Exercises

Portfolio Optimization: Theory and Application
Appendix A – Convex Optimization Theory

Daniel P. Palomar (2025). *Portfolio Optimization: Theory and Application.* Cambridge University Press.

portfoliooptimizationbook.com

Exercise A.1: Concepts on convexity

a. Define a convex set and provide an example.
b. Define a convex function and provide an example.
c. Explain the concept of convex optimization problems and provide an example.
d. What is the difference between active and inactive constraints in an optimization problem?
e. What is the difference between a locally optimal point and a globally optimal point?
f. Define a feasibility problem and provide an example.
g. Explain the concept of least squares problems and provide an example.
h. Explain the concept of linear programming and provide an example.
i. Explain the concept of nonconvex optimization and provide an example.
j. Explain the difference between a convex and a nonconvex optimization problem.

Solution

a. A convex set is a set where for any two points in the set, the line connecting them is also contained within the set. An example of a convex set is a closed interval on the real line, such as $[a, b]$.

b. A convex function is a function where for any two points in its domain, the line connecting them lies above or on the graph of the function. An example of a convex function is $f(x) = x^2$.

c. Convex optimization problems are optimization problems where the objective function is convex and the constraints are convex sets. An example of a convex optimization problem is minimizing $f(x) = x^2 + 1$ subject to $x \leq 2$.

d. Active constraints are constraints that are satisfied with equality at the optimal solution, while inactive constraints are satisfied with inequality. In other words, active constraints play a role in determining the optimal solution, while inactive constraints do not.

e. A locally optimal point is a point that is optimal within a neighborhood, meaning that there is no other feasible point nearby that has a better objective value. A globally optimal point is a point that is optimal over the entire feasible set, meaning that there is no other feasible point with a better objective value.

f. The feasibility problem is to find any feasible solutions for an optimization problem without regard to the objective value. An example of a feasibility problem is finding a feasible solution to a system of linear equations.

g. Least squares problems are optimization problems where the objective is to minimize the sum of squared differences between observed data and a mathematical model. An example of a least squares problem is fitting a line to a set of data points using the method of least squares.

h. Linear programming is a type of optimization problem where the objective function and constraints are all linear. An example of a linear programming problem is maximizing profit by determining the optimal production levels of different products subject to resource constraints.

i. Nonconvex optimization refers to optimization problems where the objective function or constraints are not convex. An example of a nonconvex optimization problem is minimizing a function with multiple local minima.

j. The main difference between a convex and a nonconvex optimization problem is the nature of the objective function and constraints. In a convex optimization problem, the objective function and constraints are convex, allowing for efficient solution methods and guaranteeing global optimality. In a nonconvex optimization problem, the objective function or constraints (or both) are not convex, making the problem more challenging to solve and potentially leading to multiple local optima.

Exercise A.2: Convexity of sets

Determine the convexity of the following sets:

a. $\mathcal{X} = \left\{x \in \mathbb{R} \mid x^2 - 3x + 2 \geq 0\right\}$.
b. $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \max\{x_1, x_2, \ldots, x_n\} \leq 1\}$.
c. $\mathcal{X} = \left\{\boldsymbol{x} \in \mathbb{R}^n \mid \alpha \leq \boldsymbol{c}^\mathsf{T}\boldsymbol{x} \leq \beta\right\}$.
d. $\mathcal{X} = \left\{\boldsymbol{x} \in \mathbb{R}^2 \mid x_1 \geq 1,\ x_2 \geq 2,\ x_1x_2 \geq 1\right\}$.
e. $\mathcal{X} = \left\{(x, y) \in \mathbb{R}^2 \mid y \geq x^2\right\}$.
f. $\mathcal{X} = \left\{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x} - \boldsymbol{c}\| \leq \boldsymbol{a}^\mathsf{T}\boldsymbol{x} + b\right\}$.

g. $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n \mid (\boldsymbol{a}^\top \boldsymbol{x} + b)/(\boldsymbol{c}^\top \boldsymbol{x} + d) \geq 1,\ \boldsymbol{c}^\top \boldsymbol{x} + d \geq 1\}$.
h. $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}^\top \boldsymbol{x} \geq b \text{ or } \|\boldsymbol{x} - \boldsymbol{c}\| \leq 1\}$.
i. $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x}^\top \boldsymbol{y} \leq 1 \text{ for all } \boldsymbol{y} \in \mathcal{S}\}$, where $\mathcal{S}$ is an arbitrary set.

Solution

a. This set is convex as the feasible set is $\mathcal{X} = \{x \in \mathbb{R} \mid x \in [1, 2]\}$.

b. This set is convex. For any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ and $0 \leq \theta \leq 1$, $\boldsymbol{z} = \theta \boldsymbol{x} + (1 - \theta)\boldsymbol{y}$, we have $z_i <= \max(x_i, y_i) \leq 1, \forall i$.

c. This set is convex. It is an intersection of convex halfspaces.

d. This set is convex. It is an intersection of three sets. The first two are convex halfspaces. The last one is a convex set. Therefore, the intersection of the three convex sets is convex.

e. This set is convex. It is a norm cone.

f. This set is convex. For any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ and $0 \leq \theta \leq 1$, we have

$$\begin{aligned}
\|\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y} - \boldsymbol{c}\| \quad &= \|\theta(\boldsymbol{x} - \boldsymbol{c}) + (1-\theta)(\boldsymbol{y} - \boldsymbol{c})\| \\
&\leq \theta\|\boldsymbol{x} - \boldsymbol{c}\| + (1-\theta)\|\boldsymbol{y} - \boldsymbol{c}\| \\
&\leq \theta(\boldsymbol{a}^\top \boldsymbol{x} + b) + (1-\theta)(\boldsymbol{a}^\top \boldsymbol{y} + b) \\
&= \boldsymbol{a}^\top(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) + b
\end{aligned}$$

g. This set is convex. It can be written as $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n \mid (\boldsymbol{a}^\top \boldsymbol{x} + b) \geq (\boldsymbol{c}^\top \boldsymbol{x} + d),\ \boldsymbol{c}^\top \boldsymbol{x} + d \geq 1\}$, which is the intersection of two halfspaces.

h. This set is not convex. It is an union of a halfspace and a norm ball. A special case is $\mathcal{X} = \{x \in \mathbb{R} \mid x \geq 2 \text{ or } |x| \leq 1\}$. The two sets are disconnected, and thus the union cannot be convex.

i. This set is convex. It is the intersection of a batch of convex halfspaces.

Exercise A.3: Convexity of functions

Determine the convexity of the following functions:

a. $f(\boldsymbol{x}) = \alpha g(\boldsymbol{x}) + \beta$, where g is a convex function, and α and β are scalars with $\alpha > 0$.
b. $f(\boldsymbol{x}) = \|\boldsymbol{x}\|^p$ with $p \geq 1$.
c. $f(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2$.
d. The difference between the maximum and minimum value of a polynomial on a given interval,

as a function of its coefficients:

$$f(\boldsymbol{x}) = \sup_{t\in[0,1]} p_{\boldsymbol{x}}(t) - \inf_{t\in[0,1]} p_{\boldsymbol{x}}(t),$$

where $p_{\boldsymbol{x}}(t) = x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1}$.

e. $f(\boldsymbol{x}) = \boldsymbol{x}^\mathsf{T}\boldsymbol{Y}^{-1}\boldsymbol{x}$ (with $\boldsymbol{Y} \succ \boldsymbol{0}$).

f. $f(\boldsymbol{Y}) = \boldsymbol{x}^\mathsf{T}\boldsymbol{Y}^{-1}\boldsymbol{x}$ (with $\boldsymbol{Y} \succ \boldsymbol{0}$).

g. $f(\boldsymbol{x},\boldsymbol{Y}) = \boldsymbol{x}^\mathsf{T}\boldsymbol{Y}^{-1}\boldsymbol{x}$ (with $\boldsymbol{Y} \succ \boldsymbol{0}$). Hint: Use the Schur complement.

h. $f(\boldsymbol{x}) = \sqrt{\sqrt{\boldsymbol{a}^\mathsf{T}\boldsymbol{x} + b}}$.

i. $f(\boldsymbol{X}) = \log\det(\boldsymbol{X})$ on $\mathbb{S}^n_{++}$.

j. $f(\boldsymbol{X}) = \det(\boldsymbol{X})^{1/n}$ on $\mathbb{S}^n_{+}$.

k. $f(\boldsymbol{X}) = \mathrm{Tr}\left(\boldsymbol{X}^{-1}\right)$ on $\mathbb{S}^n_{++}$.

l. $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{x} - \boldsymbol{b}^\mathsf{T}\log(\boldsymbol{x})$, where $\boldsymbol{\Sigma} \succ \boldsymbol{0}$ and the log function is applied elementwise.

Solution

a. f can be viewed as a composition $u(g(\boldsymbol{x}))$ of the scalar function $u(t) = \alpha t + \beta, t \in \mathbb{R}$, and the function $g(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^n$. In this case, u is convex and monotonically increasing over $\mathbb{R}$ (since $\alpha > 0$), while g is convex over $\mathbb{R}^n$. Hence, f is convex over $\mathbb{R}^n$.

b. f can be viewed as a composition $g(h(\boldsymbol{x}))$ of the scalar function $g(t) = t^p$, $p \geq 1$ and the function $h(\boldsymbol{x}) = \|\boldsymbol{x}\|$. In this case, g is convex and monotonically increasing over the nonnegative octant, which is the set of values that h can take,while h is convex over $\mathbb{R}^n$ (since any vector norm is convex). Hence f is convex over $\mathbb{R}^n$.

c. Since $\nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A}^\mathsf{T}\boldsymbol{A} \succeq 0$, f is convex.

d. $p_{\boldsymbol{x}}(t)$ is an affine function of $\boldsymbol{x}$. Therefore $\sup_{t\in[a,b]} p_{\boldsymbol{x}}(t)$ is convex in $\boldsymbol{x}$. Similarly, $\inf_{t\in[a,b]} p(t)$ is concave in $\boldsymbol{x}$. Therefore $f(\boldsymbol{x})$ is convex in $\boldsymbol{x}$.

e. The function $f(\boldsymbol{Y}) = \boldsymbol{x}^\mathsf{T}\boldsymbol{Y}^{-1}\boldsymbol{x}$ is convex. To see this, suppose $\boldsymbol{Y} = \boldsymbol{Z} + t\boldsymbol{V}$, where $\boldsymbol{Z} \succ \boldsymbol{0}$, $\boldsymbol{V} \in \mathbb{S}^n$. We define $g(t) = \boldsymbol{x}^\mathsf{T}\left(\boldsymbol{Z} + t\boldsymbol{V}\right)^{-1}\boldsymbol{x} = \mathrm{Tr}\left(\boldsymbol{X}\left(\boldsymbol{Z} + t\boldsymbol{V}\right)^{-1}\right)$ with $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^\mathsf{T}$.

$$\begin{aligned}
g(t) &= \mathrm{Tr}\left(\boldsymbol{X}\left(\boldsymbol{Z} + t\boldsymbol{V}\right)^{-1}\right)\\
&= \mathrm{Tr}\left(\boldsymbol{Z}^{-1/2}\boldsymbol{X}\boldsymbol{Z}^{-1/2}\left(\boldsymbol{I} + t\boldsymbol{Z}^{-1/2}\boldsymbol{V}\boldsymbol{Z}^{-1/2}\right)^{-1}\right)\\
&= \mathrm{Tr}\left(\boldsymbol{Z}^{-1/2}\boldsymbol{X}\boldsymbol{Z}^{-1/2}\boldsymbol{Q}\left(\boldsymbol{I} + t\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{Q}^\mathsf{T}\right)\\
&= \mathrm{Tr}\left(\boldsymbol{Q}^\mathsf{T}\boldsymbol{Z}^{-1/2}\boldsymbol{X}\boldsymbol{Z}^{-1/2}\boldsymbol{Q}\left(\boldsymbol{I} + t\boldsymbol{\Lambda}\right)^{-1}\right)\\
&= \textstyle\sum_{i=1}^n \left(\boldsymbol{Q}^\mathsf{T}\boldsymbol{Z}^{-1/2}\boldsymbol{X}\boldsymbol{Z}^{-1/2}\boldsymbol{Q}\right)_{ii}\left(1 + t\lambda_i\right)^{-1},
\end{aligned}$$

where we used the eigenvalue decomposition $\boldsymbol{Z}^{-1/2}\boldsymbol{V}\boldsymbol{Z}^{-1/2} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^\mathsf{T}$. Since $\boldsymbol{Q}^\mathsf{T}\boldsymbol{Z}^{-1/2}\boldsymbol{X}\boldsymbol{Z}^{-1/2}\boldsymbol{Q}$ is symmetric and positive semidefinite, $\left(\boldsymbol{Q}^\mathsf{T}\boldsymbol{Z}^{-1/2}\boldsymbol{X}\boldsymbol{Z}^{-1/2}\boldsymbol{Q}\right)_{ii} \geq 0$, $i = 1, \ldots, n$. The function g is thus a nonnegative weighted sum of convex functions $1/(1 + t\lambda_i)$, hence it is convex.

f. The function $f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{Y}^{-1}\boldsymbol{x}$ is convex since it is quadratic with $\boldsymbol{Y}^{-1} \succ \boldsymbol{0}$.

g. The function $f(\boldsymbol{x}, \boldsymbol{Y}) = \boldsymbol{x}^\top \boldsymbol{Y}^{-1}\boldsymbol{x}$ is convex. The epigraph of f is

$$\text{epi}f = \{(\boldsymbol{x}, \boldsymbol{Y}, t) \mid \boldsymbol{Y} \succ \boldsymbol{0},\ \boldsymbol{x}^\top \boldsymbol{Y}^{-1}\boldsymbol{x} \leq t\}.$$

Using the Schur complement condition for positive semidefiniteness of a block matrix,

$$\text{epi}f = \left\{(\boldsymbol{x}, \boldsymbol{Y}, t) \mid \boldsymbol{Y} \succ \boldsymbol{0},\ \begin{bmatrix} \boldsymbol{Y} & \boldsymbol{x} \\ \boldsymbol{x}^\top & t \end{bmatrix} \succeq \boldsymbol{0}\right\}.$$

Since the condition is a linear matrix inequality in $(\boldsymbol{x}, \boldsymbol{Y}, t)$, epi$f$ is convex, and the function f is also convex. Since $f(\boldsymbol{x}, \boldsymbol{Y}) = \boldsymbol{x}^\top \boldsymbol{Y}^{-1}\boldsymbol{x}$ is jointly convex on $\boldsymbol{x}$ and $\boldsymbol{Y}$, it is convex in each variable separately. Thus, from the convexity of $f(\boldsymbol{x}, \boldsymbol{Y})$ we can also infer the convexity of (5) and (6).

h. Since $\boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{b}$ is affine and $\sqrt{\sqrt{\cdot}}$ is a concave function, it follows that $\sqrt{\sqrt{\boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{b}}}$ is concave.

i. The concavity of logdet$(\boldsymbol{X})$ on $\mathbb{S}^n_{++}$ can be verified by restricting to an arbitrary line. Suppose $\boldsymbol{X} = \boldsymbol{Z} + t\boldsymbol{V}$ where $\boldsymbol{Z}, \boldsymbol{V} \in \mathbb{S}^n$. We define $g(t) = \text{logdet}(\boldsymbol{Z} + t\boldsymbol{V})$ and restrict g to the values of t for which $\boldsymbol{Z} + t\boldsymbol{V} \succ 0$. Without loss of generality, we assume $t = 0$ is inside this interval, i.e., $\boldsymbol{Z} \succ \boldsymbol{0}$. Then

$$\begin{aligned} g(t) &= \text{logdet}\left(\boldsymbol{Z} + t\boldsymbol{V}\right) \\ &= \text{logdet}\left(\boldsymbol{Z}^{1/2}\left(\boldsymbol{I} + t\boldsymbol{Z}^{-1/2}\boldsymbol{V}\boldsymbol{Z}^{-1/2}\right)\boldsymbol{Z}^{1/2}\right) \\ &= \textstyle\sum_{i=1}^n \log(1 + t\lambda_i) + \text{logdet}(\boldsymbol{Z}) \end{aligned}$$

where $\lambda_1, ..., \lambda_n$ are the eigenvalues of $\boldsymbol{Z}^{-1/2}\boldsymbol{V}\boldsymbol{Z}^{-1/2}$. It can be easily checked that $g''(t) = -\sum_{i=1}^n \frac{\lambda_i^2}{(1+t\lambda_i)^2} \leq 0$ and we conclude that logdet$(\boldsymbol{X})$ is concave on $\mathbb{S}^n_{++}$

j. Define $g(t) = f(\boldsymbol{Z} + t\boldsymbol{V})$, where $\boldsymbol{Z} \succ 0$ and$\boldsymbol{V} \in \mathbb{S}^n$.

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

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\boldsymbol{Z}^{-1/2}\boldsymbol{V}\boldsymbol{Z}^{-1/2}$. From the last equality we see that g is a concave function of t on $\{t|\boldsymbol{Z} + t\boldsymbol{V} \succ 0\}$, since $\det(\boldsymbol{Z}) > 0$ and the geometric mean $(\prod_{i=1}^n x_i)^{1/n}$ is concave on $\mathbb{R}^n_{++}$.

k. Define $g\left(t\right) = f\left(\boldsymbol{Z} + t\boldsymbol{V}\right)$, where $\boldsymbol{Z} \succ \boldsymbol{0}$ and $\boldsymbol{V} \in \mathbb{S}^n$,

$$\begin{aligned} g\left(t\right) &= \text{Tr}\left(\left(\boldsymbol{Z} + t\boldsymbol{V}\right)^{-1}\right) \\ &= \text{Tr}\left(\boldsymbol{Z}^{-1}\left(\boldsymbol{I} + t\boldsymbol{Z}^{-1/2}\boldsymbol{V}\boldsymbol{Z}^{-1/2}\right)^{-1}\right) \\ &= \text{Tr}\left(\boldsymbol{Z}^{-1}\boldsymbol{Q}\left(\boldsymbol{I} + t\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{Q}^\top\right) \\ &= \textstyle\sum_{i=1}^n \left(\boldsymbol{Q}^\top \boldsymbol{Z}^{-1}\boldsymbol{Q}\right)_{ii}\left(1 + t\lambda_i\right)^{-1}, \end{aligned}$$

where $\boldsymbol{Z}^{-1/2}\boldsymbol{V}\boldsymbol{Z}^{-1/2} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^\mathsf{T}$. In the last equality, g is expressed as a positive weighted sum of convex functions $(1+t\lambda_i)^{-1}$, hence it is convex.

l. Not convex in general. For example, let $x \in \mathbb{R}$ with $\boldsymbol{\Sigma} = 1$ and $b = -1$, it's clear that $f(x) = x^2 + \log(x)$ is not convex.

Exercise A.4: Reformulation of problems

a. Rewrite the following optimization problem as an LP (assuming $d > \|\boldsymbol{c}\|_1$):

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \dfrac{\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_1}{\boldsymbol{c}^\mathsf{T}\boldsymbol{x} + d} \\ \text{subject to} & \|\boldsymbol{x}\|_\infty \leq 1. \end{array}$$

b. Rewrite the following optimization problem as an LP:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \frac{\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_1}{1 - \|\boldsymbol{x}\|_\infty}.$$

c. Rewrite the following constraint as an SOC constraint:

$$\left\{(\boldsymbol{x}, y, z) \in \mathbb{R}^{n+2} \mid \|\boldsymbol{x}\|^2 \leq yz, y \geq 0, z \geq 0\right\}.$$

Hint: You may need the equality $yz = \frac{1}{4}\left((y+z)^2 - (y-z)^2\right)$.

d. Rewrite the following problem as an SOCP:

$$\begin{array}{ll} \underset{\boldsymbol{x}, y\geq 0, z \geq 0}{\text{minimize}} & \boldsymbol{a}^\mathsf{T}\boldsymbol{x} + \kappa\sqrt{\boldsymbol{x}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{x}} \\ \text{subject to} & \|\boldsymbol{x}\|^2 \leq yz, \end{array}$$

where $\boldsymbol{\Sigma} \succeq \boldsymbol{0}$.

e. Rewrite the following problem as an SOCP:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{x}^\mathsf{T}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{a}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} & \boldsymbol{B}\boldsymbol{x} \leq \boldsymbol{b}, \end{array}$$

where $\boldsymbol{A} \succeq \boldsymbol{0}$.

f. Rewrite the following problem as an SDP:

$$\underset{\boldsymbol{X} \succeq \boldsymbol{0}}{\text{minimize}} \quad \text{Tr}\left((\boldsymbol{I} + \boldsymbol{X})^{-1}\right) + \text{Tr}\left(\boldsymbol{A}\boldsymbol{X}\right).$$

Solution

a. First we show the original problem and the following one

$$\begin{array}{ll} \underset{\boldsymbol{y},t}{\text{minimize}} & \|\boldsymbol{A}\boldsymbol{y}-\boldsymbol{b}t\|_1 \\ \text{subject to} & \|\boldsymbol{y}\|_\infty \leq t, \\ & \boldsymbol{c}^\mathsf{T}\boldsymbol{y}+dt=1 \end{array}$$

is equivalent. Define $\boldsymbol{y} = \boldsymbol{x}/\left(\boldsymbol{c}^\mathsf{T}\boldsymbol{x}+d\right)$ and $t = 1/\left(\boldsymbol{c}^\mathsf{T}\boldsymbol{x}+d\right)$. Then $\|\boldsymbol{A}\boldsymbol{y}-\boldsymbol{b}t\|_1 = \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_1/\left(\boldsymbol{c}^\mathsf{T}\boldsymbol{x}+d\right)$, $\|\boldsymbol{y}\|_\infty = t\|\boldsymbol{x}\|_\infty \leq t$, and $\boldsymbol{c}^\mathsf{T}\boldsymbol{y}+dt=1$. Conversely, suppose $\boldsymbol{y}$ and t are feasible for the above problem, then $t>0$. Define $\boldsymbol{x}=\boldsymbol{y}/t$, then we have $\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_1/\left(\boldsymbol{c}^\mathsf{T}\boldsymbol{x}+d\right) = \|\boldsymbol{A}\boldsymbol{y}-\boldsymbol{b}t\|_1$ and $\|\boldsymbol{x}\|_\infty \leq 1$. The above problem is equivalent to the following LP:

$$\begin{array}{ll} \underset{\boldsymbol{y},t,\boldsymbol{s}}{\text{minimize}} & \boldsymbol{s}^\mathsf{T}\mathbf{1} \\ \text{subject to} & -t\mathbf{1} \leq \boldsymbol{y} \leq t\mathbf{1}, \\ & -\boldsymbol{s} \leq \boldsymbol{A}\boldsymbol{y}-\boldsymbol{b}t \leq \boldsymbol{s} \\ & \boldsymbol{c}^\mathsf{T}\boldsymbol{y}+dt=1. \end{array}$$

b. We first note that by introducing an auxiliary scalar variable t we can formulate the problem as

$$\begin{array}{ll} \underset{\boldsymbol{x},t}{\text{minimize}} & \dfrac{\|\ \boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\ \|_1}{t} \\ \text{subject to} & t+\|\ \boldsymbol{x}\ \|_\infty \leq 1 \end{array}$$

with an implicit constraint $t>0$. A change of variables $\boldsymbol{y}=\boldsymbol{x}/t, z=1/t$ gives a convex problem

$$\begin{array}{ll} \underset{\boldsymbol{y},z}{\text{minimize}} & \|\ \boldsymbol{A}\boldsymbol{y}-\boldsymbol{b}z\ \|_1 \\ \text{subject to} & 1+\|\ \boldsymbol{y}\ \|_\infty \leq z \end{array}$$

(Note that the constraint implies $z>0$.) This problem now reduces to an LP

$$\begin{array}{ll} \underset{\boldsymbol{x},\boldsymbol{y},\boldsymbol{u},z}{\text{maximize}} & \mathbf{1}^\mathsf{T}\boldsymbol{u} \\ \text{subject to} & -\boldsymbol{u} \preceq \boldsymbol{A}\boldsymbol{y}-\boldsymbol{b}z \preceq \boldsymbol{u} \\ & -v\mathbf{1} \preceq \boldsymbol{y} \preceq v\mathbf{1} \\ & 1+v \leq z \end{array}$$

with variables $\boldsymbol{u}\in\mathbb{R}^m, \boldsymbol{y}\in\mathbb{R}^n, z\in\mathbb{R}, v\in\mathbb{R}$.

c. Since

$$yz = \frac{1}{4}\left((y+z)^2-(y-z)^2\right),$$

the constraint $\|\boldsymbol{x}\|^2 \leq yz, y\geq 0, z\geq 0$ can be rewritten as

$$\left\|\begin{bmatrix} 2\boldsymbol{x} \\ y-z \end{bmatrix}\right\| \leq y+z.$$

Thus, the set can be rewritten as the following SOC constraint:

$$\left\{(\boldsymbol{x},y,z)\in\mathbb{R}^{n+2} \,\middle|\, \left\|\begin{bmatrix} 2\boldsymbol{x} \\ y-z \end{bmatrix}\right\| \leq y+z\right\}.$$

d. The epigraph of the problem has the following formulation:

$$\begin{aligned} \underset{\boldsymbol{x},y\geq 0,z\geq 0,t}{\text{minimize}} \quad & \boldsymbol{a}^\mathsf{T}\boldsymbol{x} + \kappa t \\ \text{subject to} \quad & \|\boldsymbol{x}\|^2 \leq yz \\ & \sqrt{\boldsymbol{x}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{x}} \leq t. \end{aligned}$$

The constraint $\|\boldsymbol{x}\|^2 \leq yz$ can be written as

$$\left\| \begin{bmatrix} 2\boldsymbol{x} \\ y - z \end{bmatrix} \right\| \leq y + z,$$

while $\sqrt{\boldsymbol{x}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{x}} \leq t$ can be written as

$$\|\boldsymbol{L}^\mathsf{T}\boldsymbol{x}\| \leq t,$$

where $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\mathsf{T}$ (Cholesky decomposition). Therefore, the problem is equivalent to an SOCP given by

$$\begin{aligned} \underset{\boldsymbol{x},y\geq 0,z\geq 0,t}{\text{minimize}} \quad & \boldsymbol{a}^\mathsf{T}\boldsymbol{x} + \kappa t \\ \text{subject to} \quad & \left\| \begin{bmatrix} 2\boldsymbol{x} \\ y - z \end{bmatrix} \right\| \leq y + z \\ & \|\boldsymbol{L}^\mathsf{T}\boldsymbol{x}\| \leq t. \end{aligned}$$

e. By introducing a variable t, then the problem can be reformulated as

$$\begin{aligned} \underset{\boldsymbol{x},t}{\text{minimize}} \quad & t + \boldsymbol{a}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{B}\boldsymbol{x} \leq \boldsymbol{b}, \\ & \boldsymbol{x}^\mathsf{T}\boldsymbol{A}\boldsymbol{x} \leq t. \end{aligned}$$

Notice that $\boldsymbol{x}^\mathsf{T}\boldsymbol{A}\boldsymbol{x} \leq t$ is equivalent to $\|\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{x}\|^2 \leq t$. The above problem can be further written in an SOCP:

$$\begin{aligned} \underset{\boldsymbol{x},t}{\text{minimize}} \quad & t + \boldsymbol{a}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{B}\boldsymbol{x} \leq \boldsymbol{b}, \\ & \left\| \begin{bmatrix} 2\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{x} \\ t - 1 \end{bmatrix} \right\| \leq t + 1. \end{aligned}$$

f. The epigraph of the problem is given by

$$\begin{aligned} \underset{\boldsymbol{X}\succ\boldsymbol{0},\boldsymbol{M}}{\text{minimize}} \quad & \text{Tr}\left(\boldsymbol{M}\right) + \text{Tr}\left(\boldsymbol{A}\boldsymbol{X}\right) \\ \text{subject to} \quad & \left(\boldsymbol{I} + \boldsymbol{X}\right)^{-1} \preceq \boldsymbol{M}. \end{aligned}$$

Applying the Schur complement, we get

$$\begin{aligned} \underset{\boldsymbol{X}\succ\boldsymbol{0},\boldsymbol{M}}{\text{minimize}} \quad & \text{Tr}\left(\boldsymbol{M}\right) + \text{Tr}\left(\boldsymbol{A}\boldsymbol{X}\right) \\ \text{subject to} \quad & \begin{bmatrix} \boldsymbol{I} + \boldsymbol{X} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{M} \end{bmatrix} \succeq \boldsymbol{0}, \end{aligned}$$

which is an SDP.

Exercise A.5: Concepts on problem resolution

a. How would you determine if a convex problem is feasible or infeasible?
b. How would you determine if a convex problem has a unique solution or multiple solutions?
c. What are the main ways to solve a convex problem?
d. Given a nonconvex optimization problem, what strategies can be used to find an approximate solution?

Solution

a. Check the feasibility of constraints. If there exists a point that satisfies all the constraints, then the problem is feasible.

b. If the feasible set is not empty, and the objective function is strictly convex, then the convex problem has a unique solution, otherwise there exists multiple solutions.

c. We can solve the convex problem by working out the KKT conditions. If the convex problem are in the classical classes like LP, QP, SOCP, and SDP, there are corrsponding solvers that help to solve the problem efficently. If the convex problem are more complex, then we can use the general nonlinear solvers like gradient-based methods and metaheuristic algorithms.

d. First, we can try to transform the nonconvex problem into a convex one by relaxing or approximating certain constraints or objectives, and then take the optimal point as an approximated solution. Besides, local optimization methods such as gradient-based methods help us to find a local optimum of the nonconvex problem. Randomized search via such as metaheuristic algorithms can also be an option.

Exercise A.6: Linear regression

a. Consider the line equation $y = \alpha x + \beta$. Choose some values for α and β, and generate 100 noisy pairs $(x_i, y_i),\ i = 1, \dots, 100$ (i.e., add some random noise to each observation y_i).
b. Formulate a regression problem to fit the 100 data points with a line based on least squares. Plot the true and estimated lines along with the points.
c. Repeat the regression using several other definitions of error in the problem formulation. Plot and compare all the estimated lines.

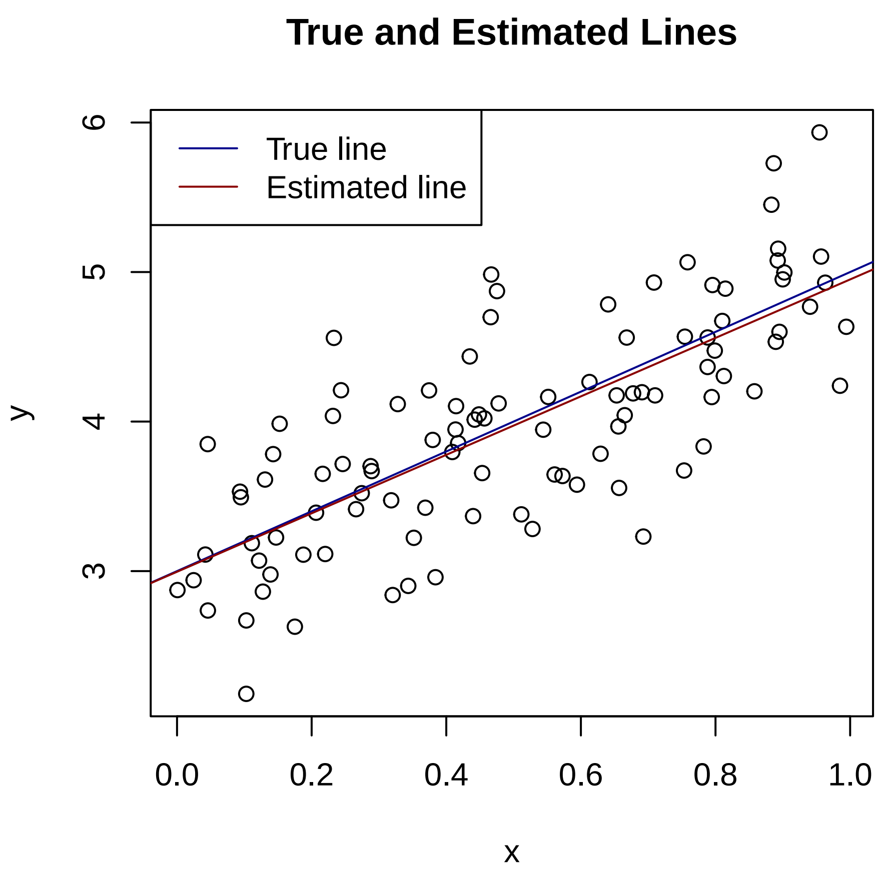
Solution

a. The R code is shown below:

```r
set.seed(123)
alpha <- 2
beta <- 3
x <- runif(100)
y_true <- alpha * x + beta
y <- y_true + rnorm(100, sd = 0.5)
```

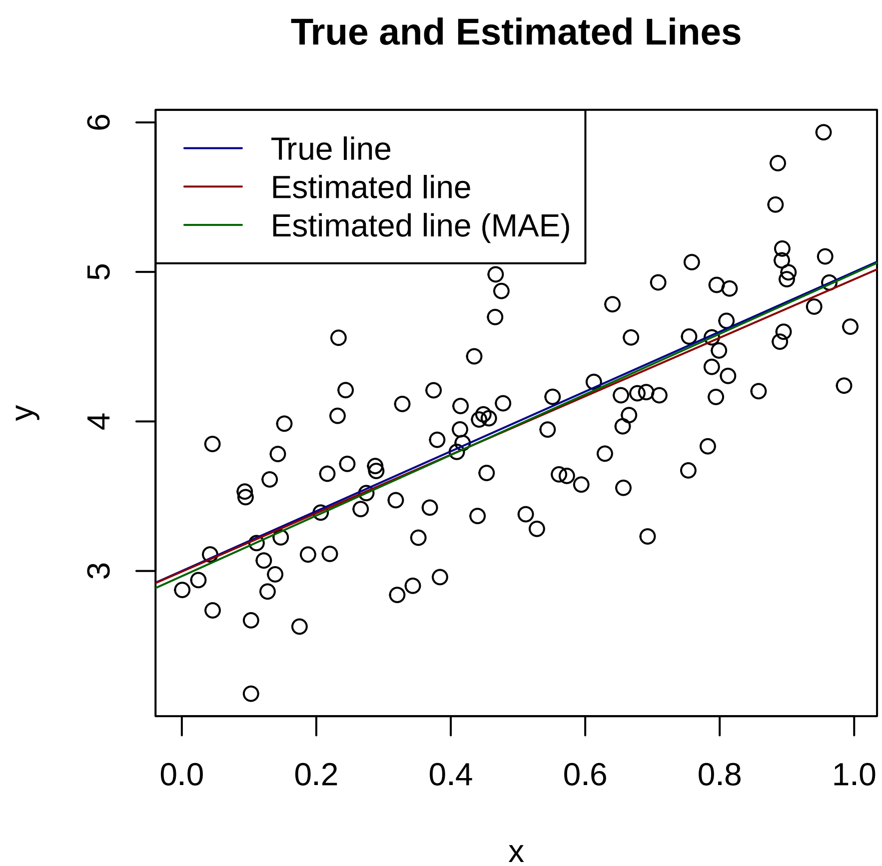
b. The R code is shown below:

```r
model <- lm(y ~ x)  # fit a linear model
plot(x, y, main = "True and Estimated Lines")
abline(a = beta, b = alpha, col = "darkblue")  # true line
abline(model, col = "darkred")  # estimated line
legend("topleft", legend = c("True line", "Estimated line"),
col = c("darkblue", "darkred"), lty = 1)
```



c. The R code is shown below:

```
# Using Mean Absolute Error (MAE) as the error definition
model_mae <- optim(par = c(0, 0), fn = function(par) mean(abs(y -
(par[1] * x + par[2]))))
plot(x, y, main = "True and Estimated Lines")
abline(a = beta, b = alpha, col = "darkblue")  # true line
abline(model, col = "darkred")  # estimated line
# estimated line using MAE
abline(a = model_mae$par[2], b = model_mae$par[1], col = "darkgreen")
legend("topleft", legend = c("True line", "Estimated line",
"Estimated line (MAE)"), col = c("darkblue", "darkred", "darkgreen"), lty = 1)
```



Exercise A.7: Concepts on Lagrange duality

a. Define Lagrange duality and explain its significance in convex optimization.
b. Give an example of a problem and its dual.
c. List the KKT conditions and explain their role in convex optimization.
d. Provide an example of a problem with its KKT conditions.
e. Try to find a solution that satisfies the previous KKT conditions. Is this always possible?

Solution

a. Lagrange duality is a fundamental concept in convex optimization that involves transforming a given optimization problem, known as the primal problem, into its dual problem. The Lagrange duality provides a lower bound on the optimal value of the primal problem and helps in characterizing the optimality conditions. It is significant as it allows us to solve a difficult primal problem by solving a potentially simpler dual problem, providing insights into the problem structure and facilitating analysis.

b. Consider the following primal problem:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \boldsymbol{x}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

The dual problem associated with this primal problem is formulated as follows:

$$\underset{\nu}{\text{maximize}} \quad -\frac{1}{4}\boldsymbol{\nu}^\mathsf{T}\boldsymbol{A}A^\mathsf{T}\boldsymbol{\nu} - \boldsymbol{b}^\mathsf{T}\boldsymbol{\nu}.$$

c. The Karush-Kuhn-Tucker (KKT) conditions play a crucial role in convex optimization. They are a set of necessary conditions for optimality that apply to both the primal and dual problems. The KKT conditions ensure that a point satisfies the first-order optimality conditions and the constraints. Consider the following primal problem:

$$\begin{array}{lll} \underset{x}{\text{minimize}} & f_0(x) & \\ \text{subject to} & f_i(x) \le 0, & i = 1, 2, \ldots, m \\ & h_j(x) = 0, & j = 1, 2, \ldots, p \end{array}$$

The KKT conditions are as follows:

- Primal feasibility:

$$\begin{array}{ll} f_i(x) \le 0, & i = 1, 2, \ldots, m \\ h_j(x) = 0, & j = 1, 2, \ldots, p \end{array}$$

- Dual feasibility:

$$\lambda \ge 0$$

- Complementary slackness:

$$\lambda_i f_i(x) = 0, \quad i = 1, 2, \ldots, m$$

- Gradient of the Lagrangian:

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{j=1}^{p} \nu_j \nabla h_j(x) = 0$$

d. Let's consider an example problem and its KKT conditions. Suppose we have the following optimization problem:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) = x^2 \\ \text{subject to} & x \geq 2 \end{array}$$

The KKT conditions for this problem are:

- Primal feasibility: $x \geq 2$
- Dual feasibility: $\lambda \geq 0$
- Complementary slackness: $\lambda(x - 2) = 0$
- Gradient of the Lagrangian: $2x - \lambda = 0$

e. It's easy to verify that $x = 2$ satisfies the previous KKT conditions. It is not always possible to find a solution that satisfies all the KKT conditions. In some cases, the primal and dual problems may be infeasible, or the KKT conditions may not have a feasible solution. However, if both the primal and dual problems are feasible and satisfy strong duality, then a solution satisfying the KKT conditions exists, ensuring the optimality of the solution.

Exercise A.8: Solution via KKT conditions

For the following problems, determine the convexity, write the Lagrangian and KKT conditions, and derive a closed-form solution:

a. Risk parity portfolio:

$$\begin{array}{ll} \underset{\boldsymbol{x} \geq \boldsymbol{0}}{\text{minimize}} & \sqrt{\boldsymbol{x}^\mathsf{T} \boldsymbol{\Sigma} \boldsymbol{x}} \\ \text{subject to} & \boldsymbol{b}^\mathsf{T} \log(\boldsymbol{x}) \geq 1, \end{array}$$

where $\boldsymbol{\Sigma} \succ \boldsymbol{0}$ and the log function is applied elementwise.

b. Projection onto the simplex:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 \\ \text{subject to} & \mathbf{1}^\mathsf{T}\boldsymbol{x} = (\leq) 1,\ \boldsymbol{x} \geq \mathbf{0}. \end{array}$$

c. Projection onto a diamond:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 \\ \text{subject to} & \|\boldsymbol{x}\|_1 \leq 1. \end{array}$$

Solution

a. Let $\lambda \geq 0$ be the Lagrange multiplier associated with the constraint. The Lagrangian is given by:

$$\mathcal{L}(\boldsymbol{x}, \lambda) = \sqrt{\boldsymbol{x}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{x}} + \lambda\left(1 - \boldsymbol{b}^\mathsf{T}\log(\boldsymbol{x})\right)$$

To derive the KKT conditions, we need to take the partial derivatives of the Lagrangian with respect to $\boldsymbol{x}$ and set them to zero:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}} = \frac{1}{2}\frac{1}{\sqrt{\boldsymbol{x}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{x}}} \cdot (2\boldsymbol{\Sigma}\boldsymbol{x}) - \lambda\frac{\boldsymbol{b}}{\boldsymbol{x}} = \mathbf{0},$$

where $\frac{\boldsymbol{b}}{\boldsymbol{x}}$ represents elementwise division. The KKT conditions are given by:

- Primal feasibility: $\boldsymbol{b}^\mathsf{T}\log(\boldsymbol{x}) \geq 1$ and $\boldsymbol{x} \geq \mathbf{0}$ (original constraints)
- Dual feasibility: $\lambda \geq 0$ (Lagrange multiplier non-negativity)
- Complementary slackness: $\lambda(1 - \boldsymbol{b}^\mathsf{T}\log(\boldsymbol{x})) = 0$
- Stationarity: $\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}} = \mathbf{0}$

Defining $\tilde{\boldsymbol{x}} = \sigma^{-1/2}\lambda^{-1/2}\boldsymbol{x}$ with $\sigma = \sqrt{\boldsymbol{x}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{x}}$, we can rewrite $\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}} = \mathbf{0}$ as

$$\boldsymbol{\Sigma}\tilde{\boldsymbol{x}} = \frac{\boldsymbol{b}}{\tilde{\boldsymbol{x}}},$$

which is the desired risk parity/budgeting condition.

b. We apply the standard KKT conditions for the problem. The Lagrangian of the problem is

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \upsilon) = \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\| - \boldsymbol{\lambda}^\mathsf{T}\boldsymbol{x} - \upsilon(\mathbf{1}^\mathsf{T}\boldsymbol{x} - 1),$$

where $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_n]^\mathsf{T}$ and υ are the Lagrange multipliers associated with the inequality and equality constraints, respectively. At the optimal solution $\boldsymbol{x}$, the following KKT conditions hold:

$$\begin{array}{ll} x_i - y_i - \lambda_i - \upsilon = 0, & i = 1, \ldots, n \\ \lambda_i \geq 0, & i = 1, \ldots, n \\ \lambda_i x_i = 0, & i = 1, \ldots, n \\ x_i \geq 0, & i = 1, \ldots, n \\ \sum_{i=1}^n x_i = 1 & \end{array}$$

From the complementarity slackness, it is clear that if $x_i > 0$, we must have $\lambda_i = 0$ and $x_i = y_i + \upsilon > 0$; if $x_i = 0$, we must have $\beta_i \geq 0$ and $x_i = y_i + \lambda_i + \upsilon = 0$, whence $y_i + \upsilon = -\lambda_i \leq 0$. Obviously, the components of the optimal solution $\boldsymbol{x}$ that are zeros correspond to the smaller components of y. Without loss of generality, we assume the components of $\boldsymbol{y}$ are sorted and $\boldsymbol{x}$ uses the same ordering, i.e.,

$$\begin{aligned} y_1 \geq \cdots \geq y_\rho \geq y_{\rho+1} \geq \cdots \geq y_n, \\ x_1 \geq \cdots \geq x_\rho > x_{\rho+1} \geq \cdots \geq x_n, \end{aligned}$$

and that $x_1 \geq \cdots \geq x_\rho > 0$, $x_{\rho+1} = \cdots = x_n = 0$. In other words, ρ is the number of positive components in the solution $\boldsymbol{x}$. Now we apply the last condition and have

$$1 = \textstyle\sum_{i=1}^n x_i = \sum_{i=1}^\rho x_i = \sum_{i=1}^\rho (y_i + \upsilon),$$

which gives $\upsilon = \frac{1}{\rho}(1 - \sum_{i=1}^\rho y_i)$. Hence ρ is the key to the solution.

(i) For $j = \rho$, we have

$$y_\rho + \frac{1}{\rho}\left(1 - \sum_{i=1}^{\rho} y_i\right) = y_\rho + \upsilon = x_\rho > 0.$$

(ii) For $j < \rho$, we have

$$\begin{aligned} y_j + \frac{1}{j}\left(1 - \sum_{i=1}^{j} y_i\right) &= \frac{1}{j}\left(j y_j + \sum_{i=j+1}^{\rho} y_i + 1 - \sum_{i=1}^{\rho} y_i\right) \\ &= \frac{1}{j}\left(j y_j + \sum_{i=j+1}^{\rho} y_i + \rho\upsilon\right) \\ &= \frac{1}{j}\left(j(y_j + \upsilon) + \sum_{i=j+1}^{\rho} (y_i + \upsilon)\right) > 0, \end{aligned}$$

since $y_i + \upsilon > 0$ for $i = j, \ldots, \rho$.

(iii) For $j > \rho$, we have

$$\begin{aligned} y_j + \tfrac{1}{j}\left(1 - \textstyle\sum_{i=1}^{j} y_i\right) &= \tfrac{1}{j}\left(j y_j + 1 - \textstyle\sum_{i=1}^{\rho} y_i - \sum_{i=\rho+1}^{j} y_i\right) \\ &= \tfrac{1}{j}\left(j y_j + \rho\upsilon - \textstyle\sum_{i=\rho+1}^{j} y_i\right) \\ &= \tfrac{1}{j}\left(\rho(y_j + \upsilon) + \textstyle\sum_{i=\rho+1}^{j} (y_j - y_i)\right) \leq 0, \end{aligned}$$

since $y_j + \lambda \leq 0$ for $j > \rho$ and $y_j \leq y_i$.

Therefore, we have

$$\rho = \max\left\{1 \leq j \leq n : y_j + \tfrac{1}{j}\left(1 - \textstyle\sum_{i=1}^{j} y_i\right) > 0\right\}.$$

The optimal point is $x_i = \max\{y_i + \upsilon, 0\}$ with $\upsilon = \frac{1}{\rho}(1 - \sum_{i=1}^\rho y_{[i]})$ and $\{y_{[i]}\}$ are sorted elements of $\boldsymbol{y}$ such that $y_{[1]} \geq y_{[2]} \geq \ldots y_{[n]}$.

c. Projection onto a diamond:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 \\ \text{subject to} & \|\boldsymbol{x}\|_1 \leq 1. \end{array}$$

Let $\mathbf{x}^\star$ denote the optimal solution. Then, we first show that for $\forall i$, $x_i^\star y_i \geq 0$. Assume by contradiction that the claim does not hold. Thus, there exists i for which $x_i^\star y_i < 0$. Let $\bar{\mathbf{x}}$ be a vector such that $\bar{x}_i = 0$ and for all $j \neq i$ we have $\bar{x}_j = x_j^\star$. Therefore, $\|\bar{\mathbf{x}}\|_1 = \|\mathbf{x}\|_1 - |w_i| \leq z$ and hence $\bar{\mathbf{x}}$ is a feasible solution. In addition,

$$\begin{aligned} \|\mathbf{x}^\star - \mathbf{y}\|_2^2 - \|\bar{\mathbf{x}} - \mathbf{y}\|_2^2 &= (x_i^\star - y_i)^2 - (0 - y_i)^2 \\ &= x_i^{\star 2} - 2x_i^\star y_i > x_i^{\star 2} > 0. \end{aligned}$$

We thus constructed a feasible solution $\bar{\mathbf{x}}$ which attains an objective value smaller than that of $\mathbf{x}^\star$. This leads us to the desired contradiction. Therefore, the above claim indicates that each non-zero component of the optimal solution $\mathbf{x}^\star$ for the original problem shares the sign of its counterpart in $\mathbf{y}$. We note that if $\|\mathbf{y}\|_1 \leq 1$ then the solution of the original problem is $\mathbf{x}^\star = \mathbf{y}$. Therefore, from now on we assume that $\|\mathbf{y}\|_1 > 1$. In this case, the optimal solution must be on the boundary of the constraint set and thus we can replace the inequality constraint with $\|\mathbf{x}\| \leq 1$ with an equality constraint $\|\mathbf{x}\|_1 = 1$. Based on the above statement and the symmetry of the objective, we are ready to present our reduction. Let $\mathbf{u}$ be a vector obtained by taking the value of each component of $\mathbf{y}$, $u_i = |y_i|$. We now replace original problem with

$$\begin{array}{ll} \underset{\mathbf{z}\in\mathbb{R}^n}{\text{minimize}} & \dfrac{1}{2}\|\mathbf{z}-\mathbf{u}\|_2^2 \\ \text{subject to} & \mathbf{1}^\mathsf{T}\mathbf{z} = 1, \mathbf{z} \geq 0. \end{array}$$

Once we obtain the solution $\mathbf{z}$ for the problem above, we construct the optimal of the original problem by setting $x_i = \text{sign}(y_i) z_i$.

Exercise A.9: Dual problems

Find the dual of the following problems:

a. Vanishing maximum eigenvalue problem:

$$\begin{array}{ll} \underset{t,\boldsymbol{X}}{\text{minimize}} & t \\ \text{subject to} & t\boldsymbol{I} \succeq \boldsymbol{X}, \\ & \boldsymbol{X} \succeq \mathbf{0}. \end{array}$$

b. Matrix upper bound problem:

$$\begin{array}{ll} \underset{\boldsymbol{X}}{\text{minimize}} & \text{Tr}(\boldsymbol{X}) \\ \text{subject to} & \boldsymbol{X} \succeq \boldsymbol{A}, \\ & \boldsymbol{X} \succeq \boldsymbol{B} \end{array}$$

where $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{S}_+^n$.

c. Log det problem:

$$\begin{array}{lll} \underset{\boldsymbol{X} \succeq \boldsymbol{0}}{\text{minimize}} & \text{Tr}(\boldsymbol{C}\boldsymbol{X}) + \log\det(\boldsymbol{X}^{-1}) & \\ \text{subject to} & \boldsymbol{A}_i^\mathsf{T} \boldsymbol{X} \boldsymbol{A}_i \preceq \boldsymbol{B}_i, & i = 1, \ldots, m, \end{array}$$

where $\boldsymbol{C} \in \mathbb{S}_+^n$ and $\boldsymbol{B}_i \in \mathbb{S}_{++}^n$ for $i = 1, \ldots, m,$.

Solution

a. The Lagrangian function is

$$\begin{aligned} L(t, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\Lambda}) &= t + \text{Tr}\left((\boldsymbol{X} - t\boldsymbol{I})\,\boldsymbol{Z}\right) - \text{Tr}\left(\boldsymbol{X}\boldsymbol{\Lambda}\right) \\ &= t\left(1 - \text{Tr}\left(\boldsymbol{Z}\right)\right) + \text{Tr}\left(\boldsymbol{X}\left(\boldsymbol{Z} - \boldsymbol{\Lambda}\right)\right), \end{aligned}$$

where $\boldsymbol{Z} \succeq \boldsymbol{0}$ and $\boldsymbol{\Lambda} \succeq \boldsymbol{0}$ are the dual variables. The dual function is given by

$$g(\boldsymbol{Z}, \boldsymbol{\Lambda}) = \inf_{t, \boldsymbol{X}} L(t, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\Lambda}) = \begin{cases} 0 & \text{Tr}\left(\boldsymbol{Z}\right) = 1,\ \boldsymbol{Z} = \boldsymbol{\Lambda} \\ -\infty & \text{otherwise.} \end{cases}$$

The dual problem can be written as

$$\begin{array}{ll} \underset{\boldsymbol{Z}}{\text{maximize}} & 0 \\ \text{subject to} & \text{Tr}\left(\boldsymbol{Z}\right) = 1 \\ & \boldsymbol{Z} \succeq \boldsymbol{0}. \end{array}$$

b. The Lagrangian function is

$$\begin{aligned} L(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\Lambda}) = &\quad \text{Tr}\left(\boldsymbol{X}\right) + \text{Tr}\left((\boldsymbol{A} - \boldsymbol{X})\,\boldsymbol{Z}\right) + \text{Tr}\left((\boldsymbol{B} - \boldsymbol{X})\,\boldsymbol{\Lambda}\right) \\ = &\quad \text{Tr}\left(\boldsymbol{X}\left(\boldsymbol{I} - \boldsymbol{Z} - \boldsymbol{\Lambda}\right)\right) + \text{Tr}\left(\boldsymbol{A}\boldsymbol{Z}\right) + \text{Tr}\left(\boldsymbol{B}\boldsymbol{\Lambda}\right) \end{aligned}$$

where $\boldsymbol{Z} \succeq \boldsymbol{0}$ and $\boldsymbol{\Lambda} \succeq \boldsymbol{0}$ are the dual variables. The dual function is given by

$$g(\boldsymbol{Z}, \boldsymbol{\Lambda}) = \inf_{\boldsymbol{X}} L(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\Lambda}) = \begin{cases} \text{Tr}\left(\boldsymbol{A}\boldsymbol{Z}\right) + \text{Tr}\left(\boldsymbol{B}\boldsymbol{\Lambda}\right) & \boldsymbol{Z} + \boldsymbol{\Lambda} = \boldsymbol{I} \\ -\infty & \text{otherwise.} \end{cases}$$

The dual problem can therefore be expressed as

$$\begin{array}{ll} \underset{\boldsymbol{Z}, \boldsymbol{\Lambda}}{\text{maximize}} & \text{Tr}\left(\boldsymbol{A}\boldsymbol{Z}\right) + \text{Tr}\left(\boldsymbol{B}\boldsymbol{\Lambda}\right) \\ \text{subject to} & \boldsymbol{Z} + \boldsymbol{\Lambda} = \boldsymbol{I} \\ & \boldsymbol{Z} \succeq \boldsymbol{0},\ \boldsymbol{\Lambda} \succeq \boldsymbol{0}, \end{array}$$

or more simply

$$\begin{array}{ll}\underset{\boldsymbol{Z}}{\text{maximize}} & \operatorname{Tr}\left(\left(\boldsymbol{A}-\boldsymbol{B}\right)\boldsymbol{Z}\right) \\ \text{subject to} & \mathbf{0} \preceq \boldsymbol{Z} \preceq \boldsymbol{I}.\end{array}$$

c. The Lagrangian is:

$$\mathcal{L}\left(\boldsymbol{X}, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m\right) = \operatorname{Tr}\left(\boldsymbol{C}\boldsymbol{X}\right) - \log\det\boldsymbol{X} + \sum_{i=1}^{m} \operatorname{Tr}\left[\boldsymbol{Z}_i\left(\boldsymbol{A}_i^{\mathsf{T}}\boldsymbol{X}\boldsymbol{A}_i - \boldsymbol{B}_i\right)\right].$$

As a function of $\boldsymbol{X}$, this is strictly convex, and therefore $\boldsymbol{X}$ minimizes the Lagrangian if and only if the gradient of the Lagrangian with respect to $\boldsymbol{X}$ is zero:

$$\boldsymbol{C} - \boldsymbol{X}^{-1} + \sum_{i=1}^{m} \boldsymbol{A}_i\boldsymbol{Z}_i\boldsymbol{A}_i^{\mathsf{T}} = 0.$$

This equation has a solution if and only if $\boldsymbol{C} + \sum_{i=1}^{m} \boldsymbol{A}_i\boldsymbol{Z}_i\boldsymbol{A}_i^{\mathsf{T}} \succ \mathbf{0}$. Thus we find that the $\boldsymbol{X}$ that minimizes the Lagrangian is:

$$\boldsymbol{X}^{\star} = \left(\boldsymbol{C} + \sum_{i=1}^{m} \boldsymbol{A}_i\boldsymbol{Z}_i\boldsymbol{A}_i^{\mathsf{T}}\right)^{-1}.$$

Therefore, the dual function is:

$$\begin{aligned}
g\left(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m\right) &= L\left(\boldsymbol{X}^{\star}, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m\right) \\
&= \log\det\left(\boldsymbol{C} + \sum_{i=1}^{m} \boldsymbol{A}_i\boldsymbol{Z}_i\boldsymbol{A}_i^{\mathsf{T}}\right) + \operatorname{Tr}\left[\boldsymbol{C}\left(\boldsymbol{C} + \sum_{i=1}^{m} \boldsymbol{A}_i\boldsymbol{Z}_i\boldsymbol{A}_i^{\mathsf{T}}\right)^{-1}\right] \\
&\qquad + \sum_{i=1}^{m} \operatorname{Tr}\left[\boldsymbol{Z}_i\left(\boldsymbol{A}_i^{\mathsf{T}}\left(\boldsymbol{C} + \sum_{i=1}^{m} \boldsymbol{A}_i\boldsymbol{Z}_i\boldsymbol{A}_i^{\mathsf{T}}\right)^{-1}\boldsymbol{A}_i - \boldsymbol{B}_i\right)\right] \\
&= \log\det\left(\boldsymbol{C} + \sum_{i=1}^{m} \boldsymbol{A}_i\boldsymbol{Z}_i\boldsymbol{A}_i^{\mathsf{T}}\right) \\
&\qquad + \operatorname{Tr}\left[\boldsymbol{I}_n - \left(\sum_{i=1}^{m} \boldsymbol{A}_i\boldsymbol{Z}_i\boldsymbol{A}_i^{\mathsf{T}}\right)\left(\boldsymbol{C} + \sum_{i=1}^{m} \boldsymbol{A}_i\boldsymbol{Z}_i\boldsymbol{A}_i^{\mathsf{T}}\right)^{-1}\right] \\
&\qquad - \sum_{i=1}^{m} \operatorname{Tr}\left(\boldsymbol{Z}_i\boldsymbol{B}_i\right) + \operatorname{Tr}\left[\left(\sum_{i=1}^{m} \boldsymbol{A}_i\boldsymbol{Z}_i\boldsymbol{A}_i^{\mathsf{T}}\right)\left(\boldsymbol{C} + \sum_{i=1}^{m} \boldsymbol{A}_i\boldsymbol{Z}_i\boldsymbol{A}_i^{\mathsf{T}}\right)^{-1}\right] \\
&= \log\det\left(\boldsymbol{C} + \sum_{i=1}^{m} \boldsymbol{A}_i\boldsymbol{Z}_i\boldsymbol{A}_i^{\mathsf{T}}\right) - \sum_{i=1}^{m} \operatorname{Tr}\left(\boldsymbol{Z}_i\boldsymbol{B}_i\right) + n
\end{aligned}$$

with the domain $\operatorname{dom} g = \left\{ (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m) | \, \boldsymbol{C} + \sum_{i=1}^{m} \boldsymbol{A}_i \boldsymbol{Z}_i \boldsymbol{A}_i^\mathsf{T} \succ \boldsymbol{0} \right\}$. So the dual problem is

$$\begin{array}{ll} \text{maximize} & \log\det\left(\boldsymbol{C} + \sum_{i=1}^{m} \boldsymbol{A}_i \boldsymbol{Z}_i \boldsymbol{A}_i^\mathsf{T}\right) - \sum_{i=1}^{m} \operatorname{Tr}(\boldsymbol{Z}_i \boldsymbol{B}_i) + n \\ \text{subject to} & \boldsymbol{C} + \sum_{i=1}^{m} \boldsymbol{A}_i \boldsymbol{Z}_i \boldsymbol{A}_i^\mathsf{T} \succ \boldsymbol{0} \\ & \boldsymbol{Z}_i \succ \boldsymbol{0}, \quad i = 1, \ldots, m. \end{array}$$

Exercise A.10: Multi-objective optimization

a. Explain the concept of multi-objective optimization problems.
b. What is the significance of the weights in the scalarization of a multi-objective problem?
c. Provide an example of a bi-objective convex optimization problem and its scalarization.
d. Solve this scalarized bi-objective problem for different values of the weight and plot the optimal trade-off curve.

Solution

a. Multi-objective optimization (MOO) involves the simultaneous optimization of multiple conflicting objectives. Unlike traditional optimization problems with a single objective, MOO aims to find a set of solutions that represent the best trade-offs between different objectives. These objectives may include cost, efficiency, quality, or reliability, and improving one objective may come at the expense of others. MOO algorithms, such as evolutionary algorithms or mathematical programming techniques, explore the solution space to identify the Pareto optimal front set of solutions where no other solution can improve one objective without worsening another. Decision-makers can then select from the Pareto optimal solutions based on their preferences and priorities. MOO provides a valuable framework for tackling real-world problems with competing objectives and enables informed decision-making.

b. Scalarization involves combining multiple objectives into a single-objective problem by applying weights to each objective. These weights reflect the decision-maker's preferences and guide the trade-offs made between different objectives. By adjusting the weights, decision-makers can express their preferences and influence the resulting solution.

c. In mean-variance portfolio optimization, a bi-objective convex problem, the scalarization involves assigning weights to expected return and portfolio variance to create a single objective. By adjusting the weight, investors can balance risk and return trade-offs. By assigning a higher weight on portfolio variance, investors can prioritize risk reduction in the scalarization process, leading to more conservative portfolio allocations with lower volatility.

d. TBD.