

INFERENCE OVER NETWORKS

LECTURE #8: Useful Matrix Results

Professor Ali H. Sayed
UCLA Electrical Engineering





References

[Appendix F \(Useful Matrix and Convergence Results, pp. 761-776\):](#)

A. H. Sayed, "Adaptation, learning, and optimization over networks," ***Foundations and Trends in Machine Learning***, vol. 7, issue 4-5, pp. 311-801, NOW Publishers, 2014.

[Appendices C \(Stochastic Matrices\) and D \(Block Maximum Norm\):](#)

A. H. Sayed, "Diffusion adaptation over networks," in ***Academic Press Library in Signal Processing***, vol. 3, R. Chellapa and S. Theodoridis, editors, pp. 323-454, Academic Press, Elsevier, 2014.

Kronecker Products

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.



Kronecker Products

Traditional Kronecker Form

Let $A = [a_{ij}]_{i,j=1}^n$ and $B = [b_{ij}]_{i,j=1}^m$ be $n \times n$ and $m \times m$ possibly complex-valued matrices, respectively, whose individual (i, j) -th entries are denoted by a_{ij} and b_{ij} . Their Kronecker product is denoted by $K = A \otimes B$ and is defined as the $nm \times nm$ matrix whose entries are given by [104, 113]:

$$K \triangleq A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & & \ddots & \\ a_{n1}B & a_{n2}B & \dots & a_{nn}B \end{bmatrix} \quad (\text{F.1})$$



Kronecker Products

Let $\{\lambda_i(A), i = 1, \dots, n\}$ and $\{\lambda_j(B), j = 1, \dots, m\}$ denote the eigenvalues of A and B , respectively. Then, the eigenvalues of $A \otimes B$ will consist of all nm product combinations $\{\lambda_i(A)\lambda_j(B)\}$. A similar conclusion holds for the singular values of $A \otimes B$ in relation to the singular values of the individual matrices A and B , which we denote by $\{\sigma_i(A), \sigma_j(B)\}$. [Table F.1](#) lists some well-known properties of Kronecker products for matrices $\{A, B, C, D\}$ of compatible dimensions and column vectors $\{x, y\}$.

Kronecker Products



Table F.1: Properties of the traditional Kronecker product definition (F.1).

-
1. $(A + B) \otimes C = (A \otimes C) + (B \otimes C)$
 2. $(A \otimes B)(C \otimes D) = (AC \otimes BD)$
 3. $(A \otimes B)^\top = A^\top \otimes B^\top$
 4. $(A \otimes B)^* = A^* \otimes B^*$
 5. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
 6. $(A \otimes B)^\ell = A^\ell \otimes B^\ell$
 7. $\{\lambda(A \otimes B)\} = \{\lambda_i(A)\lambda_j(B)\}_{i=1, j=1}^{n,m}$
 8. $\{\sigma(A \otimes B)\} = \{\sigma_i(A)\sigma_j(B)\}_{i=1, j=1}^{n,m}$
 9. $\det(A \otimes B) = (\det A)^m (\det B)^n$
 10. $\text{Tr}(A \otimes B) = \text{Tr}(A)\text{Tr}(B)$
 11. $\text{Tr}(AB) = [\text{vec}(B^\top)]^\top \text{vec}(A) = [\text{vec}(B^*)]^* \text{vec}(A)$
 12. $\text{vec}(ACB) = (B^\top \otimes A)\text{vec}(C)$
 13. $\text{vec}(xy^\top) = y \otimes x$
-



Block Kronecker Products

Block Kronecker Form

Let \mathcal{A} now denote a *block* matrix of size $np \times np$ with each block having size $p \times p$. We denote the (i, j) —th sub-matrix of \mathcal{A} by the notation A_{ij} ; it is a block of size $p \times p$. Likewise, we let \mathcal{B} denote a second *block* matrix of size $mp \times mp$ with each of its blocks having the same size $p \times p$. We denote the (i, j) —th sub-matrix of \mathcal{B} by the notation B_{ij} ; it is a block of size $p \times p$. The *block* Kronecker product of these two matrices is denoted by $\mathcal{K} = \mathcal{A} \otimes_b \mathcal{B}$ and is defined as the following block matrix of dimensions $nmp^2 \times mnp^2$ [146]:



Block Kronecker Products

$$\mathcal{K} \triangleq \mathcal{A} \otimes_b \mathcal{B} = \begin{bmatrix} K_{11} & K_{12} & \dots & K_{1n} \\ K_{21} & K_{22} & \dots & K_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ K_{n1} & K_{n2} & \dots & K_{nn} \end{bmatrix} \quad (\text{F.2})$$

where each block K_{ij} is $mp^2 \times mp^2$ and is constructed as follows:



Block Kronecker Products

$$K_{ij} = \begin{bmatrix} A_{ij} \otimes B_{11} & A_{ij} \otimes B_{12} & \dots & A_{ij} \otimes B_{1m} \\ A_{ij} \otimes B_{21} & A_{ij} \otimes B_{22} & \dots & A_{ij} \otimes B_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{ij} \otimes B_{m1} & A_{ij} \otimes B_{m2} & \dots & A_{ij} \otimes B_{mm} \end{bmatrix} \quad (\text{F.3})$$

Table F.2 lists some useful properties of block Kronecker products for matrices $\{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}\}$ with blocks of size $p \times p$. The last three properties involve the block vectorization operation denoted by bvec : it vectorizes each block entry of the matrix and then stacks the resulting columns on top of each other, i.e.,

Block Kronecker Products



$$\begin{aligned} \text{bvec}(\mathcal{A}) &\triangleq \text{col} \{ \text{vec}(A_{11}), \text{vec}(A_{21}), \dots, \text{vec}(A_{n1}), \\ &\quad \text{vec}(A_{21}), \text{vec}(A_{22}), \dots, \text{vec}(A_{n2}), \\ &\quad \vdots \\ &\quad \text{vec}(A_{1n}), \text{vec}(A_{2n}), \dots, \text{vec}(A_{nn}) \} \end{aligned} \tag{F.4}$$



Block Kronecker Products

Table F.2: Properties of the block Kronecker product definition (F.2).

-
1. $(\mathcal{A} + \mathcal{B}) \otimes_b \mathcal{C} = (\mathcal{A} \otimes_b \mathcal{C}) + (\mathcal{B} \otimes_b \mathcal{C})$
 2. $(\mathcal{A} \otimes_b \mathcal{B})(\mathcal{C} \otimes_b \mathcal{D}) = (\mathcal{AC} \otimes_b \mathcal{BD})$
 3. $(\mathcal{A} \otimes \mathcal{B}) \otimes_b (\mathcal{C} \otimes \mathcal{D}) = (\mathcal{A} \otimes \mathcal{C}) \otimes (\mathcal{B} \otimes \mathcal{D})$
 4. $(\mathcal{A} \otimes_b \mathcal{B})^\top = \mathcal{A}^\top \otimes_b \mathcal{B}^\top$
 5. $(\mathcal{A} \otimes_b \mathcal{B})^* = \mathcal{A}^* \otimes_b \mathcal{B}^*$
 6. $\{\lambda(\mathcal{A} \otimes_b \mathcal{B})\} = \{\lambda_i(\mathcal{A})\lambda_j(\mathcal{B})\}_{i=1, j=1}^{np, mp}$
 7. $\text{Tr}(\mathcal{AB}) = [\text{bvec}(\mathcal{B}^\top)]^\top \text{bvec}(\mathcal{A}) = [\text{bvec}(\mathcal{B}^*)]^* \text{bvec}(\mathcal{A})$
 8. $\text{bvec}(\mathcal{ACB}) = (\mathcal{B}^\top \otimes_b \mathcal{A}) \text{bvec}(\mathcal{C})$
 9. $\text{bvec}(xy^\top) = y \otimes_b x$
-



Block Kronecker Products

Figure F.1 illustrates one of the advantages of working with the `bvec` operation for block matrices [279]. The figure compares the effect of the block vectorization operation to that of the regular `vec` operation. It is seen that `bvec` preserves the locality of the blocks from the original matrix: entries arising from the same block appear together followed by entries of the other successive blocks. In contrast, in the `vec` construction, entries from different blocks are blended together.



Block Kronecker Products

13

Lecture #8: Useful Matrix Results

EE210B: Inference over Networks (A. H. Sayed)

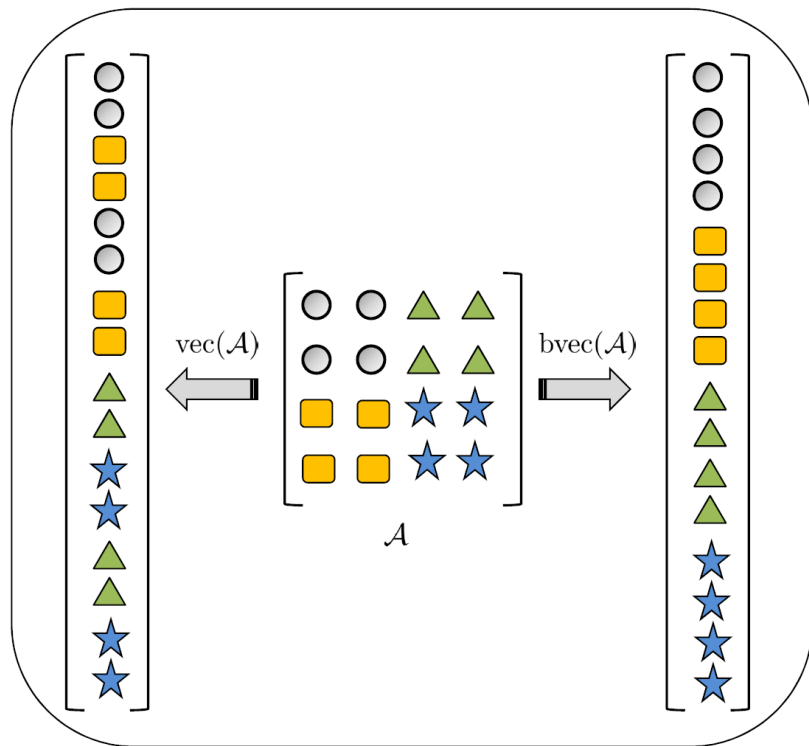


Figure F.1: Schematic comparison of the regular and block vectorization operations. It is seen that the bvec operation preserves the locality of the blocks from the original matrix, while the entries of the blocks get mixed up in the regular vec operation.

Vector and Matrix Norms

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.



Vector Norms

For any vector x of size $N \times 1$ and entries $\{x_k\}$, any of the definitions listed in [Table F.3](#) constitutes a valid vector norm.

Table F.3: Useful vector norms, where the $\{x_k\}$ denote the entries of $x \in \mathbb{C}^N$.

$\ x\ _1 \triangleq \sum_{k=1}^N x_k $	(1-norm)
$\ x\ _\infty \triangleq \max_{1 \leq k \leq N} x_k $	(∞ -norm)
$\ x\ _2 \triangleq \left(\sum_{k=1}^N x_k ^2 \right)^{1/2}$	(Euclidean norm)
$\ x\ _p \triangleq \left(\sum_{k=1}^N x_k ^p \right)^{1/p}$	(p -norm, for any real $p \geq 1$)

Matrix Norms



There are similarly many useful matrix norms. For any matrix A of dimensions $N \times N$ and entries $\{a_{\ell k}\}$, any of the definitions listed in [Table F.4](#) constitutes a valid matrix norm. In particular, the 2-induced norm of A is equal to its largest singular value:

$$\|A\|_2 = \sigma_{\max}(A) \tag{F.5}$$

Matrix Norms



Table F.4: Useful matrix norms, where the $\{a_{\ell k}\}$ denote the entries of $A \in \mathbb{C}^{N \times N}$.

$\ A\ _1 \triangleq \max_{1 \leq k \leq N} \left(\sum_{\ell=1}^N a_{\ell k} \right)$	(1–norm, or maximum absolute column sum)
$\ A\ _\infty \triangleq \max_{1 \leq \ell \leq N} \left(\sum_{k=1}^N a_{\ell k} \right)$	(∞ –norm, or maximum absolute row sum)
$\ A\ _F \triangleq \sqrt{\text{Tr}(A^* A)}$	(Frobenius norm)
$\ A\ _p \triangleq \max_{x \neq 0} \left(\frac{\ Ax\ _p}{\ x\ _p} \right)$	(p –induced norm for any real $p \geq 1$)



Equivalent Matrix Norms

A fundamental result in matrix theory is that all matrix norms in finite dimensional spaces are *equivalent*. Specifically, if $\|A\|_a$ and $\|A\|_b$ denote two generic matrix norms, then there exist positive constants c_ℓ and c_u that bound one norm by the other from above and from below such as [104, 113]:

$$c_\ell \|A\|_b \leq \|A\|_a \leq c_u \|A\|_b \quad (\text{F.6})$$

The values of $\{c_\ell, c_u\}$ are independent of the matrix entries though they may be dependent on the matrix dimensions. Vector norms are also equivalent to each other.



One Useful Matrix Norm

Let B denote an $N \times N$ matrix with eigenvalues $\{\lambda_k\}$. The spectral radius of B , denoted by $\rho(B)$, is defined as

$$\rho(B) \triangleq \max_{1 \leq k \leq N} |\lambda_k| \quad (\text{F.7})$$

We introduce the Jordan canonical decomposition of B and write $B = TJT^{-1}$, where T is an invertible transformation and J is a block diagonal matrix, say, with q blocks:

$$J = \text{diag}\{J_1, J_2, \dots, J_q\} \quad (\text{F.8})$$



One Useful Matrix Norm

Each block J_q has a Jordan structure with an eigenvalue λ_q on its diagonal entries, unit entries on the first sub-diagonal, and zeros everywhere else. For example, for a block of size 4×4 :

$$J_q = \begin{bmatrix} \lambda_q & & & \\ 1 & \lambda_q & & \\ & 1 & \lambda_q & \\ & & 1 & \lambda_q \end{bmatrix} \quad (\text{F.9})$$



One Useful Matrix Norm

Let ϵ denote an arbitrary positive scalar that we are free to choose and define the $N \times N$ diagonal scaling matrix:

$$D \triangleq \text{diag} \left\{ \epsilon, \epsilon^2, \dots, \epsilon^N \right\} \quad (\text{F.10})$$

Following Lemma 5.6.10 from [113] and Problem 14.19 from [133], we can use the quantity T originating from B to define the following matrix norm, denoted by $\|\cdot\|_\rho$, for any matrix A of size $N \times N$:

$$\|A\|_\rho \triangleq \left\| DT^{-1}ATD^{-1} \right\|_1 \quad (\text{F.11})$$



One Useful Matrix Norm

in terms of the 1–norm (i.e., maximum absolute column sum) of the matrix product on the right-hand side. It is not difficult to verify that the transformation (F.11) is a valid matrix norm, namely, that it satisfies the following properties, for any matrices A and C of compatible dimensions and for any complex scalar α :



One Useful Matrix Norm

24

Lecture #8: Useful Matrix Results

EE210B: Inference over Networks (A. H. Sayed)

$$\left\{ \begin{array}{l} \text{(a) } \|A\|_\rho \geq 0 \text{ with } \|A\|_\rho = 0 \text{ if, and only if, } A = 0 \\ \text{(b) } \|\alpha A\|_\rho = |\alpha| \|A\|_\rho \\ \text{(c) } \|A + C\|_\rho \leq \|A\|_\rho + \|C\|_\rho \text{ (triangular inequality)} \\ \text{(d) } \|AC\|_\rho \leq \|A\|_\rho \|C\|_\rho \text{ (sub-multiplicative property)} \end{array} \right. \quad (\text{F.12})$$



One Useful Matrix Norm

25

Lecture #8: Useful Matrix Results

EE210B: Inference over Networks (A. H. Sayed)

One important property of the ρ -norm defined by (F.11) is that when it is applied to the matrix B itself, it will hold that:

$$\rho(B) \leq \|B\|_\rho \leq \rho(B) + \epsilon \quad (\text{F.13})$$

That is, the ρ -norm of B will be sandwiched between two bounds defined by its spectral radius. It follows that if the matrix B is stable to begin with, so that $\rho(B) < 1$, then we can always select ϵ small enough to ensure $\|B\|_\rho < 1$.



Statement

The matrix norm defined by (F.11) is also an induced norm relative to the following vector norm:

$$\|x\|_\rho \triangleq \|DT^{-1}x\|_1 \quad (\text{F.14})$$

That is, for any matrix A , it holds that

$$\|A\|_\rho = \max_{x \neq 0} \left(\frac{\|Ax\|_\rho}{\|x\|_\rho} \right) \quad (\text{F.15})$$

Proof



Proof. Indeed, using (F.14), we first note that for any vector $x \neq 0$:

$$\begin{aligned}\|Ax\|_\rho &= \|DT^{-1}Ax\|_1 \\ &= \|DT^{-1}A \cdot TD^{-1}DT^{-1} \cdot x\|_1 \\ &\leq \|DT^{-1}ATD^{-1}\|_1 \cdot \|DT^{-1}x\|_1 \\ &= \|A\|_\rho \cdot \|x\|_\rho\end{aligned}\tag{F.16}$$

so that

$$\max_{x \neq 0} \left(\frac{\|Ax\|_\rho}{\|x\|_\rho} \right) \leq \|A\|_\rho\tag{F.17}$$

Proof



To show that equality holds in (F.17), it is sufficient to exhibit one nonzero vector x_o that attains equality. Let k_o denote the index of the column that attains the maximum absolute column sum in the matrix product $DT^{-1}ATD^{-1}$. Let e_{k_o} denote the column basis vector of size $N \times 1$ with one at location k_o and zeros elsewhere. Select

$$x_o \triangleq TD^{-1}e_{k_o} \tag{F.18}$$

Then, it is straightforward to verify that

$$\|x_o\|_\rho \triangleq \|DT^{-1}x_o\|_1 \stackrel{\text{(F.18)}}{=} \|e_{k_o}\|_1 = 1 \tag{F.19}$$

Proof



and

$$\begin{aligned}\|Ax_o\|_\rho &\stackrel{\Delta}{=} \|DT^{-1}Ax_o\|_1 \\ &= \|DT^{-1}A \cdot TD^{-1}DT^{-1} \cdot x_o\|_1 \\ &\stackrel{\text{(F.18)}}{=} \|DT^{-1}ATD^{-1}e_{k_o}\|_1 \\ &= \|A\|_\rho\end{aligned}\tag{F.20}$$

so that, for this particular vector, we have

$$\frac{\|Ax_o\|_\rho}{\|x_o\|_\rho} = \|A\|_\rho\tag{F.21}$$

as desired.





Block Maximum Norm

Let $x = \text{col}\{x_1, x_2, \dots, x_N\}$ now denote an $N \times 1$ *block* column vector whose individual entries are themselves vectors of size $M \times 1$ each. Following [32, 209, 231, 233], the block maximum norm of x is denoted by $\|x\|_{b,\infty}$ and is defined as

$$\|x\|_{b,\infty} \triangleq \max_{1 \leq k \leq N} \|x_k\| \quad (\text{F.22})$$



Block Maximum Norm

That is, $\|x\|_{b,\infty}$ is equal to the largest Euclidean norm of its block components. This vector norm induces a block maximum matrix norm. Let \mathcal{A} denote an arbitrary $N \times N$ block matrix with individual block entries of size $M \times M$ each. Then, the block maximum norm of \mathcal{A} is defined as

$$\|\mathcal{A}\|_{b,\infty} \triangleq \max_{x \neq 0} \left(\frac{\|\mathcal{A}x\|_{b,\infty}}{\|x\|_{b,\infty}} \right) \quad (\text{F.23})$$

The block maximum norm has several useful properties — see [\[209\]](#).

Properties



Lemma F.1 (Some useful properties of the block maximum norm). The block maximum norm satisfies the following properties:

(a) Let $\mathcal{U} = \text{diag}\{U_1, U_2, \dots, U_N\}$ denote an $N \times N$ block diagonal matrix with $M \times M$ unitary blocks $\{U_k\}$. Then, the block maximum norm is unitary-invariant, i.e., $\|\mathcal{U}x\|_{b,\infty} = \|x\|_{b,\infty}$ and $\|\mathcal{U}\mathcal{A}\mathcal{U}^*\|_{b,\infty} = \|\mathcal{A}\|_{b,\infty}$.

(b) Let $\mathcal{D} = \text{diag}\{D_1, D_2, \dots, D_N\}$ denote an $N \times N$ block diagonal matrix with $M \times M$ Hermitian blocks $\{D_k\}$. Then, $\rho(\mathcal{D}) = \|\mathcal{D}\|_{b,\infty}$.



Properties

(c) Let A be an $N \times N$ matrix and define $\mathcal{A} = A \otimes I_M$ whose blocks are therefore of size $M \times M$ each. If A is left-stochastic (as defined further ahead by (F.46)), then $\|\mathcal{A}^\top\|_{b,\infty} = 1$.

(d) Consider a block diagonal matrix \mathcal{D} as in part (b) and any left-stochastic matrices \mathcal{A}_1 and \mathcal{A}_2 constructed as in part (c). Then, it holds that

$$\rho(\mathcal{A}_2^\top \mathcal{D} \mathcal{A}_1^\top) \leq \rho(\mathcal{D}) \quad (\text{F.24})$$

Jensen's Inequality

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.



Jensen's Inequality

There are several variations and generalizations of Jensen's inequality. One useful form for our purposes is the following. Let $\{w_k\}$ denote a collection of N possibly complex-valued column vectors for $k = 1, 2, \dots, N$. Let $\{\alpha_k\}$ denote a collection of nonnegative real coefficients that add up to one:

$$\sum_{k=1}^N \alpha_k = 1, \quad 0 \leq \alpha_k \leq 1 \quad (\text{F.25})$$



Jensen's Inequality

Jensen's inequality states that for any real-valued convex function $f(x) \in \mathbb{R}$, it holds [45, 126, 172]:

$$f\left(\sum_{k=1}^N \alpha_k w_k\right) \leq \sum_{k=1}^N \alpha_k f(w_k) \quad (\text{F.26})$$

In particular, let

$$z \triangleq \sum_{k=1}^N \alpha_k w_k \quad (\text{F.27})$$



Jensen's Inequality

If we select the function $f(z) = \|z\|^2$ in terms of the squared Euclidean norm of the vector z , then it follows from (F.26) that

$$\left\| \sum_{k=1}^N \alpha_k w_k \right\|^2 \leq \sum_{k=1}^N \alpha_k \|w_k\|^2 \quad (\text{F.28})$$



Jensen's Inequality

There is also a useful stochastic version of Jensen's inequality. If $\mathbf{a} \in \mathbb{R}^M$ is a real-valued random variable, then it holds that

$$f(\mathbb{E} \mathbf{a}) \leq \mathbb{E}(f(\mathbf{a})) \quad (\text{when } f(x) \in \mathbb{R} \text{ is convex}) \quad (\text{F.29})$$

$$f(\mathbb{E} \mathbf{a}) \geq \mathbb{E}(f(\mathbf{a})) \quad (\text{when } f(x) \in \mathbb{R} \text{ is concave}) \quad (\text{F.30})$$

where it is assumed that \mathbf{a} and $f(\mathbf{a})$ have bounded expectations. We remark that a function $f(x)$ is said to be concave if, and only if, $-f(x)$ is convex.

Perturbation Bounds On Eigenvalues

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.



Weyl's Theorem

The first result, known as Weyl's Theorem [113, 260], shows how the eigenvalues of a Hermitian matrix are disturbed through additive perturbations to the entries of the matrix. Thus, let $\{A', A, \Delta A\}$ denote arbitrary $N \times N$ Hermitian matrices with ordered eigenvalues $\{\lambda_m(A'), \lambda_m(A), \lambda_m(\Delta A)\}$, i.e.,

$$\lambda_1(A) \geq \lambda_2(A) \geq \dots \geq \lambda_N(A) \quad (\text{F.31})$$



Weyl's Theorem

and similarly for the eigenvalues of $\{A', \Delta A\}$, with the subscripts 1 and N representing the largest and smallest eigenvalues, respectively. Weyl's Theorem states that if A is perturbed to

$$A' = A + \Delta A \quad (\text{F.32})$$

then the eigenvalues of the new matrix are bounded as follows:

$$\lambda_n(A) + \lambda_N(\Delta A) \leq \lambda_n(A') \leq \lambda_n(A) + \lambda_1(\Delta A) \quad (\text{F.33})$$



Weyl's Theorem

for $1 \leq n \leq N$. In particular, it follows that the maximum eigenvalue is perturbed as follows:

$$\lambda_{\max}(A) + \lambda_{\min}(\Delta A) \leq \lambda_{\max}(A') \leq \lambda_{\max}(A) + \lambda_{\max}(\Delta A) \quad (\text{F.34})$$

In the special case when $\Delta A \geq 0$, we conclude from (F.33) that $\lambda_n(A') \geq \lambda_n(A)$ for all $n = 1, 2, \dots, N$.



Gershgorin's Theorem

The second result, known as Gershgorin's Theorem [48, 94, 101, 104, 113, 254, 264], specifies circular regions within which the eigenvalues of a matrix are located. Thus, consider an $N \times N$ matrix A with scalar entries $\{a_{\ell k}\}$. With each diagonal entry $a_{\ell\ell}$ we associate a disc in the complex plane centered at $a_{\ell\ell}$ and with

$$r_{\ell} \triangleq \sum_{k \neq \ell, k=1}^N |a_{\ell k}| \quad (\text{F.35})$$



Gershgorin's Theorem

That is, r_ℓ is equal to the sum of the magnitudes of the non-diagonal entries on the same row as $a_{\ell\ell}$. We denote the disc by D_ℓ ; it consists of all points that satisfy

$$D_\ell = \left\{ z \in \mathbb{C}^N \text{ such that } |z - a_{\ell\ell}| \leq r_\ell \right\} \quad (\text{F.36})$$

The theorem states that the spectrum of A (i.e., the set of all its eigenvalues, denoted by $\lambda(A)$) is contained in the union of all N Gershgorin discs:

$$\lambda(A) \subset \bigcup_{\ell=1}^N D_\ell \quad (\text{F.37})$$



Gershgorin's Theorem

A stronger statement of the Gershgorin theorem covers the situation in which some of the Gershgorin discs happen to be disjoint. Specifically, if the union of L of the discs is disjoint from the union of the remaining $N - L$ discs, then the theorem further asserts that L eigenvalues of A will lie in the first union of L discs and the remaining $N - L$ eigenvalues of A will lie in the second union of $N - L$ discs.

Lyapunov Equations

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.

Setting



In this section, we introduce two particular Lyapunov equations and list some of their properties. We only list results that are used in the text. There are many other insightful results on Lyapunov equations. Interested readers may consult the works [132, 133, 149, 150] and the many references therein for additional information.



Discrete-Time Lyapunov Eqs

48

Lecture #8: Useful Matrix Results

EE210B: Inference over Networks (A. H. Sayed)

Given $N \times N$ matrices X , A , and Q , where Q is Hermitian and non-negative definite, we consider first discrete-time Lyapunov equations, also called Stein equations, of the following form:

$$X - A^* X A = Q \tag{F.38}$$

Let $\lambda_k(A)$ denote any of the eigenvalues of A . In the discrete-time case, a stable matrix A is one whose eigenvalues lie inside the unit disc (i.e., their magnitudes are strictly less than one).



Statement

Lemma F.2 (Discrete-time Lyapunov equation). Consider the Lyapunov equation (F.38). The following facts hold:

(a) The solution X is unique if, and only if, $\lambda_k(A)\lambda_\ell^*(A) \neq 1$ for all $k, \ell = 1, 2, \dots, N$. In this case, the unique solution X is Hermitian.

(b) When A is stable (i.e., all its eigenvalues are inside the unit disc), the solution X is unique, Hermitian, and nonnegative-definite. Moreover, it admits the series representation:

$$X = \sum_{n=0}^{\infty} (A^*)^n Q A^n \quad (\text{F.39})$$

Proof



Proof. We call upon property 12 from Table F.1 for Kronecker products and apply the vec operation to both sides of (F.38) to get

$$(I - A^T \otimes A^*)\text{vec}(X) = \text{vec}(Q) \quad (\text{F.40})$$

This linear system of equations has a unique solution, $\text{vec}(X)$, if, and only if, the coefficient matrix, $I - A^T \otimes A^*$, is nonsingular. This latter condition requires $\lambda_k(A)\lambda_\ell^*(A) \neq 1$ for all $k, \ell = 1, 2, \dots, N$. When A is stable, all of its eigenvalues lie inside the unit disc and this uniqueness condition is automatically satisfied. If we conjugate both sides of (F.38) we find that X^*

Proof



satisfies the same Lyapunov equation as X and, hence, by uniqueness, we must have $X = X^*$. Finally, let $F = A^\top \otimes A^*$. When A is stable, the matrix F is also stable since $\rho(F) = [\rho(A)]^2 < 1$. In this case, the matrix inverse $(I - F)^{-1}$ admits the series expansion

$$(I - F)^{-1} = I + F + F^2 + F^3 + \dots \quad (\text{F.41})$$

so that using (F.40) we have

Proof



$$\begin{aligned}\text{vec}(X) &= (I - F)^{-1} \text{vec}(Q) \\ &= \sum_{n=0}^{\infty} F^n \text{vec}(Q) \\ &= \sum_{n=0}^{\infty} ((A^T)^n \otimes (A^*)^n) \text{vec}(Q) \\ &= \sum_{n=0}^{\infty} \text{vec}((A^*)^n Q A^n)\end{aligned}\tag{F.42}$$

from which we deduce the series representation (F.39).



Continuous-Time Lyapunov Eqs



A similar analysis applies to the following continuous-time Lyapunov equation (also called a Sylvester equation):

$$XA^* + AX + Q = 0 \quad (\text{F.43})$$

In the continuous-time case, a stable matrix A is one whose eigenvalues lie in the open left-half plane (i.e., they have strictly negative real parts).



Statement

Lemma F.3 (Continuous-time Lyapunov equation). Consider the Lyapunov equation (F.43). The following facts hold:

- (a) The solution X is unique if, and only if, $\lambda_k(A) + \lambda_\ell^*(A) \neq 0$ for all $k, \ell = 1, 2, \dots, N$. In this case, the unique solution X is Hermitian.
 - (b) When A is stable (i.e., all its eigenvalues lie in the open left-half plane), the solution X is unique, Hermitian, and nonnegative-definite.
-

Proof



Proof. We call again upon property 12 from [Table F.1](#) for Kronecker products and apply the vec operation to both sides of (F.43) to get

$$[(A^* \otimes I) + (I \otimes A)] \text{vec}(X) = -\text{vec}(Q) \quad (\text{F.44})$$

This linear system of equations has a unique solution, $\text{vec}(X)$, if, and only if, the coefficient matrix, $(A^* \otimes I) + (I \otimes A)$, is nonsingular. This latter condition requires $\lambda_k(A) + \lambda_\ell^*(A) \neq 0$ for all $k, \ell = 1, 2, \dots, N$. To see this, let $F = (A^* \otimes I) + (I \otimes A)$ and let us verify that the eigenvalues of F are given by all linear combinations $\lambda_k(A) + \lambda_\ell^*(A)$. Consider the eigenvalue-eigenvector pairs $Ax_k = \lambda_k(A)x_k$ and $A^*y_\ell = \lambda_\ell^*(A)y_\ell$. Then, using property 2 from [Table F.1](#) for Kronecker products we get

Proof



$$\begin{aligned} F(y_\ell \otimes x_k) &= [(A^* \otimes I) + (I \otimes A)](y_\ell \otimes x_k) \\ &= (A^* y_\ell \otimes x_k) + (y_\ell \otimes A x_k) \\ &= \lambda_\ell^*(A)(y_\ell \otimes x_k) + \lambda_k(A)(y_\ell \otimes x_k) \\ &= (\lambda_k(A) + \lambda_\ell^*(A))(y_\ell \otimes x_k) \end{aligned} \tag{F.45}$$

so that the vector $(y_\ell \otimes x_k)$ is an eigenvector for F with eigenvalue $\lambda_k(A) + \lambda_\ell^*(A)$, as claimed. If we now conjugate both sides of (F.43) we find that X^* satisfies the same Lyapunov equation as X and, hence, by uniqueness, we must have $X = X^*$.



Stochastic Matrices

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.



Stochastic Matrices

Consider $N \times N$ matrices A with nonnegative entries, $\{a_{\ell k} \geq 0\}$. The matrix $A = [a_{\ell k}]$ is said to be left-stochastic if it satisfies

$$A^T \mathbf{1} = \mathbf{1} \quad (\text{left-stochastic}) \quad (\text{F.46})$$

where $\mathbf{1}$ denotes the column vector whose entries are all equal to one. It follows that the entries on each column of A add up to one. The matrix A is said to be doubly-stochastic if the entries on each of its columns and on each of its rows add up to one, i.e., if

$$A\mathbf{1} = \mathbf{1}, \quad A^T \mathbf{1} = \mathbf{1} \quad (\text{doubly-stochastic}) \quad (\text{F.47})$$

Properties



Lemma F.4 (Properties of stochastic matrices). Let A be an $N \times N$ left or doubly-stochastic matrix:

(a) The spectral radius of A is equal to one, $\rho(A) = 1$. It follows that all eigenvalues of A lie inside the unit disc, i.e., $|\lambda(A)| \leq 1$. The matrix A may have multiple eigenvalues with magnitude equal to one.



Properties

(b) If A is additionally a primitive matrix (cf. [definition \(6.1\)](#)), then A will have a single eigenvalue at one (i.e., the eigenvalue at one will have multiplicity one). All other eigenvalues of A will lie strictly inside the unit circle. Moreover, with proper sign scaling, all entries of the right-eigenvector of A corresponding to the single eigenvalue at one will be strictly positive, namely, if we let p denote this right-eigenvector with entries $\{p_k\}$ and normalize the entries to add up to one, then

$$Ap = p, \quad \mathbf{1}^\top p = 1, \quad p_k > 0, \quad k = 1, 2, \dots, N \quad (\text{F.48})$$

We refer to p as the *Perron eigenvector* of A . All other eigenvectors of A associated with the other eigenvalues will have at least one negative or complex entry.

Proof



We prove the result for right stochastic matrices; a similar argument applies to left or doubly stochastic matrices. Let A be a right-stochastic matrix. Then, $A\mathbf{1} = \mathbf{1}$, so that $\lambda = 1$ is one of the eigenvalues of A . Moreover, for any matrix A , it holds that $\rho(A) \leq \|A\|_\infty$, where $\|\cdot\|_\infty$ denotes the maximum absolute row sum of its matrix argument. But since all rows of A add up to one, we have $\|A\|_\infty = 1$. Therefore, $\rho(A) \leq 1$. And since we already know that A has an eigenvalue at $\lambda = 1$, we conclude that $\rho(A) = 1$.

The above result asserts that the spectral radius of a stochastic matrix is unity and that A has an eigenvalue at $\lambda = 1$. The result, however, does not rule out the possibility of multiple eigenvalues at $\lambda = 1$, or even other eigenvalues with magnitude equal to one.

→ Strong primitiveness ensures a unique eigenvalue at one by the
Perron Frobenius Theorem.



Inequality Recursions

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.



Deterministic Recursion

Lemma F.5 (Deterministic recursion). Let $u(i) \geq 0$ denote a scalar deterministic (i.e., non-random) sequence that satisfies the inequality recursion:

$$u(i+1) \leq [1 - a(i)]u(i) + b(i), \quad i \geq 0 \quad (\text{F.49})$$

(a) When the scalar sequences $\{a(i), b(i)\}$ satisfy the four conditions:

$$0 \leq a(i) < 1, \quad b(i) \geq 0, \quad \sum_{i=0}^{\infty} a(i) = \infty, \quad \lim_{i \rightarrow \infty} \frac{b(i)}{a(i)} = 0 \quad (\text{F.50})$$

it holds that $\lim_{i \rightarrow \infty} u(i) = 0$.



Deterministic Recursion

(b) When the scalar sequences $\{a(i), b(i)\}$ are of the form

$$a(i) = \frac{c}{i+1}, \quad b(i) = \frac{d}{(i+1)^{p+1}}, \quad c > 0, \quad d > 0, \quad p > 0 \quad (\text{F.51})$$

it holds that, for large enough i , the sequence $u(i)$ converges to zero at one of the following rates depending on the value of c :

$$\begin{cases} u(i) \leq \left(\frac{d}{c-p}\right) \frac{1}{i^p} + o(1/i^p), & c > p \\ u(i) = O(\log i / i^p), & c = p \\ u(i) = O(1/i^c), & c < p \end{cases} \quad (\text{F.52})$$

The fastest convergence rate occurs when $c > p$ and is in the order of $1/i^p$.



Big and Little-O Notation

68

Lecture #8: Useful Matrix Results

EE210B: Inference over Networks (A. H. Sayed)

$a(i) = O(b(i))$ means $|a(i)| \leq c|b(i)|$ for some constant c and large i . Example:

$a(i) = O(1/i) \implies a(i)$ decays asymptotically at a rate comparable to $1/i$

$a(i) = o(b(i))$ means that asymptotically the sequence $a(i)$ decays faster than $b(i)$, or $|a(i)|/|b(i)| \rightarrow 0$ as $i \rightarrow \infty$. Example:

$a(i) = o(1/i) \implies a(i)$ decays asymptotically at a faster rate than $1/i$

$$\begin{cases} a = O(\mu) \implies |a| \text{ is in the order of } \mu \\ a = o(\mu) \implies |a| \text{ is some higher power in } \mu \end{cases}$$

Stochastic Recursion



Lemma F.6 (Stochastic recursion). Let $\mathbf{u}(i) \geq 0$ denote a scalar sequence of nonnegative random variables satisfying $\mathbb{E} \mathbf{u}(0) < \infty$ and the stochastic recursion:

$$\mathbb{E} [\mathbf{u}(i+1) | \mathbf{u}(0), \mathbf{u}(1), \dots, \mathbf{u}(i)] \leq [1 - a(i)]\mathbf{u}(i) + b(i), \quad i \geq 0 \quad (\text{F.53})$$

in terms of the conditional expectation on the left-hand side, and where the scalar and nonnegative deterministic sequences $\{a(i), b(i)\}$ satisfy the five conditions:

$$0 \leq a(i) < 1, \quad b(i) \geq 0, \quad \sum_{i=0}^{\infty} a(i) = \infty, \quad \sum_{i=0}^{\infty} b(i) < \infty, \quad \lim_{i \rightarrow \infty} \frac{b(i)}{a(i)} = 0 \quad (\text{F.54})$$

Then, it holds that $\lim_{i \rightarrow \infty} \mathbf{u}(i) = 0$ almost surely, and $\lim_{i \rightarrow \infty} \mathbb{E} \mathbf{u}(i) = 0$.

End of Lecture

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.