Support Vector Machines - I

Prof. Dan A. Simovici

UMB

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 $f(x_1) \leqslant M$ 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 $f(x_1) \geqslant m$ for every $x_1 \in E$. The element x_0 is a minimizer of f on E.

- If $f : [a, b] \longrightarrow \mathbb{R}$ 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 ∇f notation

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

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

Example

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

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

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

Example

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

$$f(\boldsymbol{x}) = \sum_{j=1}^{n} (\boldsymbol{b}_j' \boldsymbol{x} - c_j)^2.$$

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

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

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.

Example

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

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

Definition

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

Definition

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

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

Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be a function in $C^2(B(\boldsymbol{x}_0, r))$. The Hessian matrix in $\boldsymbol{x}_0$ is

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

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

$$\begin{aligned} \boldsymbol{h}' H_f(\boldsymbol{x}_0)\boldsymbol{h} &= a_{11}h_1^2 + 2a_{12}h_1h_2 + a_{22}h_2^2 \\ &= h_2^2\left(a_{11}\xi^2 + 2a_{12}\xi + a_{22}\right), \end{aligned}$$

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

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

- $\boldsymbol{h}' H_f(\boldsymbol{x}_0)\boldsymbol{h} \geqslant 0$ for every $\boldsymbol{h}$ if $a_{11} > 0$ and $a_{12}^2 - a_{11}a_{22} < 0$; in this case, $H_f(\boldsymbol{x}_0)$ is positive semidefinite and $\boldsymbol{x}_0$ is a local minimum;
- $\boldsymbol{h}' H_f(\boldsymbol{x}_0)\boldsymbol{h} \leqslant 0$ for every $\boldsymbol{h}$ if $a_{11} < 0$ and $a_{12}^2 - a_{11}a_{22} < 0$; in this case, $H_f(\boldsymbol{x}_0)$ is negative semidefinite and $\boldsymbol{x}_0$ is a local maximum;
- if $a_{12}^2 - a_{11}a_{22} \geqslant 0$; in this case, $H_f(\boldsymbol{x}_0)$ is neither positive nor negative definite, so $\boldsymbol{x}_0$ 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 $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ be m points in $\mathbb{R}^n$. The function $f(\boldsymbol{x}) = \sum_{i=1}^{m} \| \boldsymbol{x} - \boldsymbol{a}_i \|^2$ gives the sum of squares of the distances between $\boldsymbol{x}$ and the points $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$. We will prove that this sum has a global minimum obtained when $\boldsymbol{x}$ is the barycenter of the set $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\}$.

Example (cont'd)

We have

$$\begin{aligned} f(\boldsymbol{x}) &= m \parallel \boldsymbol{x} \parallel^2 -2\sum_{i=1}^{m} \boldsymbol{a}_i'\boldsymbol{x} + \sum_{i=1}^{m} \parallel \boldsymbol{a}_i \parallel^2 \\ &= m(x_1^2 + \cdots + x_n^2) - 2\sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}x_j + \sum_{i=1}^{m} \parallel \boldsymbol{a}_i \parallel^2, \end{aligned}$$

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 $1 \leqslant j \leqslant n$.

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 $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\}$.

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 $f(\boldsymbol{x})$, *where* $\boldsymbol{x} \in \mathbb{R}^n$,
subject to $\boldsymbol{c}(\boldsymbol{x}) \leqslant \mathbf{0}_m$, *where* $\boldsymbol{c} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$,
and $\boldsymbol{d}(\boldsymbol{x}) = \mathbf{0}_p$, *where* $\boldsymbol{d} : \mathbb{R}^n \longrightarrow \mathbb{R}^p$.

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

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

If the feasible region $R_{\mathbf{c},\mathbf{d}}$ is non-empty and bounded, then, under certain conditions a solution exists. If $R_{\mathbf{c},\mathbf{d}} = \emptyset$ we say that the constraints are *inconsistent*.

If only inequality constraints are present (as specified by the function $\boldsymbol{c}$) the feasible region is:

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

Let $\boldsymbol{x} \in R_{\boldsymbol{c}}$. The *set of active constraints* at $\boldsymbol{x}$ is

$$\mathrm{ACT}(R_{\boldsymbol{c}}, \boldsymbol{c}, \boldsymbol{x}) = \{i \in \{1, \dots, m\} \mid c_i(\boldsymbol{x}) = 0\}.$$

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

Definition

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

$$\begin{aligned} &\textit{minimize } f(\boldsymbol{x}), \textit{ where } \boldsymbol{x} \in \mathbb{R}^n, \\ &\qquad\qquad \textit{subject to } \boldsymbol{x} \in R_{\boldsymbol{c}}. \end{aligned}$$

If $\boldsymbol{x}_0$ exists in $R_{\boldsymbol{c}}$ that $f(\boldsymbol{x}_0) = \min\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in R_{\boldsymbol{c}}\}$ we refer to $\boldsymbol{x}_0$ as a solution of $\mathrm{MP}(f, \boldsymbol{c})$.

If $\boldsymbol{h} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ we can write

$$\boldsymbol{h}(\boldsymbol{x}) = \begin{pmatrix} h_1(\boldsymbol{x}) \\ \vdots \\ h_m(\boldsymbol{x}) \end{pmatrix},$$

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

$$(D\boldsymbol{h})(\boldsymbol{x}) = \begin{pmatrix} (\nabla h_1)(\boldsymbol{x})' \\ \vdots \\ (\nabla h_m)(\boldsymbol{x})' \end{pmatrix}.$$

Example

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

$$(D\boldsymbol{h})(\boldsymbol{x}) = \begin{pmatrix} x_2 & x_1 \\ 2x_1 & 0 \\ 0 & 2x_2 \end{pmatrix}.$$

Observe that the rows of $(D\boldsymbol{h})(\boldsymbol{x})$ are the gradients of the components of h.

Theorem

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

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

If $\boldsymbol{x}_0$ *is a regular point of* $\boldsymbol{h}$ *and a local extremum of* f *subjected to the restriction* $\boldsymbol{h}(\boldsymbol{x}_0) = \boldsymbol{0}_m$*, then* $(\nabla f)(\boldsymbol{x}_0)$ *is a linear combination of* $(\nabla h_1)(\boldsymbol{x}_0), \ldots, (\nabla h_m)(\boldsymbol{x}_0)$.

Example

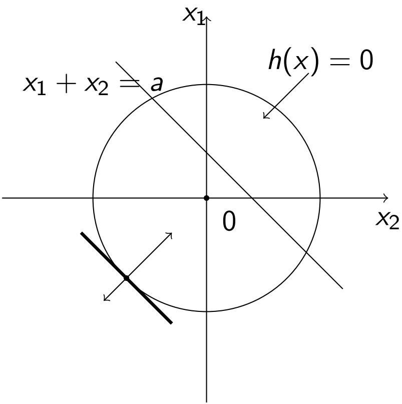
Suppose that we wish to minimize $f(\boldsymbol{x}) = x_1 + x_2$ subject to the condition

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

We have

$$\begin{aligned} (\nabla f)(\boldsymbol{x}) &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ (\nabla h)(\boldsymbol{x}) &= \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}. \end{aligned}$$

Example continued



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

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

To apply the Lagrange multiplier technique the constraint gradients

$$(\nabla h_1)(\boldsymbol{x}), \cdots, (\nabla h_m)(\boldsymbol{x})$$

must be linearly independent. In this case, $\boldsymbol{x}$ is said to be regular.
There may not exist Lagrange multipliers for a local minimum that is not regular.

Example

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

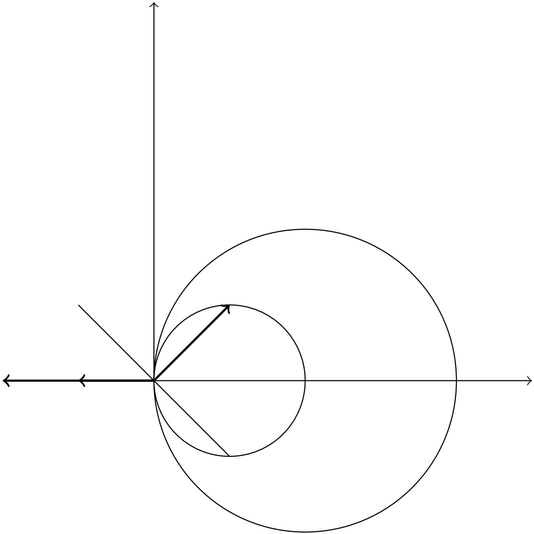
$$h_1(\boldsymbol{x}) = (x_1 - 1)^2 + x_2^2 - 1 = 0, h_2(\boldsymbol{x}) = (x_1 - 2)^2 + x_2^2 - 4 = 0.$$

We have

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

and

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



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, so $\mathbf{0}$ 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 $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be the function defined by $f(\boldsymbol{x}) = \boldsymbol{x}'A\boldsymbol{x}$.

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

Since $(\nabla f) = 2A\boldsymbol{x}$ and $(\nabla h)(\boldsymbol{x}) = 2\boldsymbol{x}$ there exists λ such that $2A\boldsymbol{x}_0 = 2\lambda\boldsymbol{x}_0$ for any extremum of f subjected to $\| \boldsymbol{x}_0 \| = 1$. Thus, $\boldsymbol{x}_0$ 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 $(\nabla c_i)(\boldsymbol{x}_0)$ for $i \in \mathrm{ACT}(S, \boldsymbol{c}, \boldsymbol{x}_0)\}$, and
- ensure that the coefficient of the gradient of the objective function $(\nabla f)(\boldsymbol{x}_0)$ 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 $\boldsymbol{c} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$. Let $\boldsymbol{x}_0$ be a local minimum in S of f subjected to the restriction $\boldsymbol{c}(\boldsymbol{x}_0) \leqslant \boldsymbol{0}_m$.*
Suppose that f is differentiable in $\boldsymbol{x}_0$, c_i are differentiable in $\boldsymbol{x}_0$ for $i \in ACT(S, \boldsymbol{c}, \boldsymbol{x}_0)$, and c_i are continuous in $\boldsymbol{x}_0$ for $i \notin ACT(S, \boldsymbol{c}, \boldsymbol{x}_0)$.
If $\{(\nabla c_i)(\boldsymbol{x}_0) \mid i \in ACT(S, \boldsymbol{c}, \boldsymbol{x}_0)\}$ is a linearly independent set, then there exist non-negative numbers w_i for $i \in ACT(S, \boldsymbol{c}, \boldsymbol{x}_0)$ 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 $\boldsymbol{x}_0$ for $i \notin \mathrm{ACT}(S, \boldsymbol{c}, \boldsymbol{x}_0)$, then the previous condition can be written as:

- $(\nabla f)(\boldsymbol{x}_0) + \sum_{i=1}^{m} w_i (\nabla c_i)(\boldsymbol{x}_0) = \mathbf{0}_n$;
- $\boldsymbol{w}'\boldsymbol{c}(\boldsymbol{x}_0) = 0$;
- $\boldsymbol{w} \geqslant \mathbf{0}_m$, where $\boldsymbol{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix}$.

The Primal Problem

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

$$\begin{aligned} &\textit{minimize } f(\boldsymbol{x}), \textit{where } \boldsymbol{x} \in C, \\ &\qquad\qquad \textit{subject to } \boldsymbol{c}(\boldsymbol{x}) \leqslant \mathbf{0}_m \\ &\qquad\qquad \textit{and } \boldsymbol{d}(\boldsymbol{x}) = \mathbf{0}_p. \end{aligned}$$

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

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:

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

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 $c_i(\boldsymbol{x}) \leqslant 0$; the component v_j of $\boldsymbol{v}$ is the *Lagrangian multiplier* corresponding to the constraint $d_j(\boldsymbol{x}) = 0$.

Lemma

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

Proof: at each feasible $\boldsymbol{x}$ we have $c_i(x) \leqslant 0$ and $d_i(\boldsymbol{x}) = 0$, 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 $1 \leqslant i \leqslant m$.

Lemma

The optimal value of the primal problem f^* is

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

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

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

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

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 $\boldsymbol{u} \geqslant \mathbf{0}_m$.

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:

$$
\begin{aligned}
& g(t\boldsymbol{u}_1 + (1-t)\boldsymbol{u}_2, t\boldsymbol{v}_1 + (1-t)\boldsymbol{v}_2) \\
& \quad = \inf\{f(\boldsymbol{x}) + (t\boldsymbol{u}_1' + (1-t)\boldsymbol{u}_2')\boldsymbol{c}(\boldsymbol{x}) + (t\boldsymbol{v}_1' + (1-t)\boldsymbol{v}_2')\boldsymbol{d}(\boldsymbol{x}) \mid \boldsymbol{x} \in S\} \\
& \quad = \inf\{t(f(\boldsymbol{x}) + \boldsymbol{u}_1'\boldsymbol{c} + \boldsymbol{v}_1'\boldsymbol{d}) + (1-t)(f(\boldsymbol{x}) + \boldsymbol{u}_2'\boldsymbol{c}(\boldsymbol{x}) + \boldsymbol{v}_2'\boldsymbol{d}(\boldsymbol{x})) \mid \boldsymbol{x} \in S\} \\
& \quad \geqslant t\inf\{f(\boldsymbol{x}) + \boldsymbol{u}_1'\boldsymbol{c} + \boldsymbol{v}_1'\boldsymbol{d} \mid \boldsymbol{x} \in S\} \\
& \qquad + (1-t)\inf\{f(\boldsymbol{x}) + \boldsymbol{u}_2'\boldsymbol{c}(\boldsymbol{x}) + \boldsymbol{v}_2'\boldsymbol{d}(\boldsymbol{x}) \mid \boldsymbol{x} \in S\} \\
& \quad = tg(\boldsymbol{u}_1, \boldsymbol{v}_1) + (1-t)g(\boldsymbol{u}_2, \boldsymbol{v}_2),
\end{aligned}
$$

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 $f, \boldsymbol{c}$ or $\boldsymbol{d}$.
- 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 $f_* = f(x_*)$, $(\boldsymbol{u}_*, \boldsymbol{v}_*)$ is an optimum for g, and $g_* = g(\boldsymbol{u}_*, \boldsymbol{v}_*)$. We have $g_* \leqslant f_*$.*

Proof: Since $\boldsymbol{c}(\boldsymbol{x}_*) \leqslant \mathbf{0}_m$ and $\boldsymbol{d}(\boldsymbol{x}_*) = \mathbf{0}_p$ 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, $g(\boldsymbol{u}, \boldsymbol{v}) = \inf_{\boldsymbol{x} \in C} L(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}) \leqslant f_*$ for all $\boldsymbol{u}$ and $\boldsymbol{v}$.
Since g_* is the optimal value of g, the last inequality implies $g_* \leqslant f_*$.

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

$$\begin{aligned}\sup_{\boldsymbol{u} \geqslant \mathbf{0}_m, \boldsymbol{v}} L(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}) &= \sup_{\boldsymbol{u} \geqslant \mathbf{0}_m, \boldsymbol{v}} f(\boldsymbol{x}) + \boldsymbol{u}'\boldsymbol{c}(\boldsymbol{x}) + \boldsymbol{v}'\boldsymbol{d}(\boldsymbol{x}) \\ &= \begin{cases} f(\boldsymbol{x}) & \text{if } \boldsymbol{c}(\boldsymbol{x}) \leqslant \mathbf{0}_m, \\ \infty & \text{otherwise} \end{cases},\end{aligned}$$

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

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

Thus, the weak duality amounts to the inequality

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

and the strong duality is equivalent to the equality

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

Example

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

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

The constraint functions are $\boldsymbol{c}(\boldsymbol{x}) = -\boldsymbol{x}$ and $\boldsymbol{d}(\boldsymbol{x}) = A\boldsymbol{x} - \boldsymbol{b}$ and the Lagrangian L is

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}) &= \boldsymbol{a}'\boldsymbol{x} - \boldsymbol{u}'\boldsymbol{x} + \boldsymbol{v}'(A\boldsymbol{x} - \boldsymbol{b}) \\ &= -\boldsymbol{v}'\boldsymbol{b} + (\boldsymbol{a}' - \boldsymbol{u}' + \boldsymbol{v}'A)\boldsymbol{x}. \end{aligned}$$

Example (cont'd)

This yields the dual function

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

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

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

Thus, the dual problem is

maximize $g(\boldsymbol{u}, \boldsymbol{v})$,

subject to $\boldsymbol{u} \geqslant \boldsymbol{0}_m$.

Example (cont'd)

An equivalent of the dual problem is

$$\begin{aligned} &\textit{maximize } -\mathbf{v}'\mathbf{b}, \\ &\qquad\qquad \textit{subject to } \mathbf{a} - \mathbf{u} + A'\mathbf{v} = \mathbf{0}_n \\ &\qquad\qquad \textit{and } \mathbf{u} \geqslant \mathbf{0}_m. \end{aligned}$$

In turn, this problem is equivalent to:

$$\begin{aligned} &\textit{maximize } -\mathbf{v}'\mathbf{b}, \\ &\qquad\qquad \textit{subject to } \mathbf{a} + A'\mathbf{v} \geqslant \mathbf{0}_n. \end{aligned}$$

Example

The following optimization problem

$$\begin{aligned} &\textit{minimize } \tfrac{1}{2}\boldsymbol{x}' Q\boldsymbol{x} - \boldsymbol{r}'\boldsymbol{x}, \\ &\qquad \textit{where } \boldsymbol{x} \in \mathbb{R}^n, \\ &\qquad \textit{subject to } A\boldsymbol{x} \geqslant \boldsymbol{b}, \end{aligned}$$

where $Q \in \mathbb{R}^{n\times n}$ is a positive definite matrix, $\boldsymbol{r} \in \mathbb{R}^n$, $A \in \mathbb{R}^{p\times n}$, and $\boldsymbol{b} \in \mathbb{R}^p$ is known as a *quadratic optimization problem*.

The Lagrangian L is

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

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

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

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

Example

Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{R}^n$. We seek to determine a closed sphere $B[\boldsymbol{x}, r]$ of minimal radius that includes all points $\boldsymbol{a}_i$ for $1 \leqslant i \leqslant m$. 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* $r \geqslant 0$,

subject to $\| \boldsymbol{x} - \boldsymbol{a}_i \| \leqslant r$ *for* $1 \leqslant i \leqslant m$.

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

$$\begin{aligned} L(r, \boldsymbol{x}, \boldsymbol{u}) &= r^2 + \sum_{i=1}^{m} u_i (\| \boldsymbol{x} - \boldsymbol{a}_i \|^2 - r^2) \\ &= r^2 \left(1 - \sum_{i=1}^{m} u_i \right) + \sum_{i=1}^{m} u_i \| \boldsymbol{x} - \boldsymbol{a}_i^2 \| \end{aligned}$$

and the dual function is:

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

This leads to the following conditions:

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

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

The first equality yields $\sum_{i=1}^{m} u_i = 1$. Therefore, from the second equality we obtain $\boldsymbol{x} = \sum_{i=1}^{m} u_i \boldsymbol{a}_i$. This shows that for $\boldsymbol{x}$ is a convex combination of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$. 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 $g_i(\boldsymbol{x}, r) = \| \boldsymbol{x} - \boldsymbol{a}_i \|^2 - r^2 \leqslant 0$ are *not convex*.

Example

Consider the primal problem

minimize $x_1^2 + x_2^2$, *where* $x_1, x_2 \in \mathbb{R}$,

subject to $x_1 - 1 \geqslant 0$.

It is clear that the minimum of $f(\boldsymbol{x})$ 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 $\sup\{g(u_1) \mid u_1 \geqslant 0\} = 0$ 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:

$$\textit{minimize } f(\boldsymbol{x}) = ax_1^2 + bx_2^2$$
$$\textit{subject to } px_1 + qx_2 + r \leqslant 0 \textit{ and } x_1 \geqslant 0,\ x_2 \geqslant 0.$$

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

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

where u is a Lagrangian multiplier.

Thus, the dual problem objective function is

$$\begin{aligned} g(u) &= \inf_{\boldsymbol{x} \in C} L(\boldsymbol{x}, u) \\ &= \inf_{\boldsymbol{x} \in C} ax_1^2 + bx_2^2 + u(px_1 + qx_2 + r) \\ &= \inf_{\boldsymbol{x} \in C} \{ax_1^2 + upx_1 \mid x_1 \geqslant 0\} \\ &\quad + \inf_{\boldsymbol{x} \in C} \{bx_2^2 + uqx_2 \mid x_2 \geqslant 0\} + ur \end{aligned}$$

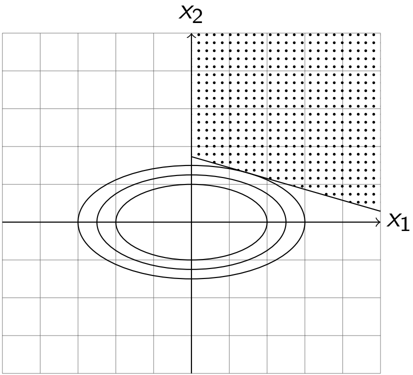
The infima are achieved when $x_1 = -\frac{up}{2a}$ and $x_2 = -\frac{uq}{2b}$ if $u \geqslant 0$ and at $\boldsymbol{x} = \mathbf{0}_2$ if $u < 0$. Thus,

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

which is a concave function.

The maximum of $g(u)$ is achieved when $u = \frac{2r}{\frac{p^2}{a}+\frac{q^2}{b}}$ 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 $\boldsymbol{x}$ is located on an ellipse $ax_1^2 + bx_2^2 - k = 0$, then $f(\boldsymbol{x}) = k$. 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 $\boldsymbol{x}_0 = \begin{pmatrix} x_{01} \\ x_{02} \end{pmatrix}$ 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.