=0}^{2k} \left( \sum_{m+n=j+2} Y_{mn} \right) t^j \end{aligned}$$

and

$$x_{j+1} = \sum_{m+n=j+2} Y_{mn}, \text{ or, } x_i = \sum_{m+n=i+1} Y_{mn}$$

On the other hand, if  $x_i = \sum_{m+n=i+1} Y_{mn}$  for some  $Y \in \mathbf{S}_+^{k+1}$ , then by similar index substitution trick,

$$p(t) = e(t)^T Y e(t) \geq 0$$

for any  $t \in \mathbb{R}$ , since  $Y$  is positive semidefinite.

(c) Note that for  $x \in K_{pol}$  that corresponds to  $Y(x) \in \mathbf{S}_+^{k+1}$ ,

$$\begin{aligned} x^T z &= \sum_{i=1}^{2k+1} x_i z_i \\ &= \text{tr}(Y(x)^T H(z)) \end{aligned}$$

Due to the fact that  $(\mathbf{S}_+^{k+1})^* = \mathbf{S}_+^{k+1}$ , we conclude  $K_{pol}^* = K_{han}$ . □

**Exercise 3.1 [Boyd & Vandenberghe, 2004]**

Suppose  $f : \mathbb{R} \rightarrow \mathbb{R}$  is convex, and  $a, b \in \text{dom } f$  with  $a < b$ .

(a) Show that

$$f(x) \leq \frac{b-x}{b-a}f(a) + \frac{x-a}{b-a}f(b)$$

for all  $x \in [a, b]$ .

(b) Show that

$$\frac{f(x) - f(a)}{x - a} \leq \frac{f(b) - f(a)}{b - a} \leq \frac{f(b) - f(x)}{b - x}$$

for all  $x \in (a, b)$ . Draw a sketch that illustrates this inequality.

(c) Suppose  $f$  is differentiable. Use the result in (b) to show that

$$f'(a) \leq \frac{f(b) - f(a)}{b - a} \leq f'(b).$$

Note that these inequalities also follow from (3.2):

$$f(b) \geq f(a) + f'(a)(b - a), f(a) \geq f(b) + f'(b)(a - b)$$

(d) Suppose  $f$  is twice differentiable. Use the result in (c) to show that  $f''(a) \geq 0$  and  $f''(b) \geq 0$ .

*Ans:*

(a) Let  $\theta = \frac{b-x}{b-a} \in [0, 1]$ , then the inequality is nothing but

$$f(\theta a + (1 - \theta)b) \leq \theta f(a) + (1 - \theta)f(b)$$

(b) The leftmost term is the slope of the secant jointing  $(a, f(a)), (x, f(x))$ , second  $(a, f(a)), (b, f(b))$ , and third  $(x, f(x)), (b, f(b))$ ; it suffices to show the slope of secants of the curve is increasing, that is, one equality of the two given ones. Take the inequality in (a) and subtract both sides with  $f(a)$

$$\begin{aligned} f(x) - f(a) &\leq \left( \frac{b-x}{b-a} - 1 \right) f(a) + \frac{x-a}{b-a} f(b) \\ &= -\frac{x-a}{b-a} f(a) + \frac{x-a}{b-a} f(b) \\ &= \frac{x-a}{b-a} (f(b) - f(a)) \\ \frac{f(x) - f(a)}{x - a} &\leq \frac{f(b) - f(a)}{b - a} \quad (\text{divide both sides with } x - a) \end{aligned}$$

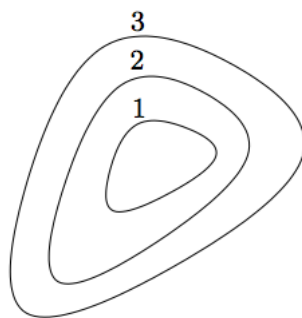
(c)  $f'(a)$  is the limit of LHS (of the inequality from (b)) as  $x \rightarrow a^+$  and  $f'(b)$  RHS as  $x \rightarrow b^-$ . By differentiability, limit taken on single side should give the derivative. Also by manipulating terms in (3.2) (subtracting  $f(a)$  or  $f(b)$  and dividing by  $b - a$ ) we

can get this inequality too.

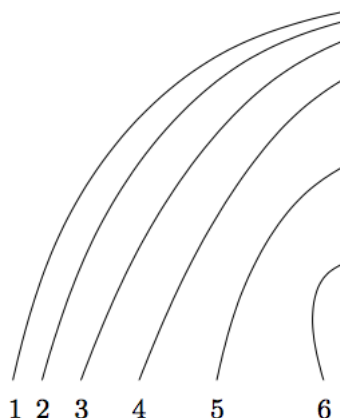
(d) The inequality  $f'(a) \leq f'(b)$  for arbitrary  $a$  and  $b$  shows the monotonicity of  $f'(x)$ . By basic calculus we know the derivative of a monotonically increasing function must be nonnegative, that is  $f''(x) \geq 0$  for all  $x \in (a, b)$ . (The proof involves Mean Value Theorem. I can elaborate if required.)  $\square$

**Exercise 3.2 [Boyd & Vandenberghe, 2004]**

*Level sets of convex, concave, quasiconvex, and quasiconcave functions.* Some level sets of a function  $f$  are shown below. The curve labeled 1 shows  $\{x \mid f(x) = 1\}$ , etc.



Could  $f$  be convex (concave, quasiconvex, quasiconcave)? Explain your answer. Repeat for the level curves shown below.



*Ans:* From the graphs we can first assert the first function is quasiconvex and second quasiconcave, due to the convexity of the level sets. We can further suggest the first function is convex; picture the epigraph by looking at the level sets, each slide of it looks rather like a parabola. We can't suggest the second function to be concave because the function doesn't seem increasing enough from the upper left corner to the lower left corner, which could make its epigraph non concave.

**Additional Exercise 2.10 [Boyd & Vandenberghe, 2017]**

*Weighted geometric mean.* The geometric mean  $f(x) = (\prod_k x_k)^{1/n}$  with  $\text{dom } f = \mathbb{R}_{++}^n$  is concave, as shown on page 74. Extend the proof to show that

$$f(x) = \prod_{k=1}^n x_k^{\alpha_k}, \text{dom } f = \mathbb{R}_{++}^n$$

is concave, where  $\alpha_k$  are positive numbers with  $\sum_k \alpha_k \leq 1$ .

*Ans:* First observe functions  $m_k(x_k) = x_k^{\alpha_k}$  are positive (we're assuming  $x_k > 0$ ) and concave since  $\alpha_k \in [0, 1]$ . ( $m_k''(x_k) = -\alpha_k(1 - \alpha_k)x_k^{\alpha_k-2} \leq 0$ .)

Take first derivative of  $f$ ,

$$\nabla f(x) = \begin{bmatrix} \alpha_1 x^{\alpha_1-1} \prod_{k \neq 1} x_k^{\alpha_k} \\ \alpha_2 x^{\alpha_2-1} \prod_{k \neq 2} x_k^{\alpha_k} \\ \vdots \\ \alpha_n x^{\alpha_n-1} \prod_{k \neq n} x_k^{\alpha_k} \end{bmatrix}$$

and then second derivative,

$$H_f(x) = \begin{bmatrix} \frac{f(x)\alpha_1(\alpha_1-1)}{x_1^2} & \frac{f(x)\alpha_1\alpha_2}{x_1x_2} & \dots & \frac{f(x)\alpha_1\alpha_n}{x_1x_n} \\ \frac{f(x)\alpha_1\alpha_2}{x_1x_2} & \frac{f(x)\alpha_2(\alpha_2-1)}{x_2^2} & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{f(x)\alpha_1\alpha_n}{x_1x_n} & \dots & \dots & \frac{f(x)\alpha_n(\alpha_n-1)}{x_n^2} \end{bmatrix}$$

$$=: f(x)A(x)$$

Observe that  $A(x) = \begin{bmatrix} \frac{\alpha_1(\alpha_1-1)}{x_1^2} & \frac{\alpha_1\alpha_2}{x_1x_2} & \dots & \frac{\alpha_1\alpha_n}{x_1x_n} \\ \frac{\alpha_1\alpha_2}{x_1x_2} & \frac{\alpha_2(\alpha_2-1)}{x_2^2} & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\alpha_1\alpha_n}{x_1x_n} & \dots & \dots & \frac{\alpha_n(\alpha_n-1)}{x_n^2} \end{bmatrix}$  is diagonally dominant with

all negative diagonal entries; e.g. for the first column/row

$$\begin{aligned} |a_{11}| - \sum_{i=2}^n |a_{i1}| &= \frac{\alpha_1}{x_1^2} ((1 - \alpha_1) - \alpha_2 - \dots - \alpha_n) \\ &= \frac{\alpha_1}{x_1^2} \left( 1 - \sum_{i=1}^n \alpha_i \right) \leq 0 \end{aligned}$$

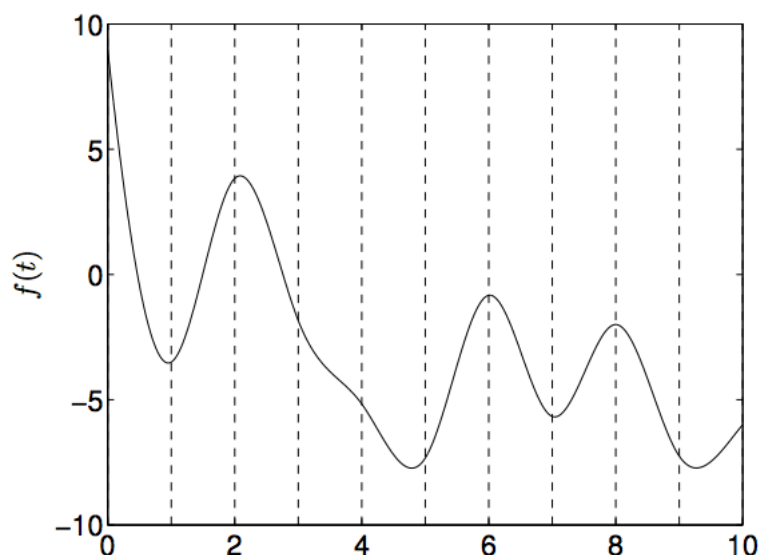
since by assumption  $\sum_i \alpha_i \leq 1$ . □

**Additional Exercise 5.8 [Boyd & Vandenberghe, 2017]**

*Least-squares fitting with convex splines.* A *cubic spline* (or *fourth-order spline*) with breakpoints  $\alpha_0, \alpha_1, \dots, \alpha_M$  (that satisfy  $\alpha_0 < \alpha_1 < \dots < \alpha_M$ ) is a piecewise-polynomial function with the following properties:

- the function is a cubic polynomial on each interval  $[\alpha_i, \alpha_{i+1}]$
- the function values, and the first and second derivatives are continuous on the interval  $(\alpha_0, \alpha_M)$ .

The figure shows an example of a cubic spline  $f(t)$  with  $M = 10$  segments and breakpoints  $\alpha_0 = 0, \alpha_1 = 1, \dots, \alpha_{10} = 10$ .

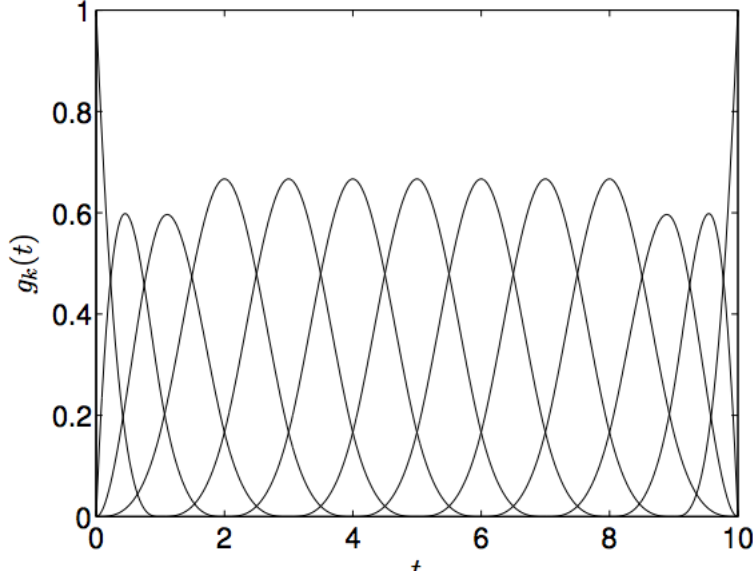


In approximation problems with splines it is convenient to parametrize a spline as a linear combination of basis functions, called *B-splines*. The precise definition of B-splines is not important for our purposes; it is sufficient to know that every cubic spline can be written as a linear combination of  $M + 3$  cubic B-splines  $g_k(t)$ , *i.e.*, in the form

$$f(t) = s_1 g_1(t) + \dots + x_{M+3} g_{M+3}(t) = x^T g(t),$$

and that there exist efficient algorithms for computing  $g(t) = (g_1(t), \dots, g_{M+3}(t))$ . The next figure shows the 13 B-splines for the breakpoints  $0, 1, \dots, 10$ .





In this exercise we study the problem of fitting a cubic spline to a set of data points, subject to the constraint that the spline is a convex function. Specifically, the breakpoints  $\alpha_0, \dots, \alpha_M$  are fixed, and we are given  $N$  data points  $(t_k, y_k)$  with  $t_k \in [\alpha_0, \alpha_M]$ . We are asked to find the convex cubic spline  $f(t)$  that minimizes the least squares criterion

$$\sum_{k=1}^N (f(t_k) - y_k)^2.$$

We will use B-splines to parametrize  $f$ , so the variables in the problem are the coefficients  $x$  in  $f(t) = x^T g(t)$ . The problem can then be written as

$$\begin{aligned} & \text{minimize} && \sum_{k=1}^N (x^T g(t_k) - y_k)^2 \\ & \text{subject to} && x^T g(t) \text{ is convex in } t \text{ on } [\alpha_0, \alpha_M] \end{aligned} \tag{1}$$

(a) Express problem (1) as a convex optimization problem of the form

$$\begin{aligned} & \text{minimize} && \|Ax - b\|_2^2 \\ & \text{subject to} && Gx \preceq h \end{aligned}$$

(b) Use CVX to solve a specific instance of the optimization problem in part (a). As in the figures above, we take  $M = 10$  and  $\alpha_0 = 0, \alpha_1 = 1, \dots, \alpha_{10} = 10$ .

Download the Matlab files `spline_data.m` and `bsplines.m`. The first m-file is used to generate the problem data. The command `[t, y] = spline_data` will generate two vectors  $t, y$  of length  $N = 51$ , with the data points  $t_k, y_k$ .

The second function can be used to compute the B-splines, and their first and second derivatives, at any given point  $u \in [0, 10]$ . The command `[g, gp, gpp] = bsplines(u)` returns three vectors of length 13 with elements  $g_k(u), g'_k(u)$ , and  $g''_k(u)$ . (The right

derivatives are returned for  $u = 0$ , and the left derivatives for  $u = 10$ .)

Solve the convex spline fitting problem (1) for this example, and plot the optimal spline.

*Ans:* (a) Let  $a_{ij} = g_j(t_i)$  and  $b_i = y_i$  then

$$\begin{aligned} \sum_{i=1}^N (x^T g(t_i) - y_i)^2 &= \sum_{i=1}^N \left( \sum_{j=1}^{M+3} x_j g_j(t_i) - y_i \right)^2 \\ &= \|Ax - b\|_2^2 \end{aligned}$$

where  $A = [a_{ij}] \in \mathbb{R}^{N \times (M+3)}$  and  $b \in \mathbb{R}^N$  and  $x \in \mathbb{R}^{M+3}$  is the minimization variable.

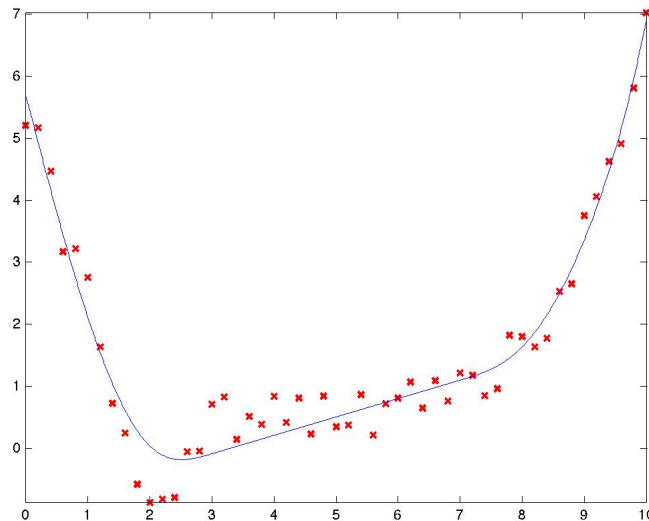
In addition to the above, the constraint  $x^T g(t)$  being convex is equivalent to  $x^T g''(t) \geq 0$ .

Let  $g_{ij} = -g_j''(t_i)$  and  $h_i = 0$ , then

$$Gx = \begin{bmatrix} -x^T g''(t_1) \\ -x^T g''(t_2) \\ \vdots \\ -x^T g''(t_N) \end{bmatrix} \preceq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = h$$

where  $G = [g_{ij}]$  and  $h = [h_i]$  ensures the convexity of  $x^T g(t)$  on the interval  $[\alpha_0, \alpha_M]$ . (The spline function  $x^T g(t)$  is piecewise-cubic-polynomial; its second derivative is therefore piecewise-linear. To ensure this piecewise-linear function is always nonnegative throughout  $[\alpha_0, \alpha_M]$ , we only need to evaluate the end points of the intervals since a linear function on an interval must be nonnegative if it's nonnegative on both of the endpoints. Note the endpoints of different intervals collide but it doesn't affect the results due to continuity of  $x^T g''(t)$ .)

(b) Here includes the graph (red crosses the data and blue curve the fitted B-spline curve) and the codes.



```
[t, y] = spline_data();
A = zeros(51,13);
G = zeros(51,13);
h = zeros(51,1);

for i=1:51
    [y1,yp,ypp] = bsplines(t(i));
    A(i,:) = y1';
    G(i,:) = -ypp';
end

cvx_begin
variable x(13)
minimize( (A*x-y)'*(A*x-y) )
subject to
    G*x <= h
cvx_end

fit_x = linspace(0,10,1000);
fit_y = zeros(1,1000);
for i=1:1000
    fit_y(i) = x' * bsplines(fit_x(i));
end

plot(t,y,'rx', 'LineWidth', 2);
hold on
handle = plot(fit_x, fit_y, 'b-');
axis equal
saveas(handle,'EE236B_hw2_p58_fit','jpg');
hold off
```