6 Positive Definite Matrices (正定矩阵)

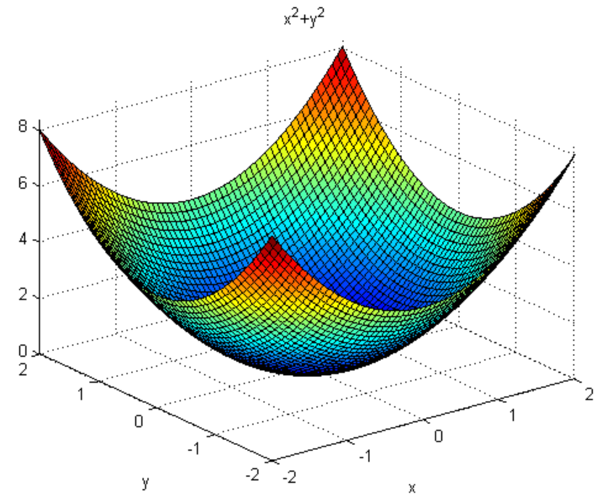
6.4

MINIMUM PRINCIPLES (最小值原理)

Minimizing without Constraints

Least Squares Again

The Rayleigh Quotient



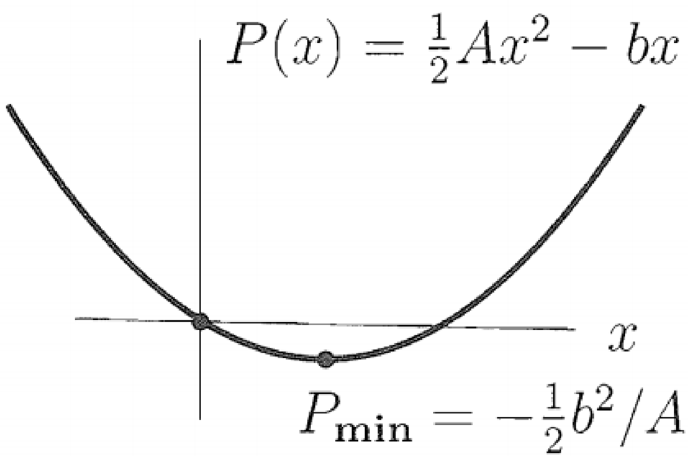
I. Minimizing without Constraints

Look at the "parabola" (抛物线) $P(x) = \frac{1}{2}Ax^2 - bx$.

If A is just a scalar, the graph of $P(x) = \frac{1}{2}Ax^2 - bx$ has zero slope when $\frac{dP}{dx} = Ax - b = 0$.

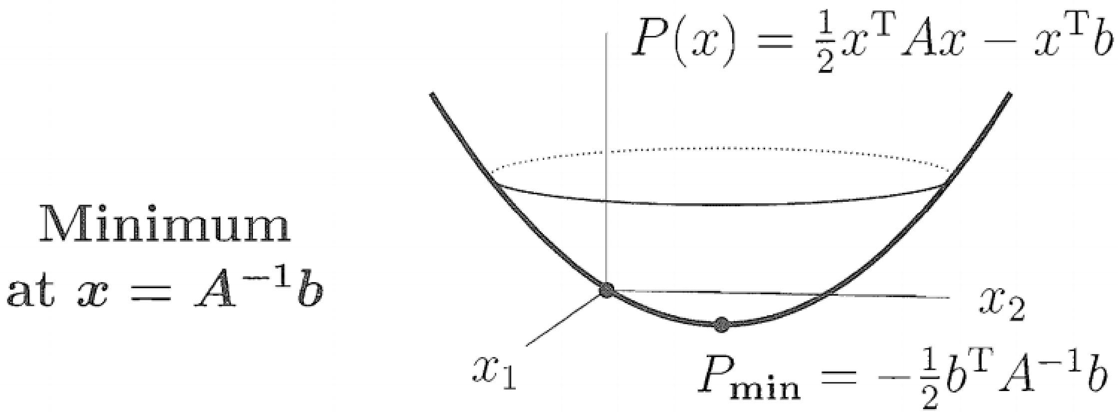
This point $x = A^{-1}b$ will be a minimum if A is positive. Then the parabola $P(x)$ opens upward.

The minimum of $P(x)$ occurs when $Ax = b$.



Minimum at $x = A^{-1}b$

In more dimensions this parabola turns into a parabolic bowl (a paraboloid, 抛物面).
To assure a *minimum* of $P(x)$, not a maximum or a saddle point, $\boldsymbol{A}$ must be *positive definite*!



The graph of a positive quadratic $P(x)$ is a parabolic bowl.

Theorem 1 If $\boldsymbol{A}$ is *real symmetric positive definite*, then $P(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{b}$ reaches its minimum at the point where $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. At that point $P_{min} = -\frac{1}{2}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b}$.

Theorem 1 If $\boldsymbol{A}$ is real symmetric positive definite, then $P(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{b}$ reaches its minimum at the point where $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. At that point $P_{min} = -\frac{1}{2}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b}$.

Proof Suppose $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. For any vector $\boldsymbol{y}$, we show that $P(\boldsymbol{y}) \geq P(\boldsymbol{x})$:

$$
\begin{aligned}
& P(\boldsymbol{y}) - P(\boldsymbol{x}) \\
&= \frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y} - \boldsymbol{y}^{\mathrm{T}}\boldsymbol{b} - \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{b} \\
&= \frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y} - \boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \qquad (set\ \boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}) \\
&= \frac{1}{2}(\boldsymbol{y} - \boldsymbol{x})^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{y} - \boldsymbol{x}).
\end{aligned}
$$

This can’t be negative since $\boldsymbol{A}$ is positive definite—and it is zero only if $\boldsymbol{y} - \boldsymbol{x} = \boldsymbol{0}$. At all other points $P(\boldsymbol{y})$ is larger than $P(\boldsymbol{x})$, so the minimum occurs at $\boldsymbol{x}$.

Substitute $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$ into $P(\boldsymbol{x})$:

Minimum value $$P_{min} = \frac{1}{2}(\boldsymbol{A}^{-1}\boldsymbol{b})^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{A}^{-1}\boldsymbol{b}) - (\boldsymbol{A}^{-1}\boldsymbol{b})^{\mathrm{T}}\boldsymbol{b}$$
$$= -\frac{1}{2}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{b}.$$

Example 1 Minimize $P(\boldsymbol{x}) = x_1^2 - x_1x_2 + x_2^2 - b_1x_1 - b_2x_2$.

(1) The usual approach, by ***calculus***, is to set the partial derivatives to zero.

$$\boxed{\begin{aligned} \frac{\partial P}{\partial x_1} &= 2x_1 - x_2 - b_1 = 0 \\ \frac{\partial P}{\partial x_2} &= -x_1 + 2x_2 - b_2 = 0 \end{aligned}}$$

(2) ***Linear algebra*** recognizes this $P(\boldsymbol{x})$ as $\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{b}$, where

$\boldsymbol{A} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$, $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$, and knows immediately that

$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ gives the minimum, i.e., $\boxed{\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}}.$

II. Least Squares Again

In minimization, our big application is least squares.

The best $\hat{\boldsymbol{x}}$ is the vector that minimizes the squared error

$$E^2 = \|\boldsymbol{Ax} - \boldsymbol{b}\|^2. \qquad \text{——} \quad \textit{quadratic}$$

Actually,

$$E^2 = (\boldsymbol{Ax} - \boldsymbol{b})^{\mathrm{T}}(\boldsymbol{Ax} - \boldsymbol{b}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} - 2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}.$$

Comparing with minimizing $\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Ax} - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{b} \Rightarrow \boldsymbol{Ax} = \boldsymbol{b}$,

Minimizing $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} - 2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{b} \Rightarrow \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.

(a new way to reach the least-squares normal equation)

III. The Rayleigh Quotient

We consider the problem of minimizing the **Rayleigh quotient**

$$R(\boldsymbol{x}) = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}$$

If $\boldsymbol{A}$ is symmetric, then there exists an orthogonal matrix $\boldsymbol{Q}$ such that

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n),$$

where $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$.

Let $\boldsymbol{y} = \boldsymbol{Q}^{-1}\boldsymbol{x} = (y_1, \ldots, y_n)^{\mathrm{T}}$. Then $\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{y}$, and

$$R(\boldsymbol{x}) = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} = \frac{(\boldsymbol{Q}\boldsymbol{y})^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{Q}\boldsymbol{y})}{(\boldsymbol{Q}\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{Q}\boldsymbol{y})} = \frac{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}} = \frac{\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2}{y_1^2 + \cdots + y_n^2} \geq \lambda_1.$$

Furthermore, if $\boldsymbol{A}\boldsymbol{x}_1 = \lambda_1 \boldsymbol{x}_1$ such that λ_1 is the *smallest* eigenvalue of $\boldsymbol{A}$, then

$$R(\boldsymbol{x}_1) = \frac{\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_1}{\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_1} = \frac{\boldsymbol{x}_1^{\mathrm{T}}\lambda_1\boldsymbol{x}_1}{\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_1} = \lambda_1.$$

$$\boxed{R(\boldsymbol{x}) = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}}$$

Theorem 2 (*Rayleigh's principle*) The minimum value of Rayleigh quotient is the smallest eigenvalue λ_1, and $R(\boldsymbol{x})$ reaches that minimum at the first eigenvector $\boldsymbol{x}_1$ of $\boldsymbol{A}$.

The Rayleigh quotient is such that

$$\lambda_1 \leq R(\boldsymbol{x}) \leq \lambda_n,$$

and the maximal value of $R(\boldsymbol{x})$ is where $\boldsymbol{A}\boldsymbol{x}_n = \lambda_n \boldsymbol{x}_n$ as

$$R(\boldsymbol{x}_n) = \frac{\boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_n}{\boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{x}_n} = \frac{\boldsymbol{x}_n^{\mathrm{T}}\lambda_n\boldsymbol{x}_n}{\boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{x}_n} = \lambda_n.$$

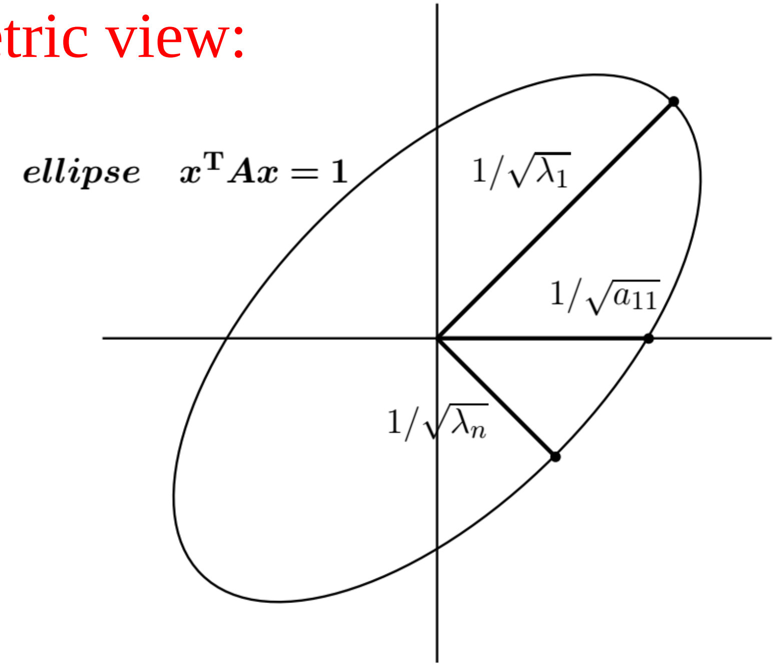
Moreover, for $\boldsymbol{x} = \boldsymbol{e}_i$, we have $R(\boldsymbol{e}_i) = a_{ii}$, and so we have the following consequence.

$$\boxed{R(\boldsymbol{x}) = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}}$$

Corollary 1 $\frac{1}{\sqrt{\lambda_n}} \le \frac{1}{\sqrt{a_{ii}}} \le \frac{1}{\sqrt{\lambda_1}}$ (or equivalently, $\lambda_1 \le a_{ii} \le \lambda_n$)

i.e., the diagonal entries of a symmetric matrix $\boldsymbol{A}$ lie between λ_1 and λ_n.

Geometric view:



The farthest $\boldsymbol{x} = \boldsymbol{x}_1/\sqrt{\lambda_1}$ and the closet $\boldsymbol{x} = \boldsymbol{x}_n/\sqrt{\lambda_n}$ both give $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\lambda\boldsymbol{x} = 1$.

These are the major axes of the ellipse.

Key words:

Minimizing without / with Constraints

Least Squares Again

The Rayleigh Quotient