Support Vector Machines - I

Prof. Dan A. Simovici

UMB
1 Preliminaries – Gradients

2 Optimization

3 General Optimization Problems

4 Lagrange Multipliers

5 Duality
Let $E$ be a subset of $\mathbb{R}$.
A function $f : E \to \mathbb{R}$ has a \textbf{maximum} $M$ on $E$ if there exists $x_0 \in E$ such that $f(x_0) = M$ and $f(x_1) \leq M$ for every $x_1 \in E$. The element $x_0$ is a \textbf{maximizer} of $f$ on $E$.
Similarly, $f : E \to \mathbb{R}$ has a \textbf{minimum} $m$ on $E$ if there exists $x_0 \in E$ such that $f(x_0) = m$ and $f(x_1) \geq m$ for every $x_1 \in E$. The element $x_0$ is a \textbf{minimizer} of $f$ on $E$. 
• If \( f : [a, b] \rightarrow \mathbb{R} \) and \( f \) is continuous, then \( f \) has a global maximum \( M \) and a global minimum \( m \) on \([a,b]\).
• If \( f \) has a derivative on \([a, b]\), and \( f'(x_0) = 0 \), then \( x_0 \) is a critical point of \( f \).
• A local extremum (minimum or maximum) can occur only at a critical point \( x_0 \). If \( f''(x_0) < 0 \), the critical point provides a local maximum; if \( f''(x_0) > 0 \) the critical point provides a local minimum.
The $\nabla f$ notation

(read “nabla f”).

Let $f : X \rightarrow \mathbb{R}$, where $X \subseteq \mathbb{R}^n$, and let $z \in X$. The **gradient** of $f$ in $z$ is the vector

$$
(\nabla f)(z) = 
\begin{pmatrix}
\frac{\partial f}{\partial x_1}(z) \\
\vdots \\
\frac{\partial f}{\partial x_n}(z)
\end{pmatrix}
\in \mathbb{R}^n.
$$
Example

Let \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) be the function \( f(x) = x_1^2 + \cdots + x_n^2 \); in other words, \( f(x) = \|x\|^2 \).

We have

\[
\frac{\partial f}{\partial x_1} = 2x_1, \ldots, \frac{\partial f}{\partial x_n} = 2x_n.
\]

Therefore, \((\nabla f)(x) = 2x\).
Example

Let \( b_j \in \mathbb{R}^n \) and \( c_j \in \mathbb{R} \) for \( 1 \leq j \leq n \), and let \( f : \mathbb{R}^n \to \mathbb{R} \) be the function

\[
\mathbf{f(x)} = \sum_{j=1}^{n} (\mathbf{b}_j^T \mathbf{x} - c_j)^2.
\]

We have \( \frac{\partial f}{\partial x_i}(\mathbf{x}) = \sum_{j=1}^{n} 2b_{ij}(\mathbf{b}_j^T \mathbf{x} - c_j) \), where \( \mathbf{b}_j = (b_{1j} \cdots b_{nj}) \) for \( 1 \leq j \leq n \). Thus, we obtain:

\[
(\nabla f)(\mathbf{x}) = 2 \begin{pmatrix} 
\sum_{j=1}^{n} 2b_{1j}(\mathbf{b}_j^T \mathbf{x} - c_j) \\
\vdots \\
\sum_{j=1}^{n} 2b_{nj}(\mathbf{b}_j^T \mathbf{x} - c_j)
\end{pmatrix} = 2(\mathbf{B}' \mathbf{x} - \mathbf{c}')\mathbf{B} = 2\mathbf{B}' \mathbf{xB} - 2\mathbf{c}' \mathbf{B},
\]

where \( \mathbf{B} = (\mathbf{b}_1 \cdots \mathbf{b}_n) \in \mathbb{R}^{n \times n} \).
The matrix-valued function $H_f : \mathbb{R}^k \rightarrow \mathbb{R}^{k \times k}$ defined by

\[ H_f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_{i_1} \partial x_{i_2}} \end{pmatrix} \]

is the \textit{Hessian matrix} of $f$. 
**Example**

For the function $f(x) = x_1^2 + \cdots + x_n^2$ discussed on Slide 6 we have

$$H_f(x) = \begin{pmatrix} 2 & 0 & \cdots & 0 \\ 0 & 2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 2 \end{pmatrix}.$$
Definition

Let $X$ be a open subset in $\mathbb{R}^n$ and let $f : X \rightarrow \mathbb{R}$ be a function. The point $x_0 \in X$ is a \textit{local minimum} for $f$ if there exists $\delta > 0$ such that $B(x_0, \delta) \subseteq X$ and $f(x_0) \leq f(x)$ for every $x \in B(x_0, \delta)$.

The point $x_0$ is a \textit{strict local minimum} if $f(x_0) < f(x)$ for every $x \in B(x_0, \delta) \setminus \{x_0\}$.
Definition

A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive semidefinite if $x'Ax \geq 0$ for all $x \in \mathbb{R}^n$.

$A$ is positive definite if $x'Ax > 0$ for all $x \in \mathbb{R}^n - \{0_n\}$.
Example

The symmetric real matrix

\[ A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \]

is positive definite if and only if \( a > 0 \) and \( b^2 - ac < 0 \). Indeed, we have \( x'Ax > 0 \) for every \( x \in \mathbb{R}^2 - \{0\} \) if and only if \( ax_1^2 + 2bx_1x_2 + cx_2^2 > 0 \), where \( x' = (x_1 \ x_2) \); elementary algebra considerations lead to \( a > 0 \) and \( b^2 - ac < 0 \).
Is the matrix

\[ A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \]

positive definite?

No, because 

\[ (x_1 \ x_2) \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 + 4x_1x_2 + x_2^2 \]

can be made negative with \( x_1 = 1 \) and \( x_2 = -1 \).
Theorem

A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive definite if and only if all its leading principal minors are positive.

The leading minors of the previous matrix are 1 and $\begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} = -3$. 
Theorem

Let $f : B(x_0, r) \rightarrow \mathbb{R}$ be a function that belongs to the class $C^2(B(x_0, r))$, where $B(x_0, r) \subseteq \mathbb{R}^k$ and $x_0$ is a critical point for $f$. If the Hessian matrix $H_f(x_0)$ is positive semidefinite, then $x_0$ is a local minimum for $f$; if $H_f(x_0)$ is negative semidefinite, then $x_0$ is a local maximum for $f$. 
Let $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ be a function in $C^2(B(x_0, r))$. The Hessian matrix in $x_0$ is

$$H_f(x_0) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix}(x_0).$$

Let $a_{11} = \frac{\partial^2 f}{\partial x_1^2}(x_0)$, $a_{12} = \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_0)$, and $a_{22} = \frac{\partial^2 f}{\partial x_2^2}(x_0)$. Note that

$$h' H_f(x_0) h = a_{11} h_1^2 + 2a_{12} h_1 h_2 + a_{22} h_2^2 = h_2^2 \left( a_{11} \xi^2 + 2a_{12} \xi + a_{22} \right),$$

where $\xi = \frac{h_1}{h_2}$. 
For a critical point $x_0$ we have:

- $h'H_f(x_0)h \geq 0$ for every $h$ if $a_{11} > 0$ and $a_{12}^2 - a_{11}a_{22} < 0$; in this case, $H_f(x_0)$ is positive semidefinite and $x_0$ is a local minimum;
- $h'H_f(x_0)h \leq 0$ for every $h$ if $a_{11} < 0$ and $a_{12}^2 - a_{11}a_{22} < 0$; in this case, $H_f(x_0)$ is negative semidefinite and $x_0$ is a local maximum;
- if $a_{12}^2 - a_{11}a_{22} \geq 0$; in this case, $H_f(x_0)$ is neither positive nor negative definite, so $x_0$ is a saddle point.

Note that in the first two previous cases we have $a_{12}^2 < a_{11}a_{22}$, so $a_{11}$ and $a_{22}$ have the same sign.
Example

Let \( a_1, \ldots, a_m \) be \( m \) points in \( \mathbb{R}^n \). The function \( f(x) = \sum_{i=1}^{m} \| x - a_i \|^2 \) gives the sum of squares of the distances between \( x \) and the points \( a_1, \ldots, a_m \). We will prove that this sum has a global minimum obtained when \( x \) is the barycenter of the set \( \{ a_1, \ldots, a_m \} \).
Example (cont’d)

We have

\[
    f(x) = m \left\| x \right\|^2 - 2 \sum_{i=1}^{m} a'_i x + \sum_{i=1}^{m} \left\| a_i \right\|^2
\]

\[
    = m(x_1^2 + \cdots + x_n^2) - 2 \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij} x_j + \sum_{i=1}^{m} \left\| a_i \right\|^2,
\]

which implies

\[
    \frac{\partial f}{\partial x_j} = 2mx_j - 2 \sum_{i=1}^{m} a_{ij}
\]

for \(1 \leq j \leq n\). Thus, there exists only one critical point given by

\[
    x_j = \frac{1}{m} \sum_{i=1}^{m} a_{ij}
\]

for \(1 \leq j \leq n\). 

The Hessian matrix $H_f = 2mI_n$ is positive definite, so the critical point is a local minimum and, in view of convexity of $f$, the global minimum. This point is the barycenter of the set $\{a_1, \ldots, a_m\}$. 
Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$, and $d : \mathbb{R}^n \rightarrow \mathbb{R}^p$ be three functions defined on $\mathbb{R}^n$. A general formulation of a constrained optimization problem is:

$$\text{minimize } f(x), \text{ where } x \in \mathbb{R}^n,$$

subject to $c(x) \leq 0_m$, where $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$,

and $d(x) = 0_p$, where $d : \mathbb{R}^n \rightarrow \mathbb{R}^p$. 
Here \( \mathbf{c} \) specifies *inequality constraints* placed on \( \mathbf{x} \), while \( \mathbf{d} \) defines *equality constraints*.

The *feasible region* of the constrained optimization problem is the set

\[
R_{c,d} = \{ \mathbf{x} \in \mathbb{R}^n \mid c(\mathbf{x}) \leq 0_m \text{ and } d(\mathbf{x}) = 0_p \}.
\]

If the feasible region \( R_{c,d} \) is non-empty and bounded, then, under certain conditions a solution exists. If \( R_{c,d} = \emptyset \) we say that the constraints are *inconsistent*. 
If only inequality constraints are present (as specified by the function $c$) the feasible region is:

$$R_c = \{ x \in \mathbb{R}^n \mid c(x) \leq 0_m \}.$$
Let $x \in R_c$. The set of active constraints at $x$ is

$$\text{ACT}(R_c, c, x) = \{ i \in \{1, \ldots, m\} \mid c_i(x) = 0 \}.$$

If $i \in \text{ACT}(R_c, c, x)$, we say that $c_i$ is an active constraint or that $c_i$ is tight on $x \in S$; otherwise, that is, if $c_i(x) < 0$, $c_i$ is an inactive constraint on $x$. 
Definition

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be two functions. The minimization problem $MP(f, c)$ is:

\[
\text{minimize } f(x), \text{ where } x \in \mathbb{R}^n, \\
\text{subject to } x \in R_c.
\]

If $x_0$ exists in $R_c$ that $f(x_0) = \min \{ f(x) \mid x \in R_c \}$ we refer to $x_0$ as a solution of $MP(f, c)$. 
If \( h : \mathbb{R}^n \rightarrow \mathbb{R}^m \) we can write

\[
\begin{pmatrix}
h_1(x) \\
\vdots \\
h_m(x)
\end{pmatrix},
\]

where \( h_j : \mathbb{R}^n \rightarrow \mathbb{R} \) are the components of \( h \) for \( 1 \leq j \leq m \). If \( h \) is a differentiable function, the function \( (Dh)(x) \) is

\[
(Dh)(x) = \begin{pmatrix}
(\nabla h_1)(x)' \\
\vdots \\
(\nabla h_m)(x)'
\end{pmatrix} \in \mathbb{R}^{m \times n}.
\]
Example

Let \( h : \mathbb{R}^2 \rightarrow \mathbb{R}^3 \) be given by

\[
h(x) = \begin{pmatrix} x_1 x_2 \\ x_1^2 \\ x_2^2 \end{pmatrix}
\]

Then

\[
(Dh)(x) = \begin{pmatrix} x_2 & x_1 \\ 2x_1 & 0 \\ 0 & 2x_2 \end{pmatrix}.
\]

Observe that the rows of \((Dh)(x)\) are the gradients of the components of \(h\).
Theorem

(Existence Theorem of Lagrange Multipliers) Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be two functions such that:

- $m < n$,
- $f \in C^1(\mathbb{R}^n)$,
- $h \in C^1(\mathbb{R}^n)$, and
- the matrix $(Dh)(x)$ is of full rank, that is, $\text{rank}((Dh)(x)) = m < n$ (which means that the gradients $(\nabla h_1)(x), \ldots , (\nabla h_m)(x)$ are linearly independent).

If $x_0$ is a local extremum of $f$ subjected to the restriction $h(x_0) = 0_m$, then $(\nabla f)(x_0)$ is a linear combination of $(\nabla h_1)(x_0), \ldots , (\nabla h_m)(x_0)$. 
Example

Suppose that we wish to minimize \( f(x) = x_1 + x_2 \) subject to the condition

\[
h(x) = x_1^2 + x_2^2 = 2.\]

We have

\[
(\nabla f)(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix},
\]

\[
(\nabla h)(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}.
\]
Example continued

At the local minimum $\mathbf{x}^* = (-1, -1)$ we have $(\nabla f)(\mathbf{x}^*) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $(\nabla h) = \begin{pmatrix} -2 \\ -2 \end{pmatrix}$, so

$$(\nabla f)(\mathbf{x}^*) + \frac{1}{2} (\nabla h) = 0.$$
To apply the Lagrange multiplier technique the constraint gradients

$$(
\nabla h_1)(x), \cdots, (\nabla h_m)(x)$$

must be linearly independent. In this case, $x$ is said to be regular.
If a local minimum is not regular Lagrange multipliers may not exist.

**Example**

Consider minimizing the function \( f(x) = x_1 + x_2 \) subject to the constraints

\[
    h_1(x) = (x_1 - 1)^2 + x_2^2 - 1 = 0, \\
    h_2(x) = (x_1 - 2)^2 + x_2^2 - 4 = 0.
\]

We have

\[
    (\nabla f)(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix},
\]

and

\[
    (\nabla h_1)(x) = \begin{pmatrix} 2(x_1 - 1) \\ 2x_2 \end{pmatrix}, \\
    (\nabla h_2)(x) = \begin{pmatrix} 2(x_1 - 2) \\ 2x_2 \end{pmatrix}.
\]
Example continued

The local minimum is at \( \begin{pmatrix} 0 \\ 0 \end{pmatrix} \). At that point, we have

\[
(\nabla f)(0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, (\nabla h_1)(0) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}, (\nabla h_2)(0) = \begin{pmatrix} -4 \\ 0 \end{pmatrix}.
\]

The gradients \((\nabla h_1)(0), (\nabla h_2)(0)\) are not linearly independent because

\[
2(\nabla h_1)(0) + (\nabla h_2)(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix},
\]

so \(0\) is not a regular point and Lagrange's multipliers do not exist.
Example

Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix and let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be the function defined by $f(x) = x'Ax$.

**Optimization problem:** minimize $f$ subjected to the restriction $\|x\| = 1$, or equivalently $h(x) = \|x\|^2 - 1 = 0$.

Since $(\nabla f) = 2Ax$ and $(\nabla h)(x) = 2x$ there exists $\lambda$ such that $2Ax_0 = 2\lambda x_0$ for any extremum of $f$ subjected to $\|x_0\| = 1$. Thus, $x_0$ must be a unit eigenvector of $A$ and $\lambda$ must be an eigenvalue of the same matrix.
The next theorem provides necessary conditions for optimality that
- include the linear independence of the gradients of the components of
  the constraint \((\nabla c_i)(x_0)\) for \(i \in \text{ACT}(S, c, x_0)\), and
- ensure that the coefficient of the gradient of the objective function
  \((\nabla f)(x_0)\) is not null.

These conditions are known as the \textit{Karush-Kuhn-Tucker conditions} or the \textit{KKT conditions}. 
Theorem

(Karush-Kuhn-Tucker Theorem) Let $S$ be a non-empty open subset of $\mathbb{R}^n$ and let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Let $x_0$ be a local minimum in $S$ of $f$ subjected to the restriction $c(x_0) \leq 0_m$.

Suppose that $f$ is differentiable in $x_0$, $c_i$ are differentiable in $x_0$ for $i \in \text{ACT}(S, c, x_0)$, and $c_i$ are continuous in $x_0$ for $i \notin \text{ACT}(S, c, x_0)$. If $\{(\nabla c_i)(x_0) \mid i \in \text{ACT}(S, c, x_0)\}$ is a linearly independent set, then there exist non-negative numbers $w_i$ for $i \in \text{ACT}(S, c, x_0)$ such that

$$(\nabla f)(x_0) + \sum \{w_i(\nabla c_i)(x_0) \mid i \in \text{ACT}(S, c, x_0)\} = 0_n.$$
Theorem continued

Furthermore, if the functions \( c_i \) are differentiable in \( x_0 \) for
\( i \notin \text{ACT}(S, c, x_0) \), then the previous condition can be written as:

- \( (\nabla f)(x_0) + \sum_{i=1}^{m} w_i(\nabla c_i)(x_0) = 0_n \);
- \( w'c(x_0) = 0 \);
- \( w \geq 0_m \), where \( w = \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix} \).
The Primal Problem

Consider the following optimization problem for an object function $f : \mathbb{R}^n \to \mathbb{R}$, a subset $C \subseteq \mathbb{R}^n$, and the constraint functions $c : \mathbb{R}^n \to \mathbb{R}^m$ and $d : \mathbb{R}^n \to \mathbb{R}^p$:

\[ \text{minimize } f(x), \text{where } x \in C, \]

\[ \text{subject to } c(x) \leq 0_m \]

\[ \text{and } d(x) = 0_p. \]

We refer to this optimization problem as the \textit{primal problem}. 
Example

Let $f : \mathbb{R}^n \to \mathbb{R}$ be the linear function $f(x) = a'x$, $A \in \mathbb{R}^{p \times n}$, and $b \in \mathbb{R}^p$. Consider the primal problem:

$$
\text{minimize } a'x, \text{ where } x \in \mathbb{R}^n,
$$

subject to $x \geq 0_n$ and

$$Ax - b = 0_p.$$

The constraint functions are $c(x) = -x$ and $d(x) = Ax - b$. 
Definition

The Lagrangian associated to the primal problem is the function
$L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$
given by:

$$L(x, u, v) = f(x) + u'c(x) + v'd(x)$$

for $x \in C$, $u \in \mathbb{R}^m$, and $v \in \mathbb{R}^p$.

The component $u_i$ of $u$ is the Lagrangian multiplier corresponding to the constraint $c_i(x) \leq 0$; the component $v_j$ of $v$ is the Lagrangian multiplier corresponding to the constraint $d_j(x) = 0$. 
Example

Let $f : \mathbb{R}^n \to \mathbb{R}$ be the linear function $f(x) = a'x$, $A \in \mathbb{R}^{p \times n}$, and $b \in \mathbb{R}^p$.

Consider the primal problem:

*minimize* $a'x$, where $x \in \mathbb{R}^n$,

*subject to* $x \geq 0^n$ and $Ax - b = 0^p$.

The constraint functions are $c(x) = -x$ and $d(x) = Ax - b$ and The Lagrangian $L$ for the primal problem considered above is:

$$L(x, u, v) = a'x - u'x + v'(Ax - b)$$

$$= -v'b + (a' - u' + v'A)x.$$
Lemma

At each feasible $x$ we have

$$f(x) = \sup\{L(x, u, v) \mid u \geq 0_m, v \in \mathbb{R}^p, u_i c_i(x) = 0 \text{ for } 1 \leq i \leq m\}.$$  

Proof: at each feasible $x$ we have $c_i(x) \leq 0$ and $d_i(x) = 0$, hence

$$L(x, u, v) = f(x) + u' c(x) + v' d(x) \leq f(x).$$

The last inequality becomes an equality if $u_i c_i(x) = 0$ for $1 \leq i \leq m$. 
Lemma

The optimal value of the primal problem $f^*$ is

$$f^* = \inf_{x} \sup_{u \geq 0_m, v} L(x, u, v).$$

Proof: Consider feasible $x$ (designated as $x \in C$). In this case we have

$$f^* = \inf_{x \in C} f(x) = \inf_{x \in C} \sup_{u \geq 0_m, v} L(x, u, v).$$

When $x$ is not feasible, since $\sup_{u \geq 0_m, v} L(x, u, v) = \infty$ for any $x \not\in C$, we have $\inf_{x \not\in C} \sup_{u \geq 0_m, v} L(x, u, v) = \infty$. Thus, in either case,

$$f^* = \inf_{x} \sup_{u \geq 0_m, v} L(x, u, v).$$
The dual optimization problem starts with the Lagrange dual function $g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ defined by

$$g(u, v) = \inf_{x \in C} L(x, u, v)$$

(1)

and consists of

 maximize $g(u, v)$, where $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^p$, subject to $u \geq 0^m$.  


Theorem

For every primal problem the Lagrange dual function \( g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R} \) defined by Equality (1) is always concave over \( \mathbb{R}^m \times \mathbb{R}^p \).
Proof

For $u_1, u_2 \in \mathbb{R}^m$ and $v_1, v_2 \in \mathbb{R}^p$ we have:

$$g(tu_1 + (1 - t)u_2, tv_1 + (1 - t)v_2)$$

$$= \inf \{ f(x) + (tu'_1 + (1 - t)u'_2)c(x) + (tv'_1 + (1 - t)v'_2)d(x) \mid x \in S \}$$

$$= \inf \{ t(f(x) + u'_1c + v'_1d) + (1 - t)(f(x) + u'_2c(x) + v'_2d(x)) \mid x \in S \}$$

$$\geq t \inf \{ f(x) + u'_1c + v'_1d \mid x \in S \}$$

$$+ (1 - t) \inf \{ f(x) + u'_2c(x) + v'_2d(x) \mid x \in S \}$$

$$= tg(u_1, v_1) + (1 - t)g(u_2, v_2),$$

which shows that $g$ is concave.
- The concavity of $g$ is significant because a local optimum of $g$ is a global optimum regardless of convexity properties of $f$, $c$ or $d$.
- Although the dual function $g$ is not given explicitly, the restrictions of the dual have a simpler form and this may be an advantage in specific cases.
- The dual function produces lower bounds for the optimal value of the primal problem.
Theorem

(The Weak Duality Theorem) Suppose that $x^*$ is an optimum of $f$ and $f^* = f(x^*)$, $(u^*, v^*)$ is an optimum for $g$, and $g^* = g(u^*, v^*)$. We have $g^* \leq f^*$.

Proof: Since $c(x^*) \leq 0_m$ and $d(x^*) = 0_p$ it follows that

$$L(x^*, u, v) = f(x^*) + u'c(x^*) + v'd(x^*) \leq f^*.$$  

Therefore, $g(u, v) = \inf_{x \in C} L(x, u, v) \leq f^*$ for all $u$ and $v$. Since $g^*$ is the optimal value of $g$, the last inequality implies $g^* \leq f^*$. 


The inequality of the previous theorem holds when \( f_* \) and \( g_* \) are finite or infinite. The difference \( f_* - g_* \) is the \textit{duality gap} of the primal problem. \textit{Strong duality} holds when the duality gap is 0.
Note that for the Lagrangian function of the primal problem we can write

\[
\sup_{u \geq 0, \nu} L(x, u, \nu) = \sup_{u \geq 0, \nu} f(x) + u'c(x) + \nu'd(x) = \begin{cases} 
\inf_{x \in \mathbb{R}^n} L(x, u, \nu) & \text{if } c(x) \leq 0, \\
\infty & \text{otherwise}
\end{cases},
\]

which implies \( f_* = \inf_{x \in \mathbb{R}^n} \sup_{u \geq 0, \nu} L(x, u, \nu). \) By the definition of \( g_* \) we also have

\[
g_* = \sup_{u \geq 0, \nu} \inf_{x \in \mathbb{R}^n} L(x, u, \nu).
\]
Thus, the weak duality amounts to the inequality

\[ \sup_{u \geq 0, \nu} \inf_{x \in \mathbb{R}^n} L(x, u, \nu) \leq \inf_{x \in \mathbb{R}^n} \sup_{u \geq 0, \nu} L(x, u, \nu), \]

and the strong duality is equivalent to the equality

\[ \sup_{u \geq 0, \nu} \inf_{x \in \mathbb{R}^n} L(x, u, \nu) = \inf_{x \in \mathbb{R}^n} \sup_{u \geq 0, \nu} L(x, u, \nu). \]
Example

Let \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) be the linear function \( f(x) = a'x \), \( A \in \mathbb{R}^{p \times n} \), and \( b \in \mathbb{R}^p \). Consider the primal problem:

\[
\text{minimize } a'x, \text{ where } x \in \mathbb{R}^n,
\text{subject to } x \geq 0_n \text{ and }
Ax - b = 0_p.
\]

The constraint functions are \( c(x) = -x \) and \( d(x) = Ax - b \) and the Lagrangian \( L \) is

\[
L(x, u, v) = a'x - u'x + v'(Ax - b)
= -v'b + (a' - u' + v'A)x.
\]
Example (cont’d)

This yields the dual function

\[ g(u, v) = \inf_{x \in C} L(x, u, v) \]
\[ = -v' b + \inf_{x \in \mathbb{R}^n} (a' - u' + v' A)x. \]

Unless \( a' - u' + v' A = 0'_n \) we have \( g(u, v) = -\infty \). Therefore, we have

\[ g(u, v) = \begin{cases} 
-v' b & \text{if } a - u + A' v = 0_n, \\
-\infty & \text{otherwise}.
\end{cases} \]

Thus, the dual problem is

\[ \text{maximize } g(u, v), \]
\[ \text{subject to } u \geq 0_m. \]
Example (cont’d)

An equivalent of the dual problem is

\[
\begin{align*}
\text{maximize } & -v' b, \\
\text{subject to } & a - u + A' v = 0_n \\
\text{and } & u \geq 0_m.
\end{align*}
\]

In turn, this problem is equivalent to:

\[
\begin{align*}
\text{maximize } & -v' b, \\
\text{subject to } & a + A' v \geq 0_n.
\end{align*}
\]
Example

The following optimization problem

\[
\text{minimize } \frac{1}{2} x' Q x - r' x,
\]

where \( x \in \mathbb{R}^n \),

subject to \( A x \geq b \),

where \( Q \in \mathbb{R}^{n \times n} \) is a positive definite matrix, \( r \in \mathbb{R}^n \), \( A \in \mathbb{R}^{p \times n} \), and \( b \in \mathbb{R}^p \) is known as a \textit{quadratic optimization problem}.
The Lagrangian $L$ is

$$L(x, u) = \frac{1}{2} x' Q x - r' x + u'(A x - b) = \frac{1}{2} x' Q x + (u' A - r') x - u' b$$

and the dual function is $g(u) = \inf_{x \in \mathbb{R}^n} L(x, u)$ subject to $u \geq 0_m$. Since $x$ is unconstrained in the definition of $g$, the minimum is attained when we have the equalities

$$\frac{\partial}{\partial x_i} \left( \frac{1}{2} x' Q x + (u' A - r') x - u' b \right) = 0$$

for $1 \leq i \leq n$, which amount to $x = Q^{-1}(r - Au)$. The dual optimization function is: $g(u) = -\frac{1}{2} u' P u - u' d - \frac{1}{2} r' Q r$ subject to $u \geq 0_p$, where $P = AQ^{-1} A'$, $d = b - AQ^{-1} r$. This shows that the dual problem of this quadratic optimization problem is itself a quadratic optimization problem.
Example

Let \( \mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n \). We seek to determine a closed sphere \( B[\mathbf{x}, r] \) of minimal radius that includes all points \( \mathbf{a}_i \) for \( 1 \leq i \leq m \). This is the minimum bounding sphere problem, formulated by J. J. Sylvester. This problem amounts to solving the following primal optimization problem:

\[
\begin{align*}
\text{minimize} & \quad r, \quad \text{where} \quad r \geq 0, \\
\text{subject to} & \quad \| \mathbf{x} - \mathbf{a}_i \| \leq r \quad \text{for} \quad 1 \leq i \leq m.
\end{align*}
\]
An equivalent formulation requires minimizing \( r^2 \) and stating the restrictions as \( \| x - a_i \|^2 - r^2 \leq 0 \) for \( 1 \leq i \leq m \). The Lagrangian of this problem is:

\[
L(r, x, u) = r^2 + \sum_{i=1}^{m} u_i (\| x - a_i \|^2 - r^2)
\]

\[
= r^2 \left( 1 - \sum_{i=1}^{m} u_i \right) + \sum_{i=1}^{m} u_i \| x - a_i \|^2
\]

and the dual function is:

\[
g(u) = \inf_{r \in \mathbb{R} \geq 0, x \in \mathbb{R}^n} L(r, x, u)
\]

\[
= \inf_{r \in \mathbb{R} \geq 0, x \in \mathbb{R}^n} r^2 \left( 1 - \sum_{i=1}^{m} u_i \right) + \sum_{i=1}^{m} u_i \| x - a_i \|^2.
\]
This leads to the following conditions:

\[
\frac{\partial L(r, \mathbf{x}, \mathbf{u})}{\partial r} = 2r \left(1 - \sum_{i=1}^{m} u_i \right) = 0
\]

\[
\frac{\partial L(r, \mathbf{x}, \mathbf{u})}{\partial x_p} = 2 \sum_{i=1}^{m} u_i (\mathbf{x} - \mathbf{a}_i)_p = 0 \text{ for } 1 \leq p \leq n.
\]

The first equality yields \(\sum_{i=1}^{m} u_i = 1\). Therefore, from the second equality we obtain \(\mathbf{x} = \sum_{i=1}^{m} u_i \mathbf{a}_i\). This shows that for \(\mathbf{x}\) is a convex combination of \(\mathbf{a}_1, \ldots, \mathbf{a}_m\). The dual function is

\[
g(\mathbf{u}) = \sum_{i=1}^{m} u_i \left(\sum_{h=1}^{m} u_h \mathbf{a}_h - \mathbf{a}_i \right) = 0
\]

because \(\sum_{i=1}^{m} u_i = 1\).

Note that the restriction functions \(g_i(\mathbf{x}, r) = \| \mathbf{x} - \mathbf{a}_i \|^2 - r^2 \leq 0\) are not convex.
Example

Consider the primal problem

\[
\text{minimize } x_1^2 + x_2^2, \text{ where } x_1, x_2 \in \mathbb{R},
\]

\[
\text{subject to } x_1 - 1 \geq 0.
\]

It is clear that the minimum of \( f(x) \) is obtained for \( x_1 = 1 \) and \( x_2 = 0 \) and this minimum is 1. The Lagrangian is

\[
L(u) = x_1^2 + x_2^2 + u_1(x_1 - 1)
\]

and the dual function is

\[
g(u) = \inf_{x} \{x_1^2 + x_2^2 + u_1(x_1 - 1) \mid x \in \mathbb{R}^2\} = -\frac{u_1^2}{4}.
\]

Then \( \sup \{g(u_1) \mid u_1 \geq 0\} = 0 \) and a gap exists between the minimal value of the primal function and the maximal value of the dual function.
Example

Let $a, b > 0$, $p, q < 0$ and let $r > 0$. Consider the following primal problem:

$$
\text{minimize } f(x) = ax_1^2 + bx_2^2 \\
\text{subject to } px_1 + qx_2 + r \leq 0 \text{ and } x_1 \geq 0, x_2 \geq 0.
$$

The set $C$ is $\{x \in \mathbb{R}^2 \mid x_1 \geq 0, x_2 \geq 0\}$. The constraint function is $c(x) = px_1 + qx_2 + r \leq 0$ and the Lagrangian of the primal problem is

$$
L(x, u) = ax_1^2 + bx_2^2 + u(px_1 + qx_2 + r),
$$

where $u$ is a Lagrangian multiplier.
Thus, the dual problem objective function is

\[
g(u) = \inf_{x \in C} L(x, u)
\]

\[
= \inf_{x \in C} ax_1^2 + bx_2^2 + u(px_1 + qx_2 + r)
\]

\[
= \inf_{x \in C} \{ax_1^2 + upx_1 \mid x_1 \geq 0\}
\]

\[
+ \inf_{x \in C} \{bx_2^2 + uqx_2 \mid x_2 \geq 0\} + ur
\]

The infima are achieved when \(x_1 = -\frac{up}{2a}\) and \(x_2 = -\frac{uq}{2b}\) if \(u \geq 0\) and at \(x = 0_2\) if \(u < 0\). Thus,

\[
g(u) = \begin{cases} 
- \left(\frac{p^2}{4a} + \frac{q^2}{4b}\right) u^2 + ru & \text{if } u \geq 0, \\
u & \text{if } u < 0
\end{cases}
\]

which is a concave function.
The maximum of $g(u)$ is achieved when $u = \frac{2r}{\frac{p^2}{a} + \frac{q^2}{b}}$ and equals

$$\frac{r^2}{\left(\frac{p^2}{a} + \frac{q^2}{b}\right)}$$

Family of Concentric Ellipses; the ellipse that “touches” the line $px_1 + qx_2 + r = 0$ gives the optimum value for $f$. The dotted area is the feasible region.
Note that if $\mathbf{x}$ is located on an ellipse $ax_1^2 + bx_2^2 - k = 0$, then $f(\mathbf{x}) = k$. Thus, the minimum of $f$ is achieved when $k$ is chosen such that the ellipse is tangent to the line $px_1 + qx_2 + r = 0$. In other words, we seek to determine $k$ such that the tangent of the ellipse at $\mathbf{x}_0 = \begin{pmatrix} x_{01} \\ x_{02} \end{pmatrix}$ located on the ellipse coincides with the line given by $px_1 + qx_2 + r = 0$. The equation of the tangent is

$$ax_1 x_{01} + bx_2 x_{02} - k = 0.$$ 

Therefore, we need to have:

$$\frac{ax_{01}}{p} = \frac{bx_{02}}{q} = \frac{-k}{r},$$

hence $x_{01} = -\frac{kp}{ar}$ and $x_{02} = -\frac{kq}{br}$. Substituting back these coordinates in the equation of the ellipse yields $k_1 = 0$ and $k_2 = \frac{r^2}{\frac{p^2}{a} + \frac{q^2}{b}}$. In this case no duality gap exists.