Linear Algebra & Properties of the Covariance Matrix

October 3rd, 2012
Let $r_{n}^{1}, r_{n}^{t}, \ldots, r_{n}^{T}$ be the historical return rates on the $n^{th}$ asset.

$$r_{n} = \begin{pmatrix} r_{n}^{1} \\ r_{n}^{2} \\ \vdots \\ r_{n}^{T} \end{pmatrix} \quad n = 1, 2, \ldots, N.$$
Estimation of $\bar{r}$ and $C$

The expected return is approximated by

$$\hat{r}_n = \frac{1}{T} \sum_{t=1}^{T} r_n^t$$

and the covariance is approximated by

$$\hat{c}_{mn} = \frac{1}{T - 1} \sum_{t=1}^{T} (r_n^t - \hat{r}_n)(r_m^t - \hat{r}_m),$$

or in matrix/vector notation

$$\hat{C} = \frac{1}{T - 1} \sum_{t=1}^{T} (r^t - \hat{r})(r^t - \hat{r})^\top$$

an outer product.
The expected return $\bar{r}$ is often estimated exogenously (e.g. not statistically estimated). But

- statistical estimation of $C$ is an important topic,
- is used in practice,
- and is the backbone of many finance strategies.

A major part of Markowitz theory was the assumption for $C$ to be a covariance matrix it must be symmetric positive definite (SPD).
Let’s focus on real matrices, and only use complex numbers if necessary. A vector $x \in \mathbb{R}^N$ and a matrix $A \in \mathbb{R}^{M \times N}$ are

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{pmatrix}$$
Matrix/Vector Multiplication

The matrix $A$ times a vector $x$ is a new vector

$$Ax = \begin{pmatrix}
a_{11} & a_{12} & \ldots & a_{1N} \\
a_{21} & a_{22} & \ldots & a_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
a_{M1} & a_{M2} & \ldots & a_{MN}
\end{pmatrix} \begin{pmatrix}
x_1 \\
x_2 \\
\vdots \\
x_N
\end{pmatrix}$$

$$= \begin{pmatrix}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1N}x_N \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2N}x_N \\
\vdots \\
a_{M1}x_1 + a_{M2}x_2 + \cdots + a_{MN}x_N
\end{pmatrix} \in \mathbb{R}^M .$$

Matrix/matrix multiplication is an extension of this.
Their transposes are

\[ x^\top = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}^\top = (x_1, x_2, \ldots, x_N) \]

and

\[ A^\top = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1N} \\ a_{21} & a_{22} & \ldots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M2} & \ldots & a_{MN} \end{pmatrix}^\top = \begin{pmatrix} a_{11} & a_{21} & \ldots & a_{M1} \\ a_{12} & a_{22} & \ldots & a_{M2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1N} & a_{2N} & \ldots & a_{MN} \end{pmatrix}. \]

Note that \((Ax)^\top = x^\top A^\top\).
The vector $x \in \mathbb{R}^N$ inner product with another vector $y \in \mathbb{R}^N$ is

$$x^\top y = y^\top x = x_1y_1 + x_2y_2 + \ldots + x_Ny_N.$$ 

There is also the outer product,

$$xy^\top = \begin{pmatrix} x_1y_1 & x_1y_2 & \ldots & x_1y_N \\ x_2y_1 & x_2y_2 & \ldots & x_2y_N \\ \vdots & \vdots & \ddots & \vdots \\ x_Ny_1 & x_Ny_2 & \ldots & x_Ny_N \end{pmatrix} \neq yx^\top.$$
The norm on the vector space $\mathbb{R}^N$ is $\| \cdot \|$, and for any $x \in \mathbb{R}^N$ we have

$$\|x\| = \sqrt{x^\top x}.$$ 

Properties of the norm

- $\|x\| \geq 0$ for any $x$,
- $\|x\| = 0$ iff $x = 0$,
- $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality),
- $|x^\top y| \leq \|x\| \|y\|$ with equality iff $x = ay$ for some $a \in \mathbb{R}$ (Cauchy-Schwartz),
- $\|x + y\|^2 = \|x\|^2 + \|y\|^2$ if $x^\top y = 0$ (Pythagorean thm).
Two vectors $x, y$ are linear independent if there is no scalar constant $a \in \mathbb{R}$ s.t.

$$y = ax.$$ 

These two vectors are orthogonal if

$$x^\top y = 0.$$ 

Clearly, orthogonality $\Rightarrow$ linear independent.
A set of vectors \( x_1, x_2, \ldots, x_n \) is said to span a subspace \( V \subset \mathbb{R}^N \) if for any \( y \in V \) there are constants \( a_1, a_2, \ldots, a_n \) such

\[
y = a_1 x_2 + a_2 x_2 + \cdots + a_n x_n.
\]

We write, span\((x_1, \ldots, x_n) = V\). In particular, a set \( x_1, x_2, \ldots, x_N \) will span \( \mathbb{R}^N \) iff they are linearly independent:

\[
\text{span}(x_1, x_2, \ldots, x_N) = \mathbb{R}^N \iff x_i \text{ linearly independent of } x_j \text{ for all } i, j.
\]

If so, then \( x_1, x_2, \ldots, x_N \) are a basis for \( \mathbb{R}^N \).
Orthogonal Basis

A basis of $\mathbb{R}^N$ is orthogonal if all its elements are orthogonal,

$$\text{span}(x_1, x_2, \ldots, x_N) = \mathbb{R}^N \quad \text{and} \quad x_i^\top x_j = 0 \ \forall i, j.$$
Invertibility

### Definition

A square matrix \( A \in \mathbb{R}^{N \times N} \) is invertible if for any \( b \in \mathbb{R}^N \) there exists \( x \) s.t.

\[ Ax = b. \]

If so, then \( x = A^{-1}b \).

\( A \) is invertible if any of the following hold:

- \( A \) has rows that are linearly independent
- \( A \) has columns that are linearly independent
- \( \det(A) \neq 0 \)
- there is no non-zero vector \( x \) s.t. \( Ax = 0 \).

In fact, these 4 statements are equivalent.
Let $A$ be an $N \times N$ matrix. A scalar $\lambda \in \mathbb{C}$ is an eigenvalue of $A$ if

$$A x = \lambda x$$

for some vector $x = re(x) + \sqrt{-1} \ im(x) \in \mathbb{C}^N$.

If so $x$ is an eigenvector.

By 4th statement of previous slide, $A^{-1}$ exists $\iff$ zero is not an eigenvalue.
Eigenvalue Diagonalization

Let $A$ be an $N \times N$ matrix with $N$ linearly independent eigenvectors. There is an $N \times N$ matrix $X$ and an $N \times N$ diagonal matrix $\Lambda$ such that

$$A = X\Lambda X^{-1},$$

where $\Lambda_{ii}$ is an eigenvalue of $A$, and the columns of $X = [x_1, x_2, \ldots, x_N]$ are eigenvectors,

$$Ax_i = \Lambda_{ii}x_i.$$
A matrix is symmetric if $A^\top = A$.

**Proposition**

*A symmetric matrix has real eigenvalues.*

**pf:** If $Ax = \lambda x$ then
\[
\lambda^2 \|x\|^2 = \lambda^2 x^* x = x^* A^2 x = x^* A^\top A x = \|Ax\|^2,
\]
which is real and nonnegative ($x^* = \text{re}(x)^\top - \sqrt{-1} \text{im}(x)^\top$ i.e the conjugate transpose). Hence, $\lambda$ is cannot be imaginary or complex. ■
Proposition

Let $A = A^\top$ be a real matrix. $A$’s eigenvectors form an orthogonal basis of $\mathbb{R}^N$, and the diagonalization is simplified:

$$A = X\Lambda X^\top,$$

i.e. $X^{-1} = X^\top$.

**Basic Idea of Proof:** Suppose that $Ax = \lambda x$ and $Ay = \alpha y$ with $\alpha \neq \lambda$. Then

$$\lambda x^\top y = x^\top A^\top y = x^\top Ay = \alpha x^\top y,$$

and so it must be that $x^\top y = 0$. 

Real Symmetric Matrices
A matrix is skew symmetric if $A^\top = -A$.

**Proposition**

A skew symmetric matrix has purely imaginary eigenvalues.

**pf:** If $Ax = \lambda x$ then

\[
\lambda^2 \|x\|^2 = \lambda^2 x^*x = x^*A^2x = -x^*A^\top Ax = -\|Ax\|^2
\]

which is less than zero. Hence, $\lambda^2 < 0$ must be purely imaginary. ■
Positive Definiteness

**Definition**

A matrix $A \in \mathbb{R}^{N \times N}$ is positive definite iff

$$x^T Ax > 0 \quad \forall x \in \mathbb{R}^N .$$

It is only positive semi-definite iff

$$x^T Ax \geq 0 \quad \forall x \in \mathbb{R}^N .$$
Positive Definiteness $\Rightarrow$ Symmetric

**Proposition**

If $A \in \mathbb{R}^{N \times N}$ is positive definite, then $A^\top = A$, i.e. it is symmetric positive definite (SPD).

**pf:** For any $x \in \mathbb{R}^N$ we have

$$0 < x^\top Ax = \frac{1}{2} \left( x^\top (A + A^\top) x + x^\top (A - A^\top) x \right),$$

but if $A - A^\top \neq 0$ then there exists eigenvector $x$ s.t.

$(A - A^\top) x = \alpha x$ where $\alpha \in \mathbb{C} \setminus \mathbb{R}$, which contradicts the inequality. ■
Eigenvalue of an SPD Matrix

Proposition

If A is SPD, then it has all positive eigenvalues.

pf: By definition for any \( y \in \mathbb{R}^N \setminus \{0\} \) it holds that

\[
y^\top A y > 0.
\]

In particular, for any \( \lambda \) and \( x \) s.t. \( Ax = \lambda x \), we have

\[
\lambda \|x\|^2 = x^\top A x > 0
\]

and so \( \lambda > 0 \). ■

Consequence: for a SPD there’s no \( x \) s.t. \( Ax = 0 \), and so \( A^{-1} \) exists.
The invertibility of the covariance matrix is very important:

- we’ve seen a need for it in the efficient frontiers and Markowitz problem,
- if it’s not invertible then there are redundant assets,
- we’ll need it to be invertible when we discuss multivariate Gaussian distributions.

If there is no risk-free, our covariance matrix is practically invertible by definition:

\[
0 < \text{var} \left( \sum_i w_i r_i \right) = \sum_{i,j} w_i w_j C_{ij} = w^\top C w
\]

where \( w \in \mathbb{R}^N \setminus \{0\} \) is any allocation vector (\( w \) may not sum to 1). Hence, \( C \) is SPD \( \Rightarrow C^{-1} \text{ exists.} \)
Recall, \( \hat{C} = \frac{1}{T-1} \sum_{t=1}^{T} (r^t - \bar{r})(r^t - \bar{r})^\top \). Each summand is symmetric, 
\[ (r^t - \hat{r})(r^t - \hat{r})^\top \] is symmetric, and sum of symmetric matrices is symmetric. But each summand is not invertible because there exists a vector \( y \) s.t. 
\[ y^\top (r^t - \hat{r}) = 0 \], which means that zero is an eigenvalue, 
\[ (r^t - \hat{r})(r^t - \hat{r})^\top y = 0 \].
So How Do We Know $\hat{C}$ is Covariance Matrix?

- If
  \[ \text{span}(r^1 - \hat{r}, r^2 - \hat{r}, \ldots, r^T - \hat{r}) = \mathbb{R}^N \]
  then $\hat{C}$ is invertible. So need $T \geq N$.
- In order for $\hat{C}$ to be close $C$ will need $T \gg N$.
- $\hat{C}$ is also SPD:
  \[
y^\top \hat{C} y = y^\top \left( \frac{1}{T-1} \sum_{t=1}^{T} (r^t - \hat{r})(r^t - \hat{r})^\top \right) y
  \]
  \[
  = \frac{1}{T-1} \sum_{t=1}^{T} y^\top (r^t - \hat{r})(r^t - \hat{r})^\top y
  = \frac{1}{T-1} \sum_{t=1}^{T} (y^\top (r^t - \hat{r}))^2 > 0, 
  \]
  because at least 1 $t$ s.t. $y^\top (r^t - \hat{r}) \neq 0$. 

Linear Algebra & Properties of the Covariance Matrix
A Simple Covariance Matrix

\[ C = (1 - \rho)I + \rho U \text{ where } I \text{ is the identity and } U_{ij} = 1 \text{ for all } i, j. \]

\[ U \mathbb{1} = N \mathbb{1} \]
\[ Ux = 0 \quad \text{if } x^\top \mathbb{1} = 0 \]

where \( \mathbb{1} = (1, 1, \ldots, 1)^\top \). Hence, for any \( x \) we have

\[ Cx = (1 - \rho)Ix + \rho Ux = (1 - \rho)x + \rho(\mathbb{1}^\top x)\mathbb{1}, \]

which means \( x \) is an eigenvector iff \( \mathbb{1}^\top x = 0 \) or \( x = a\mathbb{1} \). The eigenvalues of \( C \) are \( 1 - \rho + N\rho \) and \( 1 - \rho \), and the eigenvectors are \( \mathbb{1} \) and the \( N - 1 \) vectors orthogonal to \( \mathbb{1} \), respectively.

Linear Algebra & Properties of the Covariance Matrix
Positive Definite (Complex Hermitian Matrices)

Let $C$ be a square matrix, i.e. $C \in \mathbb{C}^{N \times N}$.

- $C$ is positive definite if
  
  $$z^* C z > 0 \quad \forall z \in \mathbb{C}^N \setminus \{0\}$$

  where $\mathbb{C}^N$ is $N$-dimensional complex vectors, and $z^*$ is conjugate transpose,

  $$z^* = (z_r + iz_i)^* = z_r - iz_i.$$  

- $C$ is positive semi-definite if
  
  $$z^* C z \geq 0 \quad \forall z \in \mathbb{C}^N \setminus \{0\}$$

- If $C$ positive definite and real, then it must also be symmetric i.e. $C_{nm} = C_{mn} \quad \forall m, n$. 

Linear Algebra & Properties of the Covariance Matrix
Gershgorin Circles

Proposition

All eigenvalues of the matrix $A$ lie in one of the Gershgorin circles,

$$| \lambda - A_{ii} | \leq \sum_{j \neq i} |A_{ij}| \quad \text{for at least } 1 \ i \leq N,$$

where $\lambda$ is an eigenvalue of $A$.

pf: Let $\lambda$ be an eigenvalue and let $x$ be an eigenvector. Choose $i$ so that $|x_i| = \max_j |x_j|$. We have $\sum_j A_{ij}x_j = \lambda x_j$. W.L.O.G. we can take $|x_i| = 1$ so that $\sum_{j \neq i} A_{ij}x_j = \lambda - A_{ii}$, and taking absolute values yields the result,

$$| \lambda - A_{ii} | = \left| \sum_{j \neq i} A_{ij}x_j \right| \leq \sum_{j \neq i} |A_{ij}| \cdot |x_j| \leq \sum_{j \neq i} |A_{ij}| \quad \blacksquare$$

Linear Algebra & Properties of the Covariance Matrix