

Portfolios that Contain Risky Assets 2: Covariance Matrices

C. David Levermore

University of Maryland, College Park, MD

Math 420: *Mathematical Modeling*

February 6, 2020 version

© 2020 Charles David Levermore

Portfolios that Contain Risky Assets

Part I: Portfolio Models

1. Risk and Reward
2. Covariance Matrices
3. Markowitz Portfolios
4. Markowitz Frontiers
5. Portfolios with Risk-Free Assets
6. Long Portfolios and Their Frontiers
7. Long Portfolios with a Safe Investment
8. Limited Portfolios and Their Frontiers
9. Limited Portfolios with Risk-Free Assets
10. Bounded Portfolios and Leverage Limits

Covariance Matrices

- 1 Introduction
- 2 Symmetry and Definiteness
- 3 Eigenpairs and Diagonalization
- 4 Statistical Interpretation
- 5 Principle Component Analysis
- 6 Dimension of the Data

Introduction

Suppose that we are considering return histories $\{r_i(d)\}_{d=1}^D$ for assets $i = 1, \dots, N$ over a period of D trading days and assign day d a weight $w(d) > 0$ such that the weights $\{w(d)\}_{d=1}^D$ satisfy

$$\sum_{d=1}^D w(d) = 1.$$

Then the return means, variances, and covariances are given by

$$\begin{aligned} m_i &= \sum_{d=1}^D w(d) r_i(d), \\ v_{ij} &= \sum_{d=1}^D w(d) (r_i(d) - m_i)(r_j(d) - m_j). \end{aligned} \tag{1.1}$$

Introduction

The return history can be expressed as $\{\mathbf{r}(d)\}_{d=1}^D$ where

$$\mathbf{r}(d) = \begin{pmatrix} r_1(d) \\ \vdots \\ r_N(d) \end{pmatrix}.$$

The N -vector of return means \mathbf{m} and the $N \times N$ -matrix of return variances and covariances \mathbf{V} then can be expressed as

$$\mathbf{m} = \begin{pmatrix} m_1 \\ \vdots \\ m_N \end{pmatrix} = \sum_{d=1}^D w(d) \mathbf{r}(d),$$

$$\mathbf{V} = \begin{pmatrix} v_{11} & \cdots & v_{1N} \\ \vdots & \ddots & \vdots \\ v_{N1} & \cdots & v_{NN} \end{pmatrix} = \sum_{d=1}^D w(d) (\mathbf{r}(d) - \mathbf{m}) (\mathbf{r}(d) - \mathbf{m})^T.$$

Introduction

We call \mathbf{V} the *covariance matrix*. It also is called the *variance/covariance matrix* or the *variance matrix*.

The most important properties of \mathbf{V} :

- *it is always symmetric,*
- *it is almost always positive definite.*
- *it can be diagonalized.*

These properties are taught in most elementary linear algebra courses, but they are so important that they will be reviewed in the next two sections. In subsequent sections these properties will be used to extract statistical information from \mathbf{V} .

Symmetry and Definiteness

Here we review the notions of symmetric and definite matrices.

Definition 1. A real $N \times N$ -matrix \mathbf{A} is said to be *symmetric* if $\mathbf{A}^T = \mathbf{A}$, where \mathbf{A}^T is the transpose of \mathbf{A} . It is said to be *nonnegative definite* or *positive semidefinite* if

$$\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0 \quad \text{for every } \mathbf{x} \in \mathbb{R}^N.$$

It is said to be *positive definite* if

$$\mathbf{x}^T \mathbf{A} \mathbf{x} > 0 \quad \text{for every nonzero } \mathbf{x} \in \mathbb{R}^N.$$

Remarks. Clearly, every positive definite matrix is nonnegative definite. A nonnegative definite matrix is positive definite if and only if

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = 0 \quad \implies \quad \mathbf{x} = \mathbf{0}. \quad (2.2)$$

Symmetry and Definiteness

Fact 1. The covariance matrix \mathbf{V} is symmetric.

Proof. It is clear from (1.1) that $v_{ij} = v_{ji}$, whereby $\mathbf{V} = \mathbf{V}^T$. □

Fact 2. The covariance matrix \mathbf{V} is nonnegative definite.

Proof. Let $\mathbf{x} \in \mathbb{R}^N$ be arbitrary. Then

$$\begin{aligned}\mathbf{x}^T \mathbf{V} \mathbf{x} &= \mathbf{x}^T \left(\sum_{d=1}^D w(d) (\mathbf{r}(d) - \mathbf{m}) (\mathbf{r}(d) - \mathbf{m})^T \right) \mathbf{x} \\ &= \sum_{d=1}^D w(d) \mathbf{x}^T (\mathbf{r}(d) - \mathbf{m}) (\mathbf{r}(d) - \mathbf{m})^T \mathbf{x} \\ &= \sum_{d=1}^D w(d) \left((\mathbf{r}(d) - \mathbf{m})^T \mathbf{x} \right)^2 \geq 0.\end{aligned}$$

Symmetry and Definiteness

Fact 3. The covariance matrix \mathbf{V} is positive definite if and only if the vectors $\{\mathbf{r}(d) - \mathbf{m}\}_{d=1}^D$ span \mathbb{R}^N .

Proof. Because $w(d) > 0$, the calculation in the previous proof shows that $\mathbf{x}^T \mathbf{V} \mathbf{x} = 0$ if and only if

$$(\mathbf{r}(d) - \mathbf{m})^T \mathbf{x} = 0 \quad \text{for every } d = 1, \dots, D. \quad (2.3)$$

First, suppose that \mathbf{V} is not positive definite. Then by (2.2) there exists an $\mathbf{x} \in \mathbb{R}^N$ such that $\mathbf{x}^T \mathbf{V} \mathbf{x} = 0$ and $\mathbf{x} \neq \mathbf{0}$. This implies by (2.3) that the vectors $\{\mathbf{r}(d) - \mathbf{m}\}_{d=1}^D$ lie in the hyperplane orthogonal (normal) to \mathbf{x} . Therefore the vectors $\{\mathbf{r}(d) - \mathbf{m}\}_{d=1}^D$ do not span \mathbb{R}^N .

Conversely, suppose that the vectors $\{\mathbf{r}(d) - \mathbf{m}\}_{d=1}^D$ do not span \mathbb{R}^N . Then there must be a nonzero vector \mathbf{x} that is orthogonal to their span. This implies that \mathbf{x} satisfies (2.3), whereby $\mathbf{x}^T \mathbf{V} \mathbf{x} = 0$. Therefore \mathbf{V} is not positive definite by (2.2).

Symmetry and Definiteness

Remark. The set of vectors $\{\mathbf{r}(d) - \mathbf{m}\}_{d=1}^D$ is linearly dependent because

$$\sum_{d=1}^D w(d)(\mathbf{r}(d) - \mathbf{m}) = \sum_{d=1}^D w(d)\mathbf{r}(d) - \mathbf{m} \sum_{d=1}^D w(d) = \mathbf{0}.$$

Therefore it can span \mathbb{R}^N only if $D > N$. Hence, we require that $D > N$.

Remark. In practice D will be much larger than N . In the homework and projects for this course usually $N \leq 10$ while $D \geq 42$ (often $D = 252$). When D is so much greater than N the covariance matrix \mathbf{V} will almost always be positive definite.

Symmetry and Definiteness

Remark. If $\{\mathbf{r}(d) - \mathbf{m}\}_{d=1}^D$ spans \mathbb{R}^N then $\{\mathbf{r}(d)\}_{d=1}^D$ also spans \mathbb{R}^N . Indeed, if $\mathbf{x} \in \text{Span}\{\mathbf{r}(d) - \mathbf{m}\}_{d=1}^D$ then for some $\{a(d)\}_{d=1}^D$ we have

$$\mathbf{x} = \sum_{d=1}^D a(d)(\mathbf{r}(d) - \mathbf{m}) = \sum_{d=1}^D \left(a(d) - w(d) \sum_{d'=1}^D a(d') \right) \mathbf{r}(d).$$

Therefore $\mathbf{x} \in \text{Span}\{\mathbf{r}(d)\}_{d=1}^D$. However, the converse need not hold. A counterexample for $N = 2$ and any $D > 2$ can be constructed as follows. Let $\{\mathbf{m}, \mathbf{n}\}$ span \mathbb{R}^2 . Let $\mathbf{r}(d) = \mathbf{m} + h(d)\mathbf{n}$ where $h(d) \neq 0$ and

$$\sum_{d=1}^D w(d)h(d) = 0.$$

Then $\{\mathbf{r}(d)\}_{d=1}^D$ spans \mathbb{R}^2 while $\{\mathbf{r}(d) - \mathbf{m}\}_{d=1}^D$ does not span \mathbb{R}^2 because every $\mathbf{r}(d) - \mathbf{m}$ is proportional to \mathbf{n} .

Eigenpairs and Diagonalization

Recall from linear algebra that an **eigenpair** (λ, \mathbf{q}) of a real $N \times N$ matrix \mathbf{A} is a scalar λ (possibly complex) and a nonzero vector \mathbf{q} (possibly with complex entries) such that

$$\mathbf{A}\mathbf{q} = \lambda\mathbf{q}. \quad (3.4)$$

An eigenpair is called a **real eigenpair** when λ and every entry of \mathbf{q} is real.

Recall too that if \mathbf{A} is real and symmetric then it has N real eigenpairs

$$(\lambda_1, \mathbf{q}_1), \quad (\lambda_2, \mathbf{q}_2), \quad \cdots \quad (\lambda_N, \mathbf{q}_N), \quad (3.5)$$

such that the eigenvectors $\{\mathbf{q}_i\}_{i=1}^N$ are an **orthonormal set**. This means that they satisfy the orthonormality conditions

$$\mathbf{q}_i^T \mathbf{q}_j = \delta_{ij} \equiv \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \quad (3.6)$$

Eigenpairs and Diagonalization

Because the $\{\mathbf{q}_i\}_{i=1}^N$ satisfy the orthonormality conditions (3.6), they form an **orthonormal basis** of \mathbb{R}^N . Every $\mathbf{x} \in \mathbb{R}^N$ can be expanded as

$$\mathbf{x} = \mathbf{q}_1 \mathbf{q}_1^T \mathbf{x} + \mathbf{q}_2 \mathbf{q}_2^T \mathbf{x} + \cdots + \mathbf{q}_N \mathbf{q}_N^T \mathbf{x}. \quad (3.7)$$

The numbers $\{\mathbf{q}_i^T \mathbf{x}\}_{i=1}^N$ are called the **coordinates** of \mathbf{x} for the orthonormal basis $\{\mathbf{q}_i\}_{i=1}^N$. By multiplying (3.7) on the left by \mathbf{x}^T we see that the square of the Euclidean norm of \mathbf{x} is given by

$$\|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x} = (\mathbf{q}_1^T \mathbf{x})^2 + (\mathbf{q}_2^T \mathbf{x})^2 + \cdots + (\mathbf{q}_N^T \mathbf{x})^2. \quad (3.8)$$

Eigenpairs and Diagonalization

Because the $\{\mathbf{q}_i\}_{i=1}^N$ are eigenvectors of \mathbf{A} , we see from (3.7) that

$$\begin{aligned}\mathbf{Ax} &= \mathbf{Aq}_1 \mathbf{q}_1^T \mathbf{x} + \mathbf{Aq}_2 \mathbf{q}_2^T \mathbf{x} + \cdots + \mathbf{Aq}_N \mathbf{q}_N^T \mathbf{x} \\ &= \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T \mathbf{x} + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^T \mathbf{x} + \cdots + \lambda_N \mathbf{q}_N \mathbf{q}_N^T \mathbf{x}.\end{aligned}\tag{3.9}$$

Hence, the $\{\lambda_i \mathbf{q}_i^T \mathbf{x}\}_{i=1}^N$ are the coordinates of \mathbf{Ax} for the orthonormal basis $\{\mathbf{q}_i\}_{i=1}^N$. Therefore by (3.8) we have

$$\|\mathbf{Ax}\|^2 = \lambda_1^2 (\mathbf{q}_1^T \mathbf{x})^2 + \lambda_2^2 (\mathbf{q}_2^T \mathbf{x})^2 + \cdots + \lambda_N^2 (\mathbf{q}_N^T \mathbf{x})^2.\tag{3.10}$$

Eigenpairs and Diagonalization

Fact 4. If \mathbf{A} is a real symmetric matrix then the properties of nonnegative definite and positive definite can be characterized as follows.

- \mathbf{A} is nonnegative definite if and only if its eigenvalues are all nonnegative.
- \mathbf{A} is positive definite if and only if its eigenvalues are all positive.

Proof. The (\implies) directions of these characterizations follow from the fact that if (λ, \mathbf{q}) is an eigenpair of \mathbf{A} that is normalized so that $\mathbf{q}^T \mathbf{q} = 1$ then

$$\lambda = \lambda \mathbf{q}^T \mathbf{q} = \mathbf{q}^T (\lambda \mathbf{q}) = \mathbf{q}^T (\mathbf{A} \mathbf{q}) = \mathbf{q}^T \mathbf{A} \mathbf{q}.$$

Hence, if \mathbf{A} is nonnegative definite then $\lambda \geq 0$ and if \mathbf{A} is positive definite then $\lambda > 0$.

Eigenpairs and Diagonalization

The (\Leftarrow) directions of these characterizations use the full power of the orthonormality conditions (3.6) as embodied by expansion (3.9),

$$\mathbf{Ax} = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T \mathbf{x} + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^T \mathbf{x} + \cdots + \lambda_N \mathbf{q}_N \mathbf{q}_N^T \mathbf{x}.$$

By taking the scalar product of this expansion with \mathbf{x} we obtain

$$\mathbf{x}^T \mathbf{Ax} = \lambda_1 (\mathbf{q}_1^T \mathbf{x})^2 + \lambda_2 (\mathbf{q}_2^T \mathbf{x})^2 + \cdots + \lambda_N (\mathbf{q}_N^T \mathbf{x})^2.$$

It is thereby clear that:

- if $\lambda_i \geq 0$ for every $i = 1, \dots, N$ then \mathbf{A} is nonnegative definite;
- if $\lambda_i > 0$ for every $i = 1, \dots, N$ then \mathbf{A} is positive definite.

This proves the (\Leftarrow) directions of the characterizations. □

Eigenpairs and Diagonalization

Fact 5. Expansion (3.9) shows that if \mathbf{A} is a real symmetric $N \times N$ matrix then it can be expressed in the factored form

$$\mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}^T, \quad (3.11)$$

where \mathbf{D} and \mathbf{Q} are the real $N \times N$ matrices constructed from the real eigenpairs (3.5) as

$$\mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_N \end{pmatrix}, \quad \mathbf{Q} = (\mathbf{q}_1 \quad \mathbf{q}_2 \quad \cdots \quad \mathbf{q}_N). \quad (3.12)$$

This factorization is called a **diagonalization** of \mathbf{A} because \mathbf{D} is a **diagonal**.

Eigenpairs and Diagonalization

The matrix \mathbf{Q} constructed from the eigenvectors $\{\mathbf{q}_i\}_{i=1}^N$ by (3.12) satisfies the relations

$$\mathbf{Q}^T \mathbf{Q} = \mathbf{I} = \mathbf{Q} \mathbf{Q}^T. \quad (3.13)$$

- $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$ is a recasting of the orthonormality conditions (3.6).
- $\mathbf{I} = \mathbf{Q} \mathbf{Q}^T$ is equivalent to $\mathbf{x} = \mathbf{Q} \mathbf{Q}^T \mathbf{x}$, which is a recasting of expansion (3.7).
- Relations (3.13) show that \mathbf{Q} and \mathbf{Q}^T are inverses of each other — i.e. that $\mathbf{Q}^{-1} = \mathbf{Q}^T$ and that $\mathbf{Q}^{-T} = \mathbf{Q}$.
- They imply that the factored form (3.11) is equivalent to $\mathbf{A} \mathbf{Q} = \mathbf{Q} \mathbf{D}$.

Any matrix \mathbf{Q} satisfying relations (3.13) is called an **orthogonal matrix**.

Eigenpairs and Diagonalization

The eigenpairs of a matrix \mathbf{A} can be approximated numerically with the Matlab command “eig”. Specifically, if the $N \times N$ matrix \mathbf{A} is stored as A then the command line

$$\gt [Q,D] = \text{eig}(A);$$

will return a diagonal matrix D of eigenvalues and an orthogonal matrix Q whose columns are corresponding eigenvectors — i.e. so that $AQ = QD$.

Remark. The Matlab command “eig” computes eigenpairs of positive definite matrices using the QR algorithm. This algorithm is often covered in graduate courses on numerical methods. It will not be presented here. However, you should know that errors can arise in the smaller eigenvalues when the ratio λ_1/λ_N is large. This ratio is the *condition number* of \mathbf{A} .

Statistical Interpretation

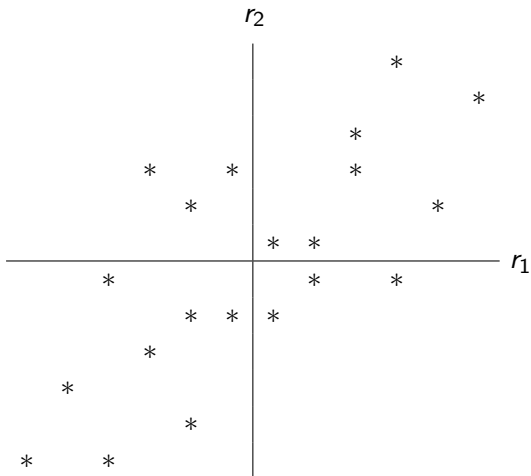
In order to illustrate the statistical information contained in the return mean vector \mathbf{m} and return covariance matrix \mathbf{V} , we consider the case $N = 2$ and $D = 21$. Given the return history $\{(r_1(d), r_2(d))\}_{d=1}^{21}$, the vector \mathbf{m} and matrix \mathbf{V} computed with uniform weights are

$$\mathbf{m} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = \frac{1}{21} \sum_{d=1}^{21} \begin{pmatrix} r_1(d) \\ r_2(d) \end{pmatrix},$$
$$\mathbf{V} = \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix} = \frac{1}{21} \sum_{d=1}^{21} \begin{pmatrix} \tilde{r}_1(d)^2 & \tilde{r}_1(d)\tilde{r}_2(d) \\ \tilde{r}_2(d)\tilde{r}_1(d) & \tilde{r}_2(d)^2 \end{pmatrix},$$

where $\tilde{r}_1(d) = r_1(d) - m_1$ and $\tilde{r}_2(d) = r_2(d) - m_2$.

Suppose that when the return history $\{(r_1(d), r_2(d))\}_{d=1}^{21}$ is plotted as points in the $r_1 r_2$ -plane we obtain the plot on the next slide.

Statistical Interpretation



Statistical Interpretation

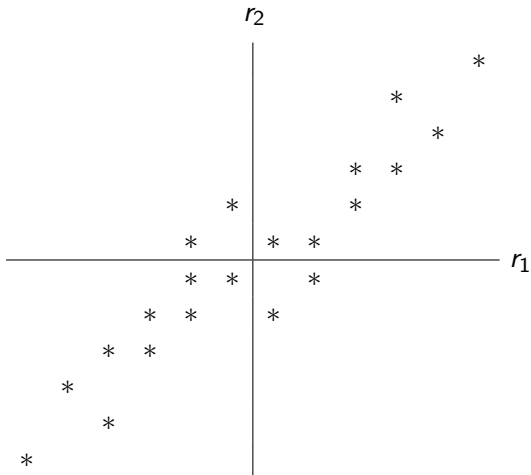
This so-called *scatter plot* shows a distribution of points clustered about the origin in a way that favors the first and third quadrants.

- The vector \mathbf{m} gives the center of the cluster. It lies in the third quadrant close to the origin.
- The matrix \mathbf{V} will have eigenvectors that are roughly parallel to \nearrow and to \nwarrow . The eigenvalue associated with \nearrow will be greater than the one associated with \nwarrow .

This is how \mathbf{m} and \mathbf{V} tell us that the points are clustered about the origin in a way that favors the first and third quadrants.

Now suppose that when the return history $\{(r_1(d), r_2(d))\}_{d=1}^{21}$ is plotted as points in the $r_1 r_2$ -plane we instead obtain the scatter plot on the next slide.

Statistical Interpretation



Statistical Interpretation

In this scatter plot r_1 and r_2 are more highly correlated than in the first.

- The vector \mathbf{m} is almost the same as it was for the first scatter plot. It lies in the third quadrant close to the origin.
- The matrix \mathbf{V} again has eigenvectors that are roughly parallel to ↗ and to ↖. However now the eigenvalue associated with ↗ is *very much greater* than the one associated with ↖.

Both the scatter plot and the analysis of \mathbf{m} and \mathbf{V} suggest that the points $\{(r_1(d), r_2(d))\}_{d=1}^{21}$ cluster along a line.

Statistical Interpretation

If eigenvectors associated with the largest eigenvalue of \mathbf{V} are proportional to $(1, b)$ then the points in the $r_1 r_2$ -plane are clustered along the line

$$r_2 - m_2 = b(r_1 - m_1).$$

This suggests $r_2(d)$ could be modeled as

$$r_2(d) - m_2 = b(r_1(d) - m_1) + z(d),$$

where $z(d)$ are small random numbers that on average sum to zero.

Remark. Scatter plots become harder to visualize as N grows beyond 3. However the eigenpair analysis of \mathbf{V} can be carried out easily for much larger N .

Principle Component Analysis

In statistics the eigenpair analysis of the covariance matrix \mathbf{V} is called *Principle Component Analysis (PCA)*.

A principle component analysis of \mathbf{V} yields N eigenpairs

$$(\lambda_1, \mathbf{q}_1), (\lambda_2, \mathbf{q}_2), \dots, (\lambda_N, \mathbf{q}_N). \quad (5.14)$$

The eigenvalues will almost always be distinct, in which case we will order them as

$$\lambda_1 > \lambda_2 > \dots > \lambda_N > 0. \quad (5.15)$$

In this case the eigenvectors will be unique up to a nonzero factor. If they are normalized so that $\|\mathbf{q}_i\| = 1$ then they are unique up to a factor of ± 1 and $\{\mathbf{q}_i\}_{i=1}^N$ will be an orthonormal basis of \mathbb{R}^N .

Principle Component Analysis

Let \mathbf{D} and \mathbf{Q} be the diagonal and orthogonal matrices constructed from the eigenpairs (5.14) as in (3.12). Then $\mathbf{V} = \mathbf{QDQ}^T$ and $\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}$.

Then the underlying return history $\{\mathbf{r}(d)\}_{d=1}^D$ can be transformed into the history $\{\mathbf{p}(d)\}_{d=1}^D$ where $\mathbf{p}(d) = \mathbf{Q}^T\mathbf{r}(d)$. The entries of $\mathbf{p}(d)$ are called the *principle components* of $\mathbf{r}(d)$. Their mean vector is given by

$$\sum_{d=1}^D w(d)\mathbf{p}(d) = \sum_{d=1}^D w(d)\mathbf{Q}^T\mathbf{r}(d) = \mathbf{Q}^T \left(\sum_{d=1}^D w(d)\mathbf{r}(d) \right) = \mathbf{Q}^T\mathbf{m}.$$

Principle Component Analysis

Similarly, their covariance matrix is given by

$$\begin{aligned} & \sum_{d=1}^D w(d) (\mathbf{p}(d) - \mathbf{Q}^T \mathbf{m}) (\mathbf{p}(d) - \mathbf{Q}^T \mathbf{m})^T \\ &= \mathbf{Q}^T \left(\sum_{d=1}^D w(d) (\mathbf{r}(d) - \mathbf{m}) (\mathbf{r}(d) - \mathbf{m})^T \right) \mathbf{Q} \\ &= \mathbf{Q}^T \mathbf{V} \mathbf{Q} = \mathbf{Q}^T (\mathbf{Q} \mathbf{D} \mathbf{Q}^T) \mathbf{Q} = (\mathbf{Q}^T \mathbf{Q}) \mathbf{D} (\mathbf{Q}^T \mathbf{Q}) = \mathbf{D}. \end{aligned}$$

Because \mathbf{D} is a diagonal matrix, the covariance of distinct entries of $\mathbf{p}(d)$ vanishes. Because the i^{th} entry of $\mathbf{p}(d)$ is $\mathbf{q}_i^T \mathbf{r}(d)$, its variance is λ_i .

Principle Component Analysis

Therefore PCA can be viewed as an orthogonal coordinate transformation that maps the data into new coordinates (the principle components) that are uncorrelated and such that the first entry has the largest variance, the second entry has the second largest variance, and so on.

Remark. The vectors \mathbf{q}_i are called the *principle component coefficients* because they are the vectors whose scalar product with the data $\mathbf{r}(d)$ gives the principle components. They are also called *loadings*.

Remark. Principle component analysis is related to a singular value decomposition of the $N \times D$ matrix

$$\mathbf{R} = \begin{pmatrix} \mathbf{r}(1) - \mathbf{m} & \mathbf{r}(2) - \mathbf{m} & \cdots & \mathbf{r}(D) - \mathbf{m} \end{pmatrix}.$$

This is because $\mathbf{V} = \mathbf{R}\mathbf{W}\mathbf{R}^T$ where \mathbf{W} is the $D \times D$ diagonal matrix with the weights $w(d)$ on the diagonal.

Dimension of the Data

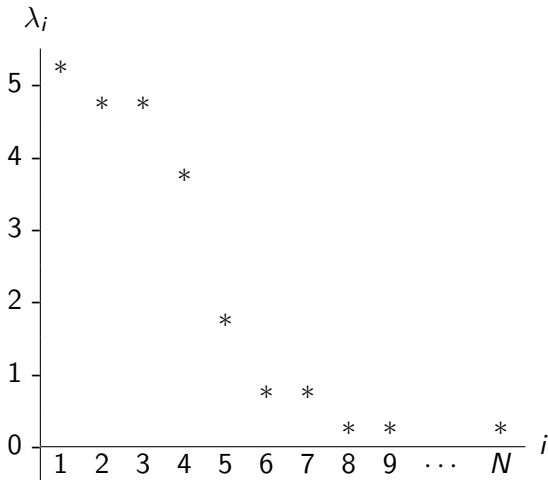
One application of PCA is to explore possible lower dimensional models that capture the bulk of the variation in the data. The dimension of such a model can be found by selecting a subset of the larger eigenvalues of \mathbf{V} . There are many ways in which this can be done.

For example, suppose that the eigenvalues of \mathbf{V} are ordered as

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N-1} \geq \lambda_N > 0,$$

and that a plot of λ_i versus i looks like the figure on the next slide.

Dimension of the Data



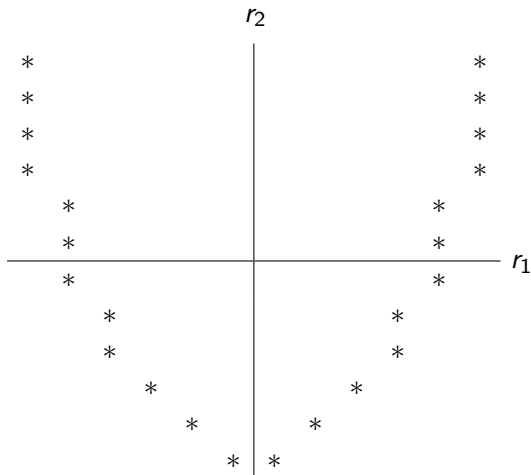
Dimension of the Data

This figure shows that the underlying data has four major dimensions and that eigenvalues are negligible for $i \geq 8$. It suggests that the data might be captured well with a 4, 5, or 7 dimensional model.

- A 4 dimensional model is suggested by the fact that the gap $\lambda_4 - \lambda_5$ is larger than the gap between any other pair of adjacent eigenvalues.
- A 7 dimension model is suggested by the fact that the ratio λ_7/λ_8 is larger than the ratio between any other pair of adjacent eigenvalues.
- A 5 or 7 dimension model might be suggested by the fact that the ratio λ_6/λ_1 is below .2 or the fact that the ratio λ_8/λ_1 is below .1. The threshold chosen is arbitrary, but should be small.

Remark. Dimensions obtained by these methods may not be the actual dimension of the data, which can be much lower when the data satisfies an approximate nonlinear relationship. Such a relationship is illustrated on the next slide for two dimensional data.

Dimension of the Data



Dimension of the Data

This figure shows that the underlying data lies along a parabola-like curve, whereby it is one dimensional. However, principle component analysis does not see this because the 2×2 matrix \mathbf{V} has two comparable eigenvalues.

Remark. This use of PCA is called *exploratory analysis*. While it can show that data behaves like its dimension is lower than N , it typically does not give a unique value for the lower dimension. However, it usually does give a small set of candidates for the lower dimension that can be evaluated by other criteria.