Positive Definite Matrices

CS6015/LARP/2018

Ack: Linear Algebra and Its Applications, Gilbert Strang

Minima, Maxima, and Saddle Points

- Up to now, we have hardly thought about the **signs of the eigenvalues**
- E*very symmetric matrix has real eigenvalues*.
- Now we will find a test that can be applied directly to A, without computing its eigenvalues, which will guarantee that **all those eigenvalues are positive**.
- The signs of the eigenvalues are often crucial.
- The highly important problem is to **recognize a minimum point**. This arises throughout science and engineering and every problem of **optimization**.
- **Examples:**

$$F(x,y) = 7 + 2(x+y)^2 - y\sin y - x^3 \qquad f(x,y) = 2x^2 + 4xy + y^2.$$

Does either $F(x,y)$ or $f(x,y)$ have a minimum at the point $x = y = 0$?

Minima, Maxima, and Saddle Points

***Remark* 3.** The *zero-order terms* $F(0,0) = 7$ and $f(0,0) = 0$ have no effect on the answer. They simply raise or lower the graphs of F and f.

***Remark* 4.** The *linear terms* give a necessary condition: To have any chance of a minimum, the first derivatives must vanish at $x = y = 0$:

$$\frac{\partial F}{\partial x} = 4(x+y) - 3x^2 = 0 \quad \text{and} \quad \frac{\partial F}{\partial y} = 4(x+y) - y\cos y - \sin y = 0$$

$$\frac{\partial f}{\partial x} = 4x + 4y = 0 \quad \text{and} \quad \frac{\partial f}{\partial y} = 4x + 2y = 0. \quad \textit{All zero.}$$

Thus $(x, y) = (0,0)$ is a stationary point for both functions. The surface $z = F(x, y)$ is tangent to the horizontal plane $z = 7$, and the surface $z = f(x, y)$ is tangent to the plane $z = 0$.

Minima, Maxima, and Saddle Points

Remark 5. The second derivatives at $(0,0)$ ***are decisive:***

$$\frac{\partial^2 F}{\partial x^2} = 4 - 6x = 4 \qquad \frac{\partial^2 f}{\partial x^2} = 4$$

$$\frac{\partial^2 F}{\partial x \partial y} = \frac{\partial^2 F}{\partial y \partial x} = 4 \qquad \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 4$$

$$\frac{\partial^2 F}{\partial y^2} = 4 + y \sin y - 2\cos y = 2 \qquad \frac{\partial^2 f}{\partial y^2} = 2.$$

- These second derivatives 4, 4, 2 contain the answer.
- Since they are the same for F and f , they must contain the same answer.
- The two functions behave in exactly the same way near the origin.
- ***F has a minimum if and only if f has a minimum***.

Minima, Maxima, and Saddle Points

***Remark* 6.** The *higher-degree terms* in F have no effect on the question of a *local* minimum, but they can prevent it from being a *global* minimum. In our example the term $-x^3$ must sooner or later pull F toward $-\infty$. For $f(x,y)$, with no higher terms, all the action is at $(0,0)$.

Every quadratic form $f = ax^2 + 2bxy + cy^2$ has a stationary point at the origin, where $\partial f/\partial x = \partial f/\partial y = 0$.

A **local minimum** would also be a **global minimum**, The surface $z = f(x,y)$ will then be shaped like a bowl, resting on the origin.

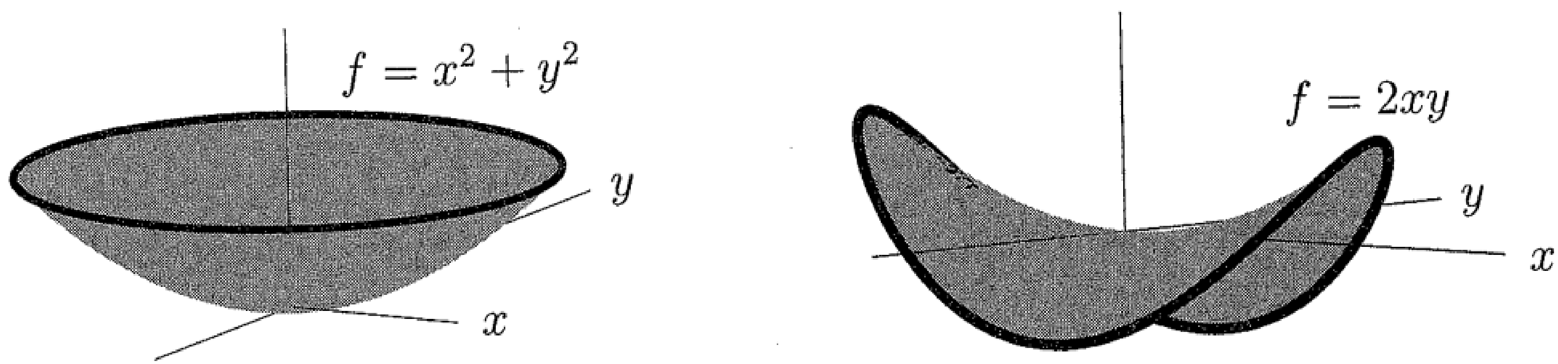

Figure 6.1: A bowl and a saddle: Definite $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and indefinite $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

Minima, Maxima, and Saddle Points

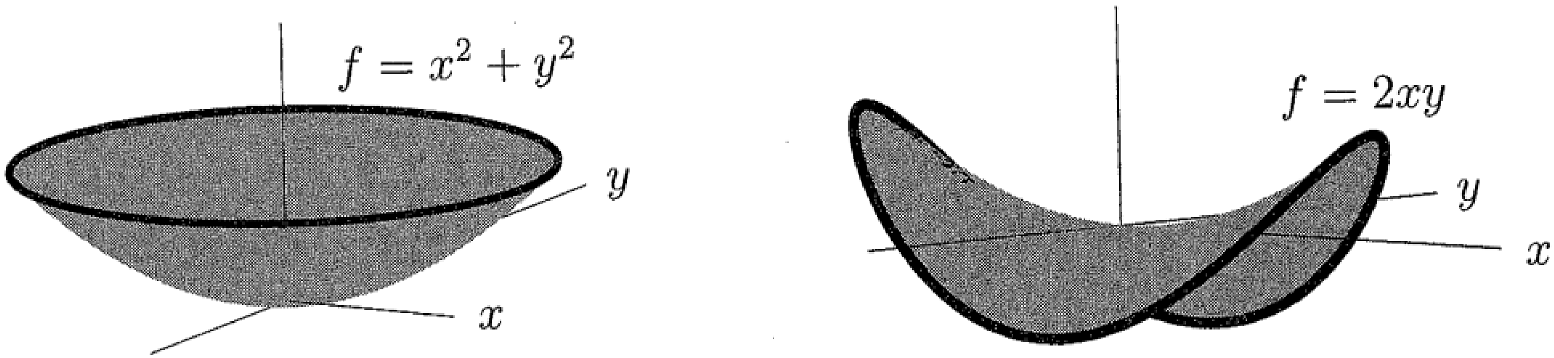

Figure 6.1: A bowl and a saddle: Definite $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and indefinite $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

If the stationary point of F is at $x = \alpha, y = \beta$, the only change would be to use the second derivatives at α, β:

Quadratic part of F

$$f(x,y) = \frac{x^2}{2}\frac{\partial^2 F}{\partial x^2}(\alpha,\beta) + xy\frac{\partial^2 F}{\partial x \partial y}(\alpha,\beta) + \frac{y^2}{2}\frac{\partial^2 F}{\partial y^2}(\alpha,\beta).$$

Minima, Maxima, and Saddle Points

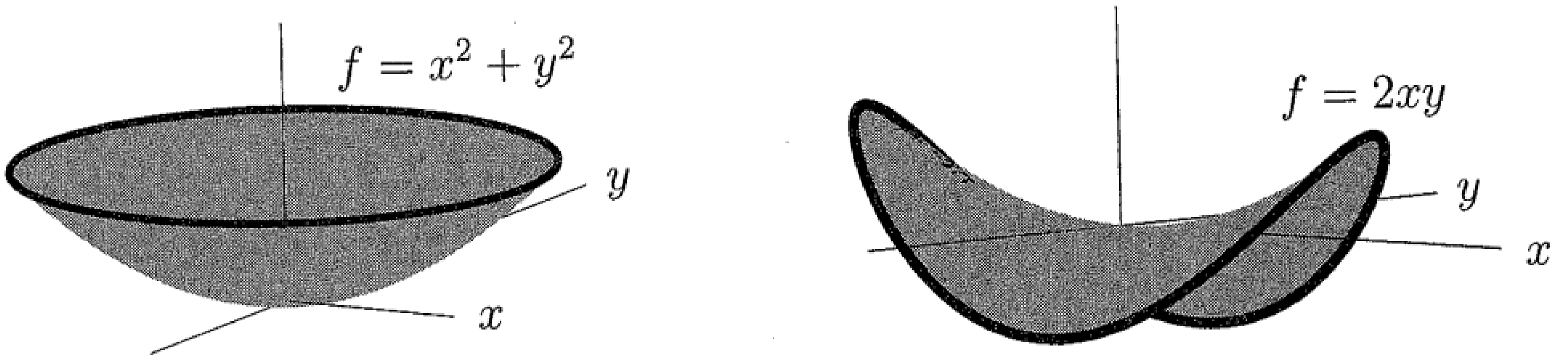

Figure 6.1: A bowl and a saddle: Definite $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and indefinite $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

- The third derivatives are drawn into the problem when the second derivatives fail to give a definite decision. That happens when the **quadratic part** is singular.
- For a true minimum, f is allowed to vanish only at $x = y = 0$.
- When $f(x, y)$ is strictly positive at all other points (the bowl goes up), it is called ***positive definite***.

Definite versus Indefinite: Bowl versus Saddle

- For a function of two variables x and y, what is the correct replacement for the condition $\frac{\partial^2 F}{\partial x^2} > 0$?

With only one variable, the sign of the second derivative decides between a minimum or a maximum.

Now we have three second derivatives: $F_{xx}, F_{xy} = F_{yx}$, and F_{yy}.

***What conditions on** a, b, **and** c **ensure that the quadratic** $f(x,y) = ax^2 + 2bxy + cy^2$ **is positive definite**?* One necessary condition is easy:

(i) *If $ax^2 + 2bxy + cy^2$ is positive definite, then necessarily $a > 0$.*

We look at $x = 1$, $y = 0$, where $ax^2 + 2bxy + cy^2$ is equal to a. This must be positive. Translating back to F, that means that $\partial^2 F/\partial x^2 > 0$. The graph must go up in the x direction. Similarly, fix $x = 0$ and look in the y direction where $f(0,y) = cy^2$:

Definite versus Indefinite: Bowl versus Saddle

(i) *If* $ax^2 + 2bxy + cy^2$ *is positive definite, then necessarily* $a > 0$.

We look at $x = 1$, $y = 0$, where $ax^2 + 2bxy + cy^2$ is equal to a. This must be positive. Translating back to F, that means that $\partial^2 F/\partial x^2 > 0$. The graph must go up in the x direction. Similarly, fix $x = 0$ and look in the y direction where $f(0, y) = cy^2$:

(ii) *If* $f(x, y)$ *is positive definite, then necessarily* $c > 0$.

Do these conditions $a > 0$ and $c > 0$ guarantee that $f(x, y)$ is always positive? The answer is **no**. A large cross term $2bxy$ can pull the graph below zero.

Definite versus Indefinite: Bowl versus Saddle

Example 1. $f(x,y) = x^2 - 10xy + y^2$. Here $a = 1$ and $c = 1$ are both positive. But f is not positive definite, because $f(1,1) = -8$. The conditions $a > 0$ and $c > 0$ ensure that $f(x,y)$ is positive on the x and y axes. But this function is negative on the line $x = y$, because $b = -10$ overwhelms a and c.

6A $ax^2 + 2bxy + cy^2$ is positive definite if and only if $a > 0$ and $ac > b^2$. Any $f(x,y)$ has a minimum at a point where $\partial F/\partial x = \partial F/\partial y = 0$ with

$$\frac{\partial F^2}{\partial x^2} > 0 \qquad \text{and} \qquad \left[\frac{\partial F^2}{\partial x^2}\right]\left[\frac{\partial F^2}{\partial y^2}\right] > \left[\frac{\partial F^2}{\partial x \partial y}\right]^2. \tag{3}$$

Definite versus Indefinite: Bowl versus Saddle

Test for a maximum: Since f has a maximum whenever $-f$ has a minimum, we just reverse the signs of a, b, and c. This actually leaves $ac > b^2$ unchanged: The quadratic form is ***negative definite*** if and only if $a < 0$ and $ac > b^2$. The same change applies for a maximum of $F(x,y)$.

Singular case $ac = b^2$**:** The second term in equation (2) disappears to leave only the first square—which is either ***positive semidefinite***, when $a > 0$, or ***negative semidefinite***, when $a < 0$. The prefix *semi* allows the possibility that f can equal zero, as it will at the point $x = b$, $y = -a$. The surface $z = f(x,y)$ degenerates from a bowl into a valley. For $f = (x+y)^2$, the valley runs along the line $x + y = 0$.

A stationary point that is neither a maximum nor a minimum is called a saddle point.

Higher Dimensions: Linear Algebra

A quadratic $f(x, y)$ comes directly from a symmetric 2 by 2 matrix!

$$x^{\mathrm{T}}Ax \textbf{ in } \mathbf{R}^2 \qquad ax^2 + 2bxy + cy^2 = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

For any symmetric matrix A, the product $x^T Ax$ is a pure quadratic form $f(x_1, \ldots, x_n)$:

$$x^{\mathrm{T}}Ax \textbf{ in } \mathbf{R}^n \qquad \begin{bmatrix} x_1 & x_2 & \cdot & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdot & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{bmatrix} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j.$$

Higher Dimensions: Linear Algebra

Example 3. $f = 2x^2 + 4xy + y^2$ and $A = \begin{bmatrix} 2 & 2 \\ 2 & 1 \end{bmatrix} \rightarrow$ *saddle point.*

Example 4. $f = 2xy$ and $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \rightarrow$ *saddle point.*

Example 5. A is 3 by 3 for $2x_1^2 - 2x_1x_2 + 2x_2^2 - 2x_2x_3 + 2x_3^2$:

$$f = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \rightarrow \textit{minimum at } (0,0,0).$$

A is the "**second derivative matrix**" with entries $a_{ij} = \dfrac{\partial^2 F}{\partial x_i \partial x_j}$.

F has a minimum when the pure quadratic $x^T A x$ is ***positive definite***.

Tests for Positive Definiteness

6B Each of the following tests is a necessary and sufficient condition for the real symmetric matrix A to be ***positive definite***:

(I) $x^{\mathrm{T}}Ax > 0$ for all nonzero real vectors x.

(II) All the eigenvalues of A satisfy $\lambda_i > 0$.

(III) All the upper left submatrices A_k have positive determinants.

(IV) All the pivots (without row exchanges) satisfy $d_k > 0$.

6C The symmetric matrix A is positive definite if and only if

(V) There is a matrix R with independent columns such that $A = R^{\mathrm{T}}R$.

Tests for Positive Definiteness

Semi-definite matrices:

The tests for semi-definiteness will relax $x^T Ax > 0, \lambda > 0, d > 0$ and $\det > 0$, to allow zeros to appear.

6D Each of the following tests is a necessary and sufficient condition for a symmetric matrix A to be ***positive semidefinite***:

(I′) $x^{\mathrm{T}}Ax \geq 0$ for all vectors x (this defines positive semidefinite).

(II′) All the eigenvalues of A satisfy $\lambda_i \geq 0$.

(III′) No principal submatrices have negative determinants.

(IV′) No pivots are negative.

(V′) There is a matrix R, possibly with dependent columns, such that $A = R^{\mathrm{T}}R$.

Tests for Positive Definiteness: Example

$$A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$ is positive *semi*definite, by all five tests:

(I′) $x^{\mathrm{T}}Ax = (x_1 - x_2)^2 + (x_1 - x_3)^2 + (x_2 - x_3)^2 \geq 0$ (zero if $x_1 = x_2 = x_3$).

(II′) The eigenvalues are $\lambda_1 = 0$, $\lambda_2 = \lambda_3 = 3$ (a zero eigenvalue).

(III′) $\det A = 0$ and smaller determinants are positive.

(IV′) $$A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \to \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{3}{2} & -\frac{3}{2} \\ 0 & -\frac{3}{2} & \frac{3}{2} \end{bmatrix} \to \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{3}{2} & 0 \\ 0 & 0 & \mathbf{0} \end{bmatrix}$$ (missing pivot).

(V′) $A = R^{\mathrm{T}}R$ with dependent columns in R:

$$\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$ $(1,1,1)$ in the nullspace.

An Ellipsoid – For a positive definite matrix A and its $x^T Ax$ the curve obtained is an ellipse in 2 dimensions and ellipsoid in n dimensions.

A = $\begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$ and $x^T Ax = 5u^2 + 8uv + 5v^2 = 1$

The ellipse is centered at $u = v = 0$, but the axes no longer line up with the coordinate axes.

It can be shown that the ***axes of the ellipse point toward the eigenvector of A***.

As $A = A^T$, those eigenvectors and axes are orthogonal.

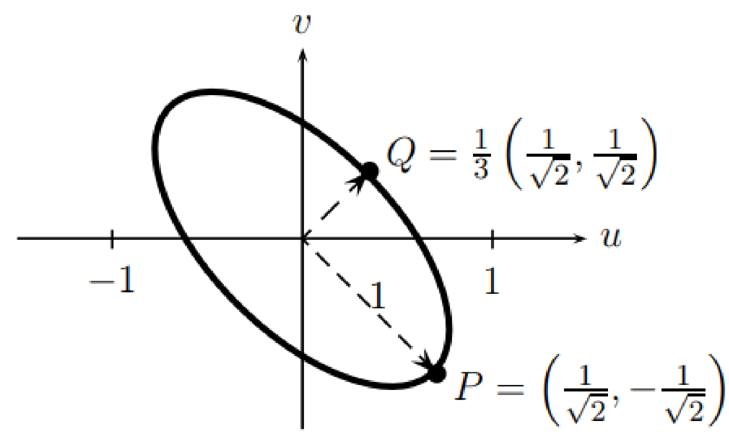

Fig: The ellipse $x^T Ax = 5u^2 + 8uv + 5v^2 = 1$ and its principal axes.

The major axis of the ellipse corresponds to the smallest eigenvalue of A.

Singular Value Decomposition

- $A = U\Sigma V^T$ is known as the "**SVD**" or the ***singular value decomposition***.
- The SVD is closely associated with the eigenvalue-eigenvector factorization $Q\Lambda Q^T$ of a positive definite matrix.
- Any $m \times n$ matrix A can be factored into

$$A = U\Sigma V^{\mathrm{T}} = (\textbf{orthogonal})(\textbf{diagonal})(\textbf{orthogonal})$$

- The columns of U $(m \times m)$ are ***eigenvectors of*** $\boldsymbol{AA^T}$, and the columns of V $(n \times n)$ are ***eigenvectors of*** $\boldsymbol{A^T A}$.
- The r singular values on the diagonal of Σ $(m \times n)$ are the ***square roots of the nonzero eigenvalues*** of both AA^T and $A^T A$.
- While eigen-value decomposition can be applied only to square matrices, SVD can be applied to any matrix (including rectangular matrix).

Singular Value Decomposition

Remark 1.

- For positive definite matrices, Σ is Λ and $U\Sigma V^T$ is identical to $Q\Lambda Q^T$.
- For other symmetric matrices, any negative eigenvalues in Λ become positive in Σ.
- For complex matrices, Σ remains real but U and V become *unitary* (the complex version of orthogonal). $\boldsymbol{A = U\Sigma V^H}$

Remark 2.

U and V give orthonormal bases for all four fundamental subspaces:

first	r	columns of U:	**column space** of A
last	$m-r$	columns of U:	**left nullspace** of A
first	r	columns of V:	**row space** of A
last	$n-r$	columns of V:	**nullspace** of A

Singular Value Decomposition

***Remark* 3.** The SVD chooses those bases in an extremely special way. They are more than just orthonormal. *When A multiplies a column v_j of V, it produces σ_j times a column of U*. That comes directly from $AV = U\Sigma$, looked at a column at a time.

Remark 4.

Eigenvectors of AA^T and A^TA must go into the columns of U and V:

$$AA^{\mathrm{T}} = (U\Sigma V^{\mathrm{T}})(V\Sigma^{\mathrm{T}}U^{\mathrm{T}}) = U\Sigma\Sigma^{\mathrm{T}}U^{\mathrm{T}} \quad \text{and, similarly,} \quad A^{\mathrm{T}}A = V\Sigma^{\mathrm{T}}\Sigma V^{\mathrm{T}}.$$

- U must be the eigenvector matrix for AA^T .
- The eigenvalue matrix in the middle is $\Sigma\Sigma^T$— which is $m \times\ m$ with σ_1^2 , … , σ_r^2 on the diagonal.
- From the $A^TA\ =\ V\Sigma^T\Sigma V^T$, the V matrix must be the eigenvector matrix for A^TA.

Singular Value Decomposition

Example 1.

This A has only one column: rank r = 1. Then Σ has only $\sigma_1 = 3$:

$$\textbf{SVD} \qquad A = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} = U_{3\times3}\Sigma_{3\times1}V^{\mathrm{T}}_{1\times1}$$

$A^{\mathrm{T}}A$ is 1 by 1, whereas AA^{T} is 3 by 3. They both have eigenvalue 9 (whose square root is the 3 in Σ). The two zero eigenvalues of AA^{T} leave some freedom for the eigenvectors in columns 2 and 3 of U. We kept that matrix orthogonal.

Singular Value Decomposition

Example 2.

Now A has rank 2, and AA^T = $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$ with $\lambda = 3$ and 1:

$$\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = U\Sigma V^{\mathrm{T}} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{matrix} /\sqrt{6} \\ /\sqrt{2} \\ /\sqrt{3} \end{matrix}$$

Notice $\sqrt{3}$ and $\sqrt{1}$. The columns of U are *left singular vectors* (unit eigenvectors of AA^T).

The columns of V are *right singular vectors* (unit eigenvectors of $A^T A$).

Applications of Singular Value Decomposition

Image Processing.

- Suppose a satellite takes a picture, and wants to send it to Earth.
- The picture may contain 1000×1000 "pixels"—a million little squares, each with a definite color.
- We can code the colors, and send back $1{,}000{,}000$ numbers.
- It is ***better to find the essential information inside the*** $\mathbf{1000} \times \mathbf{1000}$ ***matrix***, and send only that.

In SVD some σ's are significant and others are extremely small.

If we keep 20 and throw away 980, then we send only the corresponding 20 columns of U and V.

The other 980 columns are multiplied in $U\Sigma V^T$ by the small σ's that are being ignored. ***If only 20 terms are kept, we send 20 times 2000 numbers instead of a million (25 to 1 compression).***

Applications of Singular Value Decomposition

Polar decomposition.

- Every nonzero complex number z is a positive number r times a number $e^{i\theta}$ on the unit circle: $z = re^{i\theta}$.
- That expresses z in "polar coordinates."
- If we think of z as a 1×1 matrix, r corresponds to a positive definite matrix and $e^{i\theta}$ corresponds to an orthogonal matrix.
- More exactly, since $e^{i\theta}$ is complex and satisfies $e^{-i\theta}e^{i\theta} = 1$, it forms a 1×1 unitary matrix: $U^H U = I$.
- The SVD extends this "polar factorization" to matrices of any size:

> Every real square matrix can be factored into $A = QS$, where Q is ***orthogonal*** and S is ***symmetric positive semidefinite***. If A is invertible then S is positive definite.

- **Pseudo-Inverse**

Pseudo inverse is a generalization of the matrix inverse when the matrix may not be invertible.

If $A = U\Sigma V^{\mathrm{T}}$ (the SVD), then its pseudoinverse is $A^{+} = V\Sigma^{+}U^{\mathrm{T}}$.

Minimum Principles

Our goal is to find the minimum principle equivalent to $Ax = b$, and the minimization equivalent to $Ax = \lambda x$.

We want to find the "parabola" $P(x)$ whose minimum occurs when $Ax = b$.

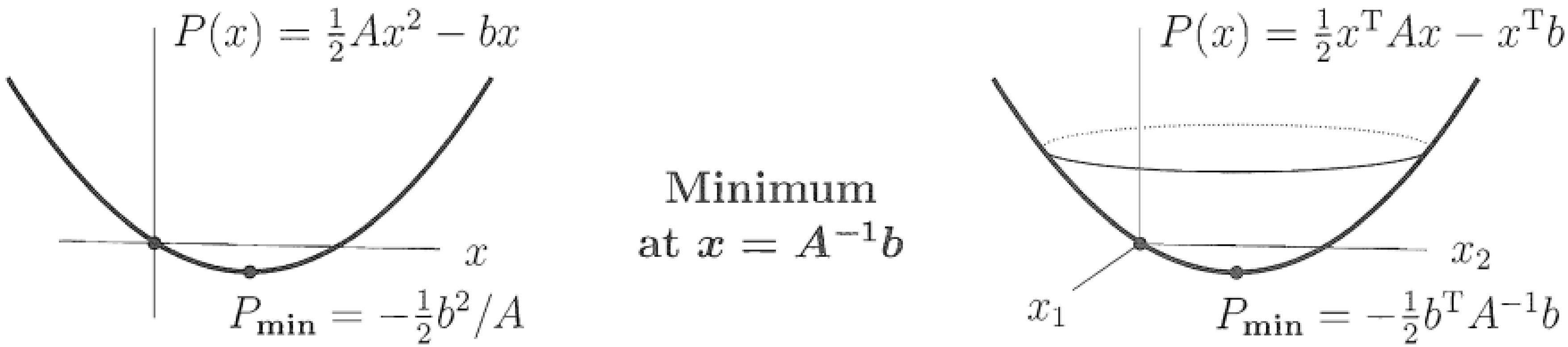

Figure 6.4: The graph of a positive quadratic $P(x)$ is a parabolic bowl.

Minimum Principles

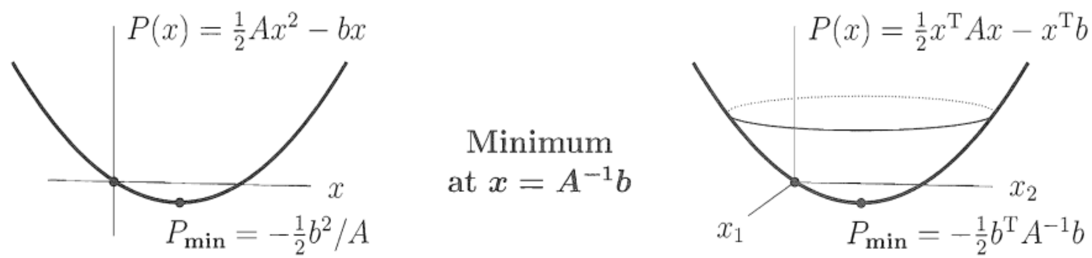

Figure 6.4: The graph of a positive quadratic $P(x)$ is a parabolic bowl.

6H If A is symmetric positive definite, then $P(x) = \frac{1}{2}x^{\mathrm{T}}Ax - x^{\mathrm{T}}b$ reaches its minimum at the point where $Ax = b$. At that point $P_{\min} = -\frac{1}{2}b^{\mathrm{T}}A^{-1}b$.

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

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

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

Minimum Principles

Example. Minimize $P(x) = x_1^2 - x_1x_2 + x_2^2 - b_1x_1 - b_2x_2$.

The usual approach, by calculus, is to set the partial derivatives to zero. This gives $Ax = b$:

$$\begin{aligned} \partial P/\partial x_1 &= 2x_1 - x_2 - b_1 = 0 \\ \partial P/\partial x_2 &= -x_1 + 2x_2 - b_2 = 0 \end{aligned} \qquad \text{means} \qquad \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Linear algebra recognizes this $P(x)$ as $\frac{1}{2} x^T Ax - x^T b$, and knows immediately that $Ax = b$ gives the minimum.

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

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

Minimizing with Constraints

- Many applications add extra equations $Cx = d$ on top of the minimization problem.
- These equations are ***constraints***. We minimize $P(x)$ subject to the extra requirement $Cx = d$.
- Usually x can't satisfy n equations $Ax = b$ and also l extra constraints $Cx = d$. We have too many equations and we need l more unknowns.

Minimizing with Constraints

Those new unknowns $y_1, \ldots, y_l$ are called ***Lagrange multipliers***.

$$L(x,y) = P(x) + y^{\mathrm{T}}(Cx - d) = \frac{1}{2}x^{\mathrm{T}}Ax - x^{\mathrm{T}}b + x^{\mathrm{T}}C^{\mathrm{T}}y - y^{\mathrm{T}}d.$$

L is chosen exactly so that $\partial L/\partial y = 0$ brings back $Cx = d$. When we set the derivatives of L to zero, we have $n + l$ equations for $n + l$ unknowns x and y:

Constrained	$\partial L/\partial x = 0:$	$Ax + C^{\mathrm{T}}y = b$
minimization	$\partial L/\partial y = 0:$	$Cx = d$

Minimizing with Constraints

***Example*.** Suppose $P(x1, x2) = \frac{1}{2}\, x_1^2 + \frac{1}{2}\, x_2^2$. Its smallest value is certainly $P_{min} \;=\; 0$.

- This unconstrained problem has $n \;=\; 2, A \;=\; I$, and $b \;=\; 0$.
- So the minimizing equation $Ax \;=\; b$ just gives $x_1 \;=\; 0$ and $x_2 \;=\; 0$.

- Now add one constraint $c_1x_1 \;+\; c_2x_2 \;=\; d$.
- This puts x on a line in the $x_1 - x_2$ plane. The old minimizer $x_1 \;=\; x_2 \;=\; 0$ is not on the line.

- The Lagrangian $L(x, y) = \frac{1}{2}\, x_1^2 \;+ \frac{1}{2}\, x_2^2 \;+ y(c_1x_1 \;+ c_2x_2 \;- d)$ has $n + l \;=\; 2 + 1$ partial derivatives

Contd.

Minimizing with Constraints

The Lagrangian $L(x, y) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 + y(c_1x_1 + c_2x_2 - d)$ has $n + l = 2 + 1$ partial derivatives:

$$\begin{aligned} \partial L/\partial x_1 &= 0 & \qquad x_1 + c_1 y &= 0 \\ \partial L/\partial x_2 &= 0 & \qquad x_2 + c_2 y &= 0 \\ \partial L/\partial y &= 0 & \qquad c_1x_1 + c_2x_2 &= d. \end{aligned}$$

Substituting $x_1 = -c_1 y$ and $x_2 = -c_2 y$ into the 3rd equation gives $-c_1^2 y - c_2^2 y = d$.

Solution $\quad y = \dfrac{-d}{c_1^2 + c_2^2} \qquad x_1 = \dfrac{c_1 d}{c_1^2 + c_2^2} \qquad x_2 = \dfrac{c_2 d}{c_1^2 + c_2^2}.$

The constrained minimum of $P = \frac{1}{2}x^{\mathrm{T}}x$ is reached at that solution point:

$$P_{C/\min} = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 = \frac{1}{2}\frac{c_1^2 d^2 + c_2^2 d^2}{(c_1^2 + c_2^2)^2} = \frac{1}{2}\frac{d^2}{c_1^2 + c_2^2}.$$

This equals $-\frac{1}{2}yd$ as predicted in equation (5), since $b = 0$ and $P_{\min} = 0$.

The Rayleigh quotient

- Goal is to find a minimization problem equivalent to $Ax = \lambda x$.
- The function to minimize cannot be a quadratic, or its derivative would be linear, and the eigenvalue problem is nonlinear (λ times x).
- The trick that succeeds is to divide one quadratic by another one:

$$\textbf{Rayleigh quotient} \qquad \text{Minimize} \quad R(x) = \frac{x^{\mathrm{T}} A x}{x^{\mathrm{T}} x}.$$

6I Rayleigh's Principle: The minimum value of the Rayleigh quotient is the smallest eigenvalue λ_1. $R(x)$ reaches that minimum at the first eigenvector x_1 of A:

$$\textbf{Minimum where } Ax_1 = \lambda x_1 \qquad R(x_1) = \frac{x_1^{\mathrm{T}} A x_1}{x_1^{\mathrm{T}} x_1} = \frac{x_1^{\mathrm{T}} \lambda_1 x_1}{x_1^{\mathrm{T}} x_1} = \lambda_1.$$

The Rayleigh quotient

6I Rayleigh's Principle: The minimum value of the Rayleigh quotient is the smallest eigenvalue λ_1. $R(x)$ reaches that minimum at the first eigenvector x_1 of A:

$$\textbf{Minimum where } Ax_1 = \lambda x_1 \qquad R(x_1) = \frac{x_1^{\mathrm{T}} A x_1}{x_1^{\mathrm{T}} x_1} = \frac{x_1^{\mathrm{T}} \lambda_1 x_1}{x_1^{\mathrm{T}} x_1} = \lambda_1 .$$

- If we keep $x^T A x = 1$, then $R(x)$ is a minimum when $x^T x = ||x||^2$ is as large as possible.
- We are looking for the point on the ellipsoid $x^T A x = 1$ farthest from the origin—the vector x of greatest length. Its longest axis points along the first eigenvector. So $R(x)$ is a minimum at x_1.

Algebraically, we can diagonalize the symmetric A by an orthogonal matrix: $Q^{\mathrm{T}} A Q = \Lambda$. Then set $x = Qy$ and the quotient becomes simple:

$$R(x) = \frac{(Qy)^{\mathrm{T}} A (Qy)}{(Qy)^{\mathrm{T}} (Qy)} = \frac{y^{\mathrm{T}} \Lambda y}{y^{\mathrm{T}} y} = \frac{\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2}{y_1^2 + \cdots + y_n^2} . \tag{11}$$

The Rayleigh quotient

Algebraically, we can diagonalize the symmetric A by an orthogonal matrix: $Q^{\mathrm{T}}AQ = \Lambda$. Then set $x = Qy$ and the quotient becomes simple:

$$R(x) = \frac{(Qy)^{\mathrm{T}}A(Qy)}{(Qy)^{\mathrm{T}}(Qy)} = \frac{y^{\mathrm{T}}\Lambda y}{y^{\mathrm{T}}y} = \frac{\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2}{y_1^2 + \cdots + y_n^2}. \tag{11}$$

The minimum of R is λ_1, at the point where $y_1 = 1$ and $y_2 = \cdots = y_n = 0$:

At all points $\quad \lambda_1(y_1^2 + y_2^2 + \cdots + y_n^2) \leq (\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2).$

The Rayleigh quotient in equation (11) is ***never below*** λ_1 and ***never above*** λ_n (the largest eigenvalue). Its minimum is at the eigenvector x_1 and its maximum is at x_n:

Maximum where $Ax_n = \lambda_n x_n \qquad R(x_n) = \dfrac{x_n^{\mathrm{T}}Ax_n}{x_n^{\mathrm{T}}x_n} = \dfrac{x_n^{\mathrm{T}}\lambda_n x_n}{x_n^{\mathrm{T}}x_n} = \lambda_n.$