Feasible directions

Consider the problem:

\[
\min_{x \in U} f(x),
\]

where \( U = \{x : g_i(x) \leq 0, \ i = 1, \ldots, m\} \subset \mathbb{R}^n \) is a nonempty set and \( \tilde{x} \in \text{cl } U \) (the closure of \( U \)).

- **The cone of feasible directions** of \( U \) at \( \tilde{x} \)

\[
D = \{d : d \neq 0, \tilde{x} + td \in U, \forall t \in [0, \delta) \text{ for some } \delta > 0\}
\]

- **The cone of improving directions**

\[
F = \{d : f(\tilde{x} + td) < f(\tilde{x}), \forall t \in [0, \delta)\}
\]

- Define \( F_0 = \{d : \nabla f(\tilde{x})^\top d < 0\} \)
**Thm. 6.1**

Assume that \( f(x) \) is differentiable at point \( \tilde{x} \in U \).

1. If \( \tilde{x} \) is a local minimum, then \( F_0 \cap D = \emptyset \).

2. Conversely, suppose that \( F_0 \cap D = \emptyset \), \( f(x) \) is convex and at \( \tilde{x} \) there exist \( \epsilon > 0 \) such that

\[
d = x - \tilde{x} \in D, \forall x \in U \cap B(\tilde{x}, \epsilon)
\]

where \( B(\tilde{x}, \epsilon) = \{ x : \| x - \tilde{x} \| < \epsilon \} \). Then \( \tilde{x} \) is a local minimum.

**Obs!** For convex functions \( F_0 = F \). Generally this does not hold.
Algebraic characterization of $D$

- Problem NLP

Minimize $f(x)$
subject to $g_i(x) \leq 0$, $i = 1, \ldots, m$.

- The index set of the active constraints at $\tilde{x}$:

$$I(\tilde{x}) = \{i : g_i(\tilde{x}) = 0\}.$$

Lemma 6.2
If the functions $g_i(x)$, $i \in I$, are differentiable at $\tilde{x}$ and the functions $g_i$ continuous for $i \notin I(\tilde{x})$, then

$$D = \{d : \nabla g_i(\tilde{x})^\top d \leq 0, i \in I(\tilde{x})\} = G_0.$$
Optimality criterion

**Thm. 6.3**

Assume that \( \tilde{x} \in U \) and that \( f(x) \) and \( g_i(x) \), \( i = 1, \ldots, m \), are differentiable at \( \tilde{x} \).

1. If \( \tilde{x} \in U \) is a local minimum, then \( F_0 \cap G_0 = \emptyset \)

2. If \( F_0 \cap G_0 = \emptyset \), \( f(x) \) is convex and \( g_i(x) \), \( i \in I(\tilde{x}) \) are differentiable at \( \tilde{x} \) and \( g_i \) continuous for \( i \notin I(\tilde{x}) \), then \( \tilde{x} \) is a local minimum.

**Example 1**

Minimize \( (x_1 - 1)^2 + (x_2 - 1)^2 \)

Subject to \( x_1 + x_2 - 1 \leq 0 \)
Farkas lemma

Let $A \in \mathbb{R}^{m \times n}$. Then only one of the following alternatives hold:

1. There exists a vector $d \in \mathbb{R}^n$ such that $Ad < 0$

2. There exists a nonzero vector $p \in \mathbb{R}_+^m$ such that $A^T p = 0$. 
Karush-Kuhn-Tucker conditions

Thm. 6.4

Let \( \tilde{x} \) be a feasible solution and \( l(\tilde{x}) \) as before. Suppose that \( f(x) \) and \( g_i(x) \) are differentiable and that the vectors \( \{\nabla g_i(\tilde{x}), i \in l(\tilde{x})\} \) at \( \tilde{x} \) are linearly independent. If \( \tilde{x} \) is a local minimum, then there exists scalars \( u_i, i \in l(\tilde{x}) \) such that

\[
\nabla f(\tilde{x}) + \sum_{i \in l(\tilde{x})} u_i \nabla g_i(\tilde{x}) = 0
\]

\[
u_i \geq 0, \ i \in l(\tilde{x}).\]
Proof of KKT-conditions

(i) Thm. 6.3 ⇒ there does not exist a vector \( d \in F_0 \cap D \), ie. the set

\[
\{ d : \nabla f(\tilde{x})^\top d < 0, \nabla g_i(\tilde{x})^\top d < 0, \ i \in I(\tilde{x}) \} = \emptyset
\]

(ii) Define matrix \( A = [\nabla f(\tilde{x}), \nabla g_I(\tilde{x})(\tilde{x})]^\top \). Hence the system \( Ad < 0 \) is inconsistent. Then by the Farkas lemma the system

\[
p \geq 0, \ A^\top p = 0, \ p \neq 0
\]

has a solution.

(iii) Denote \( p = [p_0, \ p_I(\tilde{x})] \). Then by (ii) we have

\[
p_0 \nabla f(\tilde{x}) + \sum_{i \in I(\tilde{x})} p_i \nabla g_i(\tilde{x}) = 0.
\]
Proof continues

(iv) We must have $p_0 > 0$, because otherwise we violate the linear independence of $\nabla g_i(\tilde{x}), \ i \in I(\tilde{x})$.

(v) The statement follows by setting $u_i = \frac{p_i}{p_0}$.

- The scalars $u_i$ are the Lagrangian multipliers.
- $\tilde{x} \in U$ is the primal feasibility condition.
- The dual feasibility condition:

$$\nabla f(\tilde{x}) + \sum_{i \in I(\tilde{x})} u_i \nabla g_i(\tilde{x}) = 0, \ u_i \geq 0, \ i \in I(\tilde{x})$$
Alternative form of KKT

- If \( i \not\in I(\tilde{x}) \), then set \( u_i = 0 \).

- KKT-conditions:

\[
\nabla f(\tilde{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\tilde{x}) = 0
\]

\[
u_i \geq 0, \ i = 1, \ldots, m
\]

\[
u_i g_i(\tilde{x}) = 0, \ i = 1, \ldots, m.
\]

- in vector form

\[
\nabla f(\tilde{x}) + \nabla g(\tilde{x})^\top u = 0
\]

\[
u^\top g(\tilde{x}) = 0
\]

\[
u \geq 0.
\]
KKT sufficient conditions

**Thm. 6.5**

Let $\tilde{x}$ be a KKT solution and assume that all constraints are active. Then

(a) If there exists $\epsilon > 0$ such that the functions $f$ and $g_i(x)$ are convex differentiable over $B(\tilde{x}, \epsilon) \cap U$, then $\tilde{x}$ is a local minimum.

(b) Assume that $f$ and $g_i(x)$ are differentiable. If $f$ is pseudoconvex, e.g. the condition $\nabla f(z)^\top (y - z) \geq 0$ implies that $f(y) \geq f(z)$, and the level sets of $g_i(x)$, $i = 1, \ldots, m$ are convex. Then $\tilde{x}$ is the global optimal solution.

**Obs!** For convex optimization problem the KKT conditions are also sufficient conditions.
Proof of the sufficiency

Let $x \in U$ be any point. Then

$$f(\tilde{x}) \leq f(\tilde{x}) - \sum_{i=1}^{m} \lambda_i g_i(x) \quad (\lambda_i \geq 0, \ g_i(x) \leq 0)$$

$$= f(\tilde{x}) - \sum_{i \in I(\tilde{x})} \lambda_i (g_i(x) - g_i(\tilde{x}))$$

$$\leq f(\tilde{x}) - \sum_{i \in I(\tilde{x})} \lambda_i \nabla g_i(\tilde{x})^\top (x - \tilde{x}), \ (g_i \text{ convex })$$

$$= f(\tilde{x}) + \nabla f(\tilde{x})^\top (x - \tilde{x}), \ (\text{KKT})$$

$$\leq f(x) \quad (f \text{ is convex}) \quad \square$$
Interpretation

- KKT-conditions express: There exists a function

\[ F_{\tilde{x}}(x) = f(x) + \sum_{i=1}^{m} \lambda_i(\tilde{x})g_i(x) \]

which depends on \( \tilde{x} \) such that

\[
\begin{cases}
  f(\tilde{x}) = \inf_{x \in U} f(x) \Rightarrow \nabla_x F_{\tilde{x}}(\tilde{x}) = 0 \\
  f(\tilde{x}) = F_{\tilde{x}}(\tilde{x})
\end{cases}
\]

- If \( \lambda(\tilde{x}) \) is known, we end up to the necessary condition for a unconstrained optimization.
Saddle point

Let $V$ and $M$ be any sets and

\[ L : V \times M \rightarrow \mathbb{R} \]

a function.

The point $(\tilde{x}, \lambda) \in V \times M$ is a saddle point, if

\[ \sup_{\mu \in M} L(\tilde{x}, \mu) = L(\tilde{x}, \lambda) = \inf_{x \in V} L(x, \lambda) \]
Saddle point property

**Thm. 6.6**

If $(\tilde{x}, \lambda)$ is a saddle point of a function $L : V \times M \to \mathbb{R}$, then

$$\sup_{\mu \in M} \inf_{x \in V} L(x, \mu) = L(\tilde{x}, \lambda) = \inf_{x \in V} \sup_{\mu \in M} L(x, \mu).$$

**Proof:**

- $\forall (\bar{x}, \bar{\mu}) \in V \times M : \inf_{x \in V} L(x, \bar{\mu}) \leq L(\bar{x}, \bar{\mu}) \leq \sup_{\mu \in M} L(\tilde{x}, \mu)$
- Since $(\tilde{x}, \lambda)$ is a saddle point; we have

$$\inf_{x \in V} \sup_{\mu \in M} L(x, \mu) \leq \sup_{\mu \in M} L(\tilde{x}, \mu) = L(\tilde{x}, \lambda) = \inf_{x \in V} L(x, \lambda) \leq \sup_{\mu \in M} \inf_{x \in V} L(x, \mu).$$
A general non-linear programming problem

- The cost function $f : \mathbb{R}^n \rightarrow \mathbb{R}$
- The constraint set
  \[ U = \{ x \in \mathbb{R}^n : g_i(x) \leq 0, 1 \leq i \leq m \} . \]
- The problem (P):
  \[ f(\tilde{x}) = \inf_{x \in U} f(x) . \]
- Every solution $\tilde{x}$ of the problem (P) is also the first argument of a saddle point $(\tilde{x}, \lambda)$ of a suitable Lagrangian function $L : \mathbb{R}^n \times \mathbb{M} \rightarrow \mathbb{R}$.
- Conversely; if $(\tilde{x}, \lambda)$ is a saddle point of this Lagrangian, then $\tilde{x}$ is a solution of (P).
- The second argument $\lambda$ is nothing else than the vector from $\mathbb{R}^m_+$ which appears in the KKT-conditions.
The Lagrangian

The Lagrangian associated with the problem (P)

\[ L(x, \mu) = f(x) + \sum_{i=1}^{m} \mu_i g_i(x). \]

Thm. 6.7

1. If \((\tilde{x}, \lambda) \in \mathbb{R}^n \times \mathbb{R}_+^m\) is a saddle point of the Lagrangian \(L\), then the point \(\tilde{x} \in U\) is a solution of the problem (P).

2. Suppose that the functions \(f(x)\) and \(g_i(x)\) are convex, differentiable at a point \(\tilde{x}\) in which the constraints are satisfied. Then, if \(\tilde{x}\) is a solution of (P), there exists at least one vector \(\lambda \in \mathbb{R}_+^m\) such that the pair \((\tilde{x}, \lambda)\) is a saddle point of the Lagrangian.
The associated unconstrained problem

- If one knew $\lambda$, then the constrained problem $(P)$ would be replaced by the unconstrained problem $(P_\lambda)$:

$$
\text{find } x_\lambda \text{ such that } x_\lambda \in \mathbb{R}^n : L(x_\lambda, \lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda)
$$

- How do we find $\lambda \in \mathbb{R}^m_+$?

- The dual problem $(Q)$

$$
\text{find } \lambda \text{ such that } \lambda \in \mathbb{R}^m_+ : G(\lambda) = \sup_{\mu \in \mathbb{R}^m_+} G(\mu),
$$

where

$$
G(\mu) = \inf_{x \in \mathbb{R}^n} L(x, \mu)
$$
The dual problem

- The dual problem is a constrained optimization problem;
- But the constraints are easily dealt with:
  \[ \mu_i \geq 0, \ i = 1, \ldots, m \]
- The associated projection operator is quite simple
  \[ P_+(\mu_i) = \max\{0, \mu_i\}. \]
- The constraints \( g_i(x) \leq 0 \) are, in general, impossible to deal with numerically.
1. Suppose that the constraint functions $g_i(x)$, $i = 1, \ldots, m$, are continuous and that for every $\mu \in \mathbb{R}^m_+$ the problem $(P_\mu)$ has a unique solution $x_\mu$ which depends continuously on the dual variable $\mu$. Then, if $\lambda$ is any solution of the dual problem $(Q)$, the solution $x_\lambda$ of the corresponding problem $(P_\lambda)$ is a solution of the primal problem $(P)$.

2. Suppose that the primal problem $(P)$ has at least one solution $\tilde{x}$, and that the functions $f(x)$, $g_i(x)$ are convex and differentiable and $\tilde{x} \in U$. Then the dual problem $(Q)$ has at least one solution.
Quadratic optimization problem

- The quadratic functional

\[ f(x) = \frac{1}{2} x^T A x - b^T x, \]

where \( A \) is
- symmetric
- positive definite
- \( b \) a given vector

- The constraint set

\[ U = \{ x \in \mathbb{R}^n : C x \leq d \} \]

where \( C \in \mathbb{R}^{m \times n}, \ d \in \mathbb{R}^m. \]
The primal-dual formulation

- The Lagrangian functional

\[ L(x, \mu) = \frac{1}{2}x^\top Ax - b^\top x + \mu^\top (Cx - d) \]

\[ = \frac{1}{2}x^\top Ax - (b - C^\top \mu)^\top x - \mu^\top d \]

- Thm 6.7 \( \Rightarrow \) The primal problem (P) has a unique solution \( \tilde{x} \)
- Thm 6.7 \( \Rightarrow \) it has at least one saddle point \( (\tilde{x}, \lambda) \)
- Uniqueness of the saddle point \( \Leftrightarrow \) uniqueness of the dual problem
- We have to verify: For every \( \mu \in \mathbb{R}_+^m \) the problem \( (P_\mu) \) has a unique solution.
**Uniqueness of \((P_\mu)\)**

The function \(L(\cdot, \mu)\) is quadratic

\[
L(x, \mu) = \frac{1}{2} x^\top A x - b_\mu^\top x + c_\mu
\]

and it has a unique minimum \(x_\mu\) satisfying

\[
A x_\mu = b - C^\top \mu \iff x_\mu = A^{-1}(b - C^\top \mu)
\]

which depends continuously on \(\mu\).
The dual function $G(\mu)$

- By the uniqueness of $x_\mu$ and $(Ax_\mu)^\top x = b_\mu^\top x$ we have

$$G(\mu) = L(x_\mu, \mu) = \frac{1}{2} x_\mu^\top Ax_\mu - b_\mu^\top x_\mu + c_\mu = -\frac{1}{2} b_\mu^\top x_\mu$$

$$= -\frac{1}{2} (b - C^\top \mu)^\top x_\mu - \mu^\top d$$

$$= -\frac{1}{2} (b - C^\top \mu)^\top [A^{-1}(b - C^\top \mu)] - \mu^\top d.$$ 

- After some algebraic gymnastics

$$-G(\mu) = \frac{1}{2} \mu^\top CA^{-1} C^\top \mu - (CA^{-1} b - d)^\top \mu + \frac{1}{2} b^\top A b.$$ 

- The matrix $CA^{-1} C^\top$ is symmetric and positive definite if and only if $\text{Rank}(C) = m (\leq n)$.

- When $\text{Rank} C = m \iff \text{Im} C = \mathbb{R}^m \iff \text{N}(C) = 0$, the dual problem has a unique solution.
Duality gap

Primal problem

\[ p^* = \min_{\mathbf{g}(x) \leq 0} f(x) \]

Dual problem

\[ d^* = \max_{\mathbf{u} \geq 0} G(\mathbf{u}) \]

- **Weak duality**: \( d^* \leq p^* \) is always valid, even for non-convex problems.
- **Strong duality**: \( d^* = p^* \).
- Strong duality holds for convex problems (Thm. 6.7).
- The optimal duality gap: \( p^* - d^* \geq 0 \).
Tax interpretation of duality

- The variables \( x_i, \ i = 1, \ldots, n \), describe the operation mode of a company.
- The operation cost is \( f(x) \).
- Each of the constraints \( g_j(x) \leq 0, \ j = 1, \ldots, m \), represent some resource or regulatory limit (labour, warehouse, raw materials, etc.).
- **Problem**: Find the optimal operation mode \( x \) for the company under given limitations

\[
p^* = \min_{g(x) \leq 0} f(x),
\]

where \( g(x) = [g_1(x), g_2(x), \ldots, g_m(x)] \).
Penalty for violations

- Assumption: The limits can be violated.
- The penalty is linear in the amount of violation $u_ig_i(x)$:
  
  If $u_ig_i(x) > 0$, then the company pays a penalty,
  If $u_ig_i(x) < 0$, then the company receives money for selling capacity.

- $u_i$ is the price in euros for braking the constraint $g_i(x) \leq 0$ per unit.
- Total cost
  
  $$L(x, u) = f(x) + u^\top g(x).$$
Interpretation

- The dual problem: \( d^* = \min_{u \geq 0} G(u) \).
- **Weak duality**: The optimal cost, if violations are allowed, is less than in the original situation even with the most unfavourable prices \( u_i \).
- **Strong duality**: The dual optimum \( \lambda \) is a set of prices for which there is no advantage for the company to violate the constraints. They are called the *shadow prices*. 
Equality constraints

- **Primal problem:**
  \[
  \min f(x) \\
  g(x) \leq 0 \\
  h(x) = 0
  \]
  where
  \[
  g(x) = \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix}, \quad h(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_p(x) \end{bmatrix}.
  \]

- **Lagrangian:**
  \[
  L(x, u, v) = f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{k=1}^{p} v_k h_k(x).
  \]
KKT optimality conditions

KKT-conditions

\[ \nabla_x L(x, u, v) = \nabla f(x) + \sum_{i=1}^{m} u_i \nabla g_i(x) + \sum_{k=1}^{p} v_k \nabla h_k(x) = 0 \]

\[ u_i g_i(x) = 0, \ u_1 \geq 0, \ g_i(x) \leq 0, \ i = 1, \ldots, m \]

\[ h_k(x) = 0, \ k = 1, \ldots, p \]

The dual function

\[ G(u, v) = \inf_{x \in \mathbb{R}^n} L(x, u, v) \]

is always concave.

Dual problem: \( \max_{u \geq 0, \ v} G(u, v) \).
Uzawa’s method

- The primal problem (P):
  \[
  \min_{x \in U} f(x), \quad (1)
  \]
  where
  \[
  U = \{ x \in \mathbb{R}^n : g_i(x) \leq 0, \ i = 1, \ldots, m \}. 
  \]

- The objective: *Construct an algorithm which enables us to approximate a solution of (P).*

- The dual problem (Q):
  \[
  \max_{\mu \in \mathbb{R}^m} G(\mu) = G(\lambda). 
  \]

- We apply the gradient method to the dual problem using the projection operator \((P_+ \lambda)_i = \max\{\lambda_i, 0\}, \ 1 \leq i \leq m\).

- Generate the sequence \(\{\lambda^{(k)}\}_{k \in \mathbb{N}}:\)
  \[
  \lambda^{(k+1)} = P_+ [\lambda^{(k)} + \rho_k \nabla G(\lambda^{(k)})], \ k \geq 0
  \]
The gradient of the dual function

- Let $\lambda \in \mathbb{R}_+^m$ and $\lambda + \mu \in \mathbb{R}_+^m$.
- $x_\lambda$, $x_{\lambda+\mu}$ solutions of the problem
  \[
  \min_x f(x) + \alpha^T g(x), \; \alpha = \lambda \; \text{or} \; \lambda + \mu.
  \]
- Then we have the inequalities
  \[
  L(x_\lambda, \lambda) \leq L(x_{\lambda+\mu}, \lambda), \; L(x_{\lambda+\mu}, \lambda + \mu) \leq L(x_\lambda, \lambda + \mu)
  \]
  \[
  \downarrow
  \]
  \[
  \lambda^T g(x_{\lambda+\mu}) \leq G(x_{\lambda+\mu}) - G(\lambda) \leq \mu^T g(x_\lambda).
  \]
The gradient of the dual function

Then there exists $0 \leq \theta \leq 1$ such that

$$G(\lambda + \mu) - G(\lambda) = \mu^\top g(x_\lambda) + \theta \mu^\top [g(x_{\lambda+\mu}) - g(x_\lambda)].$$

Assumption: $g_i(x)$, $1 \leq i \leq m$, are continuous. Then

$$G(\lambda + \mu) - G(\lambda) = \mu^\top g(x_\lambda) + ||\mu||\epsilon(\mu), \quad \lim_{\mu \to 0} \epsilon(\mu) = 0.$$ 

The gradient of the dual function is $\nabla G(\lambda) = g(x_\lambda)$.

$x_\lambda$ is the solution of the unconstrained problem

$$\min_x f(x) + \lambda^\top g(x)$$
Uzawa’s algorithm

- Initialization: Choose any $\lambda^{(0)} \in \mathbb{R}^m_+$
- Define a sequence of pairs $(x^{(k)}, \lambda^{(k)}) \in \mathbb{R}^n \times \mathbb{R}^m_+$ by means of the recurrence formulae:
  
  \[ \text{Calculate } x^{(k)} : \min_x f(x) + (\lambda^{(k)})^T g(x) \]
  
  \[ \text{Calculate } \lambda^{(k+1)} : \lambda_i^{(k+1)} = \max\{\lambda_i^{(k)} + \rho_k g_i(x^{(k)}), 0\} \]

- The stepsize $\rho_k$ is defined by the gradient method
The convergence of the Uzawa

- $f(x)$ elliptic: $(\nabla f(x) - \nabla f(y))^\top (x - y) \geq \alpha \|x - y\|^2$
- The constraints are linear inequalities:

$$g(x) \leq 0 \iff Cx - d \leq 0$$

- The norm of $C$: $\|C\| = \max_{x \in \mathbb{R}^n} \frac{\|Cx\|_m}{\|x\|_n}$

**Thm 6.9**

If

$$0 < \rho_k < \frac{2\alpha}{\|C\|^2},$$

then the sequence $\{x^{(k)}\}_{k \in \mathbb{N}}$ converges to the unique solution of the primal problem (P). Furthermore, if $\text{Rank}(C) = m$, then also $\lambda^{(k)} \to \lambda$ which is the unique solution of the dual problem (Q).
Example 1

Solve
\[
\min_{x_1 + x_2 \geq 2} x_1^2 + x_2^2
\]
using the Uzawa’s method.

**Solution:** Choose \( \lambda_0 = 1 \) and the stepsize \( \rho = 1 \)

1. \( \min_x x_1^2 + x_2^2 - x_1 - x_2 + 2 \Rightarrow \begin{cases} 2x_1 - 1 = 0 \\ 2x_2 - 1 = 0 \end{cases} \Rightarrow x^{(0)} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} \)

Projection phase:
\( \lambda_1 = \max\{\lambda_0 + \rho \nabla G(\lambda_0), 0\} = \max\{1 + (2 - x_1^{(0)} - x_2^{(0)}), 0\} = 2 \)

2. \( \min_x x_1^2 + x_2^2 - 2x_1 - 2x_2 + 2 \Rightarrow \begin{cases} 2x_1 - 2 = 0 \\ 2x_2 - 2 = 0 \end{cases} \Rightarrow x^{(1)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \)

Projection phase: \( \lambda_2 = \max\{2 + (2 - x_1^{(1)} - x_2^{(1)}), 0\} = 2. \) STOP!!!
The optimal solution has been found.
Example 2

\[
\begin{align*}
\min_{x_1^2 + x_2^2 - 1 \leq 0} & \quad x_1 + x_2 \\
\end{align*}
\]

The steps in Uzawa’s algorithm:

- \( \min_x x_1 + x_2 + \lambda(x_1^2 + x_2^2 - 1) \Rightarrow x_\lambda = \begin{bmatrix} -\frac{1}{2\lambda} \\ -\frac{1}{2\lambda} \end{bmatrix} \)
- The gradient of the dual function: \( \nabla G(\lambda) = g(x_\lambda) = \frac{1}{2\lambda^2} - 1 \)
- The projection: \( \lambda_{\text{next}} = \max\{\lambda + \rho(\frac{1}{2\lambda^2} - 1), 0\} \)
- What is the correct step size?
The step size $\rho = 1$

Matlab-code:

```matlab
>> r = 1; (the stepsize)
>> l(1) = 1; ($\lambda_0 = 1$)
>> for k = 1 : 20
    l(k+1) = max(l(k) + r/2 * l(k)^(-2) - r, 0);
end
```

The result

$$
\lambda = \begin{bmatrix}
1.0000 & 0.9013 & 0.5493 \\
0.5 & 0.5168 & 1.2064 \\
1.5 & 1.3888 & 0.5499 \\
0.7222 & 0.6481 & 1.2032 \\
0.6808 & 0.8386 & 0.5486 \\
0.7596 & 0.5496 & 1.21 \\
0.6262 & 1.2049 & 0.5515 \\
\end{bmatrix}
$$
$\rho = \frac{1}{2} \Rightarrow \text{Convergence}$

$$\Lambda = \begin{bmatrix}
1.000000 & 0.707271 & 0.707106 \\
0.750000 & 0.707039 & 0.707107 \\
0.694444 & 0.707135 & 0.707107 \\
0.712844 & 0.707095 & 0.707107 \\
0.704828 & 0.707112 & 0.707107 \\
0.708066 & 0.707105 & 0.707107 \\
0.706712 & 0.707108 & 0.707107
\end{bmatrix}$$
Linearly constrained quadratic optimization

- In the case of quadratic functional

\[ f(x) = \frac{1}{2} x^\top A x - b^\top x \]

subject to constraints \( U = \{ x : C x - d \leq 0 \} \).

- One step at the Uzawa’s algorithm is

  calculate \( x^{(k)} \):
  \[ Ax^{(k)} - b + C^\top \lambda^{(k)} = 0 \]

  calculate \( \lambda^{(k+1)} \):
  \[ \lambda^{(k+1)} = \max\{ \lambda^{(k)} + \rho(Cx^{(k)} - d), 0 \} \]

- If \( A \) is symmetric and positive definite, then the Uzawa’s algorithm converges provided

\[ 0 < \rho < \frac{2\sigma_1(A)}{\|C\|^2}, \]

where \( \sigma_1(A) \) is the smallest eigenvalue of \( A \).
Proof of Uzawa’s method

1. $f(x)$ elliptic;
2. $U$ is non-empty and convex

The problems

$$\min_{x \in U} f(x), \quad \min_{x \in \mathbb{R}^n} f(x) + u^\top (Cx - d)$$

have unique solutions $\tilde{x}$ and $x(u)$. Moreover, the dual problem (Q) $\min_{u \geq 0} G(u)$ has at least one solution (Thm. 6.7)
Proof continues

Some facts:

1. There exists at least one $\lambda \in \mathbb{R}^m_+$ such that the pair $(x(\lambda), \lambda)$ is a saddle point of the Lagrangian $L(x, u)$.

2. $\nabla f(x(\lambda)) + C^T \lambda = 0$.

3. $(Cx(\lambda) - d)^T (u - \lambda) \leq 0$, $\forall u \in \mathbb{R}^m_+$, since $L(x(\lambda), \lambda) = \max_{u \in \mathbb{R}^m_+} L(x(\lambda), u)$

The last relation can be written as:

$$(\lambda - (\lambda + \rho(Cx(\lambda) - d))^T (u - \lambda) \geq 0, \forall u \in \mathbb{R}^m_+.$$ 

So: $\lambda$ is a projection of $\lambda + \rho \nabla G(\lambda)$ to $\mathbb{R}^m_+$. 

Proof

In short:
\[
\begin{align*}
\nabla f(x(\lambda)) + C^\top \lambda &= 0 \\
\lambda &= P_+(\lambda + \rho(Cx(\lambda) - d)).
\end{align*}
\]

By the definition of Uzawa’s algorithm:
\[
\begin{align*}
\nabla f(x^{(k)}) + C^\top \lambda^{(k)} &= 0 \\
\lambda^{(k+1)} &= P_+(\lambda^{(k)} + \rho(Cx^{(k)} - d)).
\end{align*}
\]

Hence we obtain (projection does not increase distances)
\[
\begin{align*}
\nabla (f(x^{(k)}) - f(x(\lambda))) + C^\top (\lambda^{(k)} - \lambda) &= 0 \\
\|\lambda^{(k+1)} - \lambda\| &\leq \|\lambda^{(k)} - \lambda + \rho(C(x^{(k)} - x(\lambda)))\|.
\end{align*}
\]
Convergence of the sequence \( x^{(k)} \)

Taking the squares and using the ellipticity:

\[
\| \lambda^{(k+1)} - \lambda \|^2 \leq \| \lambda^{(k)} - \lambda \|^2 + 2\rho (C^\top (\lambda^{(k)} - \lambda)) \top (x^{(k)} - x(\lambda)) \\
+ \rho \| C(x^{(k)} - x(\lambda)) \|^2 \\
= \| \lambda^{(k)} - \lambda \|^2 \\
+ 2\rho (\nabla (f(x^{(k)}) - f(x(\lambda)))) \top (x^{(k)} - x(\lambda)) \\
+ \rho \| C(x^{(k)} - x(\lambda)) \|^2 \\
\leq \| \lambda^{(k)} - \lambda \|^2 - \rho \{2\alpha - \rho \| C \|^2 \} \| x^{(k)} - x(\lambda) \|^2
\]

\[
0 \leq \frac{2\alpha}{\| C \|^2} \Rightarrow \| \lambda^{(k+1)} - \lambda \| \leq \| \lambda^{(k)} - \lambda \|
\]
Convergence

The sequence $\{\|\lambda^{(k)} - \lambda\|\}_k$ is monotonic, decreasing and bounded below

$$\Rightarrow \lim_{k \to \infty} \{\|\lambda^{(k+1)} - \lambda\|^2 - \|\lambda^{(k)} - \lambda\|^2\} = 0$$

On the other hand

$$\rho\{2\alpha - \rho\|C\|^2\}\|x^{(k)} - x(\lambda)\|^2 \leq \|\lambda^{(k+1)} - \lambda\|^2 - \|\lambda^{(k)} - \lambda\|^2 \to 0$$

as $k \to \infty$. 

Convergence of \( \{ \lambda^{(k)} \} \) 

- \( \| \lambda^{(k)} - \lambda \| \) converges \( \Rightarrow \) \( \{ \lambda^{(k)} \} \) is bounded.
- There exist a converging subsequence \( \{ \lambda^{(l)} \} \), which converges to element \( \tilde{\lambda} \).
- \( \nabla f(x(\lambda)) + C^\top \tilde{\lambda} = 0 \)
- If \( \text{Rank}(C) = m \), then the dual problem has a unique solution \( \lambda \) for which (the primal problem is uniquely solvable) \( \nabla f(x(\lambda)) + C^\top \lambda = 0 \)
- \( \lambda = \tilde{\lambda} \), because \( \text{Ker}(C) = 0 \).
- The convergence follows for the whole sequence.