Homework 1 solutions

1. Exercise T2.9 (a).

   Solution.

$$\begin{aligned}
\|x - x_0\|_2 \leq \|x - x_i\|_2 \quad &\Longleftrightarrow \quad (x - x_0)^T(x - x_0) \leq (x - x_i)^T(x - x_i) \\
&\Longleftrightarrow \quad x^Tx - 2x_0^Tx + x_0^Tx_0 \leq x^Tx - 2x_i^Tx + x_i^Tx_i \\
&\Longleftrightarrow \quad 2(x_i - x_0)^Tx \leq x_i^Tx_i - x_0^Tx_0.
\end{aligned}$$

   This linear inequality defines a halfspace (if $x_i \neq x_0$). Therefore we can express V as a polyhedron $\{x \mid Ax \preceq b\}$ by defining

$$A = 2\begin{bmatrix} (x_1 - x_0)^T \\ (x_2 - x_0)^T \\ \vdots \\ (x_K - x_0)^T \end{bmatrix}, \qquad b = \begin{bmatrix} x_1^Tx_1 - x_0^Tx_0 \\ x_2^Tx_2 - x_0^Tx_0 \\ \vdots \\ x_K^Tx_K - x_0^Tx_0 \end{bmatrix}.$$

2. Exercise T2.12 (d, e, g).

   Solution.

   (d) Convex. For fixed y, the set $\{x \mid \|x - x_0\|_2 \leq \|x - y\|_2\}$ is a halfspace. Squaring both sides of the inequality and expanding the norms gives

$$\|x\|^2 - 2x_0^Tx + \|x_0\|_2^2 \leq \|x\|^2 - 2y^Tx + \|y\|_2^2.$$

   This is a linear inequality

$$2(y - x_0)^Tx \leq \|y\|_2^2 - \|x_0\|_2^2.$$

   Therefore the set in the assignment is an intersection of halfspaces (one for each $y \in S$).

   (e) In general this set is not convex. A simple counterexample in $\mathbf{R}$ is $S = \{-1, 1\}$ and $T = \{0\}$. We have

$$\{x \mid \mathbf{dist}(x, S) \leq \mathbf{dist}(x, T)\} = \{x \in \mathbf{R} \mid x \leq -1/2 \text{ or } x \geq 1/2\}$$

   which clearly is not convex.

(g) Convex. We have

$$\begin{aligned}\|x-a\|_2 \le \theta\|x-b\|_2 &\iff \|x-a\|_2^2 \le \theta^2\|x-b\|_2^2 \\ &\iff (1-\theta^2)x^Tx - 2(a-\theta^2 b)^Tx + (a^Ta - \theta^2 b^Tb) \le 0.\end{aligned}$$

If $\theta = 1$, this defines a halfspace (see part (d)). If $\theta < 1$, it defines a ball

$$\{x \mid (x-x_0)^T(x-x_0) \le R^2\},$$

with center x_0 and radius R given by

$$x_0 = \frac{a-\theta^2 b}{1-\theta^2}, \qquad R = \left(\frac{\theta^2\|b\|_2^2 - \|a\|_2^2}{1-\theta^2} + \|x_0\|_2^2\right)^{1/2}.$$

3. Exercise A2.10.

Solution.

(a) Define $I = \{k \mid y_k \ge 0\}$ and $J = \{k \mid y_k < 0\}$. These two sets are nonempty because y is nonzero and $\sum_k y_k = 0$. Define

$$\lambda = \sum_{k\in I} y_k = -\sum_{k\in J} y_k, \qquad \theta_k = \begin{cases} y_k/\lambda & k \in I \\ -y_k/\lambda & k \in J. \end{cases}$$

The coefficients θ_k are nonnegative and satisfy

$$\sum_{k\in I}\theta_k x_k = \sum_{k\in J}\theta_k x_k, \qquad \sum_{k\in I}\theta_k = \sum_{k\in J}\theta_k = 1.$$

This shows that the point

$$x = \sum_{k\in I}\theta_k x_k = \sum_{k\in J}\theta_k x_k$$

is in the intersection of the convex hulls of $S = \{x_k \mid k \in I\}$ and $T = \{x_k \mid k \in J\}$.

(b) We apply the result of (a) to the points $x_1, \ldots, x_m$ defined in the hint. There exists an index set $I \subseteq \{1, 2, \ldots, m\}$, with $1 \le |I| \le m-1$, and a point

$$x \in \mathbf{conv}\,\{x_k \mid k \in I\} \cap \mathbf{conv}\,\{x_k \mid k \notin I\}.$$

From the definition of the points x_k we see that if $k \in I$, then $x_k \in S_j$ for all $j \notin I$. Therefore

$$x_k \in \bigcap_{j\notin I} S_j \quad \text{for all } k \in I.$$

Since x is a convex combination of the points x_k, $k \in I$, and the intersection of convex sets is convex, it follows that

$$x \in \bigcap_{j\notin I} S_j. \tag{1}$$

Similarly, if $k \notin I$, then $x_k \in S_j$ for all $j \in I$. Therefore

$$x_k \in \bigcap_{j \in I} S_j \quad \text{for all } k \notin I.$$

Since x is also a convex combination of the points x_k, $k \notin I$, we have

$$x \in \bigcap_{j \in I} S_j. \tag{2}$$

Combining (1) and (2), we see that x is in the intersection of all sets $S_1, \ldots, S_m$.

4. *Schur complements and positive semidefinite matrices.* Let X be a symmetric matrix partitioned as

$$X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}. \tag{3}$$

If A is nonsingular, the matrix $S = C - B^T A^{-1} B$ is called the *Schur complement* of A in X. It A is positive definite, then it can be shown that $X \succeq 0$ (X is positive semidefinite) if and only if $S \succeq 0$ (see page 650 of the textbook). In this exercise we prove the extension of this result to singular A mentioned on page 651 of the textbook.

(a) Suppose $A = 0$ in (3). Show that $X \succeq 0$ if and only if $B = 0$ and $C \succeq 0$.

(b) Let A be a symmetric $n \times n$ matrix with eigendecomposition

$$A = Q \Lambda Q^T,$$

where Q is orthogonal ($Q^T Q = Q Q^T = I$) and $\Lambda = \mathbf{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Assume the first r eigenvalues λ_i are nonzero and $\lambda_{r+1} = \cdots = \lambda_n = 0$. Partition Q and Λ as

$$Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}, \qquad \Lambda = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix}$$

with Q_1 of size $n \times r$, Q_2 of size $n \times (n-r)$, and $\Lambda_1 = \mathbf{diag}(\lambda_1, \ldots, \lambda_r)$. The matrix

$$A^\dagger = Q_1 \Lambda_1^{-1} Q_1^T$$

is called the *pseudo-inverse* of A. Verify that

$$A A^\dagger = A^\dagger A = Q_1 Q_1^T, \qquad I - A A^\dagger = I - A^\dagger A = Q_2 Q_2^T.$$

The matrix–vector product $A A^\dagger x = Q_1 Q_1^T x$ is the orthogonal projection of the vector x on the range of A. The matrix–vector product $(I - A A^\dagger) x = Q_2 Q_2^T x$ is the orthogonal projection on the nullspace of A.

(c) Show that the block matrix X in (3) is positive semidefinite if and only if

$$A \succeq 0, \qquad (I - AA^\dagger)B = 0, \qquad C - B^T A^\dagger B \succeq 0.$$

The second condition means that the columns of B are in the range of A.
Hint. Let $A = Q\Lambda Q^T$ be the eigenvalue decomposition of A. Partition Q and Λ as in part (b). The matrix X in (3) is positive semidefinite if and only if the matrix

$$\begin{bmatrix} Q^T & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} \Lambda & Q^T B \\ B^T Q & C \end{bmatrix} = \begin{bmatrix} \Lambda_1 & 0 & Q_1^T B \\ 0 & 0 & Q_2^T B \\ B^T Q_1 & B^T Q_2 & C \end{bmatrix}$$

is positive semidefinite. Using the observation in part (a) we see that this matrix is positive semidefinite if and only if $Q_2^T B = 0$ and the matrix

$$\begin{bmatrix} \Lambda_1 & Q_1^T B \\ B^T Q_1 & C \end{bmatrix}$$

is positive semidefinite. Apply the Schur complement characterization for 2×2 block matrices with a positive definite 1,1 block (page 650 of the textbook) to show the result.

Solution.

(a) Suppose $A = 0$. The matrix X is positive semidefinite if and only if

$$\begin{bmatrix} u \\ v \end{bmatrix}^T \begin{bmatrix} 0 & B \\ B^T & C \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = 2u^T Bv + v^T Cv \geq 0 \quad \text{for all } u,\, v.$$

Clearly, a sufficient condition is that $B = 0$ and C is positive semidefinite. Taking $u = 0$ shows that positive semidefiniteness of C is also necessary. To see that $B = 0$ is necessary, assume $B \neq 0$. Take any v with $Bv \neq 0$ and choose $u = -tBv$. The quadratic form then reduces to

$$2u^T Bv + v^T Cv = -2tv^T B^T Bv + v^T Cv = -2t\|Bv\|_2^2 + v^T Cv,$$

which is negative for sufficiently large t.

(b) We have

$$\begin{aligned} AA^\dagger &= Q_1 \Lambda_1 Q_1^T Q_1 \Lambda_1^{-1} Q_1^T \\ &= Q_1 Q_1^T \\ I - AA^\dagger &= Q_1 Q_1^T + Q_2 Q_2^T - AA^\dagger \\ &= Q_2 Q_2^T. \end{aligned}$$

The proofs of the identities $A^\dagger A = Q_1 Q_1^T$ and $I - A^\dagger A = Q_2 Q_2^T$ are similar.

(c) Every principal submatrix of a positive semidefinite matrix is positive semidefinite. Therefore $A \succeq 0$ is a neccessary condition for X to be positive semidefinite.

Suppose we sort the eigenvalues of A so that its eigenvalue decomposition can be written as

$$A = Q\Lambda Q^T = \left[\begin{array}{cc} Q_1 & Q_2 \end{array}\right] \left[\begin{array}{cc} \Lambda_1 & 0 \\ 0 & 0 \end{array}\right] \left[\begin{array}{cc} Q_1 & Q_2 \end{array}\right]^T$$

with Λ_1 positive diagonal. Following the hint, the question can be reduced to showing that the block matrix

$$\left[\begin{array}{ccc} \Lambda_1 & 0 & Q_1^T B \\ 0 & 0 & Q_2^T B \\ B^T Q_1 & B^T Q_2 & C \end{array}\right]$$

is positive semidefinite. By the result in part (a), the matrix is positive semidefinite if and only if

$$Q_2^T B = 0, \qquad \left[\begin{array}{cc} \Lambda_1 & Q_1^T B \\ B^T Q_1 & C \end{array}\right] \succeq 0.$$

The first condition is equivalent to $(I - AA^\dagger)B = 0$. Since Λ_1 is positive definite, we can apply the Schur complement result for nonsingular 1,1 block to the 2×2 block matrix. This gives the equivalent condition

$$C - B^T Q_1 \Lambda^{-1} Q_1^T B = C - B^T A^\dagger B \succeq 0.$$

5. This problem is an introduction to the software packages CVX (cvxr.com) and CVXPY (cvxpy.org).

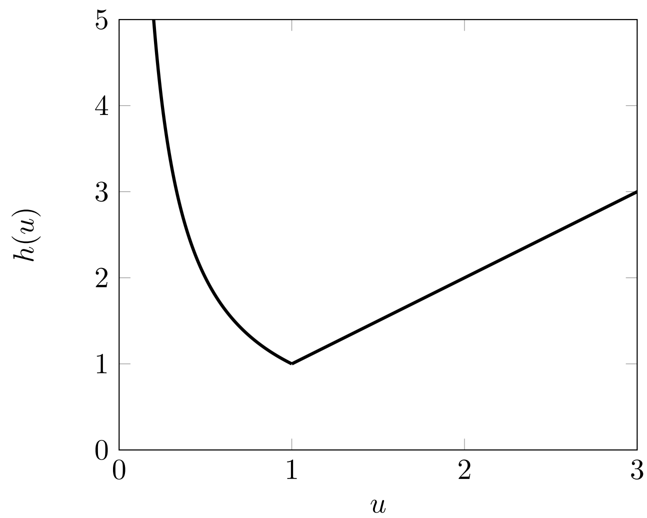
We consider the illumination problem of lecture 1. We take $I_{\text{des}} = 1$ and $p_{\max} = 1$, so the problem is

$$\begin{array}{ll} \text{minimize} & f_0(x) = \max\limits_{k=1,\ldots,m} |\log(a_k^T x)| \\ \text{subject to} & 0 \le x_j \le 1, \quad j = 1, \ldots, n, \end{array} \tag{4}$$

with variable $x \in \mathbf{R}^n$. As mentioned in the lecture, the problem is equivalent to

$$\begin{array}{ll} \text{minimize} & \max\limits_{k=1,\ldots,m} h(a_k^T p) \\ \text{subject to} & 0 \le p_j \le 1, \quad j = 1, \ldots, n, \end{array} \tag{5}$$

where $h(u) = \max\{u, 1/u\}$ for $u > 0$. The function h, shown in the figure below, is nonlinear, nondifferentiable, and convex.



To see the equivalence between (4) and (5), we note that

$$
\begin{aligned}
f_0(x) &= \max_{k=1,\ldots,m} |\log(a_k^T x)| \\
&= \max_{k=1,\ldots,m} \max\{\log(a_k^T x), \log(1/a_k^T x)\} \\
&= \log \max_{k=1,\ldots,m} \max\{a_k^T x, 1/a_k^T x\} \\
&= \log \max_{k=1,\ldots,m} h(a_k^T x),
\end{aligned}
$$

and since the logarithm is a monotonically increasing function, minimizing $f_0(x)$ is equivalent to minimizing $\max_{k=1,\ldots,m} h(a_k^T x)$.

We consider a small example with $n = 10$ lamps and $m = 20$ patches. The $m \times n$ matrix A with rows a_k^T is given in the files `illum_data.m` and `illum_data.py` on the course website (in the folder `Files/Homework/Data files`).

Use the following methods to compute three approximate solutions and the exact solution, and compare the answers (the vectors x and the corresponding values of $f_0(x)$).

(a) *Least squares with saturation.* Solve the least squares problem

$$
\text{minimize} \quad \sum_{k=1}^{m} (a_k^T x - 1)^2 = \|Ax - \mathbf{1}\|_2^2.
$$

If the solution has negative coefficients, set them to zero; if some coefficients are greater than 1, set them to 1.

(b) *Regularized least squares.* Solve the regularized least squares problem

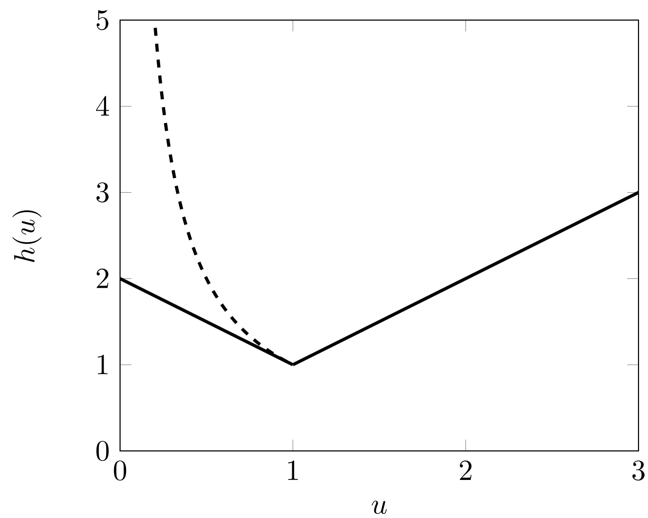
$$
\text{minimize} \quad \sum_{k=1}^{m} (a_k^T x - 1)^2 + \rho \sum_{j=1}^{n} (x_j - 0.5)^2 = \|Ax - \mathbf{1}\|_2^2 + \rho \|x - (1/2)\mathbf{1}\|_2^2,
$$

where $\rho > 0$ is a parameter. Increase ρ until all coefficients of x are in the interval $[0, 1]$.

(c) *Chebyshev approximation.* Solve the problem

$$\begin{array}{ll} \text{minimize} & \max_{k=1,\ldots,m} |a_k^T x - 1| = \|Ax - \mathbf{1}\|_\infty \\ \text{subject to} & 0 \leq x_j \leq 1, \quad j = 1, \ldots, n. \end{array}$$

We can think of this problem as obtained by approximating the nonlinear function $h(u)$ by a piecewise-linear function $|u - 1| + 1$. As shown in the figure below, this is a good approximation around $u = 1$.



(d) *Exact solution.* Solve

$$\begin{array}{ll} \text{minimize} & \max_{k=1,\ldots,m} \max(a_k^T x, 1/a_k^T x) \\ \text{subject to} & 0 \leq x_j \leq 1, \quad j = 1, \ldots, n. \end{array}$$

Use the function `inv_pos` in CVX/CVXPY to express the function $f(u) = 1/u$ with domain $\mathbf{R}_{++}$.

Solution.

(a) *Least squares with saturation.* We compute x as

```
x = A \ ones(n,1).
```

All coefficients of x are outside the feasible interval $[0, 1]$ and need to be rounded.

(b) *Regularized least squares.* We compute x by solving a least squares problem

```
x = [ A; sqrt(rho)*eye(m) ] \ [ ones(n,1); sqrt(rho)*.5*ones(m,1) ].
```

The smallest ρ that gives a feasible p is $\rho = 0.2190$.

(c) *Chebyshev approximation.* We solve this problem using CVX.

```
cvx_begin
   variable x(n)
   minimize (norm(A*x-b, inf))
   subject to
       x >= 0
       x <= 1
cvx_end
```

(d) *Exact solution.*

```
cvx_begin
   variable x(n)
   minimize (max([A*x; inv_pos(A*x)]))
   subject to
       x >= 0
       x <= 1
cvx_end
```

The results are summarized in the following table.

	Saturated LS	Weighted LS	Chebyshev	Exact
x_1	1	0.5004	1	1
x_2	0	0.4778	0.1165	0.2023
x_3	1	0.0833	0	0
x_4	0	0.0000	0	0
x_5	0	0.4561	1	1
x_6	1	0.4354	0	0
x_7	0	0.4598	1	1
x_8	1	0.4307	0.0249	0.1882
x_9	0	0.4034	0	0
x_{10}	1	0.4526	1	1
$f_0(x)$	0.8628	0.4439	0.4198	0.3575