Support Vector Machines - I

Prof. Dan A. Simovici

UMB

◆□ ▶ ◆□ ▶ ◆ 三 ▶ ◆ 三 ▶ ● ○ ● ● ● ●

1/65



2 Optimization

- General Optimization Problems
- 4 Lagrange Multipliers



Functions of One Real Variable

Let E be a subset of \mathbb{R} .

A function $f : E \longrightarrow \mathbb{R}$ has a 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 maximizer of f on E.

Similarly, $f : E \longrightarrow \mathbb{R}$ has a minimum *m* on *E* if there exists $x_0 \in E$ such that $f(x_0) = m$ and $f(x_1) \ge m$ for every $x_1 \in E$. The element x_0 is a minimizer of *f* on *E*.

- If f : [a, b] → 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'(x₀) = 0, then x₀ is a critical point of f.
- A local extremum (minimum or maximum) can occur only at a critical point x₀. If f''(x₀) < 0, the critical point provides a local maximum; if f''(x₀) > 0 the critical point provides a local minimum.

(read "nabla f"). Let $f: X \longrightarrow \mathbb{R}$, where $X \subseteq \mathbb{R}^n$, and let $z \in X$. The *gradient* of f in z is the vector

$$(
abla f)(oldsymbol{z}) = egin{pmatrix} rac{\partial f}{\partial x_1}(oldsymbol{z})\ dots\ rac{\partial f}{\partial x_n}(oldsymbol{z}) \end{pmatrix} \in \mathbb{R}^n.$$

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be the function $f(\mathbf{x}) = x_1^2 + \dots + x_n^2$; in other words, $f(\mathbf{x}) = ||\mathbf{x}||^2$. We have $\frac{\partial f}{\partial x_1} = 2x_1, \dots, \frac{\partial f}{\partial x_n} = 2x_n$.

Therefore, $(\nabla f)(\mathbf{x}) = 2\mathbf{x}$.

Let $\boldsymbol{b}_j \in \mathbb{R}^n$ and $c_j \in \mathbb{R}$ for $1 \leqslant j \leqslant n$, and let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be the function

$$f(\mathbf{x}) = \sum_{j=1}^{n} (\mathbf{b}_{j}'\mathbf{x} - c_{j})^{2}.$$

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

$$(\nabla f)(\mathbf{x}) = 2 \begin{pmatrix} \sum_{j=1}^{n} 2b_{1j}(\mathbf{b}'_{j}\mathbf{x} - c_{j}) \\ \vdots \\ \sum_{j=1}^{n} 2b_{nj}(\mathbf{b}'_{j}\mathbf{x} - c_{j}) \end{pmatrix} = 2(B'\mathbf{x} - \mathbf{c}')B = 2B'\mathbf{x}B - 2\mathbf{c}'B,$$

where $B = (\boldsymbol{b}_1 \cdots \boldsymbol{b}_n) \in \mathbb{R}^{n \times n}$.

The matrix-valued function $H_f : \mathbb{R}^k \longrightarrow \mathbb{R}^{k \times k}$ defined by

$$H_f(\boldsymbol{x}) = \left(\frac{\partial^2 f}{\partial x_{i_1} \partial x_{i_2}}\right)$$

is the *Hessian matrix* of *f*.

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

$$H_f(\mathbf{x}) = egin{pmatrix} 2 & 0 & \cdots & 0 \ 0 & 2 & \cdots & 0 \ dots & dots & \cdots & dots \ dots & dots & \cdots & dots \ 0 & 0 & \cdots & 2 \end{pmatrix}$$

Definition

Let X be a open subset in \mathbb{R}^n and let $f : X \longrightarrow \mathbb{R}$ be a function. The point $\mathbf{x}_0 \in X$ is a *local minimum* for f if there exists $\delta > 0$ such that $B(\mathbf{x}_0, \delta) \subseteq X$ and $f(\mathbf{x}_0) \leq f(\mathbf{x})$ for every $\mathbf{x} \in B(\mathbf{x}_0, \delta)$. The point \mathbf{x}_0 is a *strict local minimum* if $f(\mathbf{x}_0) < f(\mathbf{x})$ for every $\mathbf{x} \in B(\mathbf{x}_0, \delta) - \{\mathbf{x}_0\}$.

Definition

A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is *positive semidefinite* if $\mathbf{x}' A \mathbf{x} \ge 0$ for all $\mathbf{x} \in \mathbb{R}^n$. A is *positive definite* if $\mathbf{x}' A \mathbf{x} > 0$ for all $\mathbf{x} \in \mathbb{R}^n - {\mathbf{0}_n}$.

The symmetric real matrix

$$\mathsf{A} = \begin{pmatrix} \mathsf{a} & \mathsf{b} \\ \mathsf{b} & \mathsf{c} \end{pmatrix}$$

is positive definite if and only if a > 0 and $b^2 - ac < 0$. Indeed, we have $\mathbf{x}' A \mathbf{x} > 0$ for every $\mathbf{x} \in \mathbb{R}^2 - \{\mathbf{0}\}$ if and only if $ax_1^2 + 2bx_1x_2 + cx_2^2 > 0$, where $\mathbf{x}' = (x_1 \ x_2)$; 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 $f : B(\mathbf{x}_0, r) \longrightarrow \mathbb{R}$ be a function that belongs to the class $C^2(B(\mathbf{x}_0, r))$, where $B(\mathbf{x}_0, r) \subseteq \mathbb{R}^k$ and \mathbf{x}_0 is a critical point for f. If the Hessian matrix $H_f(\mathbf{x}_0)$ is positive semidefinite, then \mathbf{x}_0 is a local minimum for f; if $H_f(\mathbf{x}_0)$ is negative semidefinite, then \mathbf{x}_0 is a local maximum for f. Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be a function in $C^2(B(\mathbf{x}_0, r))$. The Hessian matrix in \mathbf{x}_0 is

$$H_f(\mathbf{x}_0) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix} (\mathbf{x}_0).$$

Let $a_{11} = \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}_0)$, $a_{12} = \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}_0)$, and $a_{22} = \frac{\partial^2 f}{\partial x_2^2}(\boldsymbol{x}_0)$. Note that

$$\begin{aligned} \mathbf{h}' H_f(\mathbf{x}_0) \mathbf{h} &= a_{11} h_1^2 + 2 a_{12} h_1 h_2 + a_{22} h_2^2 \\ &= h_2^2 \left(a_{11} \xi^2 + 2 a_{12} \xi + a_{22} \right), \end{aligned}$$

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへで

16/65

where $\xi = \frac{h_1}{h_2}$.

For a critical point \boldsymbol{x}_0 we have:

- *h*'*H_f*(*x*₀)*h* ≥ 0 for every *h* if *a*₁₁ > 0 and *a*²₁₂ *a*₁₁*a*₂₂ < 0; in this case, *H_f*(*x*₀) is positive semidefinite and *x*₀ is a local minimum;
- $h'H_f(\mathbf{x}_0)h \leq 0$ for every h if $a_{11} < 0$ and $a_{12}^2 a_{11}a_{22} < 0$; in this case, $H_f(\mathbf{x}_0)$ is negative semidefinite and \mathbf{x}_0 is a local maximum;
- if a²₁₂ − a₁₁a₂₂ ≥ 0; in this case, H_f(x₀) is neither positive nor negative definite, so x₀ 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.

Let $\mathbf{a}_1, \ldots, \mathbf{a}_m$ be *m* points in \mathbb{R}^n . The function $f(\mathbf{x}) = \sum_{i=1}^m ||\mathbf{x} - \mathbf{a}_i||^2$ gives the sum of squares of the distances between \mathbf{x} and the points $\mathbf{a}_1, \ldots, \mathbf{a}_m$. We will prove that this sum has a global minimum obtained when \mathbf{x} is the barycenter of the set $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$.

Example (cont'd) We have

$$f(\mathbf{x}) = m \| \mathbf{x} \|^2 - 2 \sum_{i=1}^m \mathbf{a}'_i \mathbf{x} + \sum_{i=1}^m \| \mathbf{a}_i \|^2$$

= $m(x_1^2 + \dots + x_n^2) - 2 \sum_{j=1}^n \sum_{i=1}^m a_{ij} x_j + \sum_{i=1}^m \| \mathbf{a}_i \|^2,$

which implies

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

for $1 \leqslant j \leqslant n$. 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 $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, $\boldsymbol{c} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, and $\boldsymbol{d} : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ be three functions defined on \mathbb{R}^n . A general formulation of a *constrained optimization problem* is:

minimize $f(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^n$, subject to $\mathbf{c}(\mathbf{x}) \leq \mathbf{0}_m$, where $\mathbf{c} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, and $\mathbf{d}(\mathbf{x}) = \mathbf{0}_p$, where $\mathbf{d} : \mathbb{R}^n \longrightarrow \mathbb{R}^p$. Here **c** specifies *inequality constraints* placed on **x**, while **d** defines *equality constraints*.

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

$$R_{\boldsymbol{c},\boldsymbol{d}} = \{ \boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{c}(\boldsymbol{x}) \leqslant \boldsymbol{0}_m \text{ and } \boldsymbol{d}(\boldsymbol{x}) = \boldsymbol{0}_p \}.$$

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 \boldsymbol{c}) the feasible region is:

$$R_{\boldsymbol{c}} = \{ \boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{c}(\boldsymbol{x}) \leqslant \boldsymbol{0}_m \}.$$

Let $x \in R_c$. The set of active constraints at x is

$$ACT(R_{c}, c, x) = \{i \in \{1, ..., m\} \mid c_{i}(x) = 0\}.$$

If $i \in ACT(R_c, c, x)$, 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 $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and $\mathbf{c} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be two functions. The minimization problem MP (f, \mathbf{c}) is: minimize $f(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^n$, subject to $\mathbf{x} \in R_{\mathbf{c}}$. If \mathbf{x}_0 exists in $R_{\mathbf{c}}$ that $f(\mathbf{x}_0) = \min\{f(\mathbf{x}) \mid \mathbf{x} \in R_{\mathbf{c}}\}$ we refer to \mathbf{x}_0 as a solution of MP (f, \mathbf{c}) . If $\boldsymbol{h}: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ we can write

$$m{h}(m{x}) = egin{pmatrix} h_1(m{x}) \ dots \ h_m(m{x}) \end{pmatrix},$$

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

$$(Dm{h})(m{x}) = egin{pmatrix} (
abla h_1)(m{x})' \\
dots \\
(
abla h_m)(m{x})' \end{pmatrix} \in \mathbb{R}^{m imes n}.$$

Let $\boldsymbol{h}:\mathbb{R}^2\longrightarrow\mathbb{R}^3$ be given by

$$\boldsymbol{h}(\boldsymbol{x}) = \begin{pmatrix} x_1 x_2 \\ x_1^2 \\ x_2^2 \end{pmatrix}$$

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 $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and $h : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ 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, rank((Dh)(x)) = m < n (which means that the gradients (∇h₁)(x),...,(∇h_m)(x) are linearly independent).

If \mathbf{x}_0 is a local extremum of f subjected to the restriction $\mathbf{h}(\mathbf{x}_0) = \mathbf{0}_m$, then $(\nabla f)(\mathbf{x}_0)$ is a linear combination of $(\nabla h_1)(\mathbf{x}_0), \dots, (\nabla h_m)(\mathbf{x}_0)$.

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

$$h(\mathbf{x}) = x_1^2 + x_2^2 = 2.$$

We have

$$(\nabla f)(\mathbf{x}) = \begin{pmatrix} 1\\ 1 \end{pmatrix}$$

 $(\nabla h)(\mathbf{x}) = \begin{pmatrix} 2x_1\\ 2x_2 \end{pmatrix}$

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 $(\nabla f)(\mathbf{x}^*) + \frac{1}{2}(\nabla h) = \mathbf{0}.$

To apply the Lagrange multiplier technique the constraint gradients $(
abla h_1)(m{x}),\cdots,(
abla h_m)(m{x})$

must be linearly independent. In this case, \boldsymbol{x} is said to be regular.

If a local minimum is not regular Lagrange multipliers may not exist.

Example

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

$$h_1(\mathbf{x}) = (x_1 - 1)^2 + x_2^2 - 1 = 0, h_2(\mathbf{x}) = (x_1 - 2)^2 + x_2^2 - 4 = 0.$$

We have

$$(\nabla f)(\boldsymbol{x}) = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

and

$$(\nabla h_1)(\mathbf{x}) = \begin{pmatrix} 2(x_1-1)\\ 2x_2 \end{pmatrix}, (\nabla h_2)(\mathbf{x}) = \begin{pmatrix} 2(x_1-2)\\ 2x_2 \end{pmatrix}.$$



Example continued

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

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

The gradients $(\nabla h_1)(\mathbf{0}), (\nabla h_2)(\mathbf{0})$ are not linearly independent because

$$2(\nabla h_1)(\mathbf{0}) + (\nabla h_2)(\mathbf{0}) = \mathbf{0}_2,$$

so **0** is not a regular point and Lagrange's multipliers do not exist.

Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix and let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be the function defined by $f(\mathbf{x}) = \mathbf{x}' A \mathbf{x}$. **Optimization problem:** minimize f subjected to the restriction $|| \mathbf{x} || = 1$, or equivalently $h(\mathbf{x}) = || \mathbf{x} ||^2 - 1 = 0$. Since $(\nabla f) = 2A\mathbf{x}$ and $(\nabla h)(\mathbf{x}) = 2\mathbf{x}$ there exists λ such that $2A\mathbf{x}_0 = 2\lambda \mathbf{x}_0$ for any extremum of f subjected to $|| \mathbf{x}_0 || = 1$. Thus, \mathbf{x}_0 must be a unit eigenvector of A and λ 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 (∇c_i)(x₀) for i ∈ ACT(S, c, x₀)}, and
- ensure that the coefficient of the gradient of the objective function $(\nabla f)(\mathbf{x}_0)$ is not null.

These conditions are known as the *Karush-Kuhn-Tucker conditions* or the *KKT conditions*.

Theorem

(Karush-Kuhn-Tucker Theorem) Let *S* be a non-empty open subset of \mathbb{R}^n and let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and $\mathbf{c} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$. Let \mathbf{x}_0 be a local minimum in *S* of *f* subjected to the restriction $\mathbf{c}(\mathbf{x}_0) \leq \mathbf{0}_m$. Suppose that *f* is differentiable in \mathbf{x}_0 , c_i are differentiable in \mathbf{x}_0 for $i \in ACT(S, \mathbf{c}, \mathbf{x}_0)$, and c_i are continuous in \mathbf{x}_0 for $i \notin ACT(S, \mathbf{c}, \mathbf{x}_0)$. If $\{(\nabla c_i)(\mathbf{x}_0) \mid i \in ACT(S, \mathbf{c}, \mathbf{x}_0)\}$ is a linearly independent set, then there exist non-negative numbers w_i for $i \in ACT(S, \mathbf{c}, \mathbf{x}_0)$ such that

$$(\nabla f)(\boldsymbol{x}_0) + \sum \{w_i(\nabla c_i)(\boldsymbol{x}_0) \mid i \in ACT(S, \boldsymbol{c}, \boldsymbol{x}_0)\} = \boldsymbol{0}_n.$$

Theorem continued

Furthermore, if the functions c_i are differentiable in \mathbf{x}_0 for $i \notin ACT(S, \mathbf{c}, \mathbf{x}_0)$, then the previous condition can be written as: • $(\nabla f)(\mathbf{x}_0) + \sum_{i=1}^m w_i (\nabla c_i)(\mathbf{x}_0) = \mathbf{0}_n$; • $\mathbf{w}' \mathbf{c}(\mathbf{x}_0) = 0$; • $\mathbf{w} \ge \mathbf{0}_m$, where $\mathbf{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix}$.

Consider the following optimization problem for an object function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, a subset $C \subseteq \mathbb{R}^n$, and the constraint functions $c : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ and $d : \mathbb{R}^n \longrightarrow \mathbb{R}^p$: minimize f(x), where $x \in C$, subject to $c(x) \leq \mathbf{0}_m$ and $d(x) = \mathbf{0}_p$.

We refer to this optimization problem as the *primal problem*.

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

minimize $\mathbf{a}'\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^n$, subject to $\mathbf{x} \ge \mathbf{0}_n$ and $A\mathbf{x} - \mathbf{b} = \mathbf{0}_p$.

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

Definition

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

$$L(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) + \mathbf{u}' \mathbf{c}(\mathbf{x}) + \mathbf{v}' \mathbf{d}(\mathbf{x})$$

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

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

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

minimize $\mathbf{a}'\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^n$, subject to $\mathbf{x} \ge \mathbf{0}_n$ and $A\mathbf{x} - \mathbf{b} = \mathbf{0}_p$.

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

$$L(\mathbf{x}, \mathbf{u}, \mathbf{v}) = \mathbf{a}'\mathbf{x} - \mathbf{u}'\mathbf{x} + \mathbf{v}'(A\mathbf{x} - \mathbf{b})$$

= $-\mathbf{v}'\mathbf{b} + (\mathbf{a}' - \mathbf{u}' + \mathbf{v}'A)\mathbf{x}.$

Lemma

At each feasible \mathbf{x} we have $f(\mathbf{x}) = \sup\{L(\mathbf{x}, \mathbf{u}, \mathbf{v})\} \mid \mathbf{u} \ge \mathbf{0}_m, \mathbf{v} \in \mathbb{R}^p, u_i \mathbf{c}_i(\mathbf{x}) = 0 \text{ for } 1 \le i \le m\}.$

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

$$L(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}) = f(\boldsymbol{x}) + \boldsymbol{u}' \boldsymbol{c}(\boldsymbol{x}) + \boldsymbol{v}' \boldsymbol{d}(\boldsymbol{x}) \leqslant f(\boldsymbol{x}).$$

The last inequality becomes an equality if $u_i \boldsymbol{c}_i(\boldsymbol{x}) = 0$ for $1 \leq i \leq m$.

Lemma

The optimal value of the primal problem f* is

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

Proof: Consider feasible x (designated as $x \in C$). In this case we have $f^* = \inf_{x \in C} f(x) = \inf_{x \in C} \sup_{u \ge 0_m, v} L(x, u, v)$. When x is not feasible, since $\sup_{u \ge 0_m, v} L(x, u, v) = \infty$ for any $x \notin C$, we have $\inf_{x \notin C} \sup_{u \ge 0_m, v} L(x, u, v) = \infty$. Thus, in either case, $f^* = \inf_x \sup_{u \ge 0_m, v} L(x, u, v)$.

The Dual Optimization Problem

The *dual optimization problem* starts with the *Lagrange dual function* $g: \mathbb{R}^m \times \mathbb{R}^p \longrightarrow \mathbb{R}$ defined by

$$g(\boldsymbol{u},\boldsymbol{v}) = \inf_{\boldsymbol{x}\in C} L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v})$$
(1)

and consists of
maximize
$$g(\boldsymbol{u}, \boldsymbol{v})$$
, where $\boldsymbol{u} \in \mathbb{R}^m$ and $\boldsymbol{v} \in \mathbb{R}^p$,
subject to $\boldsymbol{u} \ge \boldsymbol{0}_m$.

Theorem

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

Proof

For $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathbb{R}^m$ and $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^p$ we have:

$$g(tu_{1} + (1 - t)u_{2}, tv_{1} + (1 - t)v_{2})$$

$$= \inf\{f(\mathbf{x}) + (tu'_{1} + (1 - t)u'_{2})c(\mathbf{x}) + (tv'_{1} + (1 - t)v'_{2})d(\mathbf{x}) \mid \mathbf{x} \in S\}$$

$$= \inf\{t(f(\mathbf{x}) + u'_{1}c + v'_{1}d) + (1 - t)(f(\mathbf{x}) + u'_{2}c(\mathbf{x}) + v'_{2}d(\mathbf{x})) \mid \mathbf{x} \in S\}$$

$$\geq t\inf\{f(\mathbf{x}) + u'_{1}c + v'_{1}d \mid \mathbf{x} \in S\}$$

$$+ (1 - t)\inf\{f(\mathbf{x}) + u'_{2}c(\mathbf{x}) + v'_{2}d(\mathbf{x}) \mid \mathbf{x} \in S\}$$

$$= tg(u_{1}, v_{1}) + (1 - t)g(u_{2}, v_{2}),$$

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 $f_* = f(x_*)$, $(\mathbf{u}_*, \mathbf{v}_*)$ is an optimum for g, and $g_* = g(\mathbf{u}_*, \mathbf{v}_*)$. We have $g_* \leq f_*$.

Proof: Since $c(x_*) \leq \mathbf{0}_m$ and $d(x_*) = \mathbf{0}_p$ it follows that

$$L(\boldsymbol{x}_*, \boldsymbol{u}, \boldsymbol{v}) = f(\boldsymbol{x}_*) + \boldsymbol{u}' \boldsymbol{c}(\boldsymbol{x}_*) + \boldsymbol{v}' \boldsymbol{d}(\boldsymbol{x}_*) \leqslant f_*.$$

Therefore, $g(\boldsymbol{u}, \boldsymbol{v}) = \inf_{\boldsymbol{x} \in C} L(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}) \leq f_*$ for all \boldsymbol{u} and \boldsymbol{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 *duality gap* of the primal problem. *Strong duality* holds when the duality gap is 0.

Note that for the Lagrangian function of the primal problem we can write

$$\sup_{\boldsymbol{u} \ge \boldsymbol{0}_m, \boldsymbol{v}} L(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}) = \sup_{\boldsymbol{u} \ge \boldsymbol{0}_m, \boldsymbol{v}} f(\boldsymbol{x}) + \boldsymbol{u}' \boldsymbol{c}(\boldsymbol{x}) + \boldsymbol{v}' \boldsymbol{d}(\boldsymbol{x})$$

$$= \begin{cases} f(\boldsymbol{x}) & \text{if } \boldsymbol{c}(\boldsymbol{x}) \le \boldsymbol{0}_m, \\ \infty & \text{otherwise} \end{cases},$$

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

$$g_* = \sup_{oldsymbol{u} \geqslant oldsymbol{0}_m, oldsymbol{v} \, oldsymbol{x} \in \mathbb{R}^n} \inf_{oldsymbol{L}(oldsymbol{x}, oldsymbol{u}, oldsymbol{v}).$$

Thus, the weak duality amounts to the inequality

$$\sup_{\boldsymbol{u} \ge \boldsymbol{0}_m, \boldsymbol{v}} \inf_{\boldsymbol{x} \in \mathbb{R}^n} L(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}) \leqslant \inf_{\boldsymbol{x} \in \mathbb{R}^n} \sup_{\boldsymbol{u} \ge \boldsymbol{0}_m, \boldsymbol{v}} L(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}),$$

and the strong duality is equivalent to the equality

$$\sup_{\boldsymbol{u} \ge \boldsymbol{0}_m, \boldsymbol{v}} \inf_{\boldsymbol{x} \in \mathbb{R}^n} L(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}) = \inf_{\boldsymbol{x} \in \mathbb{R}^n} \sup_{\boldsymbol{u} \ge \boldsymbol{0}_m, \boldsymbol{v}} L(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}).$$

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

minimize $\mathbf{a}'\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^n$, subject to $\mathbf{x} \ge \mathbf{0}_n$ and $A\mathbf{x} - \mathbf{b} = \mathbf{0}_p$.

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

$$L(\mathbf{x}, \mathbf{u}, \mathbf{v}) = \mathbf{a}' \mathbf{x} - \mathbf{u}' \mathbf{x} + \mathbf{v}' (A\mathbf{x} - \mathbf{b})$$

= $-\mathbf{v}' \mathbf{b} + (\mathbf{a}' - \mathbf{u}' + \mathbf{v}' A) \mathbf{x}$.

Example (cont'd)

This yields the dual function

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

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

$$g(\boldsymbol{u}, \boldsymbol{v}) = egin{cases} -\boldsymbol{v}' \boldsymbol{b} & ext{if } \boldsymbol{a} - \boldsymbol{u} + A' \boldsymbol{v} = \boldsymbol{0}_n \ -\infty & ext{otherwise.} \end{cases}$$

Thus, the dual problem is

```
maximize g(\boldsymbol{u}, \boldsymbol{v}),
subject to \boldsymbol{u} \ge \boldsymbol{0}_m.
```

Example (cont'd)

An equivalent of the dual problem is

maximize
$$-\mathbf{v}'\mathbf{b}$$
,
subject to $\mathbf{a} - \mathbf{u} + A'\mathbf{v} = \mathbf{0}_n$
and $\mathbf{u} \ge \mathbf{0}_m$.

In turn, this problem is equivalent to: $maximize - \mathbf{v'b}$, $subject \ to \ \mathbf{a} + A'\mathbf{v} \ge \mathbf{0}_n$.

The following optimization problem minimize $\frac{1}{2}\mathbf{x}'Q\mathbf{x} - \mathbf{r}'\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^n$, subject to $A\mathbf{x} \ge \mathbf{b}$, where $Q \in \mathbb{R}^{n \times n}$ is a positive definite matrix, $\mathbf{r} \in \mathbb{R}^n$, $A \in \mathbb{R}^{p \times n}$, and $\mathbf{b} \in \mathbb{R}^p$ is known as a *quadratic optimization problem*. The Lagrangian L is

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

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

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

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

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: *minimize r, where r* ≥ 0 ,

subject to $\| \mathbf{x} - \mathbf{a}_i \| \leq r$ for $1 \leq i \leq m$.

An equivalent formulation requires minimizing r^2 and stating the restrictions as $\| \mathbf{x} - \mathbf{a}_i \|^2 - r^2 \leq 0$ for $1 \leq i \leq m$. The Lagrangian of this problem is:

$$L(r, \mathbf{x}, \mathbf{u}) = r^{2} + \sum_{i=1}^{m} u_{i}(||\mathbf{x} - \mathbf{a}_{i}||^{2} - r^{2})$$

= $r^{2} \left(1 - \sum_{i=1}^{m} u_{i}\right) + \sum_{i=1}^{m} u_{i} ||\mathbf{x} - \mathbf{a}_{i}^{2}||$

and the dual function is:

$$g(\boldsymbol{u}) = \inf_{\substack{r \in \mathbb{R}_{\geq 0}, \boldsymbol{x} \in \mathbb{R}^n \\ r \in \mathbb{R}_{\geq 0}, \boldsymbol{x} \in \mathbb{R}^n}} L(r, \boldsymbol{x}, \boldsymbol{u})$$
$$= \inf_{\substack{r \in \mathbb{R}_{\geq 0}, \boldsymbol{x} \in \mathbb{R}^n \\ r \in \mathbb{R}_{\geq 0}, \boldsymbol{x} \in \mathbb{R}^n}} r^2 \left(1 - \sum_{i=1}^m u_i \right) + \sum_{i=1}^m u_i \| \boldsymbol{x} - \boldsymbol{a}_i \|^2 \|.$$

This leads to the following conditions:

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

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

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

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

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.

Consider the primal problem minimize $x_1^2 + x_2^2$, where $x_1, x_2 \in \mathbb{R}$, subject to $x_1 - 1 \ge 0$. It is clear that the minimum of $f(\mathbf{x})$ is obtained for $x_1 = 1$ and $x_2 = 0$ and this minimum is 1. The Lagrangian is

$$L(\boldsymbol{u}) = x_1^2 + x_2^2 + u_1(x_1 - 1)$$

and the dual function is

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

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

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

minimize
$$f(\mathbf{x}) = ax_1^2 + bx_2^2$$

subject to $px_1 + qx_2 + r \leq 0$ and $x_1 \geq 0$, $x_2 \geq 0$.

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

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

where u is a Lagrangian multiplier.

Thus, the dual problem objective function is

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

= $\inf_{\mathbf{x} \in C} ax_1^2 + bx_2^2 + u(px_1 + qx_2 + r)$
= $\inf_{\mathbf{x} \in C} \{ax_1^2 + upx_1 \mid x_1 \ge 0\}$
+ $\inf_{\mathbf{x} \in C} \{bx_2^2 + uqx_2 \mid x_2 \ge 0\} + ur$

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

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

which is a concave function.

The maximum of g(u) is achieved when $u = \frac{2r}{\frac{p^2}{p_{\perp}^2} + \frac{q^2}{L}}$ and equals



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 **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

$$ax_1x_{01} + bx_2x_{02} - k = 0.$$

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.