Positive Definite Matrices

Notes on Linear Algebra

Chia-Ping Chen

Department of Computer Science and Engineering
National Sun Yat-Sen University
Kaohsiung, Taiwan ROC

Introduction

- Recall that the eigenvalues of Hermitian matrices are real.
- What about the signs?
- The signs of eigenvalues are important in some cases. For example, in differential equation, we need eigenvalues to be negative so the system is stable.
- Matrices whose eigenvalues are all positive are called positive definite. They are related to positive definite functions.

Quadratic Form

- A function f of two variables x, y is said to have a quadratic form if

$$f(x,y) = ax^2 + 2bxy + cy^2.$$

- f is related to a real symmetric matrix,

$$f(x,y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

- The partial derivatives are 0 at the origin. It is called a stationary point of f. In addition, $f(0,0) = 0$.

Positive Definite

- A quadratic form is said to be *positive definite* if f is positive for all points except for the origin.
- For $f(x, y) = ax^2 + 2bxy + cy^2$ to be p.d., $a > 0$ and $c > 0$. This can be shown by looking at points $(1, 0)$ and $(0, 1)$.
- But these are merely necessary conditions. For example, $f(x, y) = x^2 - 4xy + y^2 < 0$ at $(1, 1)$.

Sufficient Condition

- We can express f using squares by

$$f(x,y) = ax^2 + 2bxy + cy^2 = a(x + \frac{b}{a}y)^2 + (c - \frac{b^2}{a})y^2,$$

- From above we see that f is positive definite if

$$a > 0, \quad ac > b^2.$$

Negative Definite

- $f(x,y) = ax^2 + 2bxy + cy^2$ is negative definite if $f(x,y) < 0$ except for $(x,y) = (0,0)$.
- f is negative definite iff $-f$ is positive definite. So the sufficient condition for negative definiteness is

$$(-a) > 0, \;\; (-a)(-c) > (-b)^2.$$

- Equivalently, getting rid of negative signs,

$$a < 0, \;\; ac > b^2.$$

Singular Case

- We have a singular case if

$$ac = b^2.$$

- If $a > 0$, f is still non-negative everywhere, since

$$f(x, y) = a(x + \frac{b}{a}y)^2.$$

  The surface $z = f(x, y)$ degenerates from a bowl to a valley, along the line $ax + by = 0$.

- f is said to be positive semidefinite (psd) if $a > 0$ and negative semidefinite if $a < 0$.

Saddle Point

- The remaining case is when

$$ac < b^2.$$

- $(0, 0)$ is a saddle point. We can find two directions. Along one direction $(0, 0)$ is a minimum, and along the other direction $(0, 0)$ is a maximum.
- For example, $f(x, y) = x^2 - y^2$. $(0, 0)$ is a minimum along the x-axis, a maximum along the y-axis.
- f is said to be indefinite. It has a stationary point that is neither a minimum nor a maximum.

Quadratic Form and Matrix

- A quadratic form can be represented by a real symmetric matrix A, where $a_{ij} + a_{ji}$ equals the coefficient of term $x_i x_j$.

$$x^T A x = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i,j=1}^{n} a_{ij} x_i x_j$$

- 0 is a stationary point of $x^T A x$. The signs of eigenvalues of A determine whether 0 is a maximum, minimum or saddle point.

Hessian Matrix

- The Taylor series near $x = 0$ is

$$F(x) = F(0) + x^T \nabla F + \frac{1}{2} x^T A x + \text{higher-order terms.}$$

- A, the second-order derivative matrix, is called the Hessian matrix.

$$a_{ij} = \frac{\partial^2 F}{\partial x_i \partial x_j}.$$

- Whether a stationary point (with $\nabla F = 0$) is a minimum, maximum or saddle point is determined by A.

Positive Definite Matrices

- A real symmetric matrix A is said to be positive definite if $x^T Ax > 0$ except for $x = 0$.
- In the two-dimensional case,

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

  is positive definite iff $a > 0$ and $ac > b^2$.
- We want to generalize to n-dimensional case.

Conditions for Positive Definiteness

- Note that a and $ac - b^2$ are the determinants of the principle submatrices.
- We can generalize the above condition

  "*all principle submatrices have positive determinants.*"

- The following conditions are also sufficient and necessary for positive definiteness:
  - For all eigenvalues, $\lambda_i > 0$.
  - For all pivots, $d_i > 0$.

Positive Eigenvalues

- Suppose A is positive definite, then the eigenvalues $\lambda_i's$ are positive. Let x_i be an eigenvector of A with eigenvalue λ_i.

$$x_i^T A x_i = \lambda_i (x_i^T x_i) > 0 \ \Rightarrow \ \lambda_i > 0.$$

- Conversely, if $\lambda_i > 0$ for all λ_i, then A is positive definite. Since A is Hermitian, it has a complete set of orthonormal eigenvectors, and

$$x^T A x = x^T Q \Lambda Q^T x = \sum_i c_i^2 \lambda_i > 0.$$

Positive Determinants

- Suppose A is positive definite, then the determinants of all principle submatrices are positive.
- We first show any principle submatrix A_k is positive definite. Let x be a non-zero vector with the last $n-k$ components being 0,

$$x^T Ax = \begin{bmatrix} x_k^T & 0 \end{bmatrix} \begin{bmatrix} A_k & * \\ * & * \end{bmatrix} \begin{bmatrix} x_k \\ 0 \end{bmatrix} = x_k^T A_k x_k > 0.$$

- since all eigenvalues of A_k are positive,

$$|A_k| = \prod_i \lambda_i > 0.$$

Positive Pivots

- Suppose A is positive definite, then all pivots d_i are positive.

  $$d_k = \frac{|A_k|}{|A_{k-1}|} > 0,$$

  since $|A_k| > 0$ for all k.

- Conversely, if $d_i > 0$ for all d_i, then A is positive definite.

  $$A = LDL^T \Rightarrow x^T A x = \sum_i d_i (L^T x)_i^2 > 0.$$

Relation to Least Squares

- In a least squares problem $Rx = b$ we solve the normal equation

$$R^T R\bar{x} = R^T b.$$

- Note that the matrix $A = R^T R$ is symmetric.
- A is positive definite if R has linearly independent columns.

$$x^T Ax = x^T R^T Rx = (Rx)^T (Rx) \begin{cases} = 0, & x = 0, \\ > 0, & x \neq 0 \end{cases}$$

Cholesky Decomposition

- A is p.d. iff there exists a matrix R with independent columns such that $A = R^T R$.

  (if) $x^T A x = x^T R^T R x = |Rx|^2 > 0$ if $x \neq 0$.

  (only if) $A = LDL^T = LD^{1/2} D^{1/2} L^T = R^T R$.

  This Cholesky decomposition splits the pivots evenly between L and L^T.
- There are infinite ways to decompose a positive definite $A = R^T R$. In fact, $R' = RQ$, where Q is orthogonal, also satisfies $A = R'^T R'$.

Ellipsoids in n Dimensions

- Consider the equation $x^T Ax = 1$, where A is p.d.
- If A is diagonal, the graph is easily seen as an ellipsoid.
- If A is not diagonal, the graph is still an ellipsoid, since

$$x^T Ax = x^T Q\Lambda Q^T x = y^T \Lambda y = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2.$$

- $y_i = q_i^T x$ is the component of x along the ith eigenvector q_i.

Principle Axes

- The axes of the ellipsoid defined by $x^T Ax = 1$ point toward the eigenvectors (q_i) of A.
- They are called *principle axes*.
- The principle axes of an ellipsoid are mutually orthogonal.
- The length of the axis along q_i is $1/\sqrt{\lambda_i}$.

Semidefinite Matrices

- A matrix is said to be positive semidefinite if

$$x^T A x \geq 0 \text{ for all } x.$$

- Each of the following conditions is sufficient and necessary for positive semidefiniteness.
  - All eigenvalues are non-negative.
  - $|A_k| \geq 0$ for all principle submatrices A_k.
  - All pivots are non-negative.
  - $A = R^T R$ for some R.

An Example

- The following matrix is positive semidefinite.

$$A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

- $x^T Ax = (x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - x_1)^2 \geq 0$.
- The eigenvalues are $0, 3$, and 3.
- The pivots are $2, \frac{3}{2}$.
- $|A_1| = 2, |A_2| = 3, |A_3| = 0$.

Congruence Transform

- Suppose we have a change of variable $x = Cy$ with C nonsingular. Then $x^T A x = y^T C^T A C y$.
- The *congruence transform* is defined by

$$C^T A C$$

- C is required to be non-singular.
- $C^T A C$ is symmetric, just like A.
- For comparison, similarity transform is defined by

$$C^{-1} A C.$$

Sylvester's Law

- The signs of eigenvalues are invariant under congruence transform $A \to C^T AC$.
- Suppose A is nonsingular for simplicity. Let

$$C = QR, C(t) = tQ + (1-t)QR.$$

- The eigenvalues of $C(t)^T AC(t)$ change gradually as we vary t from 0 to 1, but they are never 0 since $C(t) = Q(tI + (1-t)R)$ is invertible. So the signs are preserved.
- Since $Q^T AQ$ and A have the same eigenvalues, the law is proved.

Signs of Pivots

- For a symmetric matrix A, the LDU-decomposition is $A = LDU = U^T DU$.
- A is a congruence transformation of D.
- According to Sylvester's law, the signs of eigenvalues of A and D (the pivots) are the same.
- Therefore the signs of the pivots agree with the signs of the eigenvalues.

Locating Eigenvalues

- The relation between pivots and eigenvalues can be used to locate eigenvalues.
- First, note that if A has an eigenvalue λ, then $A - cI$ has the eigenvalue $\lambda - c$ with the same eigenvector,

$$Ax = \lambda x \;\Rightarrow\; (A - cI)x = (\lambda - c)x.$$

Example

- Consider

$$A = \begin{bmatrix} 3 & 3 & 0 \\ 3 & 10 & 7 \\ 0 & 7 & 8 \end{bmatrix}, \; B = A - 2I = \begin{bmatrix} 1 & 3 & 0 \\ 3 & 8 & 7 \\ 0 & 7 & 6 \end{bmatrix}.$$

- B has a negative pivot, so it has a negative eigenvalue. A is positive definite.
- It follows

$$\lambda_A > 0, \lambda_B = \lambda_A - 2 < 0 \;\Rightarrow 0 < \lambda_A < 2.$$

Generalized Eigenvalue Problem

- A generalized eigenvalue problem is

$$Ax = \lambda Mx.$$

- The right-hand side has a matrix M.
- A generalized eigenvalue problem can be converted to an equivalent eigenvalue problem.

Equivalent Eigenvalue Problem

- We consider only the case that A is symmetric and M is positive definite. We can write $M = R^T R$, where R is invertible.
- Let $y = Rx$,

$$Ax = \lambda Mx = \lambda R^T Rx \Rightarrow AR^{-1}y = \lambda R^T y.$$

- Let $C = R^{-1}$ so $(R^T)^{-1} = C^T$. Then

$$C^T ACy = \lambda y.$$

This is an equivalent eigenvalue problem, with the same eigenvalues and related eigenvectors $x = Cy$.

Properties

- The eigenvalues are real since C^TAC is symmetric.
- They have the same signs as the eigenvalues of A since C^TAC is a congruence transformation of A.
- The eigenvectors y_j can be chosen orthonormal, so the eigenvectors x_j are M-orthonormal, i.e.

$$x_i^T M x_j = x_i^T R^T R x_j = y_i^T y_j = \delta_{ij}.$$

Simultaneous Diagonalization

- M and A can both be diagonalized by the eigenvectors x_i.

$$x_i^T M x_j = y_i^T y_i = \delta_{ij},$$
$$x_i^T A x_j = \lambda_j x_i^T M x_j = \lambda_j \delta_{ij}.$$

- Using x_i's as the columns of S, we have $S^T A S = \Lambda$ and $S^T M S = I$.
- Note they are congruence transforms to diagonal matrices rather than similarity transforms, as S^T is used, not S^{-1}.

Singular Value Decomposition

- The singular value decomposition, SVD, is

$$A = U\Sigma V^T.$$

- U, Σ, V are related to the matrices $A^T A$ and AA^T.
- Unlike earlier discussion, here A is not limited to be a square matrix. A is rectangular.

SVD Theorem

- Any $m \times n$ real matrix A with rank r can be factored by

$$A = U\Sigma V^T = \text{(orthogonal)(diagonal)(orthogonal)}.$$

- U is $m \times m$. Columns of U are the eigenvectors of AA^T.
- V is $n \times n$. Columns of V are the eigenvectors of A^TA.
- Σ is $m \times n$. It contains r singular values on the diagonal. A singular value is the square root of a non-zero eigenvalue of A^TA.

Proof

- Let $v_1, \dots, v_n$ be orthonormal eigenvectors of $A^T A$. We have

$$v_i^T A^T A v_j = \lambda_j v_i^T v_j = \lambda_j \delta_{ij}.$$

- Since $A^T A$ has the same nullspace as A, there are r non-zero eigenvalues. These non-zero eigenvalues are positive since $A^T A$ is p.s.d.
- Define the singular values for the positive λ_j,

$$\sigma_j = \sqrt{\lambda_j}.$$

Proof

- Define $u_j = \frac{Av_j}{\sigma_j}$. They are orthonormal eigenvectors of AA^T

$$AA^T u_j = \frac{AA^T Av_j}{\sigma_j} = \lambda_j \frac{Av_j}{\sigma_j} = \lambda_j u_j, \quad u_i^T u_j = \delta_{ij}.$$

- Construct V with v and U with u,

$$(U^T AV)_{ij} = u_i^T Av_j = \begin{cases} 0 & \text{if } j > r, \\ \sigma_j u_i^T u_j = \sigma_j \delta_{ij} & \text{if } j \le r. \end{cases}$$

That is, $U^T AV = \Sigma$. So $A = U\Sigma V^T$.

Remarks

- $AV = U\Sigma$. A multiplied by a column of V produces a multiple of column of U,

$$Av_j = \sigma_j u_j.$$

- $AA^T = U\Sigma\Sigma^T U^T$ and $A^T A = V\Sigma^T \Sigma V^T$. U is the eigenvector matrix of AA^T and V is the eigenvector matrix of $A^T A$.
- The non-zero eigenvalues of AA^T and $A^T A$ are the same. They are in $\Sigma\Sigma^T$.

Example

- Consider

$$A = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}.\ AA^T = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

- The singular values are $\sqrt{3}, 1$.
- Finding v_i and u_i, one has

$$A = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 & 1 & (/\sqrt{6}) \\ -1 & 0 & 1 & (/\sqrt{2}) \\ 1 & 1 & 1 & (/\sqrt{3}) \end{bmatrix}$$

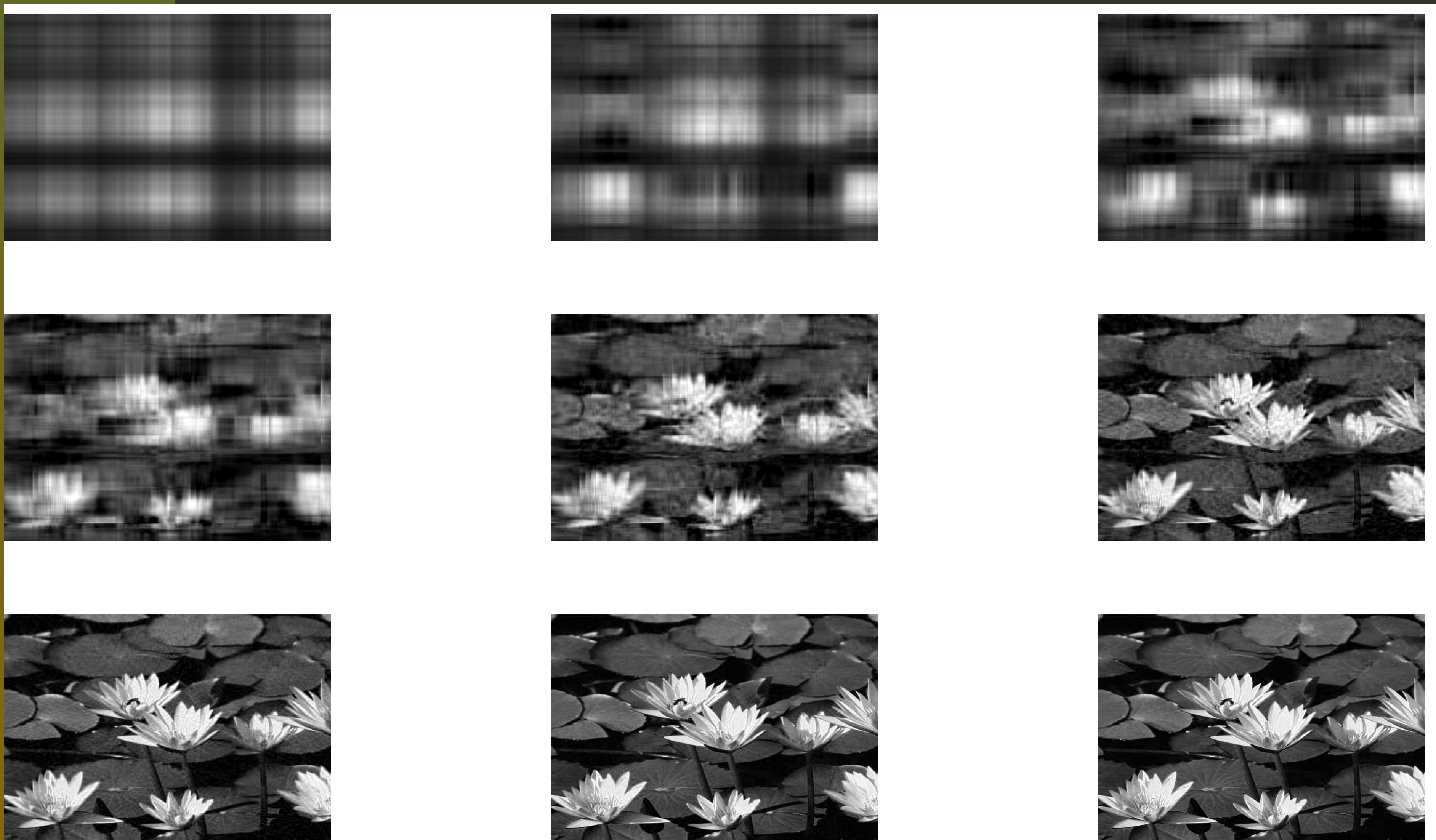
Applications of SVD

- Through SVD, we can represent a matrix as a sum of rank-one matrices

$$A = U\Sigma V^T = u_1\sigma_1 v_1^T + \cdots + u_r\sigma_r v_r^T.$$

- Suppose we have a 1000×1000 matrix, for a total of 10^6 entries. Suppose we use the above expansion and keep only the 50 most significant terms. This would require $50(1 + 1000 + 1000)$ numbers, a save of space of almost 90%.
- This is commonly used in image processing.

SVD for Image



Pseudo-Inverse

- Consider the normal equation

$$A^T A\hat{x} = A^T b.$$

- If A has dependent columns, then $A^T A$ is not invertible and $\hat{x}$ is not unique. Any vector in the nullspace of $A^T A$ can be added to $\hat{x}$.
- Among all solution, we denote the one with the minimum length by x^+.
- The matrix that produces x^+ from b is called the *pseudo-inverse* of A, denoted by A^+.

Properties of Pseudoinverse

- By our definition, $A^+b = x^+$, where x^+ is the shortest solution for

$$A^TA\hat{x} = A^Tb.$$

- A^+ is related to SVD $A = U\Sigma V^T$ by

$$A^+ = V\Sigma^+U^T,$$

where Σ^+ is $n \times m$ with diagonals $\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_r}$.

Proof of Minimum Length

- Multiplication by U^T leaves the length unchanged,

  $$|Ax-b| = |U\Sigma V^T x - b| = |\Sigma V^T x - U^T b| = |\Sigma y - U^T b|,$$

  where $y = V^T x = V^{-1} x$.
- Since Σ is a diagonal matrix, the minimum-length least-square solution is $y^+ = \Sigma^+ U^T b$.
- The minimum-length least-square solution for x is

  $$x^+ = V y^+ = V \Sigma^+ U^T b = A^+ b.$$

Rayleigh's Quotient

- A problem may have an equivalent problem in the form of minimizing an objective function.
- The eigenvalue problem

$$Ax = \lambda x$$

can be solved by looking at the Rayleigh's quotient

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

Rayleigh's Principle

- The minimum value of the Rayleigh's quotient $R(x)$ is the smallest eigenvalue λ_1 of A, achieved by the corresponding eigenvector x_1.
- This follows from

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

- Each diagonal entry is between λ_1 and λ_n since the Rayleigh quotient equals a_{ii} when $x = e_i$, the unit vector along direction x_i.

Maximin Principle

- A vector perpendicular to x_1 is in the subspace spanned by $x_2, \ldots, x_n$, so

$$\lambda_2 = \min_{x \perp x_1} R(x).$$

- For arbitrary v, the following equations hold

$$\lambda_2 \geq \min_{x \perp v} R(x), \quad \lambda_2 = \max_{v} \min_{x \perp v} R(x).$$

- Let S_j be a j-dimensional subspace, then

$$\lambda_{j+1} = \max_{S_j} \min_{x \perp S_j} R(x).$$