Delft University of Technology
Faculty of Electrical Engineering, Mathematics, and Computer Science
Circuits and Systems Group

EE 4530 APPLIED CONVEX OPTIMIZATION

8 April 2022, 09:00–12:00

Open book. Copies of the book and the course slides allowed. No other tools except a basic pocket calculator permitted.

Answer in English. Make clear in your answer how you reach the final result; the road to the answer is very important. Write your name and student number on each sheet.

Hint: Avoid losing too much time on detailed calculations, write down the general approach first.

Question 1 (10 points)

For each of the following sets or functions, explain if it is convex, concave, or neither convex nor concave. Prove it (by using the definition or some of the basic rules we have encountered in the course).

(a) The function $f(w,x,y,z) = wx + y^2 + z^2$ with $\mathrm{dom} f = \mathbf{R}^4$. *(Hint: The eigenvalues of a block diagonal matrix are the eigenvalues of each individual block.)*

(b) The intersection between the epigraph of $f(x) = x^2 - 1$ and the hypograph of $f(x) = -x^2 + 1$. Draw a sketch.

(c) The spectraplex $\mathcal{S} = \{X \in \mathbb{S}^n_+ | \mathrm{tr}(X) = 1\}$.

(d) The fantope $\mathcal{F} = \{Z \in \mathbb{S}^n : 0 \preceq Z \preceq I, \mathrm{tr}(Z) = k\}$, for some order $k > 0$.

(e) Consider the function $g(y) = \sup_{x\in\mathcal{C}}(y^\top Ay - f(x))$ with $x, y \in \mathbf{R}^n$. Under which conditions on the set $\mathcal{C}$, on the matrix A and on the function $f(\cdot)$ is the function $g(y)$ convex?

(f) The set $\mathcal{V}$ of $m \times n$ Vandermonde matrices. Recall that a $m \times n$ Vandermonde matrix is a matrix of the form $V = [1\ x\ x^2 \ldots x^{n-1}]$, where $x \in \mathbf{R}^m$ and the power is intended entry-wise.

(g) The function $f(x) = 1/(x_1x_2)$, where $x \in \mathbf{R}^2_{++}$.

Solution

(a) The function is not convex, since its Hessian has eigenvalues $\{-1, +1, 2\}$ and thus not positive semidefinite.

(b) It is the intersection of two convex sets, and thus it is convex.

(c) It is convex. It readily follows by applying the definition: given $X, Y \in \mathcal{S}$, then $\mathrm{tr}(\theta X + (1-\theta)Y) = \theta\mathrm{tr}(X) + \mathrm{tr}(Y) - \theta\mathrm{tr}(Y) = 1$.

(d) It is convex. One way to prove is to recognize that $\mathcal{F} = \{Z \in \mathbb{S}^n : Z \succeq 0\} \cap \{Z \in \mathbb{S}^n : Z \preceq I\} \cap \{Z \in \mathbb{S}^n : \mathrm{tr}(Z) = k\}$ is the intersection of linear inequality and equality constraints.

(e) The only requirement for $g(y)$ to be convex is the positive semidefinitedness of matrix A. Indeed it is the pointwise supremum of a quadratic function in y indexed by x. The set $\mathcal{C}$ and the function $f(x)$ can also be not convex.

(f) It is not convex. To show this, take two $m \times n$ Vandermonde matrices $X = [1\ x\ x^2 \ldots x^{n-1}]$ and $Y = [1\ y\ y^2 \ldots y^{n-1}]$. It suffices to show that, for instance, the convex combination $Z = (1/2)X + (1/2)Y$ is not Vandermonde. The second column of Z is the vector $z = (1/2)x + (1/2)y$, and the third column is $(1/2)x^2 + (1/2)y^2$, which is clearly different from z^2.

(g) It is convex and it can be proved through the second-order condition. The Hessian is:

$$\nabla^2 f = \begin{bmatrix} \frac{2}{x_1^3 x_2} & \frac{1}{x_1^2 x_2^2} \\ \frac{1}{x_1^2 x_2^2} & \frac{2}{x_1 x_2^3} \end{bmatrix} \tag{1}$$

and because $x \in \mathbf{R}^2_{++}$, $\frac{2}{x_1^3 x_2} > 0$, $\frac{2}{x_1 x_2^3} > 0$ the determinant $\frac{2}{x_1^3 x_2}\frac{2}{x_1 x_2^3} - \frac{1}{x_1^2 x_2^2}\frac{1}{x_1^2 x_2^2} > 0$. Thus the Hessian is PSD and we conclude that $f(x)$ is convex.

Question 2 (10 points)

Consider the following unconstrained optimization problem:

$$\underset{x\in\mathbf{R}^2}{\text{minimize}} \;\; f(x) = \max\left((x_1-1)^2 + x_2^2 \;,\; (x_1+1)^2 + x_2^2\right). \tag{2}$$

(a) Is this objective function convex? Is it differentiable?

(b) Sketch the contour plot of the function! Find a point with more than one subgradient and indicate at least 2 subgradients in that point on the sketch!

(c) Write down the equations for the basic negative subgradient step and compute $x(1)$ for $x(0) = (0, 2)$ and a constant step size of $\alpha = 1$.

(d) Does the objective function decrease?

(e) Is the method in (c) guaranteed to converge to the optimal solution? If not, what can you do to improve convergence?

Solution

(a) Convex, but not differentiable.

(b) See Figure 1 below.

(c) General subgradient step:

$$x^{(k+1)} = x^{(k)} - \alpha_k g^{(k)}$$

Note that $f(x)$ is a pointwise maximum of 2 functions and at $x^{(0)} = (0, 2)$ both functions achieve the maximum value. Therefore, we can choose any gradient of any of the 2 functions at this point. Let's choose the gradient of the second function, i.e., let $g(x) = (2x_1 + 2, 2x_2)$. Then, our subgradient step is

$$x^{(k+1)} = x^{(k)} - 1 \cdot (2x_1 + 2 \;,\; 2x_2)$$

Then,

$$\begin{aligned} x^{(1)} &= x^{(0)} - \left(2x_1^{(0)} + 2 \;,\; 2x_2^{(0)}\right) = \\ &= (0 \;,\; 2) - (2 \cdot 0 + 2 \;,\; 2 \cdot 2) = (-2 \;,\; -2) \end{aligned}$$

(d) No, the objective function value increases in this case:
$f(0) = \max\left((0-1)^2 + 2^2 \;,\; (0+1)^2 + 2^2\right) = 5$
$f(1) = \max\left((-2-1)^2 + -2^2 \;,\; (-2+1)^2 + -2^2\right) = 13$

(e) It is not. To guarantee convergence, we need to choose a different step size rule, e.g. the diminishing step size rule.

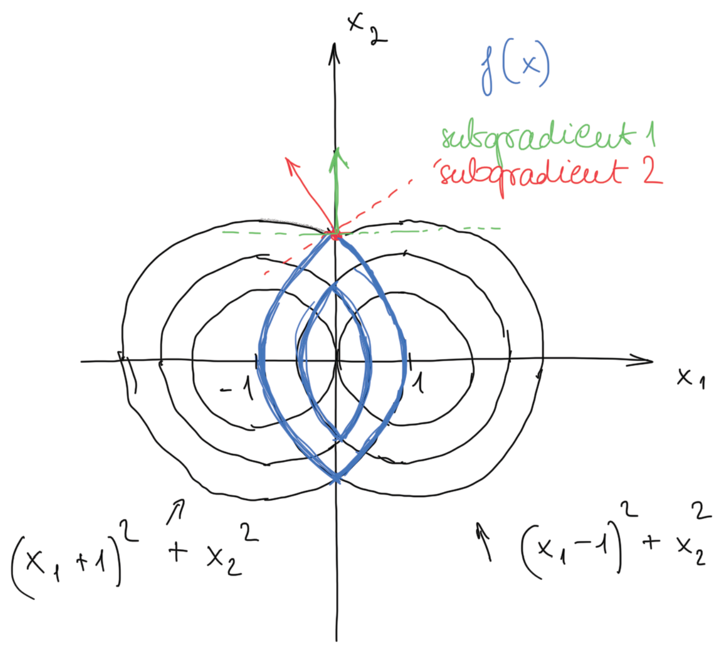

Figure 1: Figure 1

Question 3 (8 points)

Suppose we want to solve a set of noisy equations in the form

$$y = Ax + e,$$

where $A \in \mathbf{R}^{m \times n}$, $y, e \in \mathbf{R}^m$ and $x \in \mathbf{R}^n$. Here y and A are known, x is unknown, and e is unknown white Gaussian noise with zero mean and variance σ^2. The standard way to approach this problem is to minimize the least squares cost

$$\|y - Ax\|^2$$

over x. However, when we have less equations than unknowns, i.e., $m < n$, we need additional constraints to find a unique solution. Let us here for instance assume that the sum of the variables x_i is 1 and that all the variables x_i are positive.

(a) Formulate the overall problem. Is this problem convex? Why or why not?

(b) Express the Lagrangian and present the KKT conditions for this problem.

(c) Derive the dual function and formulate the dual problem.

(d) Does strong duality always hold for this problem? Why or why not? If not, give a sufficient condition for strong duality to hold.

Solution

(a) The overall problem is given by

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & \|y-Ax\|^2 \\ \text{subject to} & 1^T x = 1, \quad x \succeq 0. \end{array}$$

Since the cost function can be written as $x^T A^T A x - 2x^T A^T y + y^T y$, it is quadratic with a positive semi-definite Hessian. So the cost function is convex. Furthermore, since the constraint functions are affine, the overall problem is convex.

(b) The Lagrangian is given by

$$\begin{aligned} L(x,\nu,\lambda) &= x^T A^T A x - 2x^T A^T y + y^T y + \nu(1^T x - 1) - \lambda^T x \\ &= x^T A^T A x + x^T(\nu 1 - \lambda - 2A^T y) + y^T y - \nu. \end{aligned}$$

The KKT conditions are

- Primal constraints: $1^T x = 1$, $x \succeq 0$.
- Dual constraints: $\lambda \succeq 0$.
- Complementary slackness: $\lambda_i x_i = 0$, $\forall \in \{1,\ldots,n\}$.
- Vanishing gradient of Laplacian: $2A^T A x + \nu 1 - \lambda - 2A^T y = 0$.

(c) The Lagrangian is unbounded below if $A^T A$ is rank deficient and $\nu 1 - \lambda - 2A^T y \notin \mathcal{R}(A^T)$. If on the other hand, $\nu 1 - \lambda - 2A^T y \in \mathcal{R}(A^T)$, then the solution for minimizing the Lagrangian is given by

$$x = -\frac{1}{2}(A^T A)^\dagger(\nu 1 - \lambda - 2A^T y),$$

and plugging this into the Lagrangian we obtain

$$g(\nu,\lambda) = -\frac{1}{4}(\nu 1 - \lambda - 2A^T y)^T (A^T A)^\dagger (\nu 1 - \lambda - 2A^T y) + y^T y - \nu.$$

The dual problem then becomes

$$\begin{array}{ll} \underset{\nu\in\mathbb{R},\lambda\in\mathbb{R}^n}{\text{maximize}} & -\frac{1}{4}(\nu 1 - \lambda - 2A^T y)^T (A^T A)^\dagger (\nu 1 - \lambda - 2A^T y) + y^T y - \nu \\ \text{subject to} & \nu 1 - \lambda - 2A^T y \in \mathcal{R}(A^T), \quad \lambda \succeq 0. \end{array}$$

(d) Strong duality always holds here since we can find a feasible point $x \succ 0$ for which $1^T x = 1$. Consider for instance $x_i = 1/n$.

Question 4 (12 points)

Let us explore *optimal portfolio selection.* Consider an investor who wants to allocate his/her wealth among n assets. Each one of these assets offers random rates of return $\{r_i\}_{i=1}^n$ with the following *known* second-order statistics

$$\mathbb{E}\{r_i\} = \bar{r}_i, \quad \mathbb{E}\{(r_i - \bar{r}_i)(r_j - \bar{r}_j)\} = Q_{i,j}, \quad \forall\, i, j \in \{1, \ldots, n\}.$$

Let us also define the $n \times 1$ mean vector $\bar{r}$ with $[\bar{r}]_i = \bar{r}_i$ and the $n \times n$ covariance matrix Q with $[Q]_{i,j} = Q_{i,j}$. Throughout this question, we assume that Q is invertible, i.e., Q^{-1} exists.

(a) Consider that $x_i = [x]_i$ is the amount invested in the ith asset, i.e., $x \in \mathbf{R}^n$ is the overall investment vector. Then express the mean, $\bar{y}$, and the variance, σ^2, of the total return of the investment, y, as a function of x, $\bar{r}$ and Q. *Hint: The expression of the total return (investment) is the sum of all the asset returns.*

A moderate investor (risk-adverse) usually tends to minimize his/her risk. Therefore, the investor's problem becomes one of reducing the variability of his/her return, while guaranteeing a given level of expected return, i.e., $\bar{y} = m$.

(b) Assuming that the investor possesses a unit of wealth, i.e., $1^T x = 1$, write an optimization problem that minimizes the risk over x, and guarantees a mean return of m.

(c) Is the optimization problem obtained in (b) convex? Why or why not?

Let us now study how the solution to (b) varies with the mean return m.

(d) Write the Lagrangian for problem (b) and express the optimal investment x^* using the first order optimality condition. *Hint: First order optimality implies a vanishing Lagrangian gradient.*

(e) Using the constraints of (b) and the expression of x^* from (d), show that the Lagrange multipliers are given in the form

$$\lambda_1 = \eta_1 + \zeta_1 m, \tag{3}$$
$$\lambda_2 = \eta_2 + \zeta_2 m. \tag{4}$$

Hint: There is no need to find the exact expressions for the scalars $\eta_1, \zeta_1, \eta_2, \zeta_2$. You may also need the expression of the inverse of a 2×2 matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{5}$$

(f) Using the Langrange multipliers expression from (e), show that the risk, as a function of m, is given by

$$\sigma^2 = (\alpha m + \beta)^2 + \kappa, \tag{6}$$

where α, β and κ are scalars that depend on Q and $\bar{r}$. *Hint: First show the optimal x can be written as $x^* = mv + w$, for some vectors v and w depending on Q and $\bar{r}$.*

Consider now the case of *K-unit-commitment*, i.e., a set of K assets are selected and the unit of wealth is distributed equally among the assets.

(g) Write the new (possibly nonconvex) optimization problem and, if needed, a *convex* relaxation for this instance.

Solution

(a) The total investment is given by

$$y = \sum_{i=1}^{n} r_i x_i = r^T x. \tag{7}$$

Therefore, its mean and variance is given as

$$\bar{y} := \mathbb{E}\{y\} = \sum_{i=1}^{n} \bar{r}_i x_i = \bar{r}^T x, \tag{8}$$

$$\sigma^2 := \mathbb{E}\{(y-\bar{y})^2\} = \mathbb{E}\left\{\left(\sum_{i=1}^{n} (r_i - \bar{r}_i) x_i\right)^2\right\} = x^T Q x. \tag{9}$$

(b) The following optimization problem minimizes the risk of the investor and guarantees a given mean (expected) return

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \;\; x^T Q x \tag{10}$$

$$\text{subject to} \;\; 1^T x = 1, \;\; \bar{r}^T x = m. \tag{11}$$

(c) The problem is *always* convex. First, notice that the cost function is a quadratic form of Q. As Q is a covariance matrix, its positive-definiteness follows directly. Hence, the cost function is convex. As the equality constraints are affine functions, the convexity of the feasible region is guaranteed. Thus, the problem is convex.

(d) The Lagrangian of (b) is

$$L(x, \lambda_1, \lambda_2) = x^T Q x + \lambda_1 (1^T x - 1) + \lambda_2 (\bar{r}^T x - m). \tag{12}$$

The first order optimality condition is

$$\nabla_x L(x, \lambda_1, \lambda_2) = 0, \tag{13}$$

which leads to

$$\nabla L(x^*, \lambda_1, \lambda_2) = 2Qx^* + \lambda_1 1 + \lambda_2 \bar{r} = 0, \tag{14}$$

and

$$x^* = -\frac{1}{2} Q^{-1} (\lambda_1 1 + \lambda_2 \bar{r}). \tag{15}$$

(e) Substituting the constraints $1^T x = 1$ and $\bar{r}^T x = m$, we obtain

$$1^T x^* = 1 = a\lambda_1 + b\lambda_2 \tag{16}$$

$$\bar{r}^T x^* = m = c\lambda_1 + d\lambda_2, \tag{17}$$

for some values a, b, c, d. The previous relation leads to the system

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 1 \\ m \end{bmatrix}, \tag{18}$$

which has solution

$$\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} 1 \\ m \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}. \tag{19}$$

Hence, we can express the multipliers as

$$\lambda_1 = \eta_1 + \zeta_1 m, \tag{20}$$

$$\lambda_2 = \eta_2 + \zeta_2 m, \tag{21}$$

with $\eta_1 = d/\gamma$, $\zeta_1 = -bm/\gamma$, $\eta_2 = -c/\gamma$ and $\zeta_2 = am/\gamma$. Here, $\gamma = 1/(ad - bc)$.

(f) Recalling the solution for x^* in terms of the multipliers

$$x^* = -\frac{1}{2}Q^{-1}(\lambda_1 1 + \lambda_2 \bar{r}), \tag{22}$$

and the expressions for λ_1 and λ_2, we see that we can express the solution x^* as

$$x^* = mv + w, \tag{23}$$

for some vectors v and w that depend on $\bar{r}$ and Q. Hence, substituting this expression in the cost, we obtain

$$(mv^T + w^T)Q(mv + w) = (\alpha m + \beta)^2 + \kappa, \tag{24}$$

for appropriates values of α, β and κ.

(g) This setting can be tackled by considering x as a *Boolean* variable, i.e., $x \in \{0,1\}^n$ and modifying the constraint $1^T x = 1$ to $1^T x = K$, for a given K.

So, we write the new (nonconvex) optimization problem as

$$\underset{x\in\{0,1\}^n}{\text{minimize}} \;\; x^T Q x \tag{25}$$

$$\text{subject to} \;\; 1^T x = K, \;\; \bar{r}^T x = m. \tag{26}$$

As the optimization variable is Boolean, the problem is nonconvex. A straightforward way to convexify the problem is to relax $x \in \{0,1\}^n$ to $x \in [0,1]^n$. As the rest of the contraints are affine and the cost is convex, nothing more has to be done.