Homework 2 solutions

1. *Max–min and min–max characterization of eigenvalues.* Let A be a symmetric $n \times n$ matrix, with eigendecomposition
$$A = Q\,\mathbf{diag}(\lambda)Q^T = \sum_{i=1}^n \lambda_i q_i q_i^T.$$
The matrix Q is orthogonal ($Q^TQ = QQ^T = I$) with columns $q_1, \ldots, q_n$, and $\mathbf{diag}(\lambda)$ is the diagonal matrix with the eigenvalues $\lambda_1, \ldots, \lambda_n$ on its diagonal. We assume the eigenvalues are sorted as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

We denote by $\mu_1(X) \geq \mu_2(X) \geq \cdots \geq \mu_m(X)$ the eigenvalues of X^TAX, where X is an $n \times m$ matrix. In this problem we show that
$$\left[\begin{array}{c} \mu_1(X) \\ \mu_2(X) \\ \vdots \\ \mu_m(X) \end{array}\right] \preceq \left[\begin{array}{c} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{array}\right] \qquad (1)$$
for all matrices $X \in \mathbf{R}^{n\times m}$ with orthonormal columns ($X^TX = I$). The inequality is a component-wise vector inequality, *i.e.*, equivalent to the m scalar inequalities
$$\mu_1(X) \leq \lambda_1, \qquad \mu_2(X) \leq \lambda_2, \qquad \ldots, \qquad \mu_m(X) \leq \lambda_m.$$

(a) Suppose $X \in \mathbf{R}^{n\times m}$ is given and satisfies $X^TX = I$. We drop the argument X in $\mu_i(X)$, and write the eigendecomposition of the $m \times m$ matrix X^TAX as
$$X^TAX = \sum_{i=1}^m \mu_i v_i v_i^T.$$
The vectors $v_1, \ldots, v_m$ are orthonormal eigenvectors, and $\mu_1, \ldots, \mu_m$ are the eigenvalues, sorted as $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_m$.

Suppose $1 \leq k \leq m$. Denote by $V = \left[\begin{array}{cccc} v_1 & v_2 & \cdots & v_k \end{array}\right]$ the matrix with the first k eigenvectors $v_1, \ldots, v_k$ as its columns. Verify the following expressions for μ_k:
$$\begin{aligned} \mu_k &= \inf_{y_1^2+\cdots+y_k^2=1} (\mu_1 y_1^2 + \cdots + \mu_k y_k^2) \\ &= \inf_{y_1^2+\cdots+y_k^2=1} y^TV^T(X^TAX)Vy \\ &= \inf_{y_1^2+\cdots+y_k^2=1} y^TV^TX^TQ\,\mathbf{diag}(\lambda)Q^TXVy \\ &= \inf_{y_1^2+\cdots+y_k^2=1} \sum_{i=1}^n \lambda_i (q_i^TXVy)^2, \qquad (2) \end{aligned}$$

where $y = (y_1, \ldots, y_k) \in \mathbf{R}^k$.

(b) From the last expression (2), show that $\mu_k \leq \lambda_k$.
Hint. Consider a vector $\tilde{y} \in \mathbf{R}^k$ that satisfies $\|\tilde{y}\|_2 = 1$ and

$$q_1^T X V \tilde{y} = 0, \qquad q_2^T X V \tilde{y} = 0, \qquad \ldots, \qquad q_{k-1}^T X V \tilde{y} = 0.$$

Show that

$$\mu_k \leq \sum_{i=1}^n \lambda_i (q_i^T X V \tilde{y})^2 \leq \lambda_k.$$

Since the inequality (1) holds with equality for the matrix

$$X = [\ q_1 \quad q_2 \quad \cdots \quad q_m\ ], \tag{3}$$

we can conclude that the set

$$S = \{(\mu_1(X), \ldots, \mu_m(X)) \mid X \in \mathbf{R}^{n\times m},\ X^T X = I\}$$

has a *maximum element*, given by $(\lambda_1, \ldots, \lambda_m)$. Applying this result to $-A$, we see that the set S also has a *minimum element*, given by $(\lambda_{n-m+1}, \ldots, \lambda_n)$. This result is known as the Courant–Fischer min–max theorem.

As an application, it follows that the matrix X given in (3) is a solution of the (non-convex) optimization problem

$$\begin{array}{ll} \text{maximize} & f(X) \\ \text{subject to} & X^T X = I, \end{array}$$

with variable $X \in \mathbf{R}^{n\times m}$, for the following functions:

$$\begin{array}{rcl} f(X) & = & \lambda_{\max}(X^T A X) \\ & = & \mu_1(X), \\ f(X) & = & \lambda_{\min}(X^T A X) \\ & = & \mu_m(X), \\ f(X) & = & \mathbf{tr}(X^T A X) \\ & = & \mu_1(X) + \cdots + \mu_m(X), \end{array}$$

and, more generally, any function $f(X) = \sum_{k=1}^m h_k(\mu_k(X))$ where h_k is a nondecreasing function.

Solution.

(a) On line 1, we minimize $\mu_1 y_1^2 + \cdots + \mu_k y_k^2$ subject to the constraint $y_1^2 + \cdots + y_k^2 = 1$. Since the eigenvalues μ_i are sorted in descending order,

$$\mu_1 y_1^2 + \cdots + \mu_k y_k^2 \geq \mu_k (y_1^2 + \cdots + y_k^2) = \mu_k$$

with equality if $y = (0, \dots, 0, 1)$. Therefore the optimal value is μ_k. Line 2 follows from

$$V^T(X^TAX)V = \begin{bmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_k \end{bmatrix}.$$

On the next line we substitute the eigendecomposition of A.

(b) The conditions in the hint form a set of $k-1$ linear equations in k variables, so there exists a nonzero solution that we can normalize to satisfy $\|\tilde{y}\|_2 = 1$. We also note that

$$\sum_{i=k}^{n} (q_i^T XV\tilde{y})^2 = \sum_{i=1}^{n} (q_i^T XV\tilde{y})^2 = \|Q^TXV\tilde{y}\|_2^2 = \|\tilde{y}\|_2^2 = 1.$$

The first equality holds because $q_1^T XV\tilde{y} = \cdots = q_{k-1}^T XV\tilde{y} = 0$. For the third equality we use the fact that the matrix Q^TXV has orthonormal columns:

$$(Q^TXV)^T(Q^TXV) = V^TX^TQQ^TXV = I$$

because Q is orthogonal, and X and V have orthonormal columns.
The right-hand side of equation (2) can therefore be bounded as

$$\begin{aligned} \inf_{y^Ty=1} \sum_{i=1}^{n} \lambda_i (q_i^T XVy)^2 &\leq \sum_{i=1}^{n} \lambda_i (q_i^T XV\tilde{y})^2 \\ &= \sum_{i=k}^{n} \lambda_i (q_i^T XV\tilde{y})^2 \\ &\leq \lambda_k \sum_{i=k}^{n} (q_i^T XV\tilde{y})^2 \\ &= \lambda_k. \end{aligned}$$

On line 3 we use the fact that the eigenvalues λ_i are in descending order.

2. Exercise T3.1.

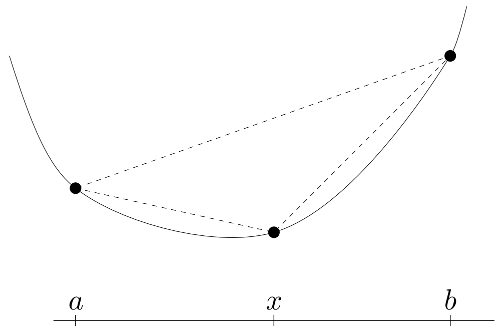
Solution.

(a) This is Jensen's inequality

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

with $\theta = (b-x)/(b-a)$.

(b) We obtain the first inequality by subtracting $f(a)$ from both sides of the inequality in (a). The second inequality follows from subtracting $f(b)$. Geometrically, the inequalities mean that the slope of the line segment between $(a, f(a))$ and $(b, f(b))$ is larger than the slope of the segment between $(a, f(a))$ and $(x, f(x))$, and smaller than the slope of the segment between $(x, f(x))$ and $(b, f(b))$.



(c) This follows from (b) by taking the limit for $x \to a$ on both sides of the first inequality, and by taking the limit for $x \to b$ on both sides of the second inequality.

(d) From part (c),
$$\frac{f'(b) - f'(a)}{b - a} \geq 0,$$
and taking the limit for $b \to a$ shows that $f''(a) \geq 0$.

3. Exercise T3.18 (b).

Solution. Define $g(t) = f(\hat{X} + tV)$, where $\hat{X} \succ 0$ and $V \in \mathbf{S}^n$.

$$\begin{aligned} g(t) &= (\det(\hat{X} + tV))^{1/n} \\ &= \left(\det \hat{X}^{1/2} \det(I + t\hat{X}^{-1/2} V Z^{-1/2}) \det \hat{X}^{1/2}\right)^{1/n} \\ &= (\det \hat{X})^{1/n} (\prod_{i=1}^n (1 + t\lambda_i))^{1/n} \end{aligned}$$

where λ_i, $i = 1, \ldots, n$, are the eigenvalues of $\hat{X}^{-1/2} V \hat{X}^{-1/2}$. From the last equality we see that g is a concave function of t on $\{t \mid \hat{X} + tV \succ 0\}$, since $\det \hat{X} > 0$ and the geometric mean $(\prod_{i=1}^n x_i)^{1/n}$ is concave on $\mathbf{R}^n_+$.

4. Exercise A3.10.

Solution. The Hessian of f is

$$\nabla^2 f(x) = f(x) \left(qq^T - \mathbf{diag}(\alpha)^{-1} \mathbf{diag}(q)^2\right)$$

where q is the vector $(\alpha_1/x_1, \ldots, \alpha_n/x_n)$. To show that $\nabla^2 f(x)$ is negative semidefinite, we verify that the inequality

$$y^T \nabla^2 f(x) y = f(x) \left((\sum_{k=1}^n \alpha_k y_k / x_k)^2 - \sum_{k=1}^n \alpha_k y_k^2 / x_k^2 \right) \leq 0$$

holds for all y. This follows from the Cauchy–Schwarz inequality

$$(\sum_{k=1}^n u_k v_k)^2 \leq (\sum_{k=1}^n u_k^2)(\sum_{k=1}^n v_k^2)$$

applied to the vectors

$$u = (\sqrt{\alpha_1}y_1/x_1, \ldots, \sqrt{\alpha_n}y_n/x_n), \qquad v = (\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_n}).$$

With this choice of u and v the Cauchy–Schwarz inequality gives

$$\left(\sum_{k=1}^n \alpha_k y_k/x_k\right)^2 \le \left(\sum_{k=1}^n \alpha_k y_k^2/x_k^2\right)\left(\sum_{k=1}^n \alpha_k\right) \le \left(\sum_{k=1}^n \alpha_k y_k^2/x_k^2\right).$$

The second inequality follows from $\sum_{k=1}^n \alpha_k \le 1$.

5. Exercise T3.19 (a).

   Solution. We can express $f(x)$ as

$$\begin{aligned} f(x) &= \alpha_r(x_{[1]} + x_{[2]} + \cdots + x_{[r]}) + (\alpha_{r-1} - \alpha_r)(x_{[1]} + x_{[2]} + \cdots + x_{[r-1]}) \\ &\quad + (\alpha_{r-2} - \alpha_{r-1})(x_{[1]} + x_{[2]} + \cdots + x_{[r-2]}) + \cdots + (\alpha_1 - \alpha_2)x_{[1]}. \end{aligned}$$

   This is a nonnegative sum of the convex functions

$$x_{[1]}, \qquad x_{[1]} + x_{[2]}, \qquad x_{[1]} + x_{[2]} + x_{[3]}, \qquad \ldots, \qquad x_{[1]} + x_{[2]} + \cdots + x_{[r]}.$$

6. Exercise A3.20 (a).

   Solution. f is the difference of a convex and a concave function. The first term is convex because it is the supremum of a family of linear functions of x. The second term is concave because it is the infimum of a family of linear functions of x.

7. Exercise A6.8.

   Solution.

   (a) The objective function is

$$\sum_{k=1}^N (x^T g(t_k) - y_k)^2 = \|Ax - b\|_2^2$$

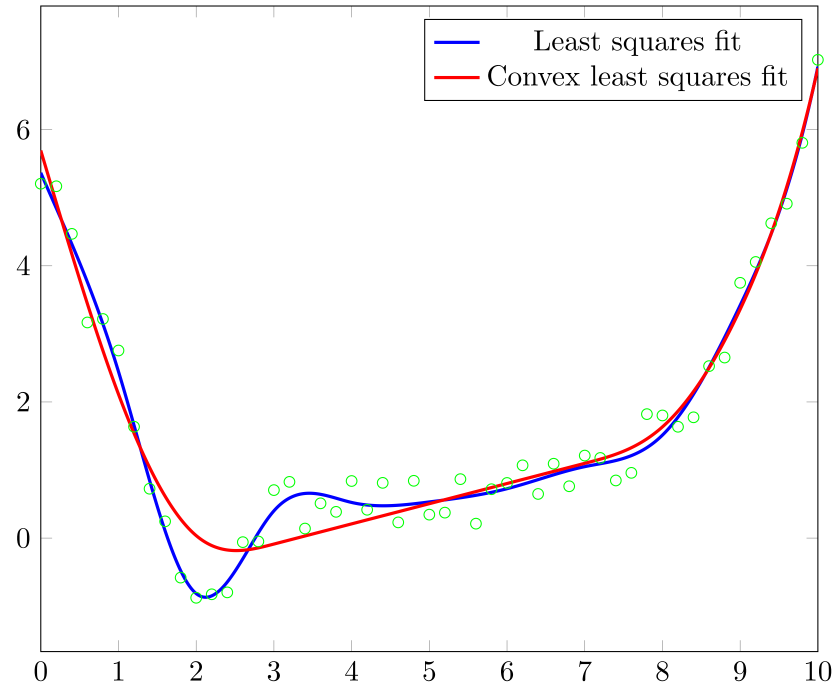
   with

$$A = \begin{bmatrix} g(t_1)^T \\ g(t_2)^T \\ \vdots \\ g(t_N)^T \end{bmatrix}, \qquad b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}.$$

   To handle the convexity constraint we note that f'' is piecewise linear in t. Therefore $f''(t) \ge 0$ for all $t \in (\alpha_0, \alpha_M)$ if and only if $f''(\alpha_k) = x^T g''(\alpha_k) \ge 0$ for $k = 0, \ldots, M$. This gives a set of linear inequalities $Gx \preceq 0$ with

$$G = -\begin{bmatrix} g''(\alpha_0)^T \\ g''(\alpha_1)^T \\ \vdots \\ g''(\alpha_M)^T \end{bmatrix}.$$

(b)
```
[u, y] = spline_data;
N = length(u);
A = zeros(N, 13);
b = y;
for k = 1:N
    [g, gp, gpp] = bsplines(u(k));
    A(k,:) = g';
end;
% Solution without convexity constraint
xls = A\b;
% Solution with convexity constraint
G = zeros(11, 13);
for k = 1:11
    [g, gp, gpp] = bsplines(k-1);
    G(k,:)=  gpp';
end;
cvx_begin
    variable x(13);
    minimize( norm(A*x - b) );
    subject to
        G*x >= 0;
cvx_end
% plot solutions
npts = 1000;
t = linspace(0, 10, npts);
fls = zeros(1, npts);
fcvx = zeros(1, npts);
for k = 1:npts
    [g, gp, gpp] = bsplines(t(k));
    fls(k) = xls' * g;
    fcvx(k) = x' * g;
end;
plot(u, y,'o', t, fls, 'b-', t, fcvx, 'r-');
```



Least squares fit
Convex least squares fit
0
2
4
6
0
1
2
3
4
5
6
7
8
9
10