Support Vector Machines-Preliminaries

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 $\mathbf{z} \in X$. The *gradient* of f in $\mathbf{z}$ is the vector

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

Example

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

We have

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

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

Example

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

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

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

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

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

is the *Hessian matrix* of f.

Example

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

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

Definition

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

Definition

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

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 $\mathbf{x}' A \mathbf{x} > 0$ 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 $a_{11} = \frac{\partial^2 f}{\partial {x_1}^2}(\mathbf{x}_0)$, $a_{12} = \frac{\partial^2 f}{\partial x_1 \, \partial x_2}(\mathbf{x}_0)$, and $a_{22} = \frac{\partial^2 f}{\partial {x_2}^2}(\mathbf{x}_0)$. Note that

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

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

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

i $\mathbf{h}' H_f(\mathbf{x}_0)\mathbf{h} \geqslant 0$ for every $\mathbf{h}$ if $a_{11} > 0$ and $a_{12}^2 - a_{11}a_{22} < 0$; in this case, $H_f(\mathbf{x}_0)$ is positive semidefinite and $\mathbf{x}_0$ is a local minimum;

ii $\mathbf{h}' H_f(\mathbf{x}_0)\mathbf{h} \leqslant 0$ for every $\mathbf{h}$ 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;

iii if $a_{12}^2 - a_{11}a_{22} \geqslant 0$; in this case, $H_f(\mathbf{x}_0)$ is neither positive nor negative definite, so $\mathbf{x}_0$ 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 $\mathbf{a}_1, \ldots, \mathbf{a}_m$ 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

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

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

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, $\mathbf{c} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, and $\mathbf{d} : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ 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 $\mathbf{c}(\mathbf{x}) \leqslant \mathbf{0}_m$, *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 $\mathbf{c}$ 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 $\mathbf{c}$) the feasible region is:

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

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

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

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

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 $\text{MP}(f, \mathbf{c})$ is:

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

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 $\text{MP}(f, \mathbf{c})$.

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

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

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

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

Example

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

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

Then

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

Theorem

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

- $m < n$,
- $f \in C^1(\mathbb{R}^n)$,
- $\mathbf{h} \in C^1(\mathbb{R}^n)$, *and*
- *the matrix* $(D\mathbf{h})(\mathbf{x})$ *is of full rank, that is,* $rank((D\mathbf{h})(\mathbf{x})) = m < n$.

If $\mathbf{x}_0$ *is a regular point of* $\mathbf{h}$ *and 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* $(\nabla h_1)(\mathbf{x}_0), \ldots, (\nabla h_m)(\mathbf{x}_0)$.

Example

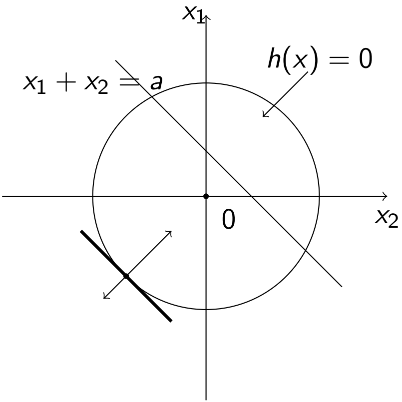
Suppose that we wish to minimize $f(\mathbf{x}) = x_1 + x_2$ subject to the condition

$$h(\mathbf{x}) = x_1^2 + x_2^2 - 2 = 0.$$

We have

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

(Example con'd)



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

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

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

Example

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

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

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 $\mathbf{0}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. At that point, we have

$$(\nabla f)(\mathbf{0}_2) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, (\nabla h_1)(\mathbf{0}_2) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}, (\nabla h_2)(\mathbf{0}_2) = \begin{pmatrix} -4 \\ 0 \end{pmatrix}.$$

The gradients $(\nabla h_1)(\mathbf{0}_2), (\nabla h_2)(\mathbf{0}_2)$ are not linearly independent because $2(\nabla h_1)(\mathbf{0}_2) + (\nabla h_2)(\mathbf{0}_2) = \mathbf{0}_2$, so $\mathbf{0}_2$ 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 $f(\mathbf{x}) = \mathbf{x}' A \mathbf{x}$.

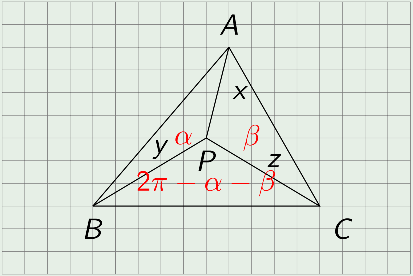
Optimization problem: minimize f subjected to the restriction $\| \mathbf{x} \| = 1$, or equivalently $h(\mathbf{x}) = \| \mathbf{x} \|^2 - 1 = 0$.

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.

Example

Let ABC be a triangle having no angle greater than $\frac{2\pi}{3}$, P a point inside ABC and let x, y, z be the lengths of the segments PA, PB and PC, respectively. The Toricelli point of the triangle is defined as the point for which $x + y + z$ is minimal.

Let the angle BPA be α, the angle APC be β and the angle CPB be $2\pi - \alpha - \beta$.



Example cont'd

The constraints are

$$
\begin{aligned}
x^2 + y^2 - 2xy\cos\alpha - c^2 &= 0,\\
x^2 + z^2 - 2xz\cos\beta - b^2 &= 0,\\
y^2 + z^2 - 2yx\cos(2\pi - \alpha - \beta) - a^2 &= 0.
\end{aligned}
$$

and the expresssion to be minimized is $x + y + z$.

Example cont'd

Thus, the Lagrangean is

$$
\begin{aligned}
L = x + y + z + \lambda_1(x^2 + y^2 - 2xy\cos\alpha - c^2) \\
+ \lambda_2(x^2 + z^2 - 2xz\cos\beta - b^2) \\
+ \lambda_3(y^2 + z^2 - 2yz\cos(2\pi - \alpha - \beta) - a^2).
\end{aligned}
$$

The optimality conditions are:

$$
\begin{aligned}
\frac{\partial L}{\partial x} = 0, \frac{\partial L}{\partial y} = 0, \frac{\partial L}{\partial z} = 0, \\
\frac{\partial L}{\partial \alpha} = 0, \frac{\partial L}{\partial \beta} = 0.
\end{aligned}
$$

Example cont'd

We have:

$$
\begin{aligned}
\frac{\partial L}{\partial x} &= 1 + 2x\lambda_1 - 2y\cos\alpha + 2x\lambda_2 - 2z\cos\beta = 0 \\
\frac{\partial L}{\partial y} &= 1 + 2y\lambda_1 - 2x\cos\alpha + 2y\lambda_3 - 2z\cos(2\pi - \alpha - \beta) = 0 \\
\frac{\partial L}{\partial z} &= 1 + 2z\lambda_2 - 2x\cos\beta + 2z\lambda_3 - 2y\cos(2\pi - \alpha - \beta) = 0 \\
\frac{\partial L}{\partial \alpha} &= 2xy\lambda_1\sin\alpha - 2yz\lambda_3\sin(2\pi - \alpha - \beta) = 0 \\
\frac{\partial L}{\partial \beta} &= 2xz\lambda_2\sin\beta - 2yz\lambda_3\sin(2\pi - \alpha - \beta) = 0.
\end{aligned}
$$

Regard the first three equations as a system in λ_1, λ_2, and λ_3:

$$\begin{aligned}
\lambda_1 + \lambda_2 &= \frac{-1 + 2y\cos\alpha + 2z\cos\beta}{2x}, \\
\lambda_1 + \lambda_3 &= \frac{-1 + 2x\cos\alpha + 2z\cos(2\pi - \alpha - \beta)}{2y}, \\
\lambda_2 + \lambda_3 &= \frac{-1 + 2x\cos\beta + 2yy\cos(2\pi - \alpha - \beta)}{2z}.
\end{aligned}$$

The last two equations are:

$$\begin{aligned}
2y(x\lambda_1 \sin\alpha - z\lambda_3 \sin(2\pi - \alpha - \beta) &= 0 \\
2z(x\lambda_2 \sin\beta - y\lambda_3 \sin(2\pi - \alpha - \beta) &= 0.
\end{aligned}$$

Example cont'd

Eliminating λ_1, λ_2 and λ_3 yields $\sin \alpha = \sin \beta$ and $\sin(\alpha + \beta) = -\sin \beta$, so $\alpha = \beta = \frac{2\pi}{3}$.

The next theorem provides necessary conditions for optimality that include the linear independence of the gradients of the components of the constraint $(\nabla c_i)(\mathbf{x}_0)$ for $i \in \mathtt{ACT}(S, \mathbf{c}, \mathbf{x}_0)\}$ 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 $\mathbf{c}(\mathbf{x}_0) \leqslant \mathbf{0}_m$. Suppose that f is differentiable in $\mathbf{x}_0$, c_i are differentiable in $\mathbf{x}_0$ for $i \in \mathtt{ACT}(S, \mathbf{c}, \mathbf{x}_0)$, and c_i are continuous in $\mathbf{x}_0$ for $i \notin \mathtt{ACT}(S, \mathbf{c}, \mathbf{x}_0)$. If $\{(\nabla c_i)(\mathbf{x}_0) \mid i \in \mathtt{ACT}(S, \mathbf{c}, \mathbf{x}_0)\}$ is a linearly independent set, then there exist non-negative numbers w_i for $i \in \mathtt{ACT}(S, \mathbf{c}, \mathbf{x}_0)$ such that*

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

Theorem continued

Furthermore, if the functions c_i are differentiable in $\mathbf{x}_0$ for $i \notin \texttt{ACT}(S, \mathbf{c}, \mathbf{x}_0)$, then the previous condition can be written as:

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

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 $\mathbf{c} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ and $\mathbf{d} : \mathbb{R}^n \longrightarrow \mathbb{R}^p$:

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

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:

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

for $\mathbf{x} \in C$, $\mathbf{u} \in \mathbb{R}^m$, and $\mathbf{v} \in \mathbb{R}^p$.
The component u_i of $\mathbf{u}$ is the *Lagrangian multiplier* corresponding to the constraint $c_i(\mathbf{x}) \leqslant 0$; the component v_j of $\mathbf{v}$ is the *Lagrangian multiplier* corresponding to the constraint $d_j(\mathbf{x}) = 0$.

Lemma

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

Proof: at each feasible $\mathbf{x}$ we have $c_i(x) \leqslant 0$ and $d_i(\mathbf{x}) = 0$, hence

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

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

Lemma

The optimal value of the primal problem f^* is

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

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

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

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

and consists of

$$\begin{aligned} &\textit{maximize } g(\mathbf{u}, \mathbf{v}), \textit{ where } \mathbf{u} \in \mathbb{R}^m \textit{ and } \mathbf{v} \in \mathbb{R}^p, \\ &\qquad\qquad \textit{subject to } \mathbf{u} \geqslant \mathbf{0}_m. \end{aligned}$$

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

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

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 $f, \mathbf{c}$ or $\mathbf{d}$.
- 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 $g_* \leqslant f_*$.*

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

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

Therefore, $g(\mathbf{u}, \mathbf{v}) = \inf_{\mathbf{x} \in C} L(\mathbf{x}, \mathbf{u}, \mathbf{v}) \leqslant f_*$ for all $\mathbf{u}$ and $\mathbf{v}$. 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

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

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

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

Thus, the weak duality amounts to the inequality

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

and the strong duality is equivalent to the equality

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

Example

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:

$$\begin{aligned} &\textit{minimize } \mathbf{a}'\mathbf{x}, \textit{ where } \mathbf{x} \in \mathbb{R}^n, \\ &\qquad \textit{subject to } \mathbf{x} \geqslant \mathbf{0}_n \textit{ and} \\ &\qquad A\mathbf{x} - \mathbf{b} = \mathbf{0}_p. \end{aligned}$$

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

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

Example (cont'd)

This yields the dual function

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

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

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

Example (cont'd)

The dual problem

$$\textit{maximize } g(\mathbf{u}, \mathbf{v}),$$
$$\textit{subject to } \mathbf{u} \geqslant \mathbf{0}_m.$$

can be expressed as:

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

In turn, this problem is equivalent to:

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

Example

The following optimization problem

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

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 $g(\mathbf{u}) = \inf_{\mathbf{x} \in \mathbb{R}^n} L(\mathbf{x}, \mathbf{u})$ subject to $\mathbf{u} \geqslant \mathbf{0}_m$. Since $\mathbf{x}$ 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 $1 \leqslant i \leqslant n$, 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 $\mathbf{u} \geqslant \mathbf{0}_p$, 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.

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 $1 \leqslant i \leqslant m$. This is the *minimum bounding sphere* problem, formulated by J. J. Sylvester. This problem amounts to solving the following primal optimization problem:

$$\begin{aligned} &\textit{minimize } r, \textit{ where } r \geqslant 0, \\ &\qquad \textit{subject to } \parallel \mathbf{x} - \mathbf{a}_i \parallel \leqslant r \textit{ for } 1 \leqslant i \leqslant m. \end{aligned}$$

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

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

and the dual function is:

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

This leads to the following conditions:

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

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 $g_i(\mathbf{x}, r) = \| \mathbf{x} - \mathbf{a}_i \|^2 - r^2 \leqslant 0$ are *not convex*.

Example

Consider the primal problem

$$\textit{minimize } x_1^2 + x_2^2, \textit{ where } x_1, x_2 \in \mathbb{R}, \\ \textit{subject to } x_1 - 1 \geqslant 0.$$

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

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

and the dual function is

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

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:

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

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

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

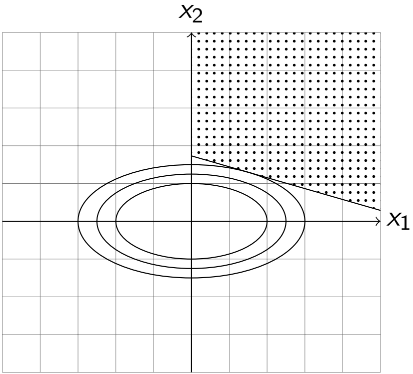
The infima are achieved when $x_1 = -\frac{up}{2a}$ and $x_2 = -\frac{uq}{2b}$ if $u \geqslant 0$ and at $\mathbf{x} = \mathbf{0}_2$ if $u < 0$. Thus,

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

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 $\mathbf{x}$ 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

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

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.