ROYAL INSTITUTE
OF TECHNOLOGY

Lecture: Lagrange relaxation

1. Lagrange relaxation
   - Global optimality conditions
   - KKT conditions for convex problems
   - Applications

Lagrange relaxation

We consider the optimization problem

$$(P) \quad \begin{bmatrix} \text{minimize} & f(\mathbf{x}) \\ \text{s.t.} & g_i(\mathbf{x}) \leq 0, \;\; i = 1, \ldots, m \end{bmatrix}$$

where $f(\mathbf{x})$ and $g_i(\mathbf{x})$ are real valued functions.

If $\mathbf{g}(\mathbf{x}) = \begin{bmatrix} g_1(\mathbf{x}) & \ldots & g_m(\mathbf{x}) \end{bmatrix}^\mathsf{T}$ then (P) can be written

$$(P) \quad \begin{bmatrix} \text{minimize} & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{g}(\mathbf{x}) \leq \mathbf{0}. \end{bmatrix}$$

The idéa behind Lagrange relaxation is to put non-negative prices $y_i \geq 0$, on the constraints and then add these to the objective function. This gives the (unconstrained) optimization problem:

$$\text{minimize} \quad f(\mathbf{x}) + \sum_{i=1}^{m} y_i g_i(\mathbf{x}) \tag{1}$$

which using $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_m \end{bmatrix}^\mathsf{T}$ can be written

$$\text{minimize} \quad f(\mathbf{x}) + \mathbf{y}^\mathsf{T}\mathbf{g}(\mathbf{x})$$

The "price" y_i is called a Lagrange multiplicator.

Definition 1. *The function* $L : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ *defined by* $L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) + \mathbf{y}^\mathsf{T}\mathbf{g}(\mathbf{x})$ *is called the Lagrange function to* (P)

Weak duality

Theorem 1 (Weak duality). *For an arbitrary $\mathbf{y} \geq \mathbf{0}$ it holds that $\min_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}) \leq f(\hat{\mathbf{x}})$, where $\hat{\mathbf{x}}$ is an optimal solution to (P).*

Proof: Since $\hat{\mathbf{x}}$ is a feasible solution to (P) it holds that $\mathbf{g}(\hat{\mathbf{x}}) \leq \mathbf{0}$. We get

$$\min_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}) \leq L(\hat{\mathbf{x}}, \mathbf{y}) = f(\hat{\mathbf{x}}) + \underbrace{\mathbf{y}^{\mathsf{T}}}_{\geq 0} \underbrace{\mathbf{g}(\hat{\mathbf{x}})}_{\leq 0} \leq f(\hat{\mathbf{x}}). \tag{2}$$

- minimizing the Lagrange function provides lower bounds to the optimization problem (P).
- By an appropriate choice of $\mathbf{y}$ a good approximation of the optimal solution to (P) is searched for. In practical algorithms one tries to solve $\max_{\mathbf{y} \geq \mathbf{0}} \min_{\mathbf{x}} L(\mathbf{x}, \mathbf{y})$. The next theorem gives conditions for the Lagrange multiplicator providing equality in (2).

Global optimality conditions

Theorem 2. *If* $(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \in \mathbf{R}^n \times \mathbf{R}^m$ *satisfies the conditions*

(1) $L(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = \min_x L(\mathbf{x}, \hat{\mathbf{y}})$,

(2) $\mathbf{g}(\hat{\mathbf{x}}) \leq \mathbf{0}$,

(3) $\hat{\mathbf{y}} \geq 0$,

(4) $\hat{\mathbf{y}}^\mathsf{T}\mathbf{g}(\hat{\mathbf{x}}) = 0$.

then $\hat{\mathbf{x}}$ *is an optimal solution to* (P).

Proof: *If* $\mathbf{x}$ *is an arbitrary feasible solution to* (P) *it holds that* $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$, *which shows that*

$$f(\mathbf{x}) \geq f(\mathbf{x}) + \hat{\mathbf{y}}^\mathsf{T}\mathbf{g}(\mathbf{x}) = L(\mathbf{x}, \hat{\mathbf{y}}) \geq L(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = f(\hat{\mathbf{x}})$$

where the first inequality follows from (3) *and* $g(\mathbf{x}) \leq \mathbf{0}$, *the second inequality follows from* (1), *and the last one from* (4).

The dual problem

The *dual objective function* $\varphi : \mathbf{R}^m_+ \to \mathbf{R}$ is defined by

$$\varphi(\mathbf{y}) = \min_x L(\mathbf{x}, \mathbf{y}) = L(\hat{\mathbf{x}}(\mathbf{y}), \mathbf{y}),$$

where $\hat{\mathbf{x}}(\mathbf{y})$ minimizes $L(\mathbf{x}, \mathbf{y})$ over $\mathbf{x}$ for a fixed $\mathbf{y} \geq 0$.

The dual problem to (P) is defined as

$$(D) \quad \left[\begin{array}{rl} \text{maximize} & \varphi(\mathbf{y}) \\ \text{s.t.} & \mathbf{y} \geq \mathbf{0}. \end{array} \right]$$

The dual problem is a convex optimization problem!

Theorem 3. *φ is a concave function on $\mathbf{R}^m_+$.*

Global optimality conditions again

Theorem 2 can be strengthened

Theorem 4. $(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \in \mathbf{R}^n \times \mathbf{R}^m$ *satisfies the global optimality conditions iff*

(1) $\hat{\mathbf{x}}$ *is an optimal solution to* (P)

(2) $\hat{\mathbf{y}}$ *is an optimal solution to* (D)

(3) $f(\hat{\mathbf{x}}) = \varphi(\hat{\mathbf{y}})$.

The proof is based on the relation

$$\varphi(\hat{\mathbf{y}}) = \min_x L(\mathbf{x}, \hat{\mathbf{y}}) = L(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = f(\hat{\mathbf{x}}) + \hat{\mathbf{y}}^T g(\hat{\mathbf{x}}) = f(\hat{\mathbf{x}})$$

Lagrange duality: An example - the primal and relaxed problems

$$(P) \quad \left[\begin{array}{rl} \text{minimize} & x_1^2 + x_2^2 - 2x_1 \\ \text{s.t.} & x_1^2 + x_2^2 - 2x_2 \leq 0, \end{array} \right]$$

Here $f(\mathbf{x}) = x_1^2 + x_2^2 - 2x_1$ and $g_1(\mathbf{x}) = x_1^2 + x_2^2 - 2x_2$.

For some arbitrary $\mathbf{y} \geq 0$, consider the Lagrange relaxed problem

$$(PR_y) \quad \left[\text{minimize}_{\mathbf{x} \in \mathbf{R}^2} \quad x_1^2 + x_2^2 - 2x_1 + y(x_1^2 + x_2^2 - 2x_2), \right]$$

with objective function

$$L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) + y g_1(\mathbf{x}) = x_1^2 + x_2^2 - 2x_1 + y(x_1^2 + x_2^2 - 2x_2).$$

Lagrange duality: An example - solving PR_y

For fixed $y \geq 0$ (PR_y) is a convex quadratic problem with

$$H = \begin{bmatrix} 2(1+y) & 0 \\ 0 & 2(1+y) \end{bmatrix}, \quad c = \begin{bmatrix} -2 \\ -2y \end{bmatrix}$$

where H is positive definite.

The problem (PR_y) has a unique solution

$$\hat{\mathbf{x}}(y) = -H(y)^{-1}c = -\begin{bmatrix} \frac{1}{2(1+y)} & 0 \\ 0 & \frac{1}{2(1+y)} \end{bmatrix} \begin{bmatrix} -2 \\ -2y \end{bmatrix} = \begin{bmatrix} \frac{1}{1+y} \\ \frac{y}{1+y} \end{bmatrix}$$

Lagrange duality: An example - the dual problem

The dual objective function is now defined as

$$\begin{aligned}\varphi(\mathbf{y}) = L(\hat{\mathbf{x}}(\mathbf{y})), \mathbf{y}) &= \left(\frac{1}{1+y}\right)^2 + \left(\frac{y}{1+y}\right)^2 - 2\left(\frac{1}{1+y}\right) + \cdots \\ &+y\left\{\left(\frac{1}{1+y}\right)^2 + \left(\frac{y}{1+y}\right)^2 - 2\left(\frac{y}{1+y}\right)\right\} \\ &= -\frac{1+y^2}{1+y}\end{aligned}$$

The dual optimization problem becomes

$$(D) \quad \left[\begin{array}{rl} \text{maximize} & -\frac{1+y^2}{1+y} \\ \text{s.t.} & y \geq 0, \end{array}\right]$$

Lagrange duality: An example - solving the dual

The dual optimization problem is convex

$$(D) \quad \begin{bmatrix} \text{minimize} & \frac{1+y^2}{1+y} \\ \text{s.t.} & y \geq 0, \end{bmatrix} \quad \sim \quad \begin{bmatrix} \text{minimize} & \frac{1+(t-1)^2}{t} \\ \text{s.t.} & t \geq 1, \end{bmatrix}$$

and with $t = 1 + y$, the dual objective function is $\phi(t) = 2/t + t - 2$, and then

$$\phi'(t) = -2/t^2 + 1, \qquad \left(\phi''(t) = 4/t^3 > 0 \text{ for } t \geq 1\right)$$

The derivative is zero for $\hat{t} = \sqrt{2}$, i.e. (D) is solved by $\hat{y} = \sqrt{2} - 1$.

Lagrange duality: An example - solving the primal

The solution to (P) is now given by

$$\hat{\mathbf{x}}(\hat{\mathbf{y}}) = \begin{bmatrix} \frac{1}{1+\hat{y}} \\ \frac{\hat{y}}{1+\hat{y}} \end{bmatrix} = \begin{bmatrix} \frac{1}{1+\sqrt{2}-1} \\ \frac{\sqrt{2}-1}{1+\sqrt{2}-1} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 1 - \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Note that the complementarity conditions holds

$$\hat{\mathbf{y}}^T \mathbf{g}_1(\hat{\mathbf{x}}) = (\sqrt{2} - 1)0 = 0$$

To show that it is the optimal solution we check the conditions for the Global optimality conditions in Theorem 4

Lagrange duality: An example - GOC

Identical optimal values

$$f(\hat{\mathbf{x}}) = \left(\frac{1}{\sqrt{2}}\right)^2 + \left(1 - \frac{1}{\sqrt{2}}\right)^2 - 2\left(\frac{1}{\sqrt{2}}\right) = -\frac{4-2\sqrt{2}}{\sqrt{2}},$$

$$\varphi(\hat{\mathbf{y}}) = -\frac{1+\hat{y}^2}{1+\hat{y}} = -\frac{1+(\sqrt{2}-1)^2}{1+\sqrt{2}-1} = -\frac{4-2\sqrt{2}}{\sqrt{2}}.$$

Primal feasibility

$$\hat{x}_1^2 + \hat{x}_2^2 - 2x_2 = \left(\frac{1}{\sqrt{2}}\right)^2 + \left(1 - \frac{1}{\sqrt{2}}\right)^2 - 2\left(1 - \frac{1}{\sqrt{2}}\right) = \cdots = 0 \leq 0.$$

Dual feasibility

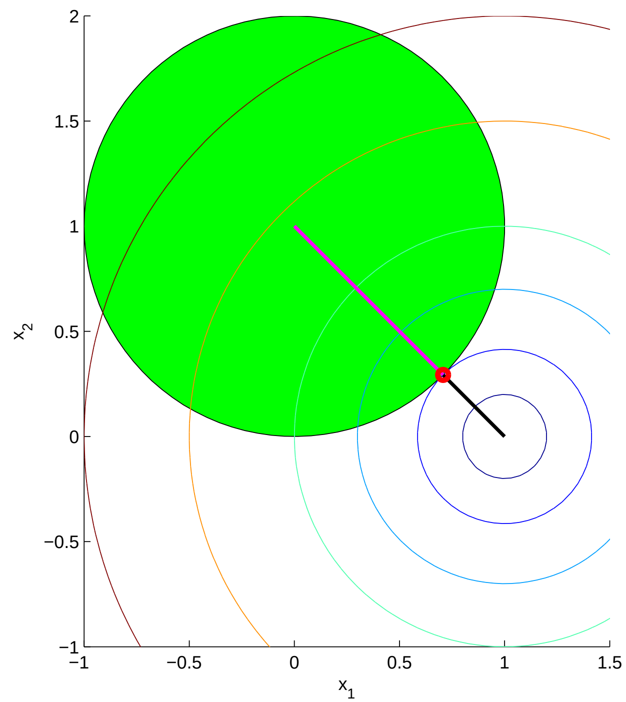
$$\hat{y} = \sqrt{2} - 1 \geq 0.$$

Lagrange duality: Example - Graphical illustration

The feasible region is depicted in green.

The level sets of the objective function are circles around the point $(1, 0)$, since $x_1^2 + x_2^2 - 2x_1 = (x_1 - 1)^2 + x_2^2 - 1$.

The red small circle denotes the optimal $\hat{\mathbf{x}}$ which lies on the line $x_1 + x_2 = 1$.



Convex optimization problems

- If the functions f and $g_1, \dots, g_m$ are convex and continuously differentiable, then condition (1) in Theorem 2 is equivalent to the condition

$$\nabla f(\hat{\mathbf{x}}) + \sum_{i=1}^{m} \hat{y}_i \nabla g_i(\hat{\mathbf{x}}) = \mathbf{0}^\mathsf{T} \tag{3}$$

  This follows since $L(\mathbf{x}, \hat{\mathbf{y}})$ is convex when $\hat{\mathbf{y}} \geq 0$ and then it holds that $\hat{\mathbf{x}}$ is a minimum point for $L(\mathbf{x}, \hat{\mathbf{y}})$ if, and only if, $\nabla_{\mathbf{x}} L(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = \mathbf{0}^\mathsf{T}$, *i.e.*, if and only if (3) is satisfied.

- The global optimality conditions in Theorem 2 are sufficient conditions for optimality, but in general not necessary. The next theorem shows that they are often also necessary conditions for convex optimization problems.

Definition 2. *The optimization problem* (P) *is a* regular *convex optimization problem if the functions* f *and* $g_1, \ldots, g_m$ *are convex and continuously differentiable and there exists a point* $\mathbf{x}_0 \in \mathbf{R}^n$ *such that* $g_i(\mathbf{x}_0) < 0$, $i = 1, \ldots, m$.

Theorem 5 (KKT for convex problems)**.** *Assume that* (P) *is a regular convex problem. Then* $\hat{\mathbf{x}}$ *is a (global) optimal solution if, and only if, there exists a vector* $\hat{\mathbf{y}} \in \mathbf{R}^m$ *such that*

(1) $\nabla f(\hat{\mathbf{x}}) + \sum_{i=1}^m \hat{y}_i \nabla g_i(\hat{\mathbf{x}}) = \mathbf{0}^\mathsf{T}$

(2) $\mathbf{g}(\hat{\mathbf{x}}) \leq \mathbf{0}$,

(3) $\hat{\mathbf{y}} \geq \mathbf{0}$,

(4) $\hat{\mathbf{y}}^\mathsf{T}\mathbf{g}(\hat{\mathbf{x}}) = 0$.

Proof: *Sufficiency was shown previously. Necessity is shown in the book.*

The conditions (2) – (4) can be made more explicit. We have that

$$\hat{\mathbf{y}}^{\mathsf{T}}\mathbf{g}(\hat{\mathbf{x}}) = \sum_{i=1}^{m} \hat{y}_i g_i(\hat{\mathbf{x}}) = 0$$

Since $g_i(\hat{\mathbf{x}}) \leq 0$ and $\hat{y}_i \geq 0$ it follows that $\hat{y}_i g_i(\hat{\mathbf{x}}) = 0$, $i = 1, \ldots, m$. We then get the equivalent conditions

(2′) $g_i(\hat{\mathbf{x}}) \leq 0$, $i = 1, \ldots, m$,

(3′) $\hat{y}_i \geq 0$, $i = 1, \ldots, m$,

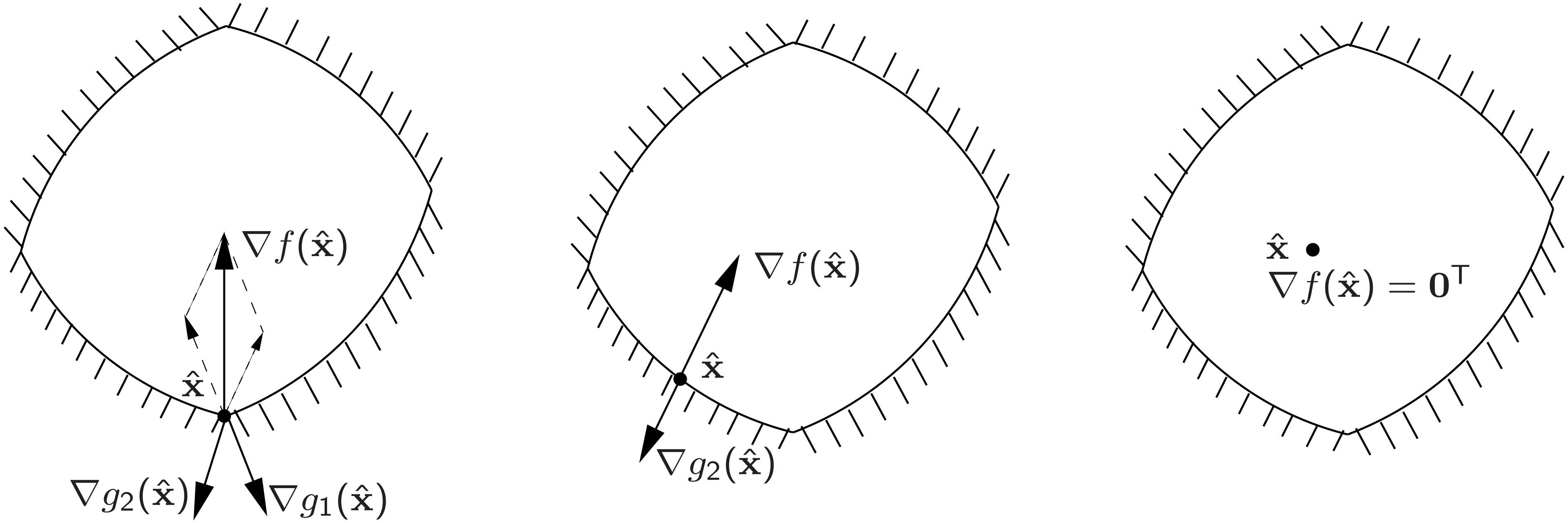
(4′) $\hat{y}_i \cdot g_i(\hat{\mathbf{x}}) = 0$, $i = 1, \ldots, m$.

Geometric interpretation

The complementarity condition ($4'$) implies that if $g_i(\hat{\mathbf{x}}) < 0$ then $y_i = 0$. Therefore, condition (1) can be written

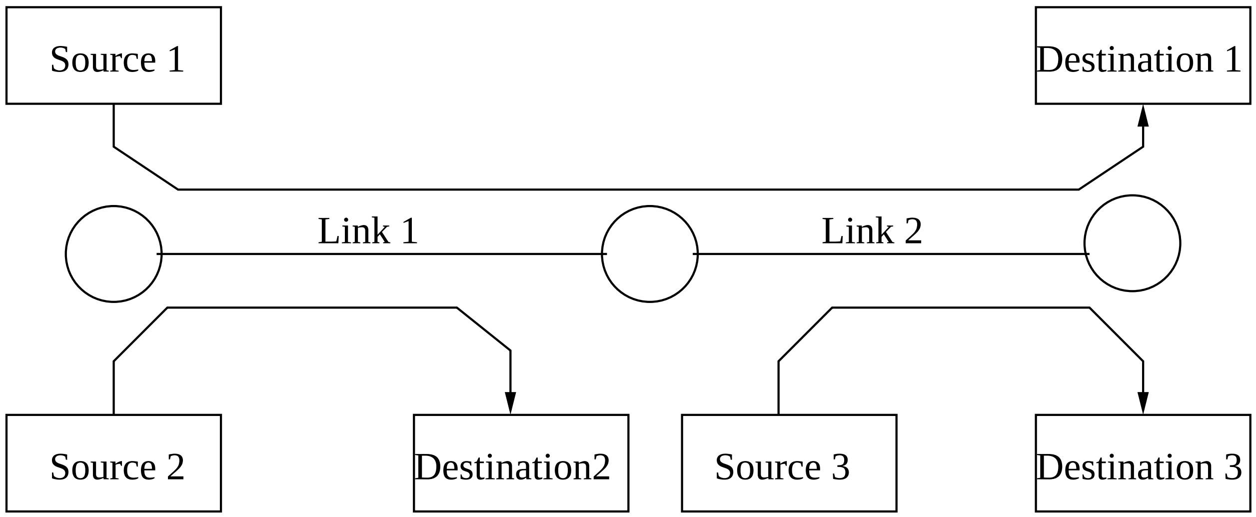
$$\nabla f(\hat{\mathbf{x}}) = -\sum_{i:g_i(\hat{\mathbf{x}})=0} \hat{y}_i \nabla g_i(\hat{\mathbf{x}})$$

this means that the gradient is a negative linear combination of the gradients of the binding (active) constraints.



Traffic control in communication systems

We consider a communication network consisting of two links. Three sources are sending data over the network to three different destinations.



- Source 1 uses both links.
- Source 2 uses link 1.
- Source 3 uses link 2.

- Link 1 has capacity 2 (normalized entity [data/s])
- Link 2 has capacity 1
- The three sources sends data with speeds x_r, $r = 1, 2, 3$.
- The three sources has each a utility function $U_r(x)$, $r = 1, 2, 3$. A common choice of the utility function is $U_r(x_r) = w_r \log(x_r)$.

For efficient and fair use of the available capacity, the data speeds are chosen using the following optimization criterion:

$$\begin{aligned} \text{maximize} \quad & U_1(x_1) + U_2(x_2) + U_3(x_3) \\ \text{s.t.} \quad & x_1 + x_2 \leq 2 \\ & x_1 + x_3 \leq 1 \\ & x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0 \end{aligned}$$

Assume $U_k(x) = \log(x_k)$, $k = 1, 2, 3$. The optimization problem can be written

$$
\begin{aligned}
-\text{minimize} \quad & -\log(x_1) - \log(x_2) - \log(x_3) \\
\text{s.t.} \quad & x_1 + x_2 \leq 2 \\
& x_1 + x_3 \leq 1
\end{aligned}
$$

We relaxed the constraints $x_k \geq 0$, $k = 1, \ldots, 3$ since they will be automatically satisfied, ($-\log(x) \to \infty$ as $x \to 0$).

The optimization problem is convex since the constraints are linear inequalities and the objective function is a sum of convex functions, and hence convex.

The optimality conditions in Theorem 5 are

$$(1) \qquad \begin{aligned} -\frac{1}{x_1} + y_1 + y_2 &= 0 \\ -\frac{1}{x_2} + y_1 &= 0 \\ -\frac{1}{x_3} + y_2 &= 0 \end{aligned}$$

$$(2) \qquad \begin{aligned} x_1 + x_2 - 2 &\leq 0 \\ x_1 + x_3 - 1 &\leq 0 \end{aligned}$$

$$(3) \qquad \begin{aligned} y_1 &\geq 0 \\ y_2 &\geq 0 \end{aligned}$$

$$(4) \qquad \begin{aligned} y_1(x_1 + x_2 - 2) &= 0 \\ y_2(x_1 + x_3 - 1) &= 0 \end{aligned}$$

from (1) we get

$$x_1 = \frac{1}{y_1 + y_2} \qquad x_2 = \frac{1}{y_1} \qquad x_3 = \frac{1}{y_2}$$

This leads to $y_1 > 0$ and $y_2 > 0$, hence the complementarity constraint (4) shows that (2) is satisfied with equality. We get

$$\begin{array}{l} \frac{1}{y_1+y_2} + \frac{1}{y_1} = 2 \\ \frac{1}{y_1+y_2} + \frac{1}{y_2} = 1 \end{array} \quad \Rightarrow \quad \begin{array}{l} y_1 = \frac{\sqrt{3}}{\sqrt{3}+1} \\ y_2 = \sqrt{3} \end{array}$$

which in turn gives the optimal data speeds

$$\hat{x}_1 = \frac{\sqrt{3}+1}{3+2\sqrt{3}} \qquad \hat{x}_2 = \frac{\sqrt{3}}{\sqrt{3}+1}, \qquad \hat{x}_3 = \frac{1}{\sqrt{3}}$$

Quadratic optimization with inequality constraints

$$
\begin{aligned}
\text{minimize} \quad & \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x} + c_0 \\
\text{s.t.} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b}.
\end{aligned} \tag{4}
$$

If $\mathbf{H}$ is positive semi-definite, then this is a convex optimization problem and we can apply Theorem 5.

Theorem 6. $\hat{\mathrm{x}}$ *is a (global) optimal solution to (4) if, and only if, there exists a vector* $\hat{\mathbf{y}} \in \mathbf{R}^m$ *such that*

(1) $\mathbf{H}\hat{\mathbf{x}} + \mathbf{c} = \mathbf{A}^\mathsf{T}\hat{\mathbf{y}}$

(2) $\mathbf{A}\hat{\mathbf{x}} \geq \mathbf{b}$,

(3) $\hat{\mathbf{y}} \geq 0$,

(4) $\hat{\mathbf{y}}^\mathsf{T}(\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}) = 0$.

Example: Continued from last time

$$\begin{aligned} \text{minimize} \quad & (x_1 - 3)^2 + (x_2 - 2)^2 \\ \text{s.t.} \quad & 2x_1 + x_2 - 6 \le 0, \\ & x_1 + 2x_2 - 6 \le 0 \end{aligned}$$

Here

$$\mathbf{H} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} -6 \\ -4 \end{bmatrix}, \quad c_0 = 13, \quad \mathbf{A} = \begin{bmatrix} -2 & -1 \\ -1 & -2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -6 \\ -6 \end{bmatrix}$$

We just check that the solution $\hat{\mathbf{x}} = (11/5, 8/5)$ and $\hat{\mathbf{y}} = (4/5, 0)$ from last time satisfies the global optimality criterum in Theorem 6.

$\mathbf{H}\hat{\mathbf{x}} + \mathbf{c} = \mathbf{A}^{\mathsf{T}}\hat{\mathbf{y}}$

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 11/5 \\ 8/5 \end{bmatrix} + \begin{bmatrix} -6 & -4 \end{bmatrix} = \begin{bmatrix} -8/5 \\ -4/5 \end{bmatrix} = \begin{bmatrix} -2 & -1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 4/5 \\ 0 \end{bmatrix}$$

$\mathbf{A}\hat{\mathbf{x}} \geq \mathbf{b}$, $$\begin{bmatrix} -2 & -1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 11/5 \\ 8/5 \end{bmatrix} = \begin{bmatrix} -6 \\ -27/5 \end{bmatrix} \geq \begin{bmatrix} -6 \\ -6 \end{bmatrix}$$

$\hat{\mathbf{y}} \geq 0$, $$\hat{y} = (4/5, 0) \geq 0$$

$\hat{\mathbf{y}}^{\mathsf{T}}(\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}) = 0.$

$$\begin{bmatrix} 4/5 & 0 \end{bmatrix} \left(\begin{bmatrix} -2 & -1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 11/5 \\ 8/5 \end{bmatrix} - \begin{bmatrix} -6 \\ -6 \end{bmatrix} \right) = \begin{bmatrix} 4/5 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 3/5 \end{bmatrix} = 0$$

Reading instructions

- Chapters 21-22 in the book.

End of the course, what's next?

Courses given at the division

- SF2863 Systems Engineering (Per2)
- SF2812 Applied Linear Optimization (Per3)
- SF2822 Applied Nonlinear Optimization (Per4)
- SF2812 Mathematical Systems Theory (Per2)
- SF2842 Geometric Control Theory (Per3)
- SF2852 Optimal Control Theory (Per4)