Week 2. Eigenvalues/eigenvectors and random vectors

Section 10.3

A nonzero vector $\mathbf{p}$ is called an eigenvector (or a characteristic vector) of a symmetric $m \times m$ matrix $\mathbf{A}$ if

$$\mathbf{Ap} = \lambda \mathbf{p}$$

for some scalar λ, which is called the eigenvalue (or the characteristic root) and $\|\mathbf{p}\| = 1$.

Characteristic equation (how to find all eigenvalues)

$$|\mathbf{A} - \lambda \mathbf{I}| = 0.$$

is a polynomial of the mth order. By the fundamental theorem of algebra there are m roots of the polynomial, i.e. m eigenvalues. For a general polynomial there may be complex roots but for a symmetric matrix $\mathbf{A}$ (this is what we assume) all roots/eigenvalues are real. Some eigenvalues may be the same as the roots of the polynomial may repeat, e.g. $(\lambda - 1)^2(\lambda - 2) = 0$ implies three roots, $\lambda_{1,2} = 1$, $\lambda_3 = 2$.

For each eigenvalue λ there is a corresponding eigenvector $\mathbf{p}$. Eigenvectors corresponding to different eigenvalues are orthogonal.

Since eigenvectors have unit length we have $\lambda = \mathbf{p}'\mathbf{Ap}$. Matrix is nonsingular if and only if none of its eigenvalues is zero.

Important to remember:

$$\begin{aligned}
\lambda_{\max} &= \max_{\mathbf{x}} \frac{\mathbf{x}'\mathbf{Ax}}{\|\mathbf{x}\|^2} = \max_{\|\mathbf{x}\|=1} \mathbf{x}'\mathbf{Ax}, \\
\lambda_{\min} &= \min_{\mathbf{x}} \frac{\mathbf{x}'\mathbf{Ax}}{\|\mathbf{x}\|^2} = \min_{\|\mathbf{x}\|=1} \mathbf{x}'\mathbf{Ax}.
\end{aligned}$$

Corollary 1 *Symmetric matrix* $\mathbf{A}$ *is positive definite if and only if*

$$\mathbf{x}'\mathbf{Ax} > 0, \quad \forall \mathbf{x} \neq \mathbf{0}$$

Example 2 *Find eigenvalues and eigenvectors of a* 2×2 *symmetric matrix*

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

Solution. The eigenvalues are the solution to the quadratic equation for λ,

$$\begin{vmatrix} a - \lambda & b \\ b & c - \lambda \end{vmatrix} = (a - \lambda)(c - \lambda) - b^2 = 0$$

with the solution

$$\lambda_{1,2} = \frac{1}{2}\left(a + c \pm \sqrt{D}\right)$$

where $D = (a - c)^2 + 4b^2$, the discriminant. If $a = c$ and $b = 0$ we have $\lambda_1 = \lambda_2$. The two eigenvectors have unit length and orthogonal. For vectors in the plane, it is convenient to express a pair of orthonormal vectors through angle θ as follows:

$$\mathbf{p}_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \quad \mathbf{p}_2 = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix},$$

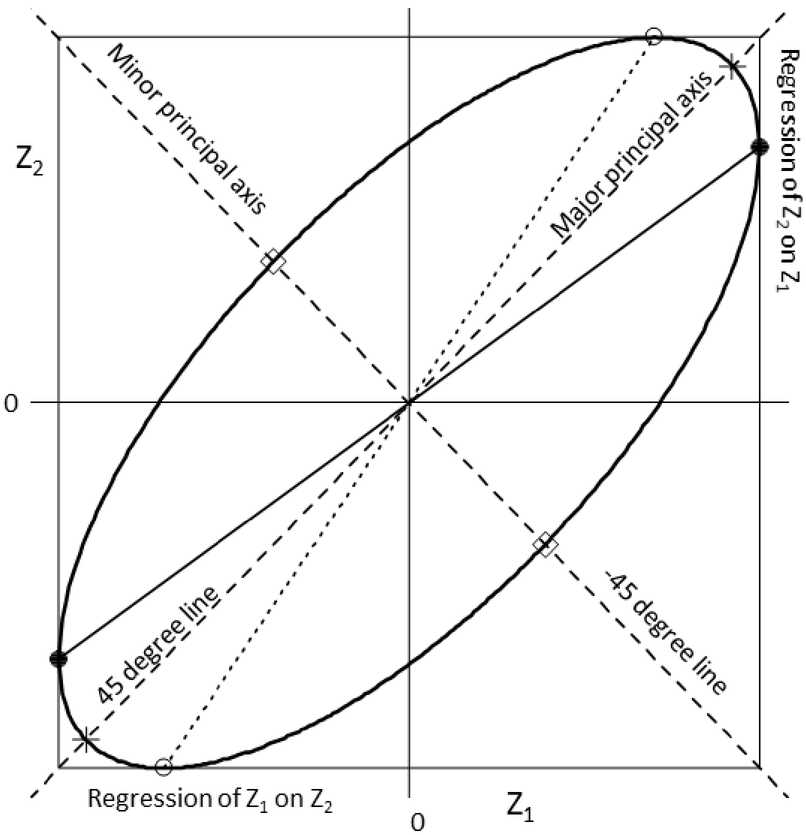
where the angle θ to be found from the vector equation $\mathbf{A}\mathbf{p}_1 = \lambda_1 \mathbf{p}_1$. The first component of this equation takes the form $a\cos\theta + b\sin\theta = \lambda_1\cos\theta$ which leads to the solution

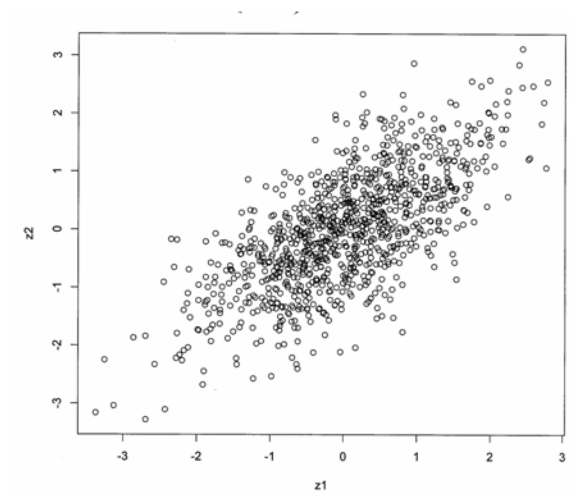
$$\theta = \arctan\frac{\lambda_1 - a}{b}.$$

$\square$

Section 3.5
Ellipse is the set of points (z_1, z_2) for which (3.29)

$$z_1^2 - 2\rho z_1 z + z_2^2 = \text{const.}$$





$$\text{z1} = \text{rnorm}(1000)$$
$$\text{z2} = \text{z1} + \text{rnorm}(1000)$$
$$\text{z2} = \text{z2}/\text{sd}(\text{z2})$$

The trace and determinant of a symmetric matrix can be expressed via eigenvalues as follows.

Theorem 3 *If* $\mathbf{A}$ *is an* $m \times m$ *symmetric matrix with eigenvalues* $\lambda_1, ..., \lambda_m$ *then*

$$tr(\mathbf{A}) = \sum_{i=1}^{m} \lambda_i, \quad |\mathbf{A}| = \prod_{i=1}^{m} \lambda_i.$$

Jordan spectral matrix decomposition

Let $\mathbf{A}$ be a $m \times m$ symmetric matrix with eigenvalues $\lambda_1, \lambda_2, ..., \lambda_m$ and corresponding $m \times 1$ eigenvectors $\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_m$. The Jordan spectral matrix decomposition is given by

$$\mathbf{A} = \sum_{i=1}^{m} \lambda_i \mathbf{p}_i \mathbf{p}_i'.$$

Combine eigenvectors into a $m \times m$ matrix

$$\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_m]$$

and let $\mathbf{\Lambda}$ $=diag(\lambda_1, \lambda_2, ..., \lambda_m)$. Then the Jordan spectral matrix decomposition (or shortly, spectral matrix decomposition) can be expressed in a compact form as

$$\mathbf{A} = \mathbf{P\Lambda P}'. \tag{1}$$

Matrix $\mathbf{P}$ is orthonormal, i.e. vector-columns are pairwise orthogonal and have unit length, or in matrix form

$$\mathbf{P}'\mathbf{P} = \mathbf{P}\mathbf{P}' = \mathbf{I},$$

where $\mathbf{I}$ is the $m \times m$ unit matrix. Multiplying by $\mathbf{P}'$ and $\mathbf{P}$ we arrive at matrix diagonalization $\mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{\Lambda}$. In words, any symmetric matrix can be reduced to a diagonal matrix upon an orthonormal transformation.

Helpful formulas:

$$tr(\mathbf{A}) = tr(\mathbf{P\Lambda P}') = tr(\mathbf{\Lambda P}'\mathbf{P}) = tr(\mathbf{\Lambda}) = \sum_{i=1}^{m} \lambda_i.$$

To prove the formula for the determinant we observe that $|\mathbf{P}| = |\mathbf{P}'| = 1$ for an orthonormal matrix $\mathbf{P}$ because $|\mathbf{P}'\mathbf{P}| = |\mathbf{P}|\,|\mathbf{P}'| = |\mathbf{P}|^2 = 1$. Therefore,

$$|\mathbf{P\Lambda P}'| = |\mathbf{P}|\,|\mathbf{\Lambda}|\,|\mathbf{P}'| = |\mathbf{\Lambda}| = \prod_{i=1}^{m} \lambda_i.$$

The spectral matrix decomposition gives rise to the definition of a matrix function.

Definition 4 ***Function of a matrix.*** *Let* $\mathbf{A}$ *be a symmetric matrix with the spectral decomposition* $\mathbf{A} = \mathbf{P\Lambda P}'$ *and* f *is a scalar function,* $R^1 \to R^1$*. Define matrix function as*

$$f(\mathbf{A}) = \mathbf{P}f(\mathbf{\Lambda})\mathbf{P}',$$

where $f(\mathbf{\Lambda})$ $=diag(f(\lambda_1), f(\lambda_2), ..., f(\lambda_m))$.

For example, we define

$$\mathbf{A}^{1/2} = \mathbf{P\Lambda}^{1/2}\mathbf{P}', \quad \mathbf{A}^{-1} = \mathbf{P\Lambda}^{-1}\mathbf{P}', \quad \mathbf{A}^{-1/2} = \mathbf{P\Lambda}^{-1/2}\mathbf{P}',$$

assuming that all eigenvalues are positive (matrix $\mathbf{A}$ is positive definite).

Corollary 5 *The eigenvalues of the inverse matrix are reciprocal of the original matrix, the eigenvalues of the squared matrix are squared of the original matrix, etc.*

R programming

To compute eigenvalues and eigenvectors in `R` use function `eigen`. This function works for any square matrix (not necessarily symmetric); `eigen$vectors` returns matrix $\mathbf{P}$ and `eigen$values` returns a vector with components $\lambda_1, ..., \lambda_m$. The eigenvalues (and corresponding eigenvectors) are returned in descending order.

In the code below we generate a random 10×10 matrix and check spectral matrix decomposition.

```
A=matrix(runif(10^2),ncol=10,nrow=10)
A=t(A)%*%A # we need a symmetric matrix
eA=eigen(A,symmetric=T)
print(eA)
AtA=A%*%t(A)
eA=eigen(AtA)
print(eA)
print(AtA)
print(eA$vectors%*%diag(eA$values,10,10)%*%t(eA$vectors))
```

The option `symmetric=T` is useful because due to round-off errors matrix $\mathbf{A}'\mathbf{A}$ may become asymmetric and then some eigenvectors may be complex.

Example 6 *(a) Generate a random 5×5 matrix. (b) Generate a symmetric positive definite matrix. (c) Compute the matrix inverse and square root using Jordan decomposition. (d) Check the result by computation.*

Vector and matrix calculus

Section 10.5

Estimation problems in statistics typically are reduced to optimization problems with vectors and/or matrices. For example, to estimate a multivariate statistical model with unknown covariance matrix we have to differentiate the log-likelihood function with respect to the covariance matrix. Moreover, to find the Fisher information matrix or Jacobian we have to differentiate a vector with respect to a vector.

Differentiation of a vector function with respect to a scalar

We have

$$\mathbf{f}(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{bmatrix}^{n\times 1} \quad \text{where } -\infty < x < \infty.$$

This function specifies a curve in R^n. Define the gradient (the vector of 1st derivatives).

$$\frac{d\mathbf{f}}{dx} = \frac{d}{dx}\mathbf{f} = \dot{\mathbf{f}} = \begin{bmatrix} f_1'(x) \\ f_2'(x) \\ \vdots \\ f_n'(x) \end{bmatrix}^{n\times 1} = \begin{bmatrix} \frac{df_1}{dx} \\ \frac{df_2}{dx} \\ \vdots \\ \frac{df_n}{dx} \end{bmatrix}^{n\times 1}.$$

If t is time and

$$\mathbf{f} = \mathbf{f}(t)$$

specifies the trajectory in R^n we interpret $\frac{d\mathbf{f}}{dx}$ as the velocity vector.

Rules of differentiation:

1. Differentiation of a linear combination

$$\frac{d}{dx}(a\mathbf{f}(x) + b\mathbf{g}(x)) = a\frac{d}{dx}\mathbf{f} + b\frac{d}{dx}\mathbf{g}.$$

2. The chain rule: let $H(\mathbf{f}(x))$ be a real-valued function or a vector argument and

$$H(\mathbf{f}) = H(f_1, f_2, ..., f_n)$$

where

$$\mathbf{f} == \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{bmatrix}^{n\times 1}.$$

Then

$$\frac{dH(\mathbf{f}(x))}{dx} = \left(\frac{\partial H}{\partial \mathbf{f}}\right)'\left(\frac{d\mathbf{f}}{dx}\right)$$

Example 7 *Find*

$$\frac{d}{dx}\left\|\mathbf{y}-x\mathbf{b}\right\|^2.$$

We have

$$\frac{d}{dx}\left\|\mathbf{y}-x\mathbf{b}\right\|^2 = -2(\mathbf{y}-x\mathbf{b})'\mathbf{b}$$

Example 8 *An asteroid approaches Earth following the trajectory $x(t) = 2+2t$, $y(t) = -2-t$, $z(t) = 1+t$ where $t \geq 0$ is time. Assuming that Earth is at location $(3, -1, 5)$, (a) does the asteroid approach or moves away from Earth at the time of the discovery, (b) at what time the distance is minimum? (c) Will it hit Earth?*

Solution.

$$\mathbf{f}(t) = \begin{bmatrix} 2+2t \\ -2-t \\ 1+t \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}$$

$$H(t) = \left\|\mathbf{y} - \mathbf{f}(t)\right\|^2.$$

(a) By the chain rule

$$\frac{d}{dt}H(t) = -2(\mathbf{y}-\mathbf{f}(t))'\frac{df}{dt} = -2(\mathbf{y}-\mathbf{f}(t))'\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = -2\begin{bmatrix} 3-(2+2t) \\ -1-(-2-t) \\ 5-(1+t) \end{bmatrix}'\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

At time 0 we have

$$\frac{d}{dt}H(0) = -2\begin{bmatrix} 3-2 \\ -1+2 \\ 5-1 \end{bmatrix}'\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = -2\begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}'\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = -10$$

Example 9 *Differentiation of a quadratic form*

$$\frac{\partial\left(\mathbf{x}'\mathbf{A}\mathbf{x}\right)}{\partial\mathbf{x}} = 2\mathbf{A}\mathbf{x}.$$

Example 10 *Calculus of OLS.*

$$\frac{\partial}{\partial\boldsymbol{\beta}}\left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|^2 = -2\mathbf{X}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}).$$

Homework 2

1. (10 points). Generate z1 and z2 as above and graph the scatter plot. Compute and show the regression lines z1 on z2, z2 on z1, and the major principle axis. Print out the four slopes and explain the results. [Hint: use `par(mfrow=c(1,2))` and plot z1 versus z2 and z2 versus z1 with respective regressions and the major principle axis.]

2. (8 points). Prove that $\mathbf{A}^{1/2}$, $\mathbf{A}^{-1}$, and $\mathbf{A}^{-1/2}$ derived through the matrix function meet there definitions.

3. (5 points). Prove that $\partial\left\|\mathbf{x}\right\|/\partial\mathbf{x} = \mathbf{x}/\left\|\mathbf{x}\right\|.$

Solutions

1. See the R code and the graph below.

```
> hw123()
[1] 0.6759418 1.0757609
[1] 0.6759418 0.9295746
```

Explanation:

1. The regression slopes in the two plots are the same in z2 on z1 and z1 on z2 because the variance of z1 and z2 is the same = 1. Therefore the slopes

$$\frac{cov(z1,z2)}{var(z1)} = \frac{cov(z2,z1)}{var(z2)}$$

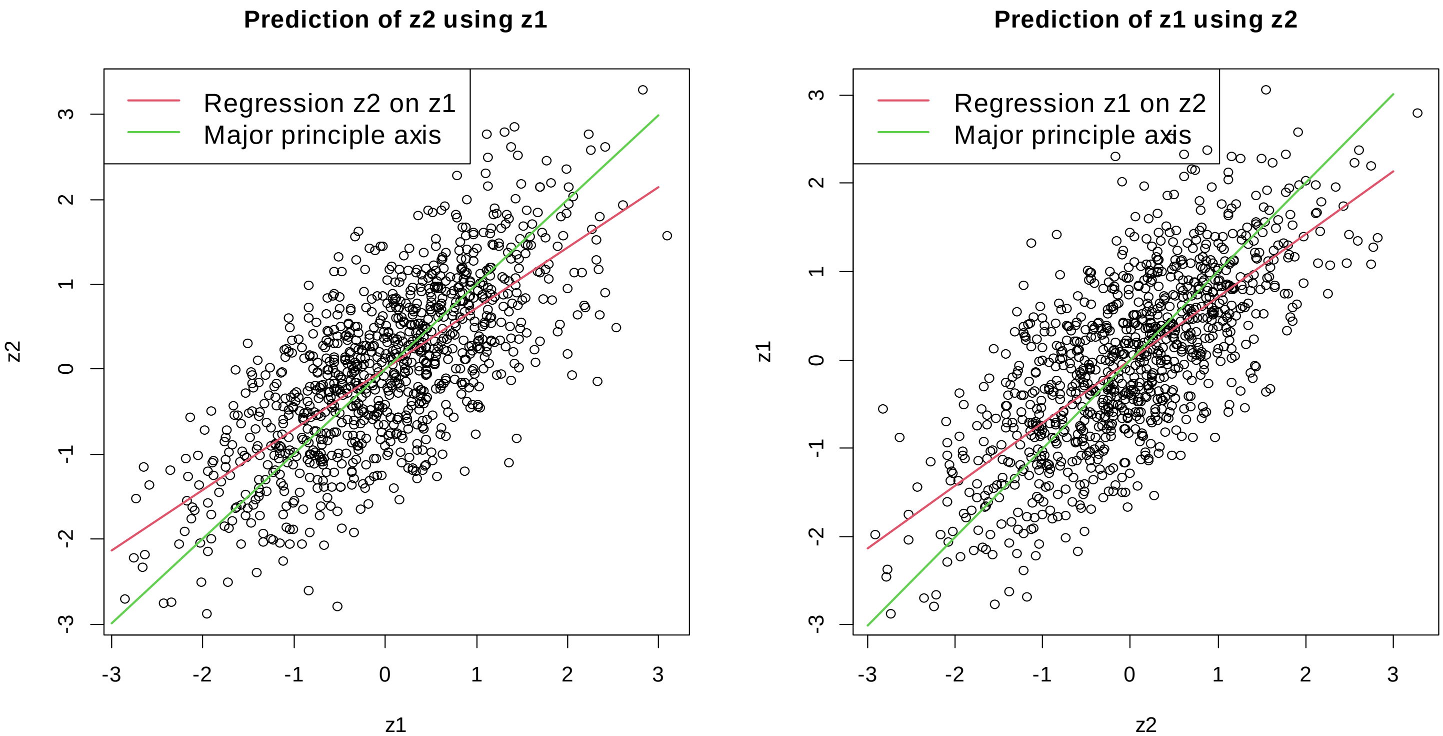
are the same.

2. The major principle axes in the two plots are the same slope because rearranging the vector-column does not change the tangent of the angles. The difference is due to a finite number of simulations - when nSim=100000 the major principal axes are almost the same.

```
hw123=function(problem=1)
{
dump("hw123","c:\\M7021\\hw123.r")
if(problem==1)
{
 z1=rnorm(1000)
 z2=z1+rnorm(1000)
 z2=z2/sd(z2)
 par(mfrow=c(1,2))
 plot(z1,z2,main="Prediction of z2 using z1")
 z1=z1-mean(z1);z2=z2-mean(z2)

 slope21=sum(z1*z2)/sum(z1^2)
 x=c(-3,3)
 lines(x,slope21*x,col=2,lwd=2)
 Z=cbind(z2,z1)
 eg=eigen(t(Z)%*%Z,sym=T)
 eg.max=eg$vectors[,1]
 lines(x,eg.max[2]/eg.max[1]*x,col=3,lwd=2)
 legend("topleft",c("Regression z2 on z1","Major principle axis"),col=2:3,lwd=2,cex=1.25)

 plot(z2,z1,main="Prediction of z1 using z2")
 slope12=sum(z1*z2)/sum(z1^2)
 x=c(-3,3)
 lines(x,slope12*x,col=2,lwd=2)
 Z=cbind(z1,z2)
 eg=eigen(t(Z)%*%Z,sym=T)
 eg.max=eg$vectors[,1]
 lines(x,eg.max[2]/eg.max[1]*x,col=3,lwd=2)
 legend("topleft",c("Regression z1 on z2","Major principle axis"),col=2:3,lwd=2,cex=1.25)
}
}
```



2. By definition, $\mathbf{A}^{1/2}$ is such that $\mathbf{A}^{1/2} \times \mathbf{A}^{1/2} = \mathbf{A}$. Indeed for $\mathbf{A}^{1/2} = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}'$ we have

$$\mathbf{A}^{1/2} \times \mathbf{A}^{1/2} = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}' \times \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}' = \mathbf{P}\mathbf{\Lambda}^{1/2} \times \mathbf{\Lambda}^{1/2}\mathbf{P}' = \mathbf{P}\mathbf{\Lambda}\mathbf{P}' = \mathbf{A}.$$

By definition, $\mathbf{A}^{-1}$ is such that $\mathbf{A}^{-1} \times \mathbf{A} = \mathbf{I}$. Indeed for $\mathbf{A}^{-1} = \mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}'$ we have

$$\mathbf{A}^{-1} \times \mathbf{A} = \mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}' \times \mathbf{P}\mathbf{\Lambda}\mathbf{P}' = \mathbf{P}\mathbf{\Lambda}^{-1} \times \mathbf{\Lambda}\mathbf{P}' = \mathbf{P}\mathbf{P}' = \mathbf{I}.$$

By definition, $\mathbf{A}^{-1/2}$ is such that $\mathbf{A}^{-1/2} \times \mathbf{A}^{-1/2} = \mathbf{A}^{-1}$. Indeed for $\mathbf{A}^{-1/2} = \mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{P}'$ we have

$$\mathbf{A}^{-1/2} \times \mathbf{A}^{-1/2} = \mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{P}' \times \mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{P}' = \mathbf{P}\mathbf{\Lambda}^{-1/2} \times \mathbf{\Lambda}^{-1/2}\mathbf{P}' = \mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}' = \mathbf{A}^{-1}.$$

3. By the chain rule

$$\frac{\partial \|\mathbf{x}\|}{\partial \mathbf{x}} = \frac{\partial \sqrt{\|\mathbf{x}\|^{\mathbf{2}}}}{\partial \mathbf{x}} = \frac{\partial \sqrt{\|\mathbf{x}\|^{\mathbf{2}}}}{\partial \|\mathbf{x}\|^{\mathbf{2}}} \times \frac{\partial \|\mathbf{x}\|^{\mathbf{2}}}{\partial \mathbf{x}} = \frac{1}{2\|\mathbf{x}\|^2} \times 2\mathbf{x} = \frac{\mathbf{x}}{\|\mathbf{x}\|^2}.$$

1 Multidimensional random vectors

Section 3.10

A random m-dimensional vector is understood as the vector-column composed of m random variables,

$$\mathbf{X}^{m\times 1}=\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix}.$$

We use bold to indicate vectors and matrices throughout the book. The cdf and density are functions of $\mathbf{x}^{m\times 1}=(x_1,x_2,...,x_m)'$ and are related via differentiation and integration as follows (here it is assumed that we deal with continuous random variables):

$$\begin{aligned} F(\mathbf{x}) &= \Pr(X_1\leq x_1, X_2\leq x_2, ..., X_m\leq x_m), \\ f(\mathbf{x}) &= \frac{\partial F}{\partial \mathbf{x}}, \\ F(\mathbf{x}) &= \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_m} f(\mathbf{u})d\mathbf{u}. \end{aligned}$$

Note that the lowercase is used since $\mathbf{x}$ is nonrandom. Many results and definitions of the bivariate distribution can be extended to multidimensional random vectors. For example, $X_1, X_2, ..., X_n$ are independent if and only if the joint cdf and density is the product of individual (marginal) cdfs and densities, respectively.

The expectation of $\mathbf{X}$ is the $m\times 1$ vector defined as the element-wise expectation vector,

$$\boldsymbol{\mu}=E(\mathbf{X})=\begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_m) \end{bmatrix}.$$

The (i,j)th element of the $m\times m$ covariance matrix is defined as $cov(X_i,X_j)$ or in the matrix form as

$$\begin{aligned} cov(\mathbf{X}) &= \begin{bmatrix} cov(X_1,X_1) & cov(X_1,X_2) & \cdots & cov(X_1,X_m) \\ cov(X_2,X_1) & cov(X_2,X_2) & \cdots & cov(X_2,X_m) \\ \vdots & \vdots & \ddots & \vdots \\ cov(X_m,X_1) & cov(X_m,X_2) & \cdots & cov(X_m,X_m) \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1m} & \sigma_{2m} & \cdots & \sigma_m^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1m}\sigma_1\sigma_m \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{2m}\sigma_2\sigma_m \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1m}\sigma_1\sigma_m & \rho_{2m}\sigma_2\sigma_m & \cdots & \sigma_m^2 \end{bmatrix}. \end{aligned}$$

Alternatively, we express $cov(\mathbf{X})=\{\rho_{ij}\sigma_i\sigma_j, i,j=1,2,..,m\}$, where

$$\rho_{ij}=\frac{\sigma_{ij}}{\sigma_i\sigma_j}$$

is the correlation coefficient between X_i and X_j. Define

$$\mathbf{D}=\mathrm{diag}(\sigma_1^2,\sigma_2^2,...,\sigma_m^2)=\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_m^2 \end{bmatrix}$$

and $\mathbf{R}^{m\times m} = \{\rho_{ij}, i,j = 1,2,...,m\}$, the correlation matrix. Then we can compactly write

$$cov(\mathbf{X}) = \mathbf{D}^{1/2}\mathbf{R}\mathbf{D}^{1/2}$$

and reversely

$$cor(\mathbf{X}) = \mathbf{R} = \mathbf{D}^{-1/2} cov(\mathbf{X})\mathbf{D}^{-1/2}. \tag{2}$$

Sometimes, we need to consider the covariance matrix between two random vectors, $\mathbf{X}^{m\times 1}$ and $\mathbf{Y}^{n\times 1}$ as a $m \times n$ matrix of cross covariances,

$$cov(\mathbf{X}, \mathbf{Y}) = \{cov(X_i, Y_j), i = 1, ..., m; j = 1, ..., n\}.$$

It is obvious that $cov(\mathbf{X}) = cov(\mathbf{X}, \mathbf{X})$. In matrix notation,

$$\begin{aligned} cov(\mathbf{X}) &= E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] \\ (cov(\mathbf{X}))_{ij} &= E[(X_i - \mu_i)(X_j - \mu_j)], \quad i = 1,2,...,m, j = 1,2,...,n \end{aligned}$$

and

$$cov(\mathbf{X}, \mathbf{Y}) = E[(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{Y} - \boldsymbol{\mu}_Y)'].$$

We say that vectors $\mathbf{X}$ and $\mathbf{Y}$ are uncorrelated if $cov(\mathbf{X}, \mathbf{Y}) = \mathbf{0}$. In this case, the correlation between each pair of the components of vectors $\mathbf{X}$ and $\mathbf{Y}$ is zero.

Sample covariance and correlation matrices

Sample covariance between variable i (the ith column of $\mathbf{X}$) and variable j (the jth column of $\mathbf{X}$)

$$cov_{ij} = \frac{1}{n}\sum_{k=1}^{n}(X_{ki} - \overline{X}_i)(X_{kj} - \overline{X}_j)$$

The sample correlation coefficient

$$r_{ij} = \frac{\sum_{k=1}^{n}(X_{ki} - \overline{X}_i)(X_{kj} - \overline{X}_j)}{\sqrt{\sum_{k=1}^{n}(X_{ki} - \overline{X}_i)^2 \sum_{k=1}^{n}(X_{kj} - \overline{X}_j)^2}}$$

They are consistent

$$p\lim_{n\to\infty} cov_{ij} = \sigma_{ij}, \quad p\lim_{n\to\infty} r_{ij} = \rho_{ij}$$

due to Slutsky theorem.

Implementation in R: If `X` is a nxm matrix then `cov(X)` (or `var(X)`) returns a mxm covariance matrix and `cor(X)` returns a mxm correlation matrix. If `Y` is a nxk matrix `cov(X,Y)` returns a mxk covariance matrix between vector-columns of X and Y (similarly `cor`).

```
> X=matrix(rnorm(100*2),ncol=2)
> Y=matrix(rnorm(100*3),ncol=3)
cov(X)
> cor(X,Y)
 [,1] [,2] [,3]
[1,] 0.019672078 0.01224822 0.05889437
[2,] -0.001996442 0.09129804 0.06821266
```

```
> cor(Y)
 [,1] [,2] [,3]
[1,] 1.00000000 0.0783955928 0.1047866184
[2,] 0.07839559 1.0000000000 0.0002170319
[3,] 0.10478662 0.0002170319 1.0000000000
```

Explain

```
> cor(X,X+X)
 [,1] [,2]
[1,] 1.0000000 -0.1784894
[2,] -0.1784894 1.0000000
```

Theorem 11 *If $\mathbf{X}$ and $\mathbf{Y}$ are uncorrelated then*

$$cov(\mathbf{X}+\mathbf{Y}) = cov(\mathbf{X}) + cov(\mathbf{Y}). \tag{3}$$

The proof follows from the component-wise examination.

Basic properties of the mean and covariance matrix are presented below.

Theorem 12 *Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be random vectors, $\mathbf{A}$ and $\mathbf{B}$ be fixed matrices, a and b be scalars. Then, assuming the dimensions comply,*

$$\begin{aligned} E(\mathbf{AX}) &= \mathbf{A}E(\mathbf{X}), \\ E(\mathbf{AX}+\mathbf{BY}) &= \mathbf{A}E(\mathbf{X})+\mathbf{B}E(\mathbf{Y}), \\ cov(\mathbf{X}+\mathbf{Y},\mathbf{Z}) &= cov(\mathbf{X},\mathbf{Z})+cov(\mathbf{Y},\mathbf{Z}), \\ cov(\mathbf{AX}) &= \mathbf{A}cov(\mathbf{X})\mathbf{A}'. \end{aligned} \tag{4}$$

The proofs are straightforward. The last identity has an important corollary: if $\mathbf{A} = \mathbf{a}'$, the vector-row, then

$$var(\mathbf{a}'\mathbf{X}) = \mathbf{a}'cov(\mathbf{X})\mathbf{a}. \tag{5}$$

The left-hand side is a linear combination of components of vector $\mathbf{X}$ and the right-hand side is a quadratic form (since $\mathbf{a}'\mathbf{X}$ is scalar we use notation $cov = var$).

Example 13 *Use (5) to prove that the variance of the sum of pairwise uncorrelated random variables is the sum of variances.*

Solution. The sum of components of $\mathbf{X}$ can be written as $\mathbf{a}'\mathbf{X}$ where $\mathbf{a} = \mathbf{1}$. If components of $\mathbf{X}$ are pairwise uncorrelated then $cov(\mathbf{X}) = \text{diag}(\sigma_1^2, \sigma_2^2, ..., \sigma_m^2)$. Applying formula (5) one obtains

$$var(\mathbf{1}'\mathbf{X}) = \mathbf{1}'cov(\mathbf{X})\mathbf{1} = \sum_{i=1}^{m} \sigma_i^2,$$

the sum of variances. □

Example 14 *Prove that covariance matrix is a nonnegative definite symmetric matrix.*

Solution. The symmetry follows from the fact that $cov(X_i, X_j) = cov(X_j, X_i)$. To prove that $cov(\mathbf{X}) = \mathbf{\Omega}$ is a nonnegative definite matrix we need to prove that $\mathbf{a}'\mathbf{\Omega}\mathbf{a} \geq 0$ for every nonrandom vector $\mathbf{a}$. Consider a random variable as a linear combination, $Y = \mathbf{a}'\mathbf{X}$. From (5) we have $var(Y) = \mathbf{a}'\mathbf{\Omega}\mathbf{a} \geq 0$ because variance is nonnegative. Therefore covariance matrix is a nonnegative definite matrix. □

Theorem 15 *The following formulas hold*

$$\begin{aligned} cov(\mathbf{X}) &= E(\mathbf{X}\mathbf{X}')-\boldsymbol{\mu}\boldsymbol{\mu}', \quad cov(\mathbf{X},\mathbf{Y}) = E(\mathbf{X}\mathbf{Y}')-\boldsymbol{\mu}_X\boldsymbol{\mu}_Y' \\ cov(\mathbf{X}) &= E[(\mathbf{X}-\boldsymbol{\mu})\mathbf{X}'] = E[\mathbf{X}(\mathbf{X}-\boldsymbol{\mu})'], \quad cov(\mathbf{X},\mathbf{Y}) = E[\mathbf{X}(\mathbf{Y}-\boldsymbol{\mu}_Y)'] = E[(\mathbf{X}-\boldsymbol{\mu}_X)\mathbf{Y}']. \end{aligned}$$

Proof. For the first formula, using the previous theorem, we have

$$\begin{aligned} cov(\mathbf{X}) &= E[(\mathbf{X}-\boldsymbol{\mu})\mathbf{X}']-E[(\mathbf{X}-\boldsymbol{\mu})\boldsymbol{\mu}'] = E(\mathbf{X}\mathbf{X}') - \boldsymbol{\mu}E(\mathbf{X}') - E(\mathbf{X})\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}' \\ &= E(\mathbf{X}\mathbf{X}') - \boldsymbol{\mu}\boldsymbol{\mu}' - \boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}' = E(\mathbf{X}\mathbf{X}') - \boldsymbol{\mu}\boldsymbol{\mu}', \end{aligned}$$

which can be viewed as a generalization of $var(X) = E(X^2) - \mu^2$. Other formulas can be proven similarly. □

Example 16 ***3-by-3 covariance matrix.*** *Let X, Y, and Z be independent (scalar) random variables. (a) Find the 3×3 covariance and correlation matrices for the random vector*

$$\mathbf{X} = \begin{bmatrix} X \\ Y - X \\ X + Y + Z \end{bmatrix}$$

by direct computation. (b) Find the covariance matrix of $\mathbf{X}$ using identity (4).

Solution. (a) To obtain the 3×3 correlation matrix we need to have $(3 \times 4)/2 = 6$ quantities: three variances and three covariances. We repeatedly use the rule that the variance of the sum of independent random variables is equal to the sum of variances. The variances of the second and the third component of vector $\mathbf{X}$ are

$$var(Y - X) = \sigma_X^2 + \sigma_Y^2, \quad var(X + Y + Z) = \sigma_X^2 + \sigma_Y^2 + \sigma_Z^2,$$

respectively. Compute covariances,

$$cov(X, Y - X) = -var(X) = -\sigma_X^2, \; cov(X, X + Y + W) = \sigma_X^2, \; cov(Y - X, X + Y + W) = \sigma_Y^2 - \sigma_X^2.$$

These imply

$$cov(\mathbf{X}) = \begin{bmatrix} \sigma_X^2 & -\sigma_X^2 & \sigma_X^2 \\ -\sigma_X^2 & \sigma_X^2 + \sigma_Y^2 & \sigma_Y^2 - \sigma_X^2 \\ \sigma_X^2 & \sigma_Y^2 - \sigma_X^2 & \sigma_X^2 + \sigma_Y^2 + \sigma_Z^2 \end{bmatrix}.$$

Since the correlation coefficient between any W and V is

$$\rho = \frac{cov(W, V)}{\sqrt{var(W)var(V)}},$$

the 3×3 correlation matrix for $(X, Y - X, X + Y + Z)$ is equal

$$\begin{bmatrix} 1 & -\frac{\sigma_X}{\sqrt{\sigma_X^2+\sigma_Y^2}} & \frac{\sigma_X}{\sqrt{\sigma_X^2+\sigma_Y^2+\sigma_W^2}} \\ -\frac{\sigma_X}{\sqrt{\sigma_X^2+\sigma_Y^2}} & 1 & \frac{\sigma_Y^2-\sigma_X^2}{\sigma_X\sqrt{\sigma_X^2+\sigma_Y^2+\sigma_W^2}} \\ \frac{\sigma_X}{\sqrt{\sigma_X^2+\sigma_Y^2+\sigma_W^2}} & \frac{\sigma_Y^2-\sigma_X^2}{\sigma_X\sqrt{\sigma_X^2+\sigma_Y^2+\sigma_W^2}} & 1 \end{bmatrix}.$$

(b) We notice that $\mathbf{X} = \mathbf{AU}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}.$$

Therefore from formula (4) we deduce

$$\begin{aligned} cov(\mathbf{X}) &= \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_X^2 & 0 & 0 \\ 0 & \sigma_Y^2 & 0 \\ 0 & 0 & \sigma_Z^2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}' \\ &= \begin{bmatrix} \sigma_X^2 & 0 & 0 \\ -\sigma_X^2 & \sigma_Y^2 & 0 \\ \sigma_X^2 & \sigma_Y^2 & \sigma_Z^2 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & -\sigma_X^2 & \sigma_X^2 \\ -\sigma_X^2 & \sigma_X^2 + \sigma_Y^2 & \sigma_Y^2 - \sigma_X^2 \\ \sigma_X^2 & \sigma_Y^2 - \sigma_X^2 & \sigma_X^2 + \sigma_Y^2 + \sigma_Z^2 \end{bmatrix}, \end{aligned}$$

the same as in (a).

Homework 2

1. (5 points). Solve Example 8. Use vector/matrx algebra and calculus.
2. (5 points). Random variables X, Y, and Z are independent with variances $1, 2$, and 3 respectively. Find $var(X + 2Y - Z)$ using formula (5).
3. (5 points). Volatility of the stock is measured as its variance. The file `compret.csv` contains the weekly return (%) of three computer industry giants. Estimate the volatility of the stock composite portfolio with 25% of DELL, 30% of ORCL and 45% MSFT and compare with the volatility of the portfolio if the proportion of stocks in the portfolio is the same (33.333%) using formula (5). [Hint: Use `cov` and `read.csv`.]
4. (5 points). Let the components of the m-dimensional random vector $\mathbf{X}$ be iid with a common variance σ^2 and $\mathbf{Y} = \mathbf{X} + Z\mathbf{1}$, where random variable Z has variance τ^2 and independent from $\mathbf{X}$. Express $cor(\mathbf{Y})$ in matrix form using $\mathbf{1}$ and $\mathbf{I}$.

Solutions

1. (a) The asteroid is approaching the Earth if and only if

$$\left.\frac{dH}{dt}\right|_{t=0} = -2(\mathbf{y} - \mathbf{f}(t))' \left.\frac{d\mathbf{f}}{dt}\right|_{t=0} < 0.$$

Compute

$$-2(\mathbf{y} - \mathbf{f}(t))' \left.\frac{d\mathbf{f}}{dt}\right|_{t=0} = -2(3-2, -1+2, 5-1) \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = -2 \times (2 - 1 + 4) = -10 < 0.$$

This means that the asteroid approaches the Earth. (b) Find the minimum of $H(t)$ by taking the derivative and setting it to zero:

$$\frac{dH}{dt} = -2(3 - 2 - 2t, -1 + 2 + t, 5 - 1 - t) \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = 0$$

or equivalently

$$2(1-2t)-(1+t)+(4-t)=0$$

with the solution

$$t_{\min}=\frac{5}{6}.$$

(c) Compute the minimum distance

$$\begin{aligned}\min_{t\geq 0} H(t) &= H(5/6)=\sqrt{(3-(2+2\times 5/6))^2+(-1-(-2-5/6))^2+(5-(1+5/6))^2}\\ &= 3.719>0\end{aligned}$$

This means that asteroid does not hit the Earth.

2. Introduce a 3×1 vector

$$\mathbf{X}=\begin{bmatrix} X\\ Y\\ Z\end{bmatrix}$$

We have

$$cov(\mathbf{X})=\begin{bmatrix} 1 & 0 & 0\\ 0 & 2 & 0\\ 0 & 0 & 3\end{bmatrix}.$$

Represent $X+2Y-Z$ as $\mathbf{a}'\mathbf{X}$ where

$$\mathbf{a}=\begin{bmatrix} 1\\ 2\\ -1\end{bmatrix}.$$

By formula (5) we obtain

$$\begin{aligned}var(X+2Y-Z) &= var(\mathbf{a}'\mathbf{X})=\mathbf{a}'cov(\mathbf{X})\mathbf{a}=\begin{bmatrix} 1\\ 2\\ -1\end{bmatrix}'\begin{bmatrix} 1 & 0 & 0\\ 0 & 2 & 0\\ 0 & 0 & 3\end{bmatrix}\begin{bmatrix} 1\\ 2\\ -1\end{bmatrix}\\ &= 1+2\times 2\times 2+1\times 3=12.\end{aligned}$$

3.

```
hw2.3=function()
{
dump("hw1.19","c:\\M7021\\hw1.19.r")
d=read.csv("c:\\M7021\\compret.csv")
co=cov(d)
a=c(.25,.3,.45)
va=t(a)%*%co%*%a
print(paste("Variance/volatility of the non-equal composite =",va))
a=c(1/3,1/3,1/3)
va=t(a)%*%co%*%a
print(paste("Variance/volatility of equal composite =",va))
}
> hw2.3()
[1] "Variance/volatility of the non-equal composite = 4.18192111854187"
```

```
[1] "Variance/volatility of equal composite = 4.09553159916657"
```

The volatility of the composite is higher compared to the equal weight volatility.

4. Since components of vector $\mathbf{X}$ and Z are independent the vector of variances of $\mathbf{Y}$ is $(\sigma^2+\tau^2)\mathbf{1}$. Therefore the diagonal matrix of variances is $\mathbf{D}=(\sigma^2+\tau^2)\mathbf{I}$. Now find the covariance matrix. Since $\mathbf{X}$ and Z are independent by Theorem 11 we have

$$cov(\mathbf{Y})=cov(\mathbf{X}+Z\mathbf{1})=cov(\mathbf{X})+cov(Z\mathbf{1})=\sigma^2\mathbf{I}+cov(Z\mathbf{1}).$$

But $cov(Z\mathbf{1})=\tau^2\mathbf{1}\mathbf{1}'$ so finally,

$$cov(\mathbf{Y})=\sigma^2\mathbf{I}+\tau^2\mathbf{1}\mathbf{1}'.$$

Now we are ready to find

$$\begin{aligned} cor(\mathbf{Y}) &= \mathbf{D}^{-1/2}cov(\mathbf{Y})\mathbf{D}^{-1/2}=\frac{1}{\sqrt{\sigma^2+\tau^2}}\mathbf{I}(\sigma^2\mathbf{I}+\tau^2\mathbf{1}\mathbf{1}')\mathbf{I}\frac{1}{\sqrt{\sigma^2+\tau^2}} \\ &= \frac{1}{\sigma^2+\tau^2}(\sigma^2\mathbf{I}+\tau^2\mathbf{1}\mathbf{1}')=\frac{\sigma^2}{\sigma^2+\tau^2}\mathbf{I}+\frac{\tau^2}{\sigma^2+\tau^2}\mathbf{1}\mathbf{1}'. \end{aligned}$$