GRAM SCHMIDT PROCESS

GRAM SCHMIDT PROCESS

Orthonormal Bases

The Gram–Schmidt Process

Given a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$\mathbf{v}_1 = \mathbf{x}_1$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2$$

$$\vdots$$

$$\mathbf{v}_p = \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\mathbf{v}_{p-1}$$

Then $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is an orthogonal basis for W. In addition

$$\text{Span}\,\{\mathbf{v}_1, \dots, \mathbf{v}_k\} = \text{Span}\,\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \qquad \text{for } 1 \le k \le p \tag{1}$$

QR Factorization

If an m*n matrix A has linearly independent columns **x1; : : : ; xn, then applying the Gram–Schmidt process (with normalizations) to x1; : : : ; xn amounts to *factoring A*,**

$$\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \qquad Q = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_n\,]$$

$$A = [\,\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n\,] = QR \;= [\,Q\mathbf{r}_1 \quad \cdots \quad Q\mathbf{r}_n\,]$$

$$\text{Span}\,\{\mathbf{x}_1, \ldots, \mathbf{x}_k\} = \text{Span}\,\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}.$$

$$\mathbf{x}_k = r_{1k}\mathbf{u}_1 + \cdots + r_{kk}\mathbf{u}_k + 0\cdot\mathbf{u}_{k+1} + \cdots + 0\cdot\mathbf{u}_n$$

R is clearly upper triangular.

QR Factorization

The QR Factorization

If A is an $m \times n$ matrix with linearly independent columns, then A can be factored as $A = QR$, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for Col A and R is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

Least-Squares Problem

If A is $m \times n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, a **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

for all $\mathbf{x}$ in $\mathbb{R}^n$.

we seek an x that makes Ax the closest point in Col A to b.

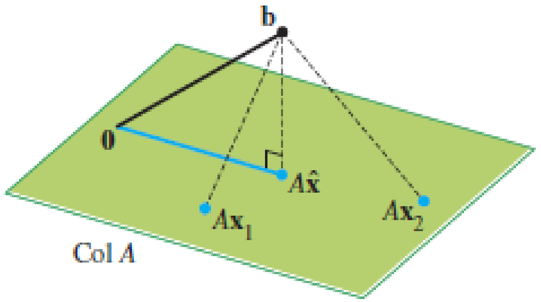


FIGURE 1 The vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other $\mathbf{x}$.

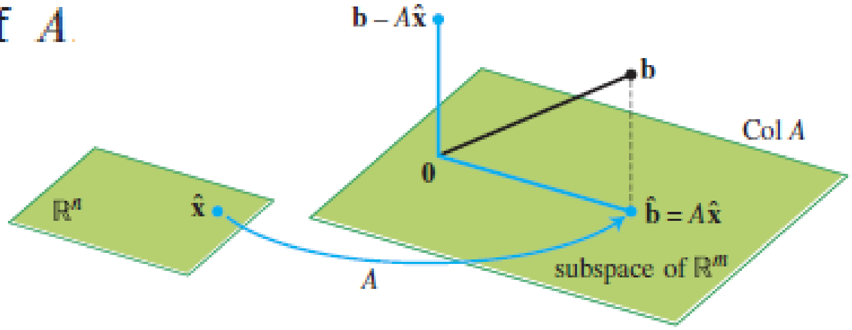
if **b happens to** be in ColA, then ***b** is Ax for some x, and such an x is a “least-squares solution***

Solution of the General Least-Squares Problem

$$\hat{\mathbf{b}} = \text{proj}_{\text{Col } A}\,\mathbf{b} \qquad\qquad A\hat{\mathbf{x}} = \hat{\mathbf{b}}$$

Since $\hat{\mathbf{b}}$ is the closest point in Col A to $\mathbf{b}$, a vector $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$

$\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to each column of A.



$\mathbf{a}_j \cdot (\mathbf{b} - A\hat{\mathbf{x}}) = 0$, and $\mathbf{a}_j^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$. Since each $\mathbf{a}_j^T$ is a row of A^T,

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$$

$$A^TA\mathbf{x} = A^T\mathbf{b}$$

normal equations

Least-Squares Problem

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$.

Least Square Solution is unique?

Let A be an $m \times n$ matrix. The following statements are logically equivalent:

a. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.

b. The columns of A are linearly independent.

c. The matrix A^TA is invertible.

When these statements are true, the least-squares solution $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} \tag{4}$$

APPLICATIONS TO LINEAR MODELS

Least Square line

❖A common task in science and engineering is to analyze and understand relationships among several quantities that vary

❖data are used to build or verify a formula that predicts the value of one variable as a function of other variables

Least-Squares Lines

$$y = \beta_0 + \beta_1 x$$

$$(x_1, y_1), \ldots, (x_n, y_n)$$

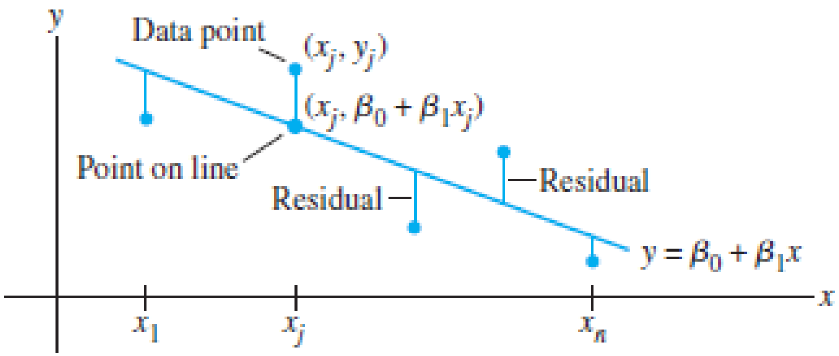


FIGURE 1 Fitting a line to experimental data.

Goal: determine the parameters B0 and B1 that make the line as "close" to the points as possible

Least Square line

There are several ways to measure how “close” the line is to the data.
Usual choice is to add the squares of the residuals

least-squares line is the line that minimizes the sum of the squares of the residuals

line of regression of y on x

Linear regression coefficients

$(x_1, y_1), \ldots, (x_n, y_n)$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$$X\beta = \mathbf{y}$$

Predicted y-value		Observed y-value
$\beta_0 + \beta_1 x_1$	$=$	y_1
$\beta_0 + \beta_1 x_2$	$=$	y_2
$\vdots$		$\vdots$
$\beta_0 + \beta_1 x_n$	$=$	y_n

Least Square line

$$X\beta = \mathbf{y}$$

*Computing the least-squares solution of XB=**y is equivalent to** finding the B that determines the least-squares line in Figure 1*

$$X^T X\beta = X^T\mathbf{y}$$

$$\mathbf{y} = X\beta + \epsilon$$

goal is to minimize the length of residual (error), which amounts to finding a least-squares solution

example

$$\boxed{y = \beta_0 + \beta_1 x + \beta_2 x^2}$$

$(x_1, y_1), \ldots, (x_n, y_n)$

$$\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon_1 \\
y_2 &= \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 + \epsilon_2 \\
\vdots & \qquad\qquad \vdots \\
y_n &= \beta_0 + \beta_1 x_n + \beta_2 x_n^2 + \epsilon_n
\end{aligned}$$

$$\begin{array}{ccccccc}
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} & = & \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} & \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} & + & \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \\
\mathbf{y} & = & X & \boldsymbol{\beta} & + & \boldsymbol{\epsilon}
\end{array}$$

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + \cdots + \beta_k f_k(x)$$

Multiple Regression

$$y = \beta_0 + \beta_1 u + \beta_2 v$$

$$y = \beta_0 + \beta_1 u + \beta_2 v + \beta_3 u^2 + \beta_4 uv + \beta_5 v^2$$

$$y = \beta_0 f_0(u, v) + \beta_1 f_1(u, v) + \cdots + \beta_k f_k(u, v)$$

$$\begin{aligned}
y_1 &= \beta_0 + \beta_1 u_1 + \beta_2 v_1 + \epsilon_1 \\
y_2 &= \beta_0 + \beta_1 u_2 + \beta_2 v_2 + \epsilon_2 \\
&\vdots \qquad\qquad\quad \vdots \\
y_n &= \beta_0 + \beta_1 u_n + \beta_2 v_n + \epsilon_n
\end{aligned}$$

$$\mathbf{y} = \overset{\substack{\text{Observation}\\ \text{vector}}}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}, \quad X = \overset{\substack{\text{Design}\\ \text{matrix}}}{\begin{bmatrix} 1 & u_1 & v_1 \\ 1 & u_2 & v_2 \\ \vdots & \vdots & \vdots \\ 1 & u_n & v_n \end{bmatrix}}, \quad \boldsymbol{\beta} = \overset{\substack{\text{Parameter}\\ \text{vector}}}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}, \quad \boldsymbol{\epsilon} = \overset{\substack{\text{Residual}\\ \text{vector}}}{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}}$$

SYMMETRIC MATRICES AND QUADRATIC FORMS

DIAGONALIZATION OF SYMMETRIC MATRICES

If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.

PROOF Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be eigenvectors that correspond to distinct eigenvalues, say, λ_1 and λ_2. To show that $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, compute

$$\begin{aligned}\lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2 &= (\lambda_1 \mathbf{v}_1)^T \mathbf{v}_2 = (A\mathbf{v}_1)^T \mathbf{v}_2 = (\mathbf{v}_1^T A^T)\mathbf{v}_2 = \mathbf{v}_1^T (A\mathbf{v}_2) = \mathbf{v}_1^T (\lambda_2 \mathbf{v}_2) \\ &= \lambda_2 \mathbf{v}_1^T \mathbf{v}_2 = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2\end{aligned}$$

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$$

DIAGONALIZATION OF SYMMETRIC MATRICES

An n* n matrix A is said to be **orthogonally diagonalizable if there** are an orthogonal matrix P and a diagonal matrix D such that

$$A = PDP^T = PDP^{-1}$$

Such a diagonalization requires n linearly independent and orthonormal eigenvectors. When is this possible?

$$A^T = (PDP^T)^T = P^{TT} D^T P^T = PDP^T = A$$

An $n \times n$ matrix A is orthogonally diagonalizable if and only if A is a symmetric matrix.

The Spectral Theorem for Symmetric Matrices

An $n \times n$ symmetric matrix A has the following properties:

a. A has n real eigenvalues, counting multiplicities.

b. The dimension of the eigenspace for each eigenvalue λ equals the multiplicity of λ as a root of the characteristic equation.

c. The eigenspaces are mutually orthogonal, in the sense that eigenvectors corresponding to different eigenvalues are orthogonal.

d. A is orthogonally diagonalizable.

Spectral Decomposition

spectral decomposition of A

$$A = PDP^T = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 \mathbf{u}_1 & \cdots & \lambda_n \mathbf{u}_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix}$$

$$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T$$

an $n \times n$ matrix of rank 1

QUADRATIC FORM

Quadratic form

A **quadratic form** on $\mathbb{R}^n$ is a function Q defined on $\mathbb{R}^n$ whose value at a vector $\mathbf{x}$ in $\mathbb{R}^n$ can be computed by an expression of the form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, where A is an $n \times n$ symmetric matrix.

$$Q(\mathbf{x}) = \mathbf{x}^T I \mathbf{x} = \|\mathbf{x}\|^2$$

Change of Variable in a Quadratic Form

$$\mathbf{x} = P\mathbf{y}, \quad \mathbf{y} = P^{-1}\mathbf{x}$$

$$\mathbf{x}^T A\mathbf{x} = (P\mathbf{y})^T A(P\mathbf{y}) = \mathbf{y}^T P^T A P\mathbf{y} = \mathbf{y}^T(P^T A P)\mathbf{y}$$

$\mathbf{y}^T D\mathbf{y}$

there is an *orthogonal* matrix P such that $P^T AP$ is a diagonal matrix D

Quadratic form

The Principal Axes Theorem

Let A be an $n \times n$ symmetric matrix. Then there is an orthogonal change of variable, $\mathbf{x} = P\mathbf{y}$, that transforms the quadratic form $\mathbf{x}^T A\mathbf{x}$ into a quadratic form $\mathbf{y}^T D\mathbf{y}$ with no cross-product term.

Classifying Quadratic Forms

A quadratic form Q is:

a. **positive definite** if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$,

b. **negative definite** if $Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$,

c. **indefinite** if $Q(\mathbf{x})$ assumes both positive and negative values.

Quadratic Forms and Eigenvalues

Let A be an $n \times n$ symmetric matrix. Then a quadratic form $\mathbf{x}^T A\mathbf{x}$ is:

a. positive definite if and only if the eigenvalues of A are all positive,

b. negative definite if and only if the eigenvalues of A are all negative, or

c. indefinite if and only if A has both positive and negative eigenvalues.

Proof

PROOF By the Principal Axes Theorem, there exists an orthogonal change of variable $\mathbf{x} = P\mathbf{y}$ such that

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

positive definite matrix A is a *symmetric* matrix for which the quadratic form $\mathbf{x}^T A\mathbf{x}$ is positive definite

CONSTRAINED OPTIMIZATION

CONSTRAINED OPTIMIZATION

Engineers, economists, scientists, and mathematicians often need to find the maximum or minimum value of a quadratic form Q(x) for **x** in some specified set.

$$\|\mathbf{x}\| = 1, \qquad \|\mathbf{x}\|^2 = 1, \qquad \mathbf{x}^T\mathbf{x} = 1$$

$$x_1^2 + x_2^2 + \cdots + x_n^2 = 1$$

When a quadratic form Q has no cross-product terms, it is easy to find the maximum and minimum of $Q(\mathbf{x})$ for $\mathbf{x}^T\mathbf{x} = 1$.

example

EXAMPLE 1 Find the maximum and minimum values of $Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$ subject to the constraint $\mathbf{x}^T\mathbf{x} = 1$.

$$\begin{aligned} Q(\mathbf{x}) &= 9x_1^2 + 4x_2^2 + 3x_3^2 \\ &\le 9x_1^2 + 9x_2^2 + 9x_3^2 \\ &= 9(x_1^2 + x_2^2 + x_3^2) \\ &= 9 \end{aligned}$$

$$\mathbf{x} = (1, 0, 0)$$

$$Q(\mathbf{x}) \ge 3x_1^2 + 3x_2^2 + 3x_3^2 = 3(x_1^2 + x_2^2 + x_3^2) = 3$$

CONSTRAINED OPTIMIZATION

$$m = \min \{\mathbf{x}^T A\mathbf{x} : \|\mathbf{x}\| = 1\}, \quad M = \max \{\mathbf{x}^T A\mathbf{x} : \|\mathbf{x}\| = 1\}$$

Let A be a symmetric matrix, and define m and M as in (2). Then M is the greatest eigenvalue λ_1 of A and m is the least eigenvalue of A. The value of $\mathbf{x}^T A\mathbf{x}$ is M when $\mathbf{x}$ is a unit eigenvector $\mathbf{u}_1$ corresponding to M. The value of $\mathbf{x}^T A\mathbf{x}$ is m when $\mathbf{x}$ is a unit eigenvector corresponding to m.

PROOF Orthogonally diagonalize A as PDP^{-1}. We know that

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} \quad \text{when } \mathbf{x} = P\mathbf{y}$$

$$\|\mathbf{x}\| = \|P\mathbf{y}\| = \|\mathbf{y}\| \quad \text{for all } \mathbf{y}$$

CONSTRAINED OPTIMIZATION

To simplify notation, suppose that A is a 3×3 matrix with eigenvalues $a \geq b \geq c$.

$$D = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$$

$$\begin{aligned} \mathbf{y}^T D\mathbf{y} &= ay_1^2 + by_2^2 + cy_3^2 \leq ay_1^2 + ay_2^2 + ay_3^2 \\ &= a(y_1^2 + y_2^2 + y_3^2) \\ &= a\|\mathbf{y}\|^2 = a \end{aligned}$$

Thus $M \leq a$, by definition of M. However, $\mathbf{y}^T D\mathbf{y} = a$ when $\mathbf{y} = \mathbf{e}_1 = (1, 0, 0)$

$$M = a$$

$$\mathbf{x} = P\mathbf{e}_1 = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \mathbf{u}_1$$

CONSTRAINED OPTIMIZATION

Let A, λ_1, and $\mathbf{u}_1$ be as in Theorem 6. Then the maximum value of $\mathbf{x}^T A\mathbf{x}$ subject to the constraints

$$\mathbf{x}^T\mathbf{x} = 1, \quad \mathbf{x}^T\mathbf{u}_1 = 0$$

is the second greatest eigenvalue, λ_2, and this maximum is attained when $\mathbf{x}$ is an eigenvector $\mathbf{u}_2$ corresponding to λ_2.

SINGULAR VALUE DECOMPOSITION

Introduction

The absolute values of the eigenvalues of a symmetric matrix A measure the amounts that A stretches or shrinks certain vectors (the eigenvectors)

$$\|A\mathbf{x}\| = \|\lambda\mathbf{x}\| = |\lambda|\,\|\mathbf{x}\| = |\lambda|$$

Example

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \qquad \mathbf{x} \mapsto A\mathbf{x}$$

$\mathbf{x}$ at which the length $\|A\mathbf{x}\|$ is maximized, and compute this maximum length $\|\mathbf{x}\| = 1$

$$\|A\mathbf{x}\|^2 = (A\mathbf{x})^T(A\mathbf{x}) = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T(A^T A)\mathbf{x}$$

$\boxed{\text{the greatest eigenvalue } \lambda_1 \text{ of } A^T A}$

example

$$A^TA = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$$

$\lambda_1 = 360, \lambda_2 = 90$, and $\lambda_3 = 0$

$$\mathbf{v}_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}$$

For $\|\mathbf{x}\| = 1$, the maximum value of $\|A\mathbf{x}\|$ is $\|A\mathbf{v}_1\| = \sqrt{360} = 6\sqrt{10}$.

Singular Values

Let A be an $m \times n$ matrix. Then A^TA is symmetric and can be orthogonally diagonalized. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A^TA, and let $\lambda_1, \ldots, \lambda_n$ be the associated eigenvalues of A^TA. Then, for $1 \le i \le n$,

$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n \ge 0$

$$\begin{aligned} \|A\mathbf{v}_i\|^2 &= (A\mathbf{v}_i)^T A\mathbf{v}_i = \mathbf{v}_i^T A^T A\mathbf{v}_i \\ &= \mathbf{v}_i^T(\lambda_i \mathbf{v}_i) \qquad \text{Since } \mathbf{v}_i \text{ is an eigenvector of } A^TA \\ &= \lambda_i \qquad \text{Since } \mathbf{v}_i \text{ is a unit vector} \end{aligned}$$

eigenvalues of A^TA are all nonnegative

The **singular values** of A are the square roots of the eigenvalues of A^TA, denoted by $\sigma_1, \ldots, \sigma_n$, and they are arranged in decreasing order. That is, $\sigma_i = \sqrt{\lambda_i}$ for $1 \le i \le n$.

the singular values of A are the lengths of the vectors $A\mathbf{v}_1, \ldots, A\mathbf{v}_n$

Theorem

THEOREM 9

Suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A^TA, arranged so that the corresponding eigenvalues of A^TA satisfy $\lambda_1 \geq \cdots \geq \lambda_n$, and suppose A has r nonzero singular values. Then $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_r\}$ is an orthogonal basis for Col A, and rank $A = r$.

$$(A\mathbf{v}_i)^T(A\mathbf{v}_j) = \mathbf{v}_i^T A^T A\mathbf{v}_j = \mathbf{v}_i^T(\lambda_j \mathbf{v}_j) = 0$$

Col A

$$\mathbf{y} = A\mathbf{x} \qquad \mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$$

$$\begin{aligned}\mathbf{y} = A\mathbf{x} &= c_1 A\mathbf{v}_1 + \cdots + c_r A\mathbf{v}_r + c_{r+1} A\mathbf{v}_{r+1} + \cdots + c_n A\mathbf{v}_n \\ &= c_1 A\mathbf{v}_1 + \cdots + c_r A\mathbf{v}_r + 0 + \cdots + 0\end{aligned}$$

SVD

THEOREM 10

The Singular Value Decomposition

Let A be an $m \times n$ matrix with rank r. Then there exists an $m \times n$ matrix Σ as in (3) for which the diagonal entries in D are the first r singular values of A, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, and there exist an $m \times m$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that

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

$$\Sigma = \begin{array}{c} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{array}{l} \\ \leftarrow m - r \text{ rows} \end{array} \\ \uparrow n - r \text{ columns} \end{array}$$

matrices U and V are not uniquely determined by A, but the diagonal entries of D are necessarily the singular values of A

columns of U **left singular vectors of A**
columns of V **right singular vectors of A**

proof

PROOF Let λ_i and $\mathbf{v}_i$ be as in Theorem 9, so that $\{A\mathbf{v}_1, \dots, A\mathbf{v}_r\}$ is an orthogonal basis for Col A. Normalize each $A\mathbf{v}_i$ to obtain an orthonormal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$, where

$$\mathbf{u}_i = \frac{1}{\|A\mathbf{v}_i\|} A\mathbf{v}_i = \frac{1}{\sigma_i} A\mathbf{v}_i \qquad\qquad A\mathbf{v}_i = \sigma_i \mathbf{u}_i$$

extend $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ to an orthonormal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ of $\mathbb{R}^m$

$$U = [\, \mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_m \,] \quad \text{and} \quad V = [\, \mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n \,]$$

$$AV = [\, A\mathbf{v}_1 \quad \cdots \quad A\mathbf{v}_r \quad \mathbf{0} \quad \cdots \quad \mathbf{0} \,] = [\, \sigma_1\mathbf{u}_1 \quad \cdots \quad \sigma_r\mathbf{u}_r \quad \mathbf{0} \quad \cdots \quad \mathbf{0} \,]$$

proof

$$AV = [\, A\mathbf{v}_1 \quad \cdots \quad A\mathbf{v}_r \quad \mathbf{0} \quad \cdots \quad \mathbf{0}\,] = [\,\sigma_1\mathbf{u}_1 \quad \cdots \quad \sigma_r\mathbf{u}_r \quad \mathbf{0} \quad \cdots \quad \mathbf{0}\,]$$

$$U\Sigma = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_m\,] \left[\begin{array}{cccc|c} \sigma_1 & & & 0 & \\ & \sigma_2 & & & 0 \\ & & \ddots & & \\ 0 & & & \sigma_r & \\ \hline & & 0 & & 0 \end{array}\right] = [\,\sigma_1\mathbf{u}_1 \quad \cdots \quad \sigma_r\mathbf{u}_r \quad \mathbf{0} \quad \cdots \quad \mathbf{0}\,] = AV$$

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

example

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$$

$$\boxed{V = [\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3\,] = \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix}}$$

$$\sigma_1 = 6\sqrt{10}, \quad \sigma_2 = 3\sqrt{10}, \quad \sigma_3 = 0$$

$$D = \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \end{bmatrix}$$

$$\boxed{\Sigma = [\,D \quad 0\,] = \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}}$$

$$\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \frac{1}{6\sqrt{10}} \begin{bmatrix} 18 \\ 6 \end{bmatrix} = \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\sigma_2} A\mathbf{v}_2 = \frac{1}{3\sqrt{10}} \begin{bmatrix} 3 \\ -9 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{bmatrix}$$

$$\boxed{A = \underset{\substack{\uparrow \\ U}}{\begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix}} \underset{\substack{\uparrow \\ \Sigma}}{\begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}} \underset{\substack{\uparrow \\ V^T}}{\begin{bmatrix} 1/3 & 2/3 & 2/3 \\ -2/3 & -1/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix}}}$$

example

Find a singular value decomposition of $A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}$ $A^TA = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}$

eigenvalues of A^TA are 18 and 0

$\sigma_1 = \sqrt{18} = 3\sqrt{2}$ and $\sigma_2 = 0$.

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$$V = [\,\mathbf{v}_1 \quad \mathbf{v}_2\,] = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

$$A\mathbf{v}_1 = \begin{bmatrix} 2/\sqrt{2} \\ -4/\sqrt{2} \\ 4/\sqrt{2} \end{bmatrix}, \quad A\mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{u}_1 = \frac{1}{3\sqrt{2}} A\mathbf{v}_1 = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix}$$

example

$\mathbf{u}_1^T \mathbf{x} = 0$ $\quad x_1 - 2x_2 + 2x_3 = 0$

$$\mathbf{w}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

Gram-Schmidt

$$\mathbf{u}_2 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -2/\sqrt{45} \\ 4/\sqrt{45} \\ 5/\sqrt{45} \end{bmatrix}$$

SVD for Image Compression

$$f_{ij} \quad \text{Where} \quad f_{ij} \equiv f(x_i, y_j)$$

Redundancy exists in Images
Size of images
Compression

$$A = USV^T = \sum_{i=1}^{r} \sigma_i \mathrm{u}_i \mathrm{v}_i^T$$

$$A_k = \sigma_1 \mathrm{u}_1 \mathrm{v}_1{}^T + \sigma_2 \mathrm{u}_2 \mathrm{v}_2{}^T + \cdots + \sigma_k \mathrm{u}_k \mathrm{v}_k{}^T$$

The total storage for A_k will be

$$k\,(m + n + 1)$$

SVD for Image Compression

$$C_R = m*n/(k\,(m+n+1))$$

To measure the quality between original image *A and the compressed image Ak, the* measurement of Mean Square Error (MSE)

$$MSE = \frac{1}{mn}\sum_{y=1}^{m}\sum_{x=1}^{n}(f_A(x,y) - f_{A_k}(x,y))$$

C_R **Comp**	**MSE** **(Quality)**
5.03	108.11
3.35	63.15
2.51	40.39
2.01	27.22
1.68	15.64
1.26	9.07
1	

Face Recognition: PCA (principle component analysis)

- SVD approach treats a set of known faces as vectors in a subspace, called "face space"
- Assume each face image has $m \times n = M$ *pixels*
- an $M \times 1$ column vector ***fi***
- A training set, *S with N* number of face images of known individuals forms an $M \times N$ *matrix:*

$$S = [\mathbf{f}_1, \mathbf{f}_2, \ldots\ldots, \mathbf{f}_N]$$

$$\bar{\mathbf{f}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{f}_i \qquad \mathbf{a}_i = \mathbf{f}_i - \bar{\mathbf{f}},,, i = 1,2,\ldots N$$

$$A = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N]$$

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

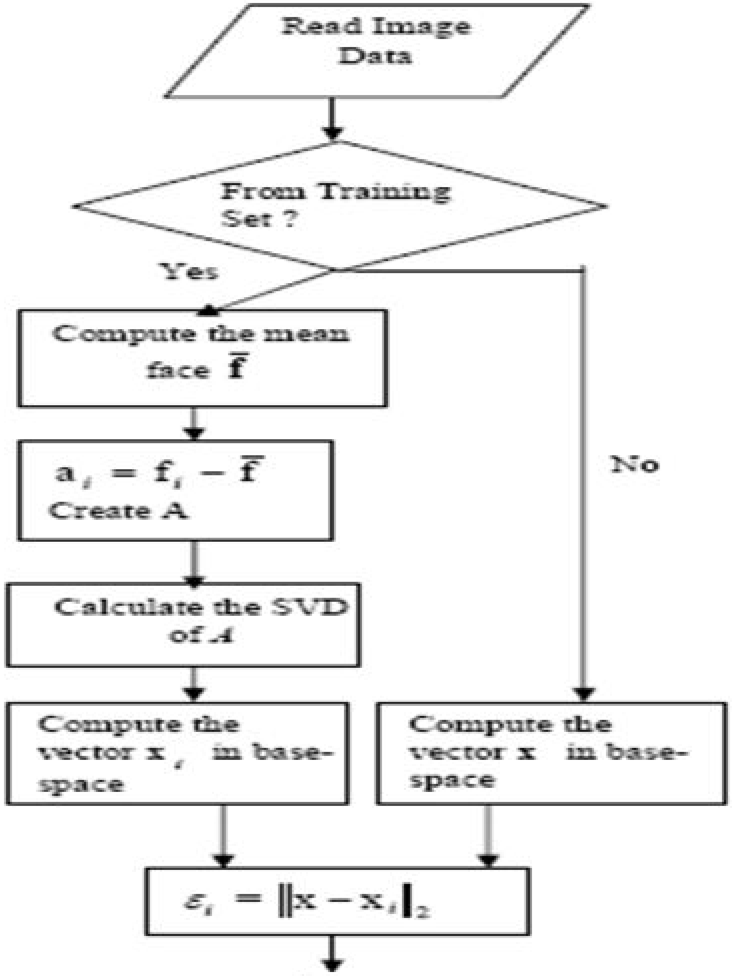
$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ form an orthonormal basis for $R(A)$

Face Recognition: PCA

$$\mathbf{x} \quad (= [x_1, x_2, \ldots, x_r]^T \ ) \qquad\qquad \mathbf{x} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r]^T (\mathbf{f} - \bar{\mathbf{f}})$$

Minimize the distance

$$\varepsilon_i = \|\mathbf{x} - \mathbf{x}_i\|_2 = [(\mathbf{x} - \mathbf{x}_i)^T (\mathbf{x} - \mathbf{x}_i)]^{1/2}$$



Read Image Data
From Training Set ?
Yes
No
Compute the mean face $\bar{\mathbf{f}}$
$\mathbf{a}_i = \mathbf{f}_i - \bar{\mathbf{f}}$
Create A
Calculate the SVD of A
Compute the vector $\mathbf{x}_i$ in base-space
Compute the vector $\mathbf{x}$ in base-space
$\varepsilon_i = \|\mathbf{x} - \mathbf{x}_i\|_2$