Quadratic Minimization under Linear Constraints

Last modified: 2026-10-31 15:22:31 (PDT)

Minimizing a quadratic under linear constraints

> **Source**
>
> This section is adapted from the Lagrange-multiplier part of Zhou [1], used under the MIT License (see the license text in Linear Algebra). The source solves the stationarity equations of the Lagrangian and reports the solution; here the solution is also proved to be the unique minimizer, directly from the positive definiteness of the matrices involved. The source assumes only that $\mathbf{S}$ is positive definite; the added hypothesis that the columns of $\mathbf{A}$ are linearly independent makes $\mathbf{A}^\top \mathbf{S}^{-1} \mathbf{A}$ invertible. The source's example with general coefficients a_1, a_2, b is specialized here to $3, 4, 5$. These parts of the source are not part of this section:
>
> - the Lagrange multiplier method for constraints that are not linear, or objectives that are not quadratic, including its multinomial maximum likelihood example
> - the eigenvalues of the block matrix of the stationarity equations, and the convexity of the Lagrangian in $\underset{\sim}{x}$ and its concavity in the multipliers
> - its homework examples

Definition 0.1 (Feasible set)**.** In a problem that looks for a best point among some allowed points of $\mathbb{R}^p$, the set $F \subseteq \mathbb{R}^p$ of allowed points is the **feasible set**, [2] and its points are feasible.

Definition 0.2 (Constraints)**.** When the feasible set F (Definition 0.1) is the set of points that satisfy some equations or inequalities, those equations or inequalities are the **constraints**.

Example 0.1 (A line as a feasible set)**.** With the single constraint $3x_1 + 4x_2 = 5$, the feasible set is the line $\{\underset{\sim}{x} \in \mathbb{R}^2 : 3x_1 + 4x_2 = 5\}$. $(1, 0.5)^\top$ is feasible, since

$$\begin{aligned} 3 \cdot 1 + 4 \cdot 0.5 &= 3 + 2 \\ &= 5; \end{aligned}$$

$\underset{\sim}{0}$ is not feasible, since $3 \cdot 0 + 4 \cdot 0 = 0 \neq 5$.

Definition 0.3 (Objective function, objective)**.** Let $f : \mathbb{R}^p \to \mathbb{R}$ and let $F \subseteq \mathbb{R}^p$ be a feasible set (Definition 0.1). In the problem of minimizing f over F, f is the **objective function** (or **objective**).

Definition 0.4 (Minimizer over a feasible set)**.** With f and F as in Definition 0.3, a **minimizer of f over F** is a feasible $\underset{\sim}{x}^*$ with $f(\underset{\sim}{x}^*) \leq f(\underset{\sim}{x})$ for every feasible $\underset{\sim}{x}$: a global minimizer of f with its domain restricted to F.

Definition 0.5 (Minimization problem)**.** With f and F as in Definition 0.3, the **minimization problem** [3] is to find a minimizer of f over F (Definition 0.4). When F is given by constraints (Definition 0.2), the problem is to minimize f subject to them.

Example 0.2 (Minimizing over a line)**.** In the problem of minimizing $x_1^2 + x_2^2$ subject to $3x_1 + 4x_2 = 5$, the objective function is $f(\underset{\sim}{x}) = x_1^2 + x_2^2$, and the feasible set is the line of Example 0.1. The feasible point $(1, 0.5)^\top$ has

$$\begin{aligned} f(1, 0.5) &= 1 + 0.25 \\ &= 1.25. \end{aligned}$$

$\underset{\sim}{0}$ minimizes f over all of $\mathbb{R}^2$, with $f(\underset{\sim}{0}) = 0$, but it is not a minimizer over the feasible set, because it is not feasible.

Definition 0.6 (Optimization algorithm)**.** Let:

- $f : \mathbb{R}^p \to \mathbb{R}$: an objective function;
- $F \subseteq \mathbb{R}^p$: a feasible set.

An **optimization algorithm** [3] is a step-by-step procedure that maps the pair (f, F) to a feasible point $\underset{\sim}{x}^\dagger$:

$$(f, F) \mapsto \underset{\sim}{x}^\dagger \in F. \tag{1}$$

In words: given (f, F), the procedure returns a feasible point $\underset{\sim}{x}^\dagger$ intended to make $f(\underset{\sim}{x}^\dagger)$ small.

Example 0.3 (Exhaustive search over a finite feasible set)**.** Let $f(\underset{\sim}{x}) = x_1^2 + x_2^2$, with the feasible set the three points $\left\{(1, 0.5)^\top, (3, -1)^\top, (-1, 2)^\top\right\}$ of the line of Example 0.1. The procedure "evaluate f at every feasible point; return the point with the smallest value" is an optimization algorithm. Computing f at each point gives

$$\begin{aligned} f(1, 0.5) &= 1 + 0.25 = 1.25, \\ f(3, -1) &= 9 + 1 = 10, \\ f(-1, 2) &= 1 + 4 = 5, \end{aligned}$$

so it returns $(1, 0.5)^\top$, a minimizer of f over this feasible set.

Here the objective (Definition 0.3) is a quadratic $\frac{1}{2}\underset{\sim}{x}^\top \mathbf{S} \underset{\sim}{x}$ with $\mathbf{S}$ positive definite, and the constraints (Definition 0.2) are m linear equations $\mathbf{A}^\top \underset{\sim}{x} = \underset{\sim}{b}$, where $\mathbf{A}$ is a $p \times m$ matrix and $\underset{\sim}{b} \in \mathbb{R}^m$. The only stationary point of the objective alone (Definition 4 in Critical Points and Optimality Conditions) is $\underset{\sim}{0}_{p\times 1}$ (its gradient is $\mathbf{S}\underset{\sim}{x}$, and $\mathbf{S}$ is invertible), which may not satisfy the constraints. A Lagrange multiplier adds one unknown per constraint, and the minimizer over the feasible set (Definition 0.1) is then found among the stationary points of a new function of the extended set of unknowns.

Definition 0.7 (Lagrangian)**.** For the problem of minimizing $f(\underset{\sim}{x})$ over $\underset{\sim}{x} \in \mathbb{R}^p$ subject to $\mathbf{A}^\top \underset{\sim}{x} = \underset{\sim}{b}$, with $\mathbf{A}$ a $p \times m$ matrix and $\underset{\sim}{b} \in \mathbb{R}^m$, the **Lagrangian** [4] is the function of $\underset{\sim}{x} \in \mathbb{R}^p$ and $\underset{\sim}{\lambda} \in \mathbb{R}^m$

$$L(\underset{\sim}{x}, \underset{\sim}{\lambda}) \stackrel{\text{def}}{=} f(\underset{\sim}{x}) + \underset{\sim}{\lambda}^\top (\mathbf{A}^\top \underset{\sim}{x} - \underset{\sim}{b}).$$

Definition 0.8 (Lagrange multipliers)**.** In the Lagrangian $L(\underset{\sim}{x}, \underset{\sim}{\lambda})$ (Definition 0.7), the entries of $\underset{\sim}{\lambda}$ are the **Lagrange multipliers** [4], one per constraint.

Example 0.4 (The gradients of the Lagrangian of a quadratic)**.** Let $f(\underset{\sim}{x}) = \frac{1}{2}\underset{\sim}{x}^\top \mathbf{S} \underset{\sim}{x}$ with $\mathbf{S}$ symmetric and constant. By the transpose of a product, $\underset{\sim}{\lambda}^\top \mathbf{A}^\top = (\mathbf{A}\underset{\sim}{\lambda})^\top$, so

$$\begin{aligned} L(\underset{\sim}{x}, \underset{\sim}{\lambda}) &= \tfrac{1}{2}\underset{\sim}{x}^\top \mathbf{S} \underset{\sim}{x} + \underset{\sim}{\lambda}^\top \mathbf{A}^\top \underset{\sim}{x} - \underset{\sim}{\lambda}^\top \underset{\sim}{b} \quad \text{(distribute)} \\ &= \tfrac{1}{2}\underset{\sim}{x}^\top \mathbf{S} \underset{\sim}{x} + (\mathbf{A}\underset{\sim}{\lambda})^\top \underset{\sim}{x} - \underset{\sim}{\lambda}^\top \underset{\sim}{b}, \quad \text{(transpose of a product)} \end{aligned}$$

and, since $\underset{\sim}{\lambda}^\top (\mathbf{A}^\top \underset{\sim}{x} - \underset{\sim}{b})$ is 1×1 and so equals its transpose,

$$L(\underset{\sim}{x}, \underset{\sim}{\lambda}) = \tfrac{1}{2}\underset{\sim}{x}^\top \mathbf{S} \underset{\sim}{x} + (\mathbf{A}^\top \underset{\sim}{x} - \underset{\sim}{b})^\top \underset{\sim}{\lambda}.$$

The partial derivatives of a sum are the sums of the partial derivatives, entry by entry. Using the first of these forms, derivative of a quadratic form for the first term, derivative of a linear combination for the second ($\mathbf{A}\underset{\sim}{\lambda}$ is constant in $\underset{\sim}{x}$), and constant with respect to a vector for the third,

$$\begin{aligned}
\frac{\partial}{\partial \underset{\sim}{x}} L(\underset{\sim}{x}, \underset{\sim}{\lambda}) &= \tfrac{1}{2}\frac{\partial}{\partial \underset{\sim}{x}}(\underset{\sim}{x}^\top \mathbf{S} \underset{\sim}{x}) + \frac{\partial}{\partial \underset{\sim}{x}}\big((\mathbf{A}\underset{\sim}{\lambda})^\top \underset{\sim}{x}\big) - \frac{\partial}{\partial \underset{\sim}{x}}(\underset{\sim}{\lambda}^\top \underset{\sim}{b}) && \text{(sum and constant multiple)} \\
&= \tfrac{1}{2}(2\mathbf{S}\underset{\sim}{x}) + \mathbf{A}\underset{\sim}{\lambda} - \underset{\sim}{0}_{p\times 1} && \text{(the three derivatives)} \\
&= \mathbf{S}\underset{\sim}{x} + \mathbf{A}\underset{\sim}{\lambda}. && \text{(simplify)}
\end{aligned}$$

Using the second form, in which $f(\underset{\sim}{x})$ and $\mathbf{A}^\top \underset{\sim}{x} - \underset{\sim}{b}$ are constant in $\underset{\sim}{\lambda}$, and derivative of a linear combination again,

$$\begin{aligned}
\frac{\partial}{\partial \underset{\sim}{\lambda}} L(\underset{\sim}{x}, \underset{\sim}{\lambda}) &= \frac{\partial}{\partial \underset{\sim}{\lambda}} f(\underset{\sim}{x}) + \frac{\partial}{\partial \underset{\sim}{\lambda}}\big((\mathbf{A}^\top \underset{\sim}{x} - \underset{\sim}{b})^\top \underset{\sim}{\lambda}\big) && \text{(sum)} \\
&= \underset{\sim}{0}_{m\times 1} + (\mathbf{A}^\top \underset{\sim}{x} - \underset{\sim}{b}) && \text{(the two derivatives)} \\
&= \mathbf{A}^\top \underset{\sim}{x} - \underset{\sim}{b}. && \text{(simplify)}
\end{aligned}$$

So $(\underset{\sim}{x}, \underset{\sim}{\lambda})$ makes the gradients equal $\underset{\sim}{0}_{p\times 1}$ and $\underset{\sim}{0}_{m\times 1}$ exactly when

$$\mathbf{S}\underset{\sim}{x} + \mathbf{A}\underset{\sim}{\lambda} = \underset{\sim}{0}_{p\times 1} \quad \text{and} \quad \mathbf{A}^\top \underset{\sim}{x} = \underset{\sim}{b}, \qquad \text{that is,} \qquad \underbrace{\begin{bmatrix} \mathbf{S} & \mathbf{A} \\ \mathbf{A}^\top & \underset{\sim}{0}_{m\times m} \end{bmatrix}}_{(p+m)\times(p+m)} \underbrace{\begin{bmatrix} \underset{\sim}{x} \\ \underset{\sim}{\lambda} \end{bmatrix}}_{(p+m)\times 1} = \underbrace{\begin{bmatrix} \underset{\sim}{0}_{p\times 1} \\ \underset{\sim}{b} \end{bmatrix}}_{(p+m)\times 1}, \tag{2}$$

the second equation being the constraint itself.

Theorem 0.1 (Minimizing a positive definite quadratic under linear constraints)**.** *Let* $\mathbf{S}$ *be a* $p \times p$ *positive definite matrix,* $\mathbf{A}$ *a* $p \times m$ *matrix with linearly independent columns, and* $\underset{\sim}{b} \in \mathbb{R}^m$*. Then the* $m \times m$ *matrix* $\mathbf{M} \overset{def}{=} \mathbf{A}^\top \mathbf{S}^{-1} \mathbf{A}$ *is invertible, and* $f(\underset{\sim}{x}) = \frac{1}{2}\underset{\sim}{x}^\top \mathbf{S} \underset{\sim}{x}$ *has exactly one minimizer over the* $\underset{\sim}{x} \in \mathbb{R}^p$ *with* $\mathbf{A}^\top \underset{\sim}{x} = \underset{\sim}{b}$*, namely*

$$\underset{\sim}{x}^* = \mathbf{S}^{-1}\mathbf{A}\mathbf{M}^{-1}\underset{\sim}{b}.$$

With $\underset{\sim}{\lambda}^* \overset{def}{=} -\mathbf{M}^{-1}\underset{\sim}{b}$*, the pair* $(\underset{\sim}{x}^*, \underset{\sim}{\lambda}^*)$ *solves Equation 2, and the minimum value is* $f(\underset{\sim}{x}^*) = \frac{1}{2}\underset{\sim}{b}^\top \mathbf{M}^{-1}\underset{\sim}{b}$*.*

Proof

Proof. The steps marked "transpose of a product" use the transpose of a product, and $\mathbf{S}$ is symmetric because it is positive definite.

$\mathbf{M}$ is invertible. $\mathbf{S}^{-1}$ exists and is positive definite (positive definite inverse), so $\mathbf{M} = \mathbf{A}^\top \mathbf{S}^{-1} \mathbf{A}$ is positive definite (operations that preserve definiteness, part 1), and so invertible (positive definite inverse again). $\mathbf{S}^{-1}$ and $\mathbf{M}^{-1}$ are symmetric (inverse of a symmetric matrix), because $\mathbf{S}$ and $\mathbf{M}$ are.

$\underset{\sim}{x}^*$ satisfies the constraint, and the pair solves Equation 2.

$$\begin{aligned}
\mathbf{A}^\top \underset{\sim}{x}^* &= \mathbf{A}^\top \mathbf{S}^{-1} \mathbf{A} \mathbf{M}^{-1} \underset{\sim}{b} && \text{(substitute } \underset{\sim}{x}^*\text{)} \\
&= \mathbf{M} \mathbf{M}^{-1} \underset{\sim}{b} && \text{(definition of } \mathbf{M}\text{)} \\
&= \underset{\sim}{b}, && (\mathbf{M}\mathbf{M}^{-1} = \mathbf{I}_m) \\
\mathbf{S} \underset{\sim}{x}^* + \mathbf{A} \underset{\sim}{\lambda}^* &= \mathbf{S}\mathbf{S}^{-1} \mathbf{A} \mathbf{M}^{-1} \underset{\sim}{b} - \mathbf{A} \mathbf{M}^{-1} \underset{\sim}{b} && \text{(substitute } \underset{\sim}{x}^* \text{ and } \underset{\sim}{\lambda}^*\text{)} \\
&= \mathbf{A} \mathbf{M}^{-1} \underset{\sim}{b} - \mathbf{A} \mathbf{M}^{-1} \underset{\sim}{b} && (\mathbf{S}\mathbf{S}^{-1} = \mathbf{I}_p) \\
&= \underset{\sim}{0}_{p \times 1}. && \text{(subtract)}
\end{aligned}$$

$\underset{\sim}{x}^*$ is the unique minimizer. Let $\underset{\sim}{x}$ satisfy $\mathbf{A}^\top \underset{\sim}{x} = \underset{\sim}{b}$, and let $\underset{\sim}{d} \stackrel{\text{def}}{=} \underset{\sim}{x} - \underset{\sim}{x}^*$, so

$$\begin{aligned}
\mathbf{A}^\top \underset{\sim}{d} &= \mathbf{A}^\top \underset{\sim}{x} - \mathbf{A}^\top \underset{\sim}{x}^* \\
&= \underset{\sim}{b} - \underset{\sim}{b} \\
&= \underset{\sim}{0}_{m \times 1}
\end{aligned}$$

(matrix products distribute; both points satisfy the constraint). First, the cross term vanishes:

$$\begin{aligned}
\underset{\sim}{d}^\top \mathbf{S} \underset{\sim}{x}^* &= \underset{\sim}{d}^\top \mathbf{S}\mathbf{S}^{-1} \mathbf{A} \mathbf{M}^{-1} \underset{\sim}{b} && \text{(substitute } \underset{\sim}{x}^*\text{)} \\
&= \underset{\sim}{d}^\top \mathbf{A} \mathbf{M}^{-1} \underset{\sim}{b} && (\mathbf{S}\mathbf{S}^{-1} = \mathbf{I}_p) \\
&= (\mathbf{A}^\top \underset{\sim}{d})^\top \mathbf{M}^{-1} \underset{\sim}{b} && \text{(transpose of a product; } (\mathbf{A}^\top)^\top = \mathbf{A}\text{)} \\
&= \underset{\sim}{0}_{m \times 1}{}^\top \mathbf{M}^{-1} \underset{\sim}{b} && (\mathbf{A}^\top \underset{\sim}{d} = \underset{\sim}{0}_{m \times 1}) \\
&= 0. && \text{(a product with a zero factor is zero)}
\end{aligned}$$

Then

$$\begin{aligned}
f(\underset{\sim}{x}) &= \tfrac{1}{2} (\underset{\sim}{x}^* + \underset{\sim}{d})^\top \mathbf{S} (\underset{\sim}{x}^* + \underset{\sim}{d}) && (\underset{\sim}{x} = \underset{\sim}{x}^* + \underset{\sim}{d}) \\
&= \tfrac{1}{2} (\underset{\sim}{x}^{*\top} + \underset{\sim}{d}^\top) \mathbf{S} (\underset{\sim}{x}^* + \underset{\sim}{d}) && \text{(the transpose of a sum is the sum of the transposes)} \\
&= \tfrac{1}{2} (\underset{\sim}{x}^{*\top} \mathbf{S} \underset{\sim}{x}^* + \underset{\sim}{x}^{*\top} \mathbf{S} \underset{\sim}{d} + \underset{\sim}{d}^\top \mathbf{S} \underset{\sim}{x}^* + \underset{\sim}{d}^\top \mathbf{S} \underset{\sim}{d}) && \text{(distribute)} \\
&= \tfrac{1}{2} (\underset{\sim}{x}^{*\top} \mathbf{S} \underset{\sim}{x}^* + \underset{\sim}{d}^\top \mathbf{S}^\top \underset{\sim}{x}^* + \underset{\sim}{d}^\top \mathbf{S} \underset{\sim}{x}^* + \underset{\sim}{d}^\top \mathbf{S} \underset{\sim}{d}) && (1 \times 1\text{: } \underset{\sim}{x}^{*\top} \mathbf{S} \underset{\sim}{d} = (\underset{\sim}{x}^{*\top} \mathbf{S} \underset{\sim}{d})^\top\text{; transpose of a product}) \\
&= \tfrac{1}{2} (\underset{\sim}{x}^{*\top} \mathbf{S} \underset{\sim}{x}^* + 2\, \underset{\sim}{d}^\top \mathbf{S} \underset{\sim}{x}^* + \underset{\sim}{d}^\top \mathbf{S} \underset{\sim}{d}) && (\mathbf{S}^\top = \mathbf{S}\text{; combine}) \\
&= \tfrac{1}{2}\, \underset{\sim}{x}^{*\top} \mathbf{S} \underset{\sim}{x}^* + \underset{\sim}{d}^\top \mathbf{S} \underset{\sim}{x}^* + \tfrac{1}{2}\, \underset{\sim}{d}^\top \mathbf{S} \underset{\sim}{d} && (\text{distribute } \tfrac{1}{2}) \\
&= f(\underset{\sim}{x}^*) + \tfrac{1}{2}\, \underset{\sim}{d}^\top \mathbf{S} \underset{\sim}{d} && \text{(the cross term is 0)} \\
&\geq f(\underset{\sim}{x}^*), && (\mathbf{S} \text{ is positive definite})
\end{aligned}$$

with equality only when $\underset{\sim}{d} = \underset{\sim}{0}_{p\times 1}$, that is, $\underset{\sim}{x} = \underset{\sim}{x}^*$.
The minimum value.

$$
\begin{aligned}
f(\underset{\sim}{x}^*) &= \tfrac{1}{2}\left(\mathbf{S}^{-1}\mathbf{A}\mathbf{M}^{-1}\underset{\sim}{b}\right)^\top \mathbf{S}\mathbf{S}^{-1}\mathbf{A}\mathbf{M}^{-1}\underset{\sim}{b} && \text{(substitute } \underset{\sim}{x}^*\text{)}\\
&= \tfrac{1}{2}\,\underset{\sim}{b}^\top \left(\mathbf{M}^{-1}\right)^\top \mathbf{A}^\top \left(\mathbf{S}^{-1}\right)^\top \mathbf{S}\mathbf{S}^{-1}\mathbf{A}\mathbf{M}^{-1}\underset{\sim}{b} && \text{(transpose of a product, applied repeatedly)}\\
&= \tfrac{1}{2}\,\underset{\sim}{b}^\top \mathbf{M}^{-1}\mathbf{A}^\top \mathbf{S}^{-1}\mathbf{S}\mathbf{S}^{-1}\mathbf{A}\mathbf{M}^{-1}\underset{\sim}{b} && (\mathbf{S}^{-1} \text{ and } \mathbf{M}^{-1} \text{ are symmetric})\\
&= \tfrac{1}{2}\,\underset{\sim}{b}^\top \mathbf{M}^{-1}\mathbf{A}^\top \mathbf{S}^{-1}\mathbf{A}\mathbf{M}^{-1}\underset{\sim}{b} && (\mathbf{S}^{-1}\mathbf{S} = \mathbf{I}_p)\\
&= \tfrac{1}{2}\,\underset{\sim}{b}^\top \mathbf{M}^{-1}\mathbf{M}\mathbf{M}^{-1}\underset{\sim}{b} && \text{(definition of } \mathbf{M}\text{)}\\
&= \tfrac{1}{2}\,\underset{\sim}{b}^\top \mathbf{M}^{-1}\underset{\sim}{b}. && (\mathbf{M}^{-1}\mathbf{M} = \mathbf{I}_m)
\end{aligned}
$$

□

Example 0.5 (The closest point on a line to the origin)**.** Minimize $x_1^2 + x_2^2$ subject to $3x_1 + 4x_2 = 5$. In the form of Theorem 0.1, $x_1^2 + x_2^2 = \frac{1}{2}\,\underset{\sim}{x}^\top (2\mathbf{I}_2)\underset{\sim}{x}$, so $\mathbf{S} = 2\mathbf{I}_2$, $\mathbf{A} = (3,4)^\top$ ($p = 2$, $m = 1$) and $\underset{\sim}{b} = [5]$. The hypotheses hold: $2\mathbf{I}_2$ is positive definite, and the single column $(3,4)^\top$ is nonzero, so it is linearly independent. Also $\mathbf{S}^{-1} = \frac{1}{2}\mathbf{I}_2$. Then, from the formulas of Theorem 0.1,

$$
\begin{aligned}
\mathbf{M} &= \mathbf{A}^\top \tfrac{1}{2}\mathbf{I}_2\,\mathbf{A}\\
&= \tfrac{1}{2}\left(3^2 + 4^2\right)\\
&= \tfrac{25}{2},\\
\underset{\sim}{x}^* &= \tfrac{1}{2}\mathbf{I}_2 \begin{bmatrix}3\\4\end{bmatrix} \left(\tfrac{2}{25}\right)(5)\\
&= \tfrac{5}{25}\begin{bmatrix}3\\4\end{bmatrix}\\
&= \begin{bmatrix}0.6\\0.8\end{bmatrix},\\
\underset{\sim}{\lambda}^* &= -\left(\tfrac{2}{25}\right)(5)\\
&= -0.4,\\
f(\underset{\sim}{x}^*) &= \tfrac{1}{2}(5)\left(\tfrac{2}{25}\right)(5)\\
&= 1.
\end{aligned}
$$

Checks:

- the constraint:

$$
\begin{aligned}
3(0.6) + 4(0.8) &= 1.8 + 3.2\\
&= 5;
\end{aligned}
$$

- the first block of Equation 2:

$$\begin{aligned}\mathbf{S}\underset{\sim}{x}^* + \mathbf{A}\underset{\sim}{\lambda}^* &= (1.2, 1.6)^\top + (3, 4)^\top(-0.4)\\ &= (1.2, 1.6)^\top - (1.2, 1.6)^\top\\ &= \underset{\sim}{0}_{2\times 1};\end{aligned}$$

- the value:

$$\begin{aligned}0.6^2 + 0.8^2 &= 0.36 + 0.64\\ &= 1.\end{aligned}$$

Geometrically, $\underset{\sim}{x}^*$ is the point of the line nearest the origin, at distance 1.

Example 0.6 (The hypotheses are needed)**.**

- If $\mathbf{S}$ is not positive definite there may be no minimizer: with $\mathbf{S} = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}$ and the constraint $x_1 = 1$ ($\mathbf{A} = (1, 0)^\top$, $\underset{\sim}{b} = [1]$), $f(1, x_2) = \frac{1}{2}(1 - x_2^2)$ decreases without bound as x_2 grows.

- If the columns of $\mathbf{A}$ are dependent, $\mathbf{M}$ is not invertible: with $\mathbf{S} = \mathbf{I}_2$ and $\mathbf{A} = \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$,

$$\begin{aligned}\mathbf{M} &= \mathbf{A}^\top\mathbf{A}\\ &= \begin{bmatrix}2 & 2\\ 2 & 2\end{bmatrix},\end{aligned}$$

  whose columns are equal. So $\mathbf{M}(1, -1)^\top = \underset{\sim}{0}_{2\times 1}$, and $\mathbf{M}$ is not invertible (a square matrix is invertible exactly when its null space is zero). The two constraints both read $x_1 + x_2 = b_i$, so they have no solution at all when $b_1 \neq b_2$. When $b_1 = b_2$ a unique minimizer still exists,

$$\begin{aligned}x_1 &= x_2\\ &= b_1/2,\end{aligned}$$

  but the formula of Theorem 0.1 cannot produce it.

Definition 0.9 (Sensitivity)**.** Suppose a minimization problem (Definition 0.5) depends on a vector $\underset{\sim}{b} \in \mathbb{R}^m$, and let $f^*(\underset{\sim}{b})$ be its minimum value, as a function of $\underset{\sim}{b}$. When the gradient $\frac{\partial}{\partial \underset{\sim}{b}} f^*(\underset{\sim}{b})$ exists (gradient), it is the **sensitivity** of the minimum value to $\underset{\sim}{b}$: by

partial derivative, its entry i is the limit, as a change in b_i alone shrinks to 0, of the resulting change in the minimum value divided by the change in b_i.

Example 0.7 (The sensitivity of a one-point problem)**.** Minimize x^2 subject to $x = b$, for a number b. The feasible set (Definition 0.1) is $\{b\}$, so the minimum value is $f^*(b) = b^2$, and its sensitivity to b is $\frac{d}{db} f^*(b) = 2b$. At $b = 3$ the sensitivity is 6: moving b from 3 to 3.1 raises the minimum value from 9 to 9.61, an increase of 0.61, close to $6 \times 0.1 = 0.6$.

Corollary 0.1 (The multiplier is the sensitivity of the minimum value)**.** *In Theorem 0.1, write the minimum value as a function of* $\underset{\sim}{b}$*,* $f^*(\underset{\sim}{b}) = \frac{1}{2}\, \underset{\sim}{b}^\top \mathbf{M}^{-1} \underset{\sim}{b}$*. Its sensitivity to* $\underset{\sim}{b}$ *(Definition 0.9) is*

$$\begin{aligned} \frac{\partial}{\partial \underset{\sim}{b}} f^*(\underset{\sim}{b}) &= \mathbf{M}^{-1} \underset{\sim}{b} \\ &= -\underset{\sim}{\lambda}^*. \end{aligned}$$

Proof

Proof. $\mathbf{M}^{-1}$ is symmetric and does not depend on $\underset{\sim}{b}$, so derivative of a quadratic form applies with $\mathbf{M}^{-1}$ in place of $\mathbf{S}$:

$$\begin{aligned} \frac{\partial}{\partial \underset{\sim}{b}} f^*(\underset{\sim}{b}) &= \tfrac{1}{2} \frac{\partial}{\partial \underset{\sim}{b}} \left(\underset{\sim}{b}^\top \mathbf{M}^{-1} \underset{\sim}{b} \right) && \text{(constant multiple)} \\ &= \tfrac{1}{2} (2 \mathbf{M}^{-1} \underset{\sim}{b}) && \text{(derivative of a quadratic form)} \\ &= \mathbf{M}^{-1} \underset{\sim}{b} && \text{(simplify)} \\ &= -\underset{\sim}{\lambda}^*. && \text{(definition of } \underset{\sim}{\lambda}^*) \end{aligned}$$

□

Example 0.8 (Moving the line)**.** In Example 0.5 with a general right side b in place of 5, $\mathbf{M}^{-1} = \frac{2}{25}$ and

$$\begin{aligned} f^*(b) &= \tfrac{1}{2} \left(\tfrac{2}{25} \right) b^2 \\ &= \tfrac{b^2}{25}, \end{aligned}$$

so $\frac{d}{db} f^*(b) = \frac{2b}{25}$. At $b = 5$ this derivative is

$$\begin{aligned} \tfrac{10}{25} &= 0.4 \\ &= -\lambda^* : \end{aligned}$$

moving the line $3x_1 + 4x_2 = b$ outward by a small step ε in b raises the minimum value

by about $0.4\,\varepsilon$.

[1] H. Zhou, "Optimization and multivariate calculus." Lecture notes for Biostat 216, Mathematical Methods for Biostatistics, University of California, Los Angeles, 2024. Accessed: Oct. 05, 2026. [Online]. Available: https://ucla-biostat-216.github.io/2024fall/slides/13-optim/13-optim.html

[2] Wikipedia contributors, "Feasible region — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Feasible_region

[3] Wikipedia contributors, "Mathematical optimization — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Mathematical_optimization

[4] Wikipedia contributors, "Lagrange multiplier — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Lagrange_multiplier