Appendix A

Probability Distributions

Name	*Probability*	*Mean*	*Variance*	*Relationships*
Discrete Uniform	$\frac{1}{N-M+1}$ $M \leq n \leq N$ 0 otherwise	$\frac{M+N}{2}$	$\frac{(N-M+2)(N-M)}{12}$	
Bernoulli or Binary	$\Pr(n=0) = 1-p$ $\Pr(n=1) = p$	p	$p(1-p)$	
Binomial	$\binom{N}{n} p^n (1-p)^{N-n}$, $n = 0, \ldots, N$	Np	$Np(1-p)$	Sum of N IID Bernoulli
Geometric	$(1-p)p^n$, $n \geq 0$	$p/1-p$	$p/(1-p)^2$	
Negative Binomial	$\binom{n-1}{N-1} p^N (1-p)^{n-N}$, $n \geq N$	N/p	$N(1-p)/p^2$	
Poisson	$\frac{\lambda^n e^{-\lambda}}{n!}$, $n \geq 0$	λ	λ	
Hypergeometric	$\frac{\binom{a}{n}\binom{b}{N-n}}{\binom{a+b}{N}}$, $n = 0, \ldots, N$; $0 \leq n \leq a+b; 0 \leq N \leq a+b$	$Na/(a+b)$	$\frac{Nab(a+b-N)}{(a+b)^2(a+b-1)}$	
Logarithmic	$\frac{-p^n}{n \log(1-p)}$	$\frac{-p}{(1-p)\log(1-p)}$	$\frac{-p[p+\log(1-p)]}{(1-p)\log q}$	

Table A.1: Discrete probability distributions.

Name	*Density*	*Mean*	*Variance*	*Relationships*
Gaussian (Normal)	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$	m	σ^2	
Bivariate Gaussian	$\frac{1}{2\pi(1-\rho^2)^{1/2}\sigma_x\sigma_y} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-m_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-m_x}{\sigma_x}\right)\left(\frac{y-m_y}{\sigma_y}\right) + \left(\frac{y-m_y}{\sigma_y}\right)^2\right]\right\}$	$\mathcal{E}[x] = m_x$, $\mathcal{E}[y] = m_y$	$\mathcal{V}[x] = \sigma_x^2$, $\mathcal{V}[y] = \sigma_y^2$, $\mathcal{E}[xy] = m_x m_y + \rho\sigma_x\rho\sigma_y$	ρ: correlation coefficient
Conditional Gaussian	$p(x\|y) = \frac{1}{\sqrt{2\pi(1-\rho)^2\sigma_x^2}} \exp\left\{-\frac{\left(x - m_x - \frac{\rho\sigma_x}{\sigma_y}(y-m_y)\right)^2}{2\sigma_x^2(1-\rho^2)}\right\}$	$m_x + \frac{\rho\sigma_x}{\sigma_y}(y-m_y)$	$\sigma_x^2(1-\rho^2)$	
Multivariate Gaussian	$\frac{1}{(\det(2\pi\mathbf{K}))^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^t\mathbf{K}^{-1}(\mathbf{x}-\mathbf{m})\right\}$	$\mathbf{m}$	$\mathbf{K}$	
Generalized Gaussian	$\frac{1}{2\Gamma(1+1/r)A(r)} e^{-\left\|\frac{x-m}{A(r)}\right\|^r}$	m	σ^2	$A(r) = \left[\frac{\sigma^2\Gamma(1/r)}{\Gamma(3/r)}\right]^{1/2}$
Chi-Squared (χ_ν^2)	$\frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\frac{\nu}{2}-1}e^{-x/2}, 0 \le x$	ν	2ν	$\chi_\nu^2 = \sum_{i=1}^{\nu} X_i^2$, X_i IID $\mathcal{N}(0,1)$
Noncentral Chi-Squared ($\chi_\nu'^2(\lambda)$)	$\frac{1}{2}(x/\lambda)^{(\nu-2)/4} I_{(\nu-2)/2}(\sqrt{\lambda}x) e^{-1/2(\lambda+x)}$	$\nu+\lambda$	$2(\nu+2\lambda)$	$\chi_\nu'^2 = \sum_{i=1}^{\nu} X_i^2$, X_i IID $\mathcal{N}(m_i,1)$, $\lambda = \sum_{i=1}^{\nu} m_i^2$
Chi χ_ν	$\frac{x^{\nu-1}e^{-x^2/2}}{2^{\nu/2-1}\Gamma\left(\frac{\nu}{2}\right)}, 0 \le x$	$\sqrt{2}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{1}{2}\nu\right)}$	$\frac{2\left[\Gamma\left(\frac{1}{2}\nu\right)\Gamma\left(\frac{1}{2}\nu+1\right) - \Gamma^2\left(\frac{\nu+1}{2}\right)\right]}{\Gamma^2\left(\frac{1}{2}\nu\right)}$	$\chi_\nu = \sqrt{\chi_\nu^2}$
Student's t	$\frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2}$	0	$\frac{\nu}{\nu-2}, 2 < \nu$	
Beta $\beta_{m,n}$	$\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} x^{m/2-1}(1-x)^{n/2-1}$, $0 < x < 1, 0 < a, b$	$\frac{m}{m+n}$	$\frac{2mn}{(m+n)^2(m+n+2)}$	$\beta_{m,n} = \frac{\chi_m^2}{\chi_m^2+\chi_n^2}$
F Distribution	$\frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2} \frac{x^{(m-2)/2}}{[1+(m/n)x]^{(m+n)/2}}, 0 \le x; 1 \le m, n$	$\frac{n}{n-2}, n > 2$	$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, n > 4$	$F_{m,n} = \frac{\chi_m^2/m}{\chi_n^2/n}$
Non-central F $F'_{m,n}(\lambda)$	$\sum_{k=0}^{\infty} \frac{\left(\frac{\lambda}{2}\right)^k}{k!} e^{-\frac{\lambda}{2}} p_{\beta\frac{m}{2}+k,\frac{n}{2}}\left(\frac{mx}{mx+n}\right)$	$\frac{n}{n-2}, n > 2$	$2\left(\frac{n}{m}\right)^2 \frac{(m+\lambda)^2+(m+2\lambda)(n-2)}{(n-2)^2(n-4)}$, $n > 4$	$F'_{m,n}(\lambda) = \frac{\chi_m'^2(\lambda)/m}{\chi_n^2/n}$
Wishart $\mathbf{W}_M(N,\mathbf{K})$	$\frac{(\det[\mathbf{w}])^{\frac{N-M-1}{2}}}{2^{\frac{NM}{2}}\Gamma_M(\frac{N}{2})(\det[\mathbf{K}])^{\frac{K}{2}}} e^{-\frac{\mathrm{tr}[\mathbf{K}^{-1}\mathbf{w}]}{2}}$, $\Gamma_M\left(\frac{N}{2}\right) = \pi^{M(M-1)/4} \times \prod_{m=0}^{M-1}\Gamma\left(\frac{N}{2}-\frac{m}{2}\right)$	$N\mathbf{K}$	$\mathrm{cov}[\mathbf{W}_{ij}\mathbf{W}_{kl}] = N\cdot(\mathbf{K}_{ik}\mathbf{K}_{jl} + \mathbf{K}_{il}\mathbf{K}_{jk})$	$\mathbf{W}_M(N,\mathbf{K}) = \sum_{n=1}^{N}\mathbf{X}_n\mathbf{X}_n'$, $\mathbf{X}_n \sim \mathcal{N}(\mathbf{0},\mathbf{K})$, $\dim[\mathbf{X}] = M$

Table A.2: Distributions related to the Gaussian.

Name	*Density*	*Mean*	*Variance*	*Relationships*
Uniform	$\frac{1}{b-a}, a \leq x \leq b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	
Triangular	$2x/a \quad 0 \leq x \leq a$ $2(1-x)/(1-a) \quad a \leq x \leq 1$	$\frac{1+a}{3}$	$\frac{1-a+a^2}{18}$	
Exponential	$\lambda e^{-\lambda x}, 0 \leq x$	$1/\lambda$	$1/\lambda^2$	
Lognormal	$\frac{1}{\sqrt{2\pi\sigma^2 x^2}} e^{-\frac{1}{2}\left(\frac{\log x - m}{\sigma}\right)^2}$, $0 < x$	$e^{m+\frac{\sigma^2}{2}}$	$e^{2m}\left(e^{2\sigma^2} - e^{\sigma^2}\right)$	
Maxwell	$\sqrt{\frac{2}{\pi}} a^{3/2} x^2 e^{-ax^2/2}, 0 < x$	$\sqrt{\frac{8}{\pi a}}$	$\left(3 - \frac{8}{\pi}\right) a^{-1}$	
Laplacian	$\frac{1}{\sqrt{2\sigma^2}} e^{-\frac{\lvert x-m \rvert}{\sqrt{\sigma^2/2}}}$	m	σ^2	
Gamma	$\frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, 0 < x, 0 < a, b$	$\frac{a}{b}$	$\frac{a}{b^2}$	
Rayleigh	$2ax e^{-ax^2}, 0 \leq x$	$\sqrt{\frac{\pi}{4a}}$	$\frac{1}{a}\left(1 - \frac{\pi}{4}\right)$	
Weibull	$abx^{b-1} e^{-ax^b}, 0 < x, 0 < a, b$	$(1/a)^{1/b} \cdot \Gamma(1 + 1/b)$	$a^{-2/b} \cdot \left[\Gamma(1 + 2/b) - \Gamma^2(1 + 1/b)\right]$	
Arc-Sine	$\frac{1}{\pi\sqrt{x(1-x)}}, 0 < x < 1$	$\frac{1}{2}$	$\frac{1}{8}$	
Sine Amplitude	$\frac{1}{\pi\sqrt{1-x^2}}, \lvert x \rvert < 1$	0	$\frac{1}{2}$	
Circular Normal	$\frac{e^{a\cos(x-m)}}{2\pi I_0(a)}, -\pi < x \leq \pi$	m		
Cauchy	$\frac{a/\pi}{(x-m)^2 + a^2}$	m (from symmetry arguments)	∞	
Logistic	$\frac{e^{-(x-m)/a}}{a\left[1 + e^{-(x-m)/a}\right]^2}, 0 < a$	m	$\frac{a^2\pi^2}{3}$	
Gumbel	$\frac{e^{-(x-m)/a}}{a} \exp\left\{-e^{-(x-m)/a}\right\}$, $0 < a$	$m + a\gamma$	$\frac{a^2\pi^2}{6}$	
Pareto	$\frac{ab^a}{x^{1-a}}, 0 < a; 0 < b \leq x$	$\frac{ab}{a-1}, a > 1$	$\frac{ab^2}{(a-2)(a-1)^2}, a > 2$	

Table A.3: Non-Gaussian distributions.

Appendix B

Matrix Theory

B.1 Basic Definitions

An $m \times n$ *matrix* $\mathbf{A}$ is a rectangular (square if $n = m$) arrangement of scalar elements A_{ij} (i^{th} row, j^{th} column).

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & & A_{2n} \\ \vdots & & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}$$

When the matrix is square, the *main diagonal* extends from the upper left corner to the lower right and consists of the elements A_{ii}. The *cross diagonal* is the opposite diagonal and consists of the elements $A_{n-i+1,i}$. A *vector* is common terminology for a column ($m \times 1$) matrix. The *dimension* of the vector equals m. Rectangular matrices are (usually) denoted by boldface uppercase letters ($\mathbf{A}, \mathbf{B}, \ldots$) and vectors by boldface lowercase letters ($\mathbf{a}, \mathbf{b}, \ldots$). The dimension of a vector is expressed as $\dim(\mathbf{a})$.

To add matrices, the number of rows in each matrix as well as the number of columns must agree. The sum of $m \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ is defined to be an $m \times n$ matrix $\mathbf{C}$ whose elements are

$$C_{ij} = A_{ij} + B_{ij}.$$

Matrix addition is commutative ($\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$). The product $\mathbf{AB}$ of two matrices $\mathbf{A}$ and $\mathbf{B}$ is only defined if the number of rows of $\mathbf{A}$ equals the number of columns of $\mathbf{B}$. Thus, if $\mathbf{A}$ represents an $l \times m$ matrix and $\mathbf{B}$ an $m \times n$ matrix, the result is an $l \times n$ matrix $\mathbf{C}$, each term of which is defined to be

$$C_{ij} = \sum_{k=1}^{m} A_{ik} B_{kj}, \quad i = 1, \ldots, l; j = 1, \ldots, n.$$

Clearly, the matrix product is not necessarily commutative ($\mathbf{AB} \neq \mathbf{BA}$), but is distributive [$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$] and associative [$\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$].

Several methods of rearranging the elements of a matrix are frequently used. The *complex conjugate* of the matrix $\mathbf{A}$ is denoted by $\mathbf{A}^*$ and consists of the complex conjugates of the elements of $\mathbf{A}$.

$$[\mathbf{A}^*]_{ij} = A^*_{ij}$$

The *transpose* of an $m \times n$ matrix $\mathbf{A}$ is an $n \times m$ matrix $\mathbf{A}^t$ whose rows are the columns of $\mathbf{A}$.

$$\left[\mathbf{A}^t\right]_{ij} = A_{ji}$$

The transpose of a product of two matrices equals the product of their transposes, but in reversed order.

$$(\mathbf{AB})^t = \mathbf{B}^t \mathbf{A}^t$$

This property applies whether the matrices are square or not. The *conjugate transpose* (sometimes known as the Hermitian transpose) of $\mathbf{A}$ is denoted by $\mathbf{A}'$ and equals $\left(\mathbf{A}^*\right)^t$.

$$[\mathbf{A}']_{ij} = A^*_{ji}$$

B.2 Basic Matrix Forms

The relations between the values of the elements of a matrix define special matrix cases. Matrices having special internal structure are important to recognize in manipulating matrix expressions.

- A *diagonal* matrix, denoted by $\mathrm{diag}[A_{11}, A_{22}, \ldots, A_{nn}]$, has non-zero entries only along the main diagonal ($i = j$) of the matrix.

$$\begin{bmatrix} A_{11} & 0 & \cdots & 0 \\ 0 & A_{22} & 0 & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & A_{nn} \end{bmatrix}$$

- a *block diagonal* matrix has square matrices $\mathbf{A}_1, \ldots, \mathbf{A}_N$ on the diagonal, with zero-valued entries elsewhere. Note that these matrices do not necessary share a common size.

$$\begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \cdots & \cdots \\ \mathbf{0} & \mathbf{A}_2 & \mathbf{0} & \cdots \\ \vdots & & \ddots & \vdots \\ \cdots & \cdots & \mathbf{0} & \mathbf{A}_N \end{bmatrix}$$

- The *identity* matrix, denoted by $\mathbf{I}$, is a special diagonal matrix having unity on the main diagonal.

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

  The identity matrix is so-named because it is the multiplicative identity for square matrices ($\mathbf{IA} = \mathbf{AI} = \mathbf{A}$).

- The ordering of the rows or columns of a square matrix can be reversed by pre- or post-multiplication by the *exchange* matrix $\mathbf{J}$.

$$\mathbf{J} = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ \vdots & .^{\cdot^{\cdot}} & 1 & 0 \\ 0 & .^{\cdot^{\cdot}} & .^{\cdot^{\cdot}} & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix}$$

- A *lower triangular* matrix $\mathbf{L}$ has non-zero values on and "below" its main diagonal.

$$\mathbf{L} = \begin{bmatrix} L_{11} & 0 & \cdots & 0 \\ L_{21} & L_{22} & 0 & \vdots \\ \vdots & \vdots & \ddots & \\ L_{n1} & L_{n2} & \cdots & L_{nn} \end{bmatrix}$$

  An upper triangular matrix can be similarly defined.

- A *Vandermonde* matrix consists of columns containing geometric sequences.

$$\mathbf{A} = \begin{bmatrix} 1 & \cdots & 1 \\ a_0 & \cdots & a_{n-1} \\ a_0^2 & \cdots & a_{n-1}^2 \\ \vdots & & \vdots \\ a_0^{m-1} & \cdots & a_{n-1}^{m-1} \end{bmatrix}$$

One special case of a square Vandermonde matrix is the DFT matrix $\mathbf{W}$ in which the elements are $a_k = \exp\{-j2\pi k/m\}$. The discrete Fourier transform of a vector $\mathbf{x}$ can be expressed as

$$\text{DFT}[\mathbf{x}] = \mathbf{W}\mathbf{x}.$$

- A *symmetric* matrix equals its transpose ($\mathbf{A}^t = \mathbf{A}$). A *conjugate symmetric* (Hermitian) matrix equals its own conjugate transpose ($\mathbf{A}' = \mathbf{A}$). Correlation matrices, which have the form $\mathbf{A} = \mathcal{E}[\mathbf{x}\mathbf{x}']$, are Hermitian.
- A *Toeplitz* matrix has constant values along each of its diagonals ($A_{ij} = a_{i-j}$).

$$\begin{bmatrix} a_0 & a_1 & \cdots & a_{n-1} \\ a_{-1} & a_0 & a_1 & \\ \vdots & \ddots & \ddots & \\ a_{-(n-1)} & a_{-(n-2)} & \cdots & a_0 \end{bmatrix}$$

In time series, correlation matrices are not only Hermitian, they are also Toeplitz. Hermitian Toeplitz matrices have no more than n unique elements in the matrix. Because of this extensive redundancy of values, many efficient techniques for computing expressions involving them have been developed.

- Constant values along the cross-diagonals of a square matrix define a *Hankel* matrix.

$$A_{ij} = a_{i+j-2} \qquad \mathbf{A} = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-1} \\ a_1 & a_2 & \cdots & \iddots & a_n \\ a_2 & \iddots & \iddots & \iddots & \vdots \\ \vdots & \iddots & \iddots & & a_{2n-3} \\ a_{n-1} & a_n & \cdots & a_{2n-3} & a_{2n-2} \end{bmatrix}$$

All Hankel matrices are symmetric. If $\mathbf{A}$ is a Hankel matrix, then both $\mathbf{JA}$ and $\mathbf{AJ}$ are Toeplitz matrices. Hankel matrices are frequently used to express in matrix form the convolution of a sequence with a finite-duration unit sample response. For example, if $y(n) = \sum_{m=n-(M-1)}^{n} h(n-m)x(m)$ for $n = M-1, \ldots, N-1$, the matrix form would be

$$\begin{bmatrix} x(0) & x(1) & \cdots & x(M-1) \\ x(1) & x(2) & \cdots & x(M) \\ \vdots & \vdots & & \vdots \\ x(N-M) & x(N-M+1) & \cdots & x(N-1) \end{bmatrix} \begin{bmatrix} h(M-1) \\ h(M-2) \\ \vdots \\ h(0) \end{bmatrix} = \begin{bmatrix} y(M-1) \\ y(M) \\ \vdots \\ y(N-1) \end{bmatrix}.$$

- A square matrix is said to be *circulant* if each row is a circular rotation of the previous one.

$$\begin{bmatrix} a_0 & a_1 & \cdots & a_{n-1} \\ a_{n-1} & a_0 & a_1 & \cdots \\ a_{n-2} & a_{n-1} & a_0 & \cdots \\ \vdots & \vdots & & \\ a_1 & a_2 & \cdots & a_0 \end{bmatrix}$$

All circulant matrices are Toeplitz; circulant matrices are symmetric only when the sequence $a_1, a_2, \ldots$ is even: $a_i = a_{n-i}, i = 1, \ldots, n/2$ (a_0 is arbitrary).

- The square matrix $\mathbf{A}$ is said to be *orthogonal* if it satisfies $\mathbf{A}^t\mathbf{A} = \mathbf{I}$. $\mathbf{A}$ is *unitary* if it satisfies $\mathbf{A}'\mathbf{A} = \mathbf{I}$.
- The square matrix $\mathbf{A}$ is said to be a *projection* matrix if $\mathbf{A}^2 = \mathbf{A}$. Because of this property, $\mathbf{A}^m = \mathbf{A}^n$ for all choices of positive integers n, m. The reason for this name will be given once eigenvalues are defined (§B.6 {188}).

- A matrix is said to have a *null space* if a set of non-zero vectors $\mathbf{x}$ exist that satisfy $\mathbf{A}\mathbf{x} = \mathbf{0}$. These vectors would thereby be orthogonal to the rows of $\mathbf{A}$. The matrix could be square or, more interestingly, it could be rectangular, having more columns than rows. The rows would then define a collection of vectors that represent all vectors for which $\mathbf{A}\mathbf{x} \neq 0$.
- The vector $\mathbf{y}$ is said to be in the *range* of the matrix $\mathbf{A}$ if it can be expressed as $\mathbf{A}\mathbf{x} = \mathbf{y}$ for some $\mathbf{x}$.

B.3 Operations on Matrices

Several operations on matrices and between them are so commonly used that specific terminology has evolved.

- The *inner product* $\mathbf{x}'\mathbf{y}$ between two vectors $\mathbf{x}$ and $\mathbf{y}$ is defined to be the scalar expressed by*

$$\mathbf{x}'\mathbf{y} = \sum_{i=1}^{n} x_i^* y_i .$$

  Since an inner product is a scalar, expressions involving inner products are frequently rearranged by noting that the transpose of a scalar is the same scalar. Therefore, $(\mathbf{x}'\mathbf{y})^t = \mathbf{y}^t\mathbf{x}^* = \mathbf{x}'\mathbf{y}$. Two vectors are said to be *orthogonal* if their inner product is zero.
- The *outer product* $\mathbf{x}\mathbf{y}'$ between two vectors x (dimension m) and y (dimension n) is an $m \times n$ matrix whose elements are

$$[\mathbf{x}\mathbf{y}']_{ij} = x_i y_j^* .$$

- The *Kronecker product* $\mathbf{A} \otimes \mathbf{B}$ between two matrices $\mathbf{A}$ ($m_\mathbf{A} \times n_\mathbf{A}$) and $\mathbf{B}$ ($m_\mathbf{B} \times n_\mathbf{B}$) is the $m_\mathbf{A} m_\mathbf{B} \times n_\mathbf{A} n_\mathbf{B}$ matrix given by

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} A_{11}\mathbf{B} & A_{12}\mathbf{B} & \cdots & A_{1n_\mathbf{A}}\mathbf{B} \\ A_{21}\mathbf{B} & A_{22}\mathbf{B} & & A_{2n_\mathbf{A}}\mathbf{B} \\ \vdots & & \ddots & \vdots \\ A_{m_\mathbf{A}1}\mathbf{B} & A_{m_\mathbf{A}2}\mathbf{B} & \cdots & A_{m_\mathbf{A}n_\mathbf{A}}\mathbf{B} \end{bmatrix} .$$

  The matrix $\mathbf{B}$ multiplied by the scalars A_{ij} is thus repeated throughout the matrix. The Kronecker product of two positive-definite matrices is also positive definite.
- The *inverse* of a matrix $\mathbf{A}$ is denoted by $\mathbf{A}^{-1}$ which satisfies $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. The inverse of a matrix is not guaranteed to exist. Numerous conditions on the inverse's existence are detailed in the following. When it does exist, the following properties hold.
  1. If $\mathbf{A}$, $\mathbf{B}$ are invertible matrices, $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
  2. If $\mathbf{A}$, $\mathbf{B}$ are invertible matrices, $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$
  3. Assuming that all inverses exist where used, the inverse of a sum of matrices can be written several useful ways.

$$\begin{aligned} (\mathbf{A} + \mathbf{B}\mathbf{C}\mathbf{D})^{-1} &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D}\mathbf{A}^{-1}\mathbf{B} + \mathbf{C}^{-1}\right)^{-1}\mathbf{D}\mathbf{A}^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\mathbf{C}\left(\mathbf{I} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B}\mathbf{C}\right)^{-1}\mathbf{D}\mathbf{A}^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{I} + \mathbf{C}\mathbf{D}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}\mathbf{D}\mathbf{A}^{-1} \end{aligned}$$

  Note that the matrices $\mathbf{B}$ and $\mathbf{D}$ need not be invertible for these relationships to be valid. In the latter two, the matrix $\mathbf{C}$ need not be invertible. Notable special cases of this result are

$$(\mathbf{A} + \mu\mathbf{x}\mathbf{y}')^{-1} = \mathbf{A}^{-1}\left(\mathbf{I} - \frac{\mathbf{x}\mathbf{y}'}{\frac{1}{\mu} + \mathbf{y}'\mathbf{A}^{-1}\mathbf{x}}\mathbf{A}^{-1}\right)$$
$$(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\left(\mathbf{A}^{-1} + \mathbf{B}^{-1}\right)^{-1}\mathbf{A}^{-1}$$

* In more abstract settings, the inner product between two vectors x and y is denoted by $\langle x, y \rangle$. There, the inner product becomes any mapping of two vectors to a scalar that satisfies three properties.

(i) $\langle y, x \rangle = \langle x, y \rangle^*$
(ii) $\langle ax + by, z \rangle = a^* \langle x, z \rangle + b^* \langle y, z \rangle$, a, b scalars
(iii) $\langle x, x \rangle > 0$, $x \neq 0$

4. If the matrix $\mathbf{A}$ is either symmetric, Hermitian, circulant, or triangular, its inverse has the same structure. The inverse of a Toeplitz matrix is *not* necessarily Toeplitz. The inverse of a Hankel matrix is symmetric.
5. The inverse of a block diagonal matrix equals the matrix consisting of the inverses of the blocks (if all of these individual inverses exist).

$$\begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \cdots & \cdots \\ \mathbf{0} & \mathbf{A}_2 & \mathbf{0} & \cdots \\ \vdots & & \ddots & \vdots \\ \cdots & \cdots & \mathbf{0} & \mathbf{A}_N \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_1^{-1} & \mathbf{0} & \cdots & \cdots \\ \mathbf{0} & \mathbf{A}_2^{-1} & \mathbf{0} & \cdots \\ \vdots & & \ddots & \vdots \\ \cdots & \cdots & \mathbf{0} & \mathbf{A}_N^{-1} \end{bmatrix}$$

- A *Givens rotation* is the multiplication of a matrix $\mathbf{A}$ by a unitary matrix $\mathbf{Q}$ that zeroes a specific element of $\mathbf{A}$ [29: pp. 43–7]. This simple operation lies at the core of many numerical algorithms for computing matrix inverses and eigensystems. The rotation matrix has the form of an identity matrix augmented by a square submatrix of sine and cosine rotation values. For example, the matrix that zeroes the second element of the fourth column of a 5×5 matrix has the form

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & c & 0 & s & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & -s & 0 & c & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

  where $c = A_{44} / \left(A_{24}^2 + A_{44}^2\right)^{1/2}$ and $s = A_{24} / \left(A_{24}^2 + A_{44}^2\right)^{1/2}$. The product $\mathbf{QA}$ performs the rotation by modifying only the second and fourth rows of $\mathbf{A}$. Column modification is possible by post-multiplying by the rotation matrix and suitably defining the rotation.
- The *determinant* of a square matrix $\mathbf{A}$ is denoted by $\det[\mathbf{A}]$ and is given by

$$\det[\mathbf{A}] = A_{11}\widetilde{A}_{11} + A_{12}\widetilde{A}_{12} + \cdots + A_{1n}\widetilde{A}_{1n},$$

  where $\widetilde{A}_{ij}$, the *cofactor* of $\mathbf{A}$, equals $(-1)^{j+1}$ times the determinant of the $(n-1) \times (n-1)$ matrix formed by deleting the i^{th} row and j^{th} column of $\mathbf{A}$. A non-zero determinant is a necessary and sufficient condition for the existence of the inverse of a matrix. The determinant of $\mathbf{A}'$ equals the conjugate of the determinant of $\mathbf{A}$: $\det[\mathbf{A}'] = \det[\mathbf{A}]^*$. The determinant of a unitary matrix has unity magnitude. The determinant of the product of two square matrices is the product of their determinants.

$$\det[\mathbf{AB}] = \det[\mathbf{A}] \det[\mathbf{B}]$$

  The determinant of a sum of matrices is *not* equal to the sum of their determinants. The determinant of a Kronecker product of square matrices equals $\det[\mathbf{A} \otimes \mathbf{B}] = \det[\mathbf{A}]^n \det[\mathbf{B}]^m$, where $\dim[\mathbf{A}] = m$ and $\dim[\mathbf{B}] = n$.
- The *trace* $\mathrm{tr}[\mathbf{A}]$ of the square matrix $\mathbf{A}$ equals the sum of its elements along the main diagonal.

$$\mathrm{tr}[\mathbf{A}] = \sum_{i=1}^{n} A_{ii}$$

  The trace of a sum of matrices equals the sum of their traces.

$$\mathrm{tr}[\mathbf{A} + \mathbf{B}] = \mathrm{tr}[\mathbf{A}] + \mathrm{tr}[\mathbf{B}]$$

  The trace of a product of two square matrices does *not* necessarily equal the product of the traces of the individual matrices; the product of two matrix's traces does equal the trace of their Kronecker product: $\mathrm{tr}[\mathbf{A} \otimes \mathbf{B}] = \mathrm{tr}[\mathbf{A}]\,\mathrm{tr}[\mathbf{B}]$. However, one of the more interesting properties of the trace is

$$\mathrm{tr}[\mathbf{AB}] = \mathrm{tr}[\mathbf{BA}].$$

Proving this property is straightforward. Using the full expression for the product of two matrices given above, $\mathrm{tr}[\mathbf{AB}] = \sum_i \sum_k A_{ik} B_{ki}$. Writing a similar expression for $\mathbf{BA}$ easily demonstrates this property. With this result, computation of the trace can be simplified in circumstances where the component matrices are not square; whichever product yields the smaller matrix can be used to compute the trace.

- The *gradient* with respect to a vector of a scalar-valued function $f(\mathbf{x})$ equals a column matrix of the partial derivatives.

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \mathrm{col}\left[\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_N}\right]$$

Examples of gradient calculation derived from this definition are:

$$\begin{aligned}
&\nabla_{\mathbf{x}}\, \mathbf{x}^t \mathbf{y} = \nabla_{\mathbf{x}}\, \mathbf{y}^t \mathbf{x} = \mathbf{y} \\
&\nabla_{\mathbf{x}}\, \mathbf{x}^t \mathbf{A} \mathbf{x} = 2\mathbf{A}\mathbf{x} \\
&\nabla_{\mathbf{x}}\, \mathrm{tr}[\mathbf{x}\mathbf{y}^t] = \nabla_{\mathbf{x}}\, \mathrm{tr}[\mathbf{y}\mathbf{x}^t] = \mathbf{y}
\end{aligned}$$

The gradient with respect to a matrix of a scalar-valued function can be defined similarly:

$$[\nabla_{\mathbf{A}} f(\mathbf{A})]_{ij} = \frac{\partial f}{\partial A_{ij}} .$$

Examples are:

$$\begin{aligned}
&\nabla_{\mathbf{A}}\, \mathrm{tr}[\mathbf{A}] = \mathbf{I} && \nabla_{\mathbf{A}}\, \mathrm{tr}[\mathbf{AB}] = \mathbf{B} \\
&\nabla_{\mathbf{A}}\, \mathrm{tr}[\mathbf{BA}] = \mathbf{B}^t && \nabla_{\mathbf{A}}\, \mathrm{tr}[\mathbf{B}\mathbf{A}^{-1}] = -[\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}]^t \\
&\nabla_{\mathbf{A}} \exp\{\mathbf{x}^t \mathbf{A} \mathbf{x}\} = \mathbf{x}\mathbf{x}^t \exp\{\mathbf{x}^t \mathbf{A} \mathbf{x}\} && \\
&\nabla_{\mathbf{A}}\, \det[\mathbf{A}] = \det[\mathbf{A}] \left(\mathbf{A}^{-1}\right)^t && \nabla_{\mathbf{A}}\, \ln \det[\mathbf{A}] = \left(\mathbf{A}^{-1}\right)^t
\end{aligned}$$

B.4 Quadratic Forms

Quadratic forms are key quantities of study in statistical signal processing. They are comprised by a Hermitian matrix $\mathbf{A}$ and a vector $\mathbf{x}$ to produce the real-valued scalar $\mathbf{x}'\mathbf{A}\mathbf{x}$. The matrix is termed the *kernel* of the quadratic form. In the special case that $\mathbf{A} = \mathbf{I}$, the quadratic form reduces to the inner product of $\mathbf{x}$ with itself; quadratic forms are thus considered generalizations of the inner product. If $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all non-zero choices of $\mathbf{x}$, the matrix $\mathbf{A}$ is said to be *positive definite*. If $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ under the same conditions, $\mathbf{A}$ is *non-negative definite*. The structure of Hermitian matrices, even if they are Toeplitz, is not sufficient to guarantee that they are non-negative definite. In contrast, Hankel matrices having no negative elements are always non-negative definite.

The argument of the exponential in the probability density function of a Gaussian random vector is a quadratic form. The kernel of that quadratic form is the covariance matrix of the random vector. Assuming zero mean for algebraic simplicity, the covariance matrix is the expected value of the outer product of the random vector $\mathbf{z}$ with itself: $\mathbf{A} = \mathcal{E}[\mathbf{z}\mathbf{z}']$. Such matrices are *always* non-negative definite. To show this result, simply consider an expanded version of the quadratic form: $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathcal{E}\left[|\mathbf{x}'\mathbf{z}|^2\right]$. Because of the squared magnitude, this quantity can never be negative.

Quadratic forms can be re-written using the trace operation. The trace of a quadratic form simply equals its value because a quadratic form is a scalar. Using the properties of the trace operation,

$$\mathrm{tr}\left[\mathbf{x}'\mathbf{A}\mathbf{x}\right] = \mathrm{tr}\left[\mathbf{A}\mathbf{x}\mathbf{x}'\right] .$$

Thus, a quadratic form equals the trace of the product between the kernel and the outer product of the vector with itself. This seemingly more complicated expression has utility in signal processing. See §2.1.9 {8} on the multivariate Gaussian distribution for the most well-known application of this formula.

Positive-definite Hermitian matrices can be expanded using the *Cholesky factorization*.

$$\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{L}'$$

where $\mathbf{D}$ is a diagonal matrix and $\mathbf{L}$ is lower-triangular with all of the diagonal elements equal to unity.

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ L_{21} & 1 & 0 & \cdots \\ L_{31} & L_{32} & \ddots & \ddots \\ \vdots & \vdots & & \ddots \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} D_{11} & 0 & 0 & \cdots \\ 0 & D_{22} & 0 & \cdots \\ \vdots & 0 & D_{33} & \ddots \\ \vdots & \vdots & & \ddots \end{bmatrix}$$

The inverse of a lower-triangular matrix is also lower-triangular; the inverse of the matrix $\mathbf{A}$ is written as $\mathbf{A}^{-1} = \mathbf{L}'^{-1}\mathbf{D}^{-1}\mathbf{L}^{-1}$.

In optimization problems, the quadratic form frequently represents a squared error cost function: the kernel imposes a shape to the range of values of $\mathbf{x}$ and the quadratic form grows quadratically with increasing the vector's length. Analytic solutions to such optimization problems are found by evaluating the *gradient* of the cost function with respect to the vector $\mathbf{x}$. The gradient with respect to $\mathbf{x}$ is evaluated by treating the vector's conjugate as a constant and *vice versa*. In this way, we find that

$$\begin{aligned} \nabla_{\mathbf{x}}\, \mathbf{x}'\mathbf{A}\mathbf{x} &= \mathbf{A}^t\mathbf{x}^* \\ \nabla_{\mathbf{x}^*}\, \mathbf{x}'\mathbf{A}\mathbf{x} &= \mathbf{A}\mathbf{x}. \end{aligned}$$

B.5 Matrix Eigenanalysis

One of the most powerful concepts in matrix algebra is eigenanalysis. Letting the matrix $\mathbf{A}$ be square, the non-zero vector $\mathbf{v}$ is said to be an *eigenvector* of $\mathbf{A}$ if it satisfies

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v},$$

where λ is the scalar termed the *eigenvalue* associated with $\mathbf{v}$. A scaled eigenvector is also an eigenvector. Because of this property, we define an eigenvector to *always* have unit inner product ($\mathbf{v}'\mathbf{v} = 1$). The *generalized eigenvector* $\mathbf{v}$ of the matrix pair $(\mathbf{A}, \mathbf{B})$ satisfies

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{B}\mathbf{v}\,.$$

Expressed in a slightly different form, the defining equation of eigenanalysis becomes $(\mathbf{A} - \lambda\mathbf{I})\,\mathbf{v} = \mathbf{0}$; for generalized eigenanalysis, $(\mathbf{A} - \lambda\mathbf{B})\,\mathbf{v} = \mathbf{0}$. If the matrices within the parentheses had inverses, the *only* solutions to these equations would be $\mathbf{v} = \mathbf{0}$, a trivial and uninteresting result. To find any non-zero solutions, the eigenvectors, these matrices must not have inverses, which implies that their determinants must be zero.

$$\det\left[\mathbf{A} - \lambda\mathbf{I}\right] = 0 \quad \det\left[\mathbf{A} - \lambda\mathbf{B}\right] = 0$$

For an $n \times n$ matrix, each equation becomes an n^{th} order polynomial. The eigenvalues of a matrix are the roots of this polynomial. Because no closed form expression exists for the roots of polynomials of greater than fourth order, the eigenvalues of a matrix must be found numerically in most cases of interest in signal processing. The n eigenvalues are conventionally labelled in decreasing order: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Eigenvalues are not necessarily unique; the number of times a value is repeated is termed its multiplicity. For example, every eigenvalue of the identity matrix equals unity and thus has multiplicity n. The eigenvalues of a sum or product of matrices are not easily expressed in terms of the eigenvalues of the component matrices. One remarkably simple result concerns Kronecker products: the eigenvalues of $\mathbf{A} \otimes \mathbf{B}$ equal $\lambda_i^{\mathbf{A}}\lambda_j^{\mathbf{B}}$, $i = 1, \ldots, m$, $j = 1, \ldots, n$. We term the set $\{\lambda_i, \mathbf{v}_i\}_{\mathbf{A}}$ of eigenvectors and associated eigenvalues the *eigensystem* of the matrix $\mathbf{A}$. For generalized systems to have n elements, the matrix $\mathbf{B}$ must be invertible. If not, the number of eigenvectors can be zero, less than n, or infinite [29: p. 252]. When invertible, the generalized eigensystem $\{\lambda_i, \mathbf{v}_i\}_{(\mathbf{A},\mathbf{B})}$ equals the eigensystem $\{\lambda_i, \mathbf{v}_i\}_{\mathbf{B}^{-1}\mathbf{A}}$.

Hermitian Matrices

In the special case where a matrix has Hermitian symmetry, several interesting properties arise. Because of the prevalence of the correlation matrix, which is Hermitian, this situation occurs often in practice.

- *The eigenvalues of a Hermitian matrix are real.* The inner product $\mathbf{v}'\mathbf{v}$ can be expressed as $\mathbf{v}'\mathbf{A}\mathbf{v} = \lambda\mathbf{v}'\mathbf{v}$. The left side of this equation equals its conjugate transpose; since the inner product $\mathbf{v}'\mathbf{v}$ is always real, we have $\lambda^* = \lambda$ and thus the eigenvalues are real.

- *If* $\mathbf{A}$ *is positive definite, all of its eigenvalues are positive.* The quadratic form $\mathbf{v}'\mathbf{A}\mathbf{v}$ is positive; as inner products are positive, $\mathbf{v}'\mathbf{A}\mathbf{v} = \lambda\mathbf{v}'\mathbf{v}$ implies that λ must also be positive.
- *The eigenvectors associated with distinct eigenvalues of a Hermitian matrix are orthogonal.* Consider $\mathbf{A}\mathbf{v}_1 = \lambda_1\mathbf{v}_1$ and $\mathbf{A}\mathbf{v}_2 = \lambda_2\mathbf{v}_2$ for $\lambda_1 \neq \lambda_2$. Noting that $\mathbf{v}_2'\mathbf{A}\mathbf{v}_1 = \lambda_1\mathbf{v}_2'\mathbf{v}_1$ and $\mathbf{v}_1'\mathbf{A}\mathbf{v}_2 = \lambda_2\mathbf{v}_1'\mathbf{v}_2$, these expressions differ on the left side only in that they are conjugate transposes of each other. Consequently, $(\lambda_2 - \lambda_1)\mathbf{v}_1'\mathbf{v}_2 = 0$, thereby indicating that $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal.

Define a matrix $\mathbf{V}$ having its columns to be the eigenvectors of the matrix $\mathbf{A}$.

$$\mathbf{V} = \mathrm{col}[\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$$

If $\mathbf{A}$ is Hermitian and has distinct eigenvalues, its eigenvectors are orthogonal, implying that $\mathbf{V}$ must satisfy $\mathbf{V}'\mathbf{V} = \mathbf{I}$ (i.e., $\mathbf{V}$ is unitary). Furthermore, the product $\mathbf{V}'\mathbf{A}\mathbf{V}$ is a diagonal matrix with the eigenvalues lying along the diagonal.

$$\mathbf{V}'\mathbf{A}\mathbf{V} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

Because $\mathbf{V}$ is unitary, we find that Hermitian matrices having distinct eigenvalues can be expressed

$$\mathbf{A} = \mathbf{V}\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}\mathbf{V}'.$$

This equation defines the matrix's *diagonal form.* From this diagonal form, the determinant and trace of $\mathbf{A}$ are easily related to the eigenvalues of $\mathbf{A}$ as

$$\begin{aligned} \det[\mathbf{A}] &= \prod_{i=1}^{n} \lambda_i \\ \mathrm{tr}[\mathbf{A}] &= \sum_{i=1}^{n} \lambda_i \end{aligned}$$

Expressing the diagonal form of $\mathbf{A}$ less concisely, an important conclusion can be drawn: *any Hermitian matrix* $\mathbf{A}$ *can be expressed with the expansion* $\mathbf{A} = \sum_i \lambda_i \mathbf{v}_i'\mathbf{v}_i$. This result is used frequently when eigenanalysis is applied to signal processing. In particular, quadratic forms can be expressed as

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{n} \lambda_i \left|\mathbf{x}'\mathbf{v}_i\right|^2 .$$

In addition, because $\mathbf{v} = \lambda\mathbf{A}^{-1}\mathbf{v}$, the eigenvectors of a matrix and its inverse are identical with the corresponding eigenvalues being reciprocals of each other. Therefore,

$$\mathbf{A}^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i}\mathbf{v}_i\mathbf{v}_i' .$$

For generalized eigenanalysis, orthogonality of eigenvectors is defined with respect to the matrix $\mathbf{B}$: $\mathbf{v}_i'\mathbf{B}\mathbf{v}_j = \delta_{ij}$. Then,*

$$\begin{aligned} \mathbf{A} &= \sum_{i=1}^{n} \lambda_i \mathbf{B}\mathbf{v}_i\mathbf{v}_i'\mathbf{B}' \\ \mathbf{A}^{-1} &= \sum_{i=1}^{n} \frac{1}{\lambda_i}\mathbf{v}_i\mathbf{v}_i' \end{aligned}$$

*Despite appearances, the second equation is correct. Test its veracity.

Rayleigh Quotient

An important quantity in matrix algebra easily manipulated in terms of eigenvectors and eigenvalues is the ratio of two quadratic forms.

$$R = \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{B}\mathbf{x}}$$

The matrices $\mathbf{A}$ and $\mathbf{B}$ are Hermitian and $\mathbf{B}$ must be positive definite. The key question asked is what vector $\mathbf{x}$ maximizes (or minimizes) this *Rayleigh quotient*. The simplest, and most prevalent case, occurs when $\mathbf{B} = \mathbf{I}$. The Rayleigh quotient is expressed in terms of the eigenvectors and eigenvalues of $\mathbf{A}$.

$$R = \frac{\sum_{i=1}^{n} \lambda_i |\mathbf{x}'\mathbf{v}_i|^2}{\mathbf{x}'\mathbf{x}}$$

The denominator is now seen to normalize the quotient: the quotient's value does not change when $\mathbf{x}$ is multiplied by a scalar. We can thus take the norm of $\mathbf{x}$ to be one and the denominator is no longer a concern. To maximize the numerator, the vector must be proportional to the eigenvector $\mathbf{v}_i$ corresponding to the maximum eigenvalue of $\mathbf{A}$. If not chosen this way, some of $\mathbf{x}$'s components would project onto other eigenvectors which necessarily have smaller eigenvalues and thus provide less weight to the sum. A similar argument applies if the minimizing vector is sought.

$$R_{\min} = \lambda_{\min}, \mathbf{x}_{\min} = \mathbf{v}_{\min} \qquad R_{\max} = \lambda_{\max}, \mathbf{x}_{\max} = \mathbf{v}_{\max}$$

Furthermore, the Rayleigh quotient is bounded by the minimum and maximum eigenvalues.

$$\lambda_{\min} \leq R \leq \lambda_{\max}$$

When $\mathbf{B}$ does not equal the identity matrix, we can use the generalized eigensystem of $(\mathbf{A}, \mathbf{B})$ to find the extrema of the Rayleigh quotient. When $\mathbf{x}$ is a generalized eigenvector, $\mathbf{A}\mathbf{x} = \lambda\mathbf{B}\mathbf{x}$, which implies that the Rayleigh quotient equal λ. Because $\mathbf{A}$ and $\mathbf{B}$ are Hermitian and positive definite, the generalized eigenvectors form a basis and the Rayleigh quotient obeys the above inequality for the generalized eigenvalues. We can express the Rayleigh quotient in other ways. First of all, we know that the generalized eigensystem equals the eigensystem of $\mathbf{B}^{-1}\mathbf{A}$. In addition, define $\tilde{\mathbf{x}} = \mathbf{B}^{1/2}\mathbf{x}$ where $\mathbf{B}^{1/2}$ is the square root of $\mathbf{B}$. Computationally, square roots may seem difficult to find; in eigenanalysis terms, a straightforward definition emerges.

$$\mathbf{B}^{1/2} \equiv \sum_{i=1}^{n} \lambda_i^{1/2} \mathbf{v}_i \mathbf{v}_i'$$

With this definition, the square root of a matrix satisfies the intuitive property $\mathbf{B}^{1/2}\mathbf{B}^{1/2} = \mathbf{B}$. The Rayleigh quotient thus becomes

$$R = \frac{\tilde{\mathbf{x}}' \left(\mathbf{B}^{-1/2}\right)' \mathbf{A} \left(\mathbf{B}^{-1/2}\right) \tilde{\mathbf{x}}}{\tilde{\mathbf{x}}'\tilde{\mathbf{x}}}$$

Thus, the vector $\tilde{\mathbf{x}}$ maximizing or minimizing the Rayleigh quotient corresponds to the eigenvector having maximum or minimum eigenvalues of the matrix product $\left(\mathbf{B}^{-1/2}\right)' \mathbf{A} \left(\mathbf{B}^{-1/2}\right)$.

Singular Value Decomposition

Related to eigenanalysis is the *singular value decomposition* or *SVD* technique. Letting $\mathbf{A}$ be an $m \times n$ matrix consisting of complex entries, it can be expressed by

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}',$$

where $\mathbf{D}$ is the $k \times k$ diagonal matrix $\mathrm{diag}[\sigma_1, \ldots, \sigma_k]$ and where $\mathbf{U}$ $(m \times k)$ and $\mathbf{V}$ $(k \times n)$ are matrices satisfying $\mathbf{U}'\mathbf{U} = \mathbf{I}_k$ and $\mathbf{V}'\mathbf{V} = \mathbf{I}_k$. Letting $\mathbf{u}_i$ and $\mathbf{v}_i$ denote the i^{th} columns of $\mathbf{U}$ and $\mathbf{V}$ respectively, then

$$\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i \quad \mathbf{u}_i'\mathbf{A} = \sigma_i \mathbf{v}_i'$$

The scalars σ_i are termed the singular values of the matrix $\mathbf{A}$ while $\mathbf{u}_i$ and $\mathbf{v}_i$ are termed the left and right (respectively) eigenvectors of $\mathbf{A}$. The number of non-zero singular values equals k, which must not exceed $\min(m, n)$. The number equals this upper limit when the side of the matrix having the smallest length (either the rows or the columns) has linearly independent components. From these expressions, the eigenvectors of $\mathbf{A}'\mathbf{A}$ and $\mathbf{A}\mathbf{A}'$ are found to be

$$\mathbf{A}'\mathbf{A}\mathbf{v}_i = \sigma_i^2\mathbf{v}_i \quad \mathbf{A}\mathbf{A}'\mathbf{u}_i = \sigma_i^2\mathbf{u}_i.$$

One of these products will not have full rank when the matrix $\mathbf{A}$ is not square. For example, if $\mathbf{A}$ is a column vector, $\mathbf{A}'\mathbf{A}$ is a scalar and thus is invertible while $\mathbf{A}\mathbf{A}'$ is an $m \times m$ matrix having only one non-zero singular value. These matrices are Hermitian; they share the same non-zero eigenvalues and these equal the *squares* of the singular values. Furthermore, the collections of vectors $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$ are each orthonormal sets. This property means that $\mathbf{U}\mathbf{U}' = \mathbf{I}_m$ and $\mathbf{V}\mathbf{V}' = \mathbf{I}_n$. The singular value decomposition suggests that all matrices have an orthonormal-like expansion of the form

$$\boxed{\mathbf{A} = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i'}.$$

No non-square matrix has an inverse. Singular value decomposition can be used to define the *pseudo inverse* of a rectangular matrix. Assuming that all of the singular values are non-zero, the pseudo inverse $\mathbf{A}^{\sim 1}$ satisfies either $\mathbf{A}\mathbf{A}^{\sim 1} = \mathbf{I}_m$ or $\mathbf{A}^{\sim 1}\mathbf{A} = \mathbf{I}_n$ according to which dimension of the matrix is the smallest. By analogy to the eigenvalue-eigenvector expansion of an invertible matrix, the singular value decomposition of a matrix's pseudo inverse would be

$$\mathbf{A}^{\sim 1} = \sum_{i=1}^{k} \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i'$$

or in matrix terms $\mathbf{A}^{\sim 1} = \mathbf{V}\mathbf{D}^{-1}\mathbf{U}'$. The pseudo inverses can be defined more directly by either $\mathbf{A}^{\sim 1} = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}$ or $\mathbf{A}^{\sim 1} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$. To be more concrete, suppose $\mathbf{A}$ has linearly independent rows. Its pseudo inverse is given by $\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}$ so that $\mathbf{A}\mathbf{A}^{\sim 1} = \mathbf{I}$: only the right inverse exists.

B.6 Projection Matrices

Eigenanalysis and singular value analysis are often used because they express a matrix's properties "naturally". One case in point is the projection matrix. A projection matrix $\mathbf{P}$ has the property $\mathbf{P}^m = \mathbf{P}^n$ for $m, n \geq 0$. The eigenvalues of a projection must therefore obey the relationship $\lambda^m = \lambda^n$. Two solutions are possible: $\lambda = 1$ and $\lambda = 0$. Thus, the eigenvectors of a projection matrix either have zero eigenvalues, in which case any vector that can be expressed by them is annihilated by the matrix ($\mathbf{P}\mathbf{x} = \mathbf{0}$), or the eigenvectors have unity eigenvalues and vectors comprised of them are unaffected by the matrix ($\mathbf{P}\mathbf{x} = \mathbf{x}$). If a vector has components belonging to both sets of eigenvectors, the part of the vector that "survives" the matrix is that portion represented by the eigenvectors having unity eigenvalues. Hence the origin of the name "projection matrix": matrices having only unity and zero eigenvalues can be used to find those components of any vector which correspond to the eigenvectors having unity eigenvalues.

For any $m \times n$ matrix $\mathbf{A}$ ($m < n$) that is used as a linear transformation, it defines a subspace for the *domain* of the transformation. In more exact terms, when the m-dimensional vectors $\mathbf{y} = \mathbf{A}\mathbf{x}$ define a vector space, what is the image of this space in the original, higher dimensional space containing the vectors $\mathbf{x}$? The most elegant response is in terms of the singular values and singular vectors of $\mathbf{A}$. The vector $\mathbf{y}$ is thereby expressed

$$\mathbf{y} = \sum_{i=1}^{m} \sigma_i (\mathbf{v}_i'\mathbf{x})\, \mathbf{u}_i \,.$$

The result of the linear transformation is always a linear combination of the left singular vectors of $\mathbf{A}$.

The projection matrix $\mathbf{P}_\mathbf{A}$ associated with a linear transformation $\mathbf{A}$ would project all n-dimensional vectors onto a subspace defined by the linear transformation. $\mathbf{P}_\mathbf{A}$ would have m unity eigenvalues for the left singular vectors of $\mathbf{A}$ and zero eigenvalues for all remaining $m - n$ eigenvectors. Defining $\mathbf{U}^\perp$ to be the matrix of eigenvectors corresponding to these zero eigenvalues, the projection matrix can be expressed as

$$\mathbf{P}_\mathbf{A} = \begin{bmatrix} \mathbf{U} & \mathbf{U}^\perp \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U} & \mathbf{U}^\perp \end{bmatrix}'.$$

This projection matrix can be expressed more directly in terms of the matrix $\mathbf{A}$ as $\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}$. Note than this expression can be written in terms of the transformation's pseudo inverse: while $\mathbf{A}\mathbf{A}^{\sim 1} = \mathbf{I}_m$, $\mathbf{P}_{\mathbf{A}} = \mathbf{A}^{\sim 1}\mathbf{A}$.

Appendix C

Ali-Silvey Distances

Ali-Silvey distances comprise a family of quantities that depend on the likelihood ratio $\Lambda(\mathbf{X})$ and on the model-describing densities p_0, p_1 in the following way.

$$d(p_0, p_1) = f\left(\mathcal{E}_0\left[c(\Lambda(\mathbf{X}))\right]\right)$$

Here, $f(\cdot)$ is a non-decreasing function, $c(\cdot)$ a convex function, and $\mathcal{E}_0[\cdot]$ means expectation with respect to p_0. Where applicable, π_0, π_1 denote the *a priori* probabilities of the models. Basseville [7] is good reference on distances in this class and many others. In all cases, the observations consist of L IID random variables.

Ali-Silvey Distances and Relation to Detection Performance			
Name	$c(\cdot)$	Performance	Comment
Kullback-Leibler $\mathcal{D}(p_1 \| p_0)$	$(\cdot)\log(\cdot)$	$\lim_{L\to\infty} -\frac{1}{L}\log P_F = d(p_0, p_1)$	Neyman-Pearson error rate under both fixed and exponentially decaying constraints on P_M (P_F)
Kullback-Leibler $\mathcal{D}(p_0 \| p_1)$	$-\log(\cdot)$	$\lim_{L\to\infty} -\frac{1}{L}\log P_M = d(p_0, p_1)$	
J-Divergence	$((\cdot)-1)\log(\cdot)$	$\pi_0\pi_1 \exp\{-d(p_0, p_1)/2\} \leq P_e$	$J(p_0, p_1) = \mathcal{D}(p_0 \| p_1) + \mathcal{D}(p_1 \| p_0)$
Chernoff	$(\cdot)^s, s \in (0,1)$	$\max \lim_{L\to\infty} -\frac{1}{L}\log P_e = \inf_{s\in(0,1)} d(p_0, p_1)$	Independent of *a priori* probabilities Not in Ali-Silvey class
M-Hypothesis Chernoff	$(\cdot)^s, s \in (0,1)$	$\max \lim_{L\to\infty} -\frac{1}{L}\log P_e = \min_{i\neq j} \inf_s d(p_i, p_j)$	
Bhattacharyya	$(\cdot)^{1/2}$	$\pi_0\pi_1[d(p_0, p_1)]^2 \leq P_e \leq \sqrt{\pi_0\pi_1}\, d(p_0, p_1)$	Minimizing $d(p_0, p_1)$ will tend to minimize P_e
Orsak	$\lvert\pi_1(\cdot) - \pi_0\rvert$	$P_e = \frac{1}{2} - \frac{1}{2}d(p_0, p_1)$	Exact formula for average error probability
Kolmogorov	$\frac{1}{2}\lvert(\cdot)-1\rvert$	If $\pi_0 = \pi_1$, $P_e = \frac{1}{2} - \frac{1}{2}d(p_0, p_1)$	
Hellinger	$((\cdot)^{1/2}-1)^2$		

Bibliography

1. M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions*. U.S. Government Printing Office, 1968.
2. H. Akaike. A new look at the statistical model identification problem. *IEEE Trans. Auto. Control*, AC-19:716–723, Dec. 1974.
3. S. T. Alexander. *Adaptive Signal Processing*. Springer-Verlag, New York, 1986.
4. B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Prentice Hall, Englewood Cliffs, NJ, 1979.
5. A. B. Baggeroer. Confidence intervals for regression (MEM) spectral estimates. *IEEE Trans. Info. Th.*, IT-22:534–545, Sept. 1976.
6. M. S. Bartlett. Smoothing periodograms from time series with continuous spectra. *Nature*, 161:686–687, 1948.
7. M. Basseville. Distance measures for signal processing and pattern recognition. *Signal Processing*, 18:349–369, 1989.
8. G. J. Bierman. *Factorization Methods for Discrete Sequential Estimation*. Academic Press, New York, 1977.
9. R. B. Blackman and J. W. Tukey. *The Measurement of Power Spectra*. Dover Press, New York, 1958.
10. J. P. Burg. *Maximum Entropy Spectral Analysis*. PhD thesis, Department of Geophysics, Stanford University, Stanford, CA, 1975.
11. J. P. Burg. The relationship between maximum entropy spectra and maximum likelihood spectra. *Geophysics*, 37:375–376, Apr. 1972.
12. D. Burshtein and E. Weinstein. Confidence intervals for the maximum entropy spectrum. *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-35:504–510, Apr. 1987.
13. J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE*, 57:1408–1418, 1969.
14. J. Capon. Maximum-likelihood spectral estimation. In S. Haykin, editor, *Nonlinear Methods of Spectral Analysis*, pages 155–179. Springer-Verlag, New York, 1979.
15. J. Capon and N. R. Goodman. Probability distributions for estimators of the frequency-wavenumber spectrum. *Proc. IEEE*, 58:1785–1786, Oct. 1970.
16. J. W. Carlyle and J. B. Thomas. On nonparametric signal detectors. *IEEE Trans. Info. Th.*, IT-10:146–152, Apr. 1964.
17. D. Chazan, M. Zakai, and J. Ziv. Improved lower bounds on signal parameter estimation. *IEEE Trans. Info. Th.*, IT-21:90–93, Jan. 1975.
18. H. Chernoff. Measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Stat.*, 23:493–507, 1952.
19. T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.
20. C. F. N. Cowan and P. M. Grant, editors. *Adaptive Filters*. Prentice Hall, Englewood Cliffs, NJ, 1985.
21. H. Cramér. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, NJ, 1946.
22. H. Cramér. *Random Variables and Probability Distributions*. Cambridge University Press, third edition, 1970.
23. J. E. Dennis, Jr. and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice Hall, Englewood Cliffs, NJ, 1983.
24. D. J. Edelblute, J. M. Fisk, and G. L. Kinnison. Criteria for optimum-signal-detection theory for arrays. *J. Acoust. Soc. Am.*, 41:199–205, Jan. 1967.
25. A. Feuer and E. Weinstein. Convergence analysis of LMS filters with uncorrelated Gaussian data. *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-33:222–229, Feb. 1985.

26. B. Friedlander. Adaptive algorithms for finite impulse response filters. In Cowan and Grant [20], pages 29–59.
27. A. Gelb, editor. *Applied Optimal Estimation*. MIT Press, Cambridge, MA, 1974.
28. J. D. Gibson and J. L. Melsa. *Introduction to Non-Parametric Detection with Applications*. Academic Press, New York, 1975.
29. G. H. Golub and C. F. van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, 1983.
30. P. Hall. *Rates of convergence in the central limit theorem*, volume 62 of *Research Notes in Mathematics*. Pitman Advanced Publishing Program, 1982.
31. F. J. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proc. IEEE*, 66:51–83, Jan. 1978.
32. S. Haykin. *Adaptive Filter Theory*. Prentice Hall, Englewood Cliffs, NJ, 1986.
33. C. W. Helstrom. *Statistical Theory of Signal Detection*. Pergamon Press, Oxford, second edition, 1968.
34. M. L. Honig and D. G. Messerschmitt. *Adaptive Filters*. Kluwer Academic Publishers, Boston, 1984.
35. E. T. Jaynes. On the rationale of the maximum entropy method. *Proc. IEEE*, 70:939–952, Sept. 1982.
36. G. M. Jenkins and D. G. Watts. *Spectral Analysis and its Applications*. Holden-Day, Inc., San Francisco, 1968.
37. D. H. Johnson. The application of spectral estimation methods to bearing estimation problems. *Proc. IEEE*, 70:1018–1028, Sept. 1982.
38. D. H. Johnson and G. C. Orsak. Relation of signal set choice to the performance of optimal non-Gaussian detectors. *IEEE Trans. Comm.*, 41:1319–1328, 1993.
39. R. E. Kalman. A new approach to linear filtering and prediction problems. *J. Basic Engineering* (ASME Trans.), 82D:35–45, 1960.
40. R. L. Kashyap. Inconsistency of the AIC rule for estimating the order of autoregressive models. *IEEE Trans. Auto. Control*, AC-25:996–998, Oct. 1980.
41. S. M. Kay. *Modern Spectral Estimation*. Prentice Hall, Englewood Cliffs, NJ, 1988.
42. E. J. Kelly, I. S. Reed, and W. L. Root. The detection of radar echoes in noise. I. *J. Soc. Indust. Appl. Math.*, 8:309–341, June 1960.
43. E. J. Kelly, I. S. Reed, and W. L. Root. The detection of radar echoes in noise. II. *J. Soc. Indust. Appl. Math.*, 8:481–507, Sept. 1960.
44. L. H. Koopmans. *The Spectral Analysis of Time Series*. Academic Press, New York, 1974.
45. R. T. Lacoss. Data adaptive spectral analysis methods. *Geophysics*, 36:661–675, Aug. 1971.
46. E. L. Lehmann. *Testing Statistical Hypotheses*. John Wiley & Sons, New York, second edition, 1986.
47. R. S. Lipster and A. N. Shiryayev. *Statistics of Random Processes I: General Theory*. Springer-Verlag, New York, 1977.
48. L. Ljung, M. Morf, and D. Falconer. Fast calculation of gain matrices for recursive estimation techniques. *Inter. J. Control*, 27:1–19, Jan. 1978.
49. J. Makhoul. Linear prediction: A tutorial review. *Proc. IEEE*, 63:561–580, Apr. 1975.
50. J. D. Markel and A. H. Gray, Jr. *Linear Prediction of Speech*. Springer-Verlag, New York, 1976.
51. S. L. Marple, Jr. *Digital Spectral Analysis*. Prentice Hall, Englewood Cliffs, NJ, 1987.
52. D. P. McGinn and D. H. Johnson. Estimation of all-pole model parameters from noise-corrupted sequences. *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-37:433–436, Mar. 1989.
53. B. R. Musicus. Fast MLM power spectrum estimation from uniformly spaced correlations. *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-33:1333–1335, Oct. 1985.
54. J. Neyman and E. S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. Roy. Soc. Ser. A*, 231:289–337, Feb. 1933.
55. A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 1989.
56. A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, second edition, 1984.
57. E. Parzen. *Stochastic Processes*. Holden-Day, San Francisco, 1962.
58. H. V. Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, New York, 1988.
59. J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, Sept. 1978.

60. H. Sakai and H. Tokumaru. Statistical analysis of a spectral estimator for ARMA processes. *IEEE Trans. Auto. Control*, AC-25:122–124, Feb. 1980.
61. G. Schwarz. Estimating the dimension of a model. *Ann. Stat.*, 6:461–464, Mar. 1978.
62. B. W. Silverman. *Density Estimation*. Chapman & Hall, London, 1986.
63. D. L. Snyder. *Random Point Processes*. Wiley, New York, 1975.
64. J. R. Thompson and R. A. Tapia. *Nonparametric Function Estimation, Modeling, and Simulation*. SIAM, Philadelphia, PA, 1990.
65. H. L. van Trees. *Detection, Estimation, and Modulation Theory, Part I*. John Wiley & Sons, New York, 1968.
66. A. Wald. *Sequential Analysis*. John Wiley & Sons, New York, 1947.
67. M. T. Wasan. *Stochastic Approximation*. Cambridge University Press, 1969.
68. A. J. Weiss and E. Weinstein. Fundamental limitations in passive time delay estimation: I. Narrow-band systems. *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-31:472–486, Apr. 1983.
69. P. D. Welch. The use of the fast Fourier transform for the estimation of power spectra: A method based on time averaging over short modified periodograms. *IEEE Trans. Audio Electroacoust.*, AU-15:70–73, Mar. 1967.
70. B. Widrow and S. D. Stearns. *Adaptive Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 1985.
71. N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. MIT Press, Cambridge, MA, 1949.
72. R. P. Wishner. Distribution of the normalized periodogram detector. *IRE Trans. Info. Th.*, IT-8:342–349, Sept. 1962.
73. P. M. Woodward. *Probability and Information Theory, with Applications to Radar*. Pergamon Press, Oxford, second edition, 1964.
74. J. Ziv and M. Zakai. Some lower bounds on signal parameter estimation. *IEEE Trans. Info. Th.*, IT-15:386–391, May 1969.