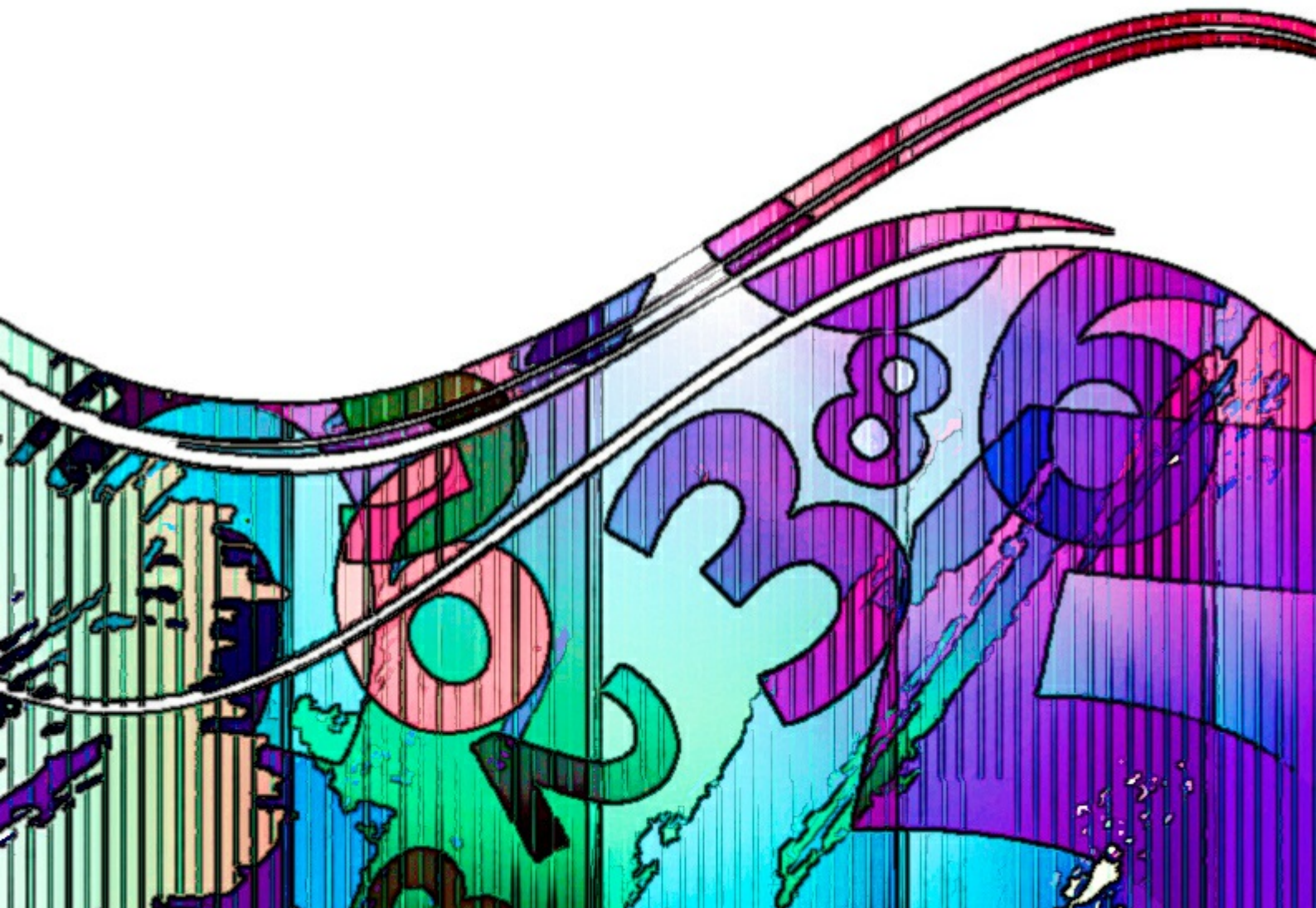


Kenneth Kuttler

Linear Algebra III

Advanced topics



KENNETH KUTTLER

LINEAR ALGEBRA III

ADVANCED TOPICS

Linear Algebra III: Advanced topics

2nd edition

© 2019 Kenneth Kuttler & bookboon.com

ISBN 978-87-403-3165-3

CONTENTS

13	Norms	327
13.1	The p Norms	333
13.2	The Condition Number	335
13.3	The Spectral Radius	337
13.4	Series And Sequences Of Linear Operators	340
13.5	Iterative Methods For Linear Systems	345
13.6	Theory Of Convergence	350
13.7	Exercises	354
14	Numerical Methods, Eigenvalues	362
14.1	The Power Method For Eigenvalues	362
14.2	The QR Algorithm	377
14.3	Exercises	392
A	Matrix Calculator On The Web	394
A.1	Use Of Matrix Calculator On Web	394

**Technical training on
WHAT you need, WHEN you need it**

At IDC Technologies we can tailor our technical and engineering training workshops to suit your needs. We have extensive experience in training technical and engineering staff and have trained people in organisations such as General Motors, Shell, Siemens, BHP and Honeywell to name a few.

Our onsite training is cost effective, convenient and completely customisable to the technical and engineering areas you want covered. Our workshops are all comprehensive hands-on learning experiences with ample time given to practical sessions and demonstrations. We communicate well to ensure that workshop content and timing match the knowledge, skills, and abilities of the participants.

We run onsite training all year round and hold the workshops on your premises or a venue of your choice for your convenience.

For a no obligation proposal, contact us today at training@idc-online.com or visit our website for more information: www.idc-online.com/onsite/

OIL & GAS ENGINEERING

ELECTRONICS

AUTOMATION & PROCESS CONTROL

MECHANICAL ENGINEERING

INDUSTRIAL DATA COMMS

ELECTRICAL POWER

Phone: +61 8 9321 1702
Email: training@idc-online.com
Website: www.idc-online.com

IDC TECHNOLOGIES

B	Positive Matrices	395
C	Functions Of Matrices	404
D	Differential Equations	409
D.1	Theory Of Ordinary Differential Equations	409
D.2	Linear Systems	410
D.3	Local Solutions	411
D.4	First Order Linear Systems	414
D.5	Geometric Theory Of Autonomous Systems	422
D.6	General Geometric Theory	426
D.7	The Stable Manifold	427
E	Compactness And Completeness	434
E.1	The Nested Interval Lemma	434
E.2	Convergent Sequences, Sequential Compactness	434
F	Some Topics Flavored With Linear Algebra	437
F.1	The Symmetric Polynomial Theorem	437
F.2	The Fundamental Theorem Of Algebra	440
F.3	Transcendental Numbers	443
F.4	More On Algebraic Field Extensions	452
	Bibliography	458
	Index	459

Chapter 13

Norms

In this chapter, X and Y are finite dimensional vector spaces which have a norm. The following is a definition.

Definition 13.0.1 A linear space X is a normed linear space if there is a norm defined on X , $\|\cdot\|$ satisfying

$$\begin{aligned} \|\mathbf{x}\| &\geq 0, \quad \|\mathbf{x}\| = 0 \text{ if and only if } \mathbf{x} = 0, \\ \|\mathbf{x} + \mathbf{y}\| &\leq \|\mathbf{x}\| + \|\mathbf{y}\|, \\ \|c\mathbf{x}\| &= |c| \|\mathbf{x}\| \end{aligned}$$

whenever c is a scalar. A set, $U \subseteq X$, a normed linear space is open if for every $p \in U$, there exists $\delta > 0$ such that

$$B(p, \delta) \equiv \{x : \|x - p\| < \delta\} \subseteq U.$$

Thus, a set is open if every point of the set is an interior point. Also, $\lim_{n \rightarrow \infty} \mathbf{x}_n = \mathbf{x}$ means $\lim_{n \rightarrow \infty} \|\mathbf{x}_n - \mathbf{x}\| = 0$. This is written sometimes as $\mathbf{x}_n \rightarrow \mathbf{x}$.

Note first that

$$\|\mathbf{x}\| = \|\mathbf{x} - \mathbf{y} + \mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y}\|$$

so

$$\|\mathbf{x}\| - \|\mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}\|.$$

Similarly

$$\|\mathbf{y}\| - \|\mathbf{x}\| \leq \|\mathbf{x} - \mathbf{y}\|$$

and so

$$\left| \|\mathbf{x}\| - \|\mathbf{y}\| \right| \leq \|\mathbf{x} - \mathbf{y}\|. \tag{13.1}$$

To begin with recall the Cauchy Schwarz inequality which is stated here for convenience in terms of the inner product space, \mathbb{C}^n .

Theorem 13.0.2 The following inequality holds for a_i and $b_i \in \mathbb{C}$.

$$\left| \sum_{i=1}^n a_i \bar{b}_i \right| \leq \left(\sum_{i=1}^n |a_i|^2 \right)^{1/2} \left(\sum_{i=1}^n |b_i|^2 \right)^{1/2}. \tag{13.2}$$

Let X be a finite dimensional normed linear space with norm $\|\cdot\|$ where the field of scalars is denoted by \mathbb{F} and is understood to be either \mathbb{R} or \mathbb{C} . Let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis for X . If $\mathbf{x} \in X$, denote by x_i the i^{th} component of \mathbf{x} with respect to this basis. Thus

$$\mathbf{x} = \sum_{i=1}^n x_i \mathbf{v}_i.$$

Definition 13.0.3 For $\mathbf{x} \in X$ and $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ a basis, define a new norm by

$$|\mathbf{x}| \equiv \left(\sum_{i=1}^n |x_i|^2 \right)^{1/2}.$$

where

$$\mathbf{x} = \sum_{i=1}^n x_i \mathbf{v}_i.$$

Similarly, for $\mathbf{y} \in Y$ with basis $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$, and y_i its components with respect to this basis,

$$|\mathbf{y}| \equiv \left(\sum_{i=1}^m |y_i|^2 \right)^{1/2}$$

For $A \in \mathcal{L}(X, Y)$, the space of linear mappings from X to Y ,

$$\|A\| \equiv \sup\{|A\mathbf{x}| : |\mathbf{x}| \leq 1\}. \tag{13.3}$$

The first thing to show is that the two norms, $\|\cdot\|$ and $|\cdot|$, are equivalent. This means the conclusion of the following theorem holds.

Theorem 13.0.4 *Let $(X, \|\cdot\|)$ be a finite dimensional normed linear space and let $|\cdot|$ be described above relative to a given basis, $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$. Then $|\cdot|$ is a norm and there exist constants $\delta, \Delta > 0$ independent of \mathbf{x} such that*

$$\delta \|\mathbf{x}\| \leq |\mathbf{x}| \leq \Delta \|\mathbf{x}\|. \tag{13.4}$$

Proof: All of the above properties of a norm are obvious except the second, the triangle inequality. To establish this inequality, use the Cauchy Schwarz inequality to write

$$\begin{aligned} |\mathbf{x} + \mathbf{y}|^2 &\equiv \sum_{i=1}^n |x_i + y_i|^2 \leq \sum_{i=1}^n |x_i|^2 + \sum_{i=1}^n |y_i|^2 + 2 \operatorname{Re} \sum_{i=1}^n x_i \bar{y}_i \\ &\leq \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2 \left(\sum_{i=1}^n |x_i|^2 \right)^{1/2} \left(\sum_{i=1}^n |y_i|^2 \right)^{1/2} \\ &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2 \|\mathbf{x}\| \|\mathbf{y}\| = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2 \end{aligned}$$

and this proves the second property above.

It remains to show the equivalence of the two norms. By the Cauchy Schwarz inequality again,

$$\begin{aligned} \|\mathbf{x}\| &\equiv \left\| \sum_{i=1}^n x_i \mathbf{v}_i \right\| \leq \sum_{i=1}^n |x_i| \|\mathbf{v}_i\| \leq \|\mathbf{x}\| \left(\sum_{i=1}^n \|\mathbf{v}_i\|^2 \right)^{1/2} \\ &\equiv \delta^{-1} |\mathbf{x}|. \end{aligned}$$

This proves the first half of the inequality.

Suppose the second half of the inequality is not valid. Then there exists a sequence $\mathbf{x}^k \in X$ such that

$$|\mathbf{x}^k| > k \|\mathbf{x}^k\|, \quad k = 1, 2, \dots$$

Then define

$$\mathbf{y}^k \equiv \frac{\mathbf{x}^k}{|\mathbf{x}^k|}.$$

It follows

$$|\mathbf{y}^k| = 1, \quad |\mathbf{y}^k| > k \|\mathbf{y}^k\|. \tag{13.5}$$

Letting y_i^k be the components of \mathbf{y}^k with respect to the given basis, it follows the vector

$$(y_1^k, \dots, y_n^k)$$

is a unit vector in \mathbb{F}^n . By the Heine Borel theorem, there exists a subsequence, still denoted by k such that

$$(y_1^k, \dots, y_n^k) \rightarrow (y_1, \dots, y_n).$$

It follows from 13.5 and this that for

$$\mathbf{y} = \sum_{i=1}^n y_i \mathbf{v}_i,$$

$$0 = \lim_{k \rightarrow \infty} \|\mathbf{y}^k\| = \lim_{k \rightarrow \infty} \left\| \sum_{i=1}^n y_i^k \mathbf{v}_i \right\| = \left\| \sum_{i=1}^n y_i \mathbf{v}_i \right\|$$

but not all the y_i equal zero. The last equation follows easily from 13.1 and

$$\left\| \sum_{i=1}^n y_i^k \mathbf{v}_i \right\| - \left\| \sum_{i=1}^n y_i \mathbf{v}_i \right\| \leq \left\| \sum_{i=1}^n (y_i^k - y_i) \mathbf{v}_i \right\| \leq \sum_{i=1}^n |y_i^k - y_i| \|\mathbf{v}_i\|$$

This contradicts the assumption that $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis and proves the second half of the inequality. ■

Definition 13.0.5 Let $(X, \|\cdot\|)$ be a normed linear space and let $\{x_n\}_{n=1}^\infty$ be a sequence of vectors. Then this is called a Cauchy sequence if for all $\varepsilon > 0$ there exists N such that if $m, n \geq N$, then

$$\|x_n - x_m\| < \varepsilon.$$

This is written more briefly as

$$\lim_{m, n \rightarrow \infty} \|x_n - x_m\| = 0.$$

Definition 13.0.6 A normed linear space, $(X, \|\cdot\|)$ is called a Banach space if it is complete. This means that, whenever, $\{\mathbf{x}_n\}$ is a Cauchy sequence there exists a unique $\mathbf{x} \in X$ such that $\lim_{n \rightarrow \infty} \|\mathbf{x} - \mathbf{x}_n\| = 0$.

Corollary 13.0.7 If $(X, \|\cdot\|)$ is a finite dimensional normed linear space with the field of scalars $\mathbb{F} = \mathbb{C}$ or \mathbb{R} , then $(X, \|\cdot\|)$ is a Banach space.

Proof: Let $\{\mathbf{x}^k\}$ be a Cauchy sequence. Then letting the components of \mathbf{x}^k with respect to the given basis be

$$x_1^k, \dots, x_n^k,$$

it follows from Theorem 13.0.4, that

$$(x_1^k, \dots, x_n^k)$$

is a Cauchy sequence in \mathbb{F}^n and so

$$(x_1^k, \dots, x_n^k) \rightarrow (x_1, \dots, x_n) \in \mathbb{F}^n.$$

Thus, letting $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{v}_i$, it follows from the equivalence of the two norms shown above that

$$\lim_{k \rightarrow \infty} \|\mathbf{x}^k - \mathbf{x}\| = \lim_{k \rightarrow \infty} \|\mathbf{x}^k - \mathbf{x}\| = 0. \quad \blacksquare$$

Corollary 13.0.8 Suppose X is a finite dimensional linear space with the field of scalars either \mathbb{C} or \mathbb{R} and $\|\cdot\|$ and $\|\cdot\|'$ are two norms on X . Then there exist positive constants, δ and Δ , independent of $\mathbf{x} \in X$ such that

$$\delta \|\mathbf{x}\| \leq \|\mathbf{x}\|' \leq \Delta \|\mathbf{x}\|.$$

Thus any two norms are equivalent.

This is very important because it shows that all questions of convergence can be considered relative to any norm with the same outcome.

Proof: Let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis for X and let $|\cdot|$ be the norm taken with respect to this basis which was described earlier. Then by Theorem 13.0.4, there are positive constants $\delta_1, \Delta_1, \delta_2, \Delta_2$, all independent of $\mathbf{x} \in X$ such that

$$\delta_2 |||\mathbf{x}||| \leq |\mathbf{x}| \leq \Delta_2 |||\mathbf{x}|||, \quad \delta_1 ||\mathbf{x}|| \leq |\mathbf{x}| \leq \Delta_1 ||\mathbf{x}||.$$

Then

$$\delta_2 |||\mathbf{x}||| \leq |\mathbf{x}| \leq \Delta_1 ||\mathbf{x}|| \leq \frac{\Delta_1}{\delta_1} |\mathbf{x}| \leq \frac{\Delta_1 \Delta_2}{\delta_1} |||\mathbf{x}|||$$

and so

$$\frac{\delta_2}{\Delta_1} |||\mathbf{x}||| \leq ||\mathbf{x}|| \leq \frac{\Delta_2}{\delta_1} |||\mathbf{x}||| \quad \blacksquare$$

Definition 13.0.9 Let X and Y be normed linear spaces with norms $||\cdot||_X$ and $||\cdot||_Y$ respectively. Then $\mathcal{L}(X, Y)$ denotes the space of linear transformations, called bounded linear transformations, mapping X to Y which have the property that

$$||A|| \equiv \sup \{ ||Ax||_Y : ||x||_X \leq 1 \} < \infty.$$

Then $||A||$ is referred to as the operator norm of the bounded linear transformation A .

It is an easy exercise to verify that $||\cdot||$ is a norm on $\mathcal{L}(X, Y)$ and it is always the case that

$$||Ax||_Y \leq ||A|| ||x||_X.$$

Furthermore, you should verify that you can replace ≤ 1 with $= 1$ in the definition. Thus

$$||A|| \equiv \sup \{ ||Ax||_Y : ||x||_X = 1 \}.$$

Theorem 13.0.10 Let X and Y be finite dimensional normed linear spaces of dimension n and m respectively and denote by $||\cdot||$ the norm on either X or Y . Then if A is any linear function mapping X to Y , then $A \in \mathcal{L}(X, Y)$ and $(\mathcal{L}(X, Y), ||\cdot||)$ is a complete normed linear space of dimension nm with

$$||A\mathbf{x}|| \leq ||A|| ||\mathbf{x}||.$$

Also if $A \in \mathcal{L}(X, Y)$ and $B \in \mathcal{L}(Y, Z)$ where X, Y, Z are normed linear spaces,

$$||BA|| \leq ||B|| ||A||$$

Proof: It is necessary to show the norm defined on linear transformations really is a norm. Again the first and third properties listed above for norms are obvious. It remains to show the second and verify $||A|| < \infty$. Letting $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis and $|\cdot|$ defined with respect to this basis as above, there exist constants $\delta, \Delta > 0$ such that

$$\delta ||\mathbf{x}|| \leq |\mathbf{x}| \leq \Delta ||\mathbf{x}||.$$

Then,

$$\begin{aligned} ||A + B|| &\equiv \sup \{ ||(A + B)(\mathbf{x})|| : ||\mathbf{x}|| \leq 1 \} \\ &\leq \sup \{ ||A\mathbf{x}|| : ||\mathbf{x}|| \leq 1 \} + \sup \{ ||B\mathbf{x}|| : ||\mathbf{x}|| \leq 1 \} \equiv ||A|| + ||B||. \end{aligned}$$

Next consider the claim that $||A|| < \infty$. This follows from

$$\begin{aligned} ||A(\mathbf{x})|| &= \left\| A \left(\sum_{i=1}^n x_i \mathbf{v}_i \right) \right\| \leq \sum_{i=1}^n |x_i| ||A(\mathbf{v}_i)|| \\ &\leq |\mathbf{x}| \left(\sum_{i=1}^n ||A(\mathbf{v}_i)||^2 \right)^{1/2} \leq \Delta ||\mathbf{x}|| \left(\sum_{i=1}^n ||A(\mathbf{v}_i)||^2 \right)^{1/2} < \infty. \end{aligned}$$

Thus $\|A\| \leq \Delta \left(\sum_{i=1}^n \|A(\mathbf{v}_i)\|^2 \right)^{1/2}$.

Next consider the assertion about the dimension of $\mathcal{L}(X, Y)$. It follows from Theorem 8.2.3. By Corollary 13.0.7 ($\mathcal{L}(X, Y), \|\cdot\|$) is complete. If $\mathbf{x} \neq \mathbf{0}$,

$$\|A\mathbf{x}\| \frac{1}{\|\mathbf{x}\|} = \left\| A \frac{\mathbf{x}}{\|\mathbf{x}\|} \right\| \leq \|A\|$$

Consider the last claim.

$$\|BA\| \equiv \sup_{\|x\| \leq 1} \|B(A(x))\| \leq \|B\| \sup_{\|x\| \leq 1} \|Ax\| = \|B\| \|A\| \blacksquare$$

Note by Corollary 13.0.8 you can define a norm any way desired on any finite dimensional linear space which has the field of scalars \mathbb{R} or \mathbb{C} and any other way of defining a norm on this space yields an equivalent norm. Thus, it doesn't much matter as far as notions of convergence are concerned which norm is used for a finite dimensional space. In particular in the space of $m \times n$ matrices, you can use the operator norm defined above, or some other way of giving this space a norm. A popular choice for a norm is the Frobenius norm discussed earlier but reviewed here.

Definition 13.0.11 Make the space of $m \times n$ matrices into a inner product space by defining

$$(A, B) \equiv \text{trace}(AB^*).$$

Another way of describing a norm for an $n \times n$ matrix is as follows.

Definition 13.0.12 Let A be an $m \times n$ matrix. Define the spectral norm of A , written as $\|A\|_2$ to be

$$\max \left\{ \lambda^{1/2} : \lambda \text{ is an eigenvalue of } A^*A \right\}.$$

That is, the largest singular value of A . (Note the eigenvalues of A^*A are all positive because if $A^*A\mathbf{x} = \lambda\mathbf{x}$, then

$$\lambda|\mathbf{x}|^2 = \lambda(\mathbf{x}, \mathbf{x}) = (A^*A\mathbf{x}, \mathbf{x}) = (A\mathbf{x}, A\mathbf{x}) \geq 0.$$

Actually, this is nothing new. It turns out that $\|\cdot\|_2$ is nothing more than the operator norm for A taken with respect to the usual Euclidean norm,

$$|\mathbf{x}| = \left(\sum_{k=1}^n |x_k|^2 \right)^{1/2}.$$

Proposition 13.0.13 The following holds.

$$\|A\|_2 = \sup \{ |A\mathbf{x}| : |\mathbf{x}| = 1 \} \equiv \|A\|.$$

Proof: Note that A^*A is Hermitian and so by Corollary 12.3.4,

$$\begin{aligned} \|A\|_2 &= \max \left\{ (A^*A\mathbf{x}, \mathbf{x})^{1/2} : |\mathbf{x}| = 1 \right\} = \max \left\{ (A\mathbf{x}, A\mathbf{x})^{1/2} : |\mathbf{x}| = 1 \right\} \\ &= \max \{ |A\mathbf{x}| : |\mathbf{x}| = 1 \} = \|A\|. \blacksquare \end{aligned}$$

Here is another proof of this proposition. Recall there are unitary matrices of the right size U, V such that $A = U \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} V^*$ where the matrix on the inside is as described in the section on the singular value decomposition. Then since unitary matrices preserve norms,

$$\begin{aligned} \|A\| &= \sup_{|\mathbf{x}| \leq 1} \left| U \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} V^* \mathbf{x} \right| = \sup_{|V^* \mathbf{x}| \leq 1} \left| U \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} V^* \mathbf{x} \right| \\ &= \sup_{|\mathbf{y}| \leq 1} \left| U \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} \mathbf{y} \right| = \sup_{|\mathbf{y}| \leq 1} \left| \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} \mathbf{y} \right| = \sigma_1 \equiv \|A\|_2 \end{aligned}$$

This completes the alternate proof.

From now on, $\|A\|_2$ will mean either the operator norm of A taken with respect to the usual Euclidean norm or the largest singular value of A , whichever is most convenient.

An interesting application of the notion of equivalent norms on \mathbb{R}^n is the process of giving a norm on a finite Cartesian product of normed linear spaces.

Definition 13.0.14 Let $X_i, i = 1, \dots, n$ be normed linear spaces with norms, $\|\cdot\|_i$. For

$$\mathbf{x} \equiv (x_1, \dots, x_n) \in \prod_{i=1}^n X_i$$

define $\theta : \prod_{i=1}^n X_i \rightarrow \mathbb{R}^n$ by

$$\theta(\mathbf{x}) \equiv (\|x_1\|_1, \dots, \|x_n\|_n)$$

Then if $\|\cdot\|$ is any norm on \mathbb{R}^n , define a norm on $\prod_{i=1}^n X_i$, also denoted by $\|\cdot\|$ by

$$\|\mathbf{x}\| \equiv \|\theta\mathbf{x}\|.$$

The following theorem follows immediately from Corollary 13.0.8.

Theorem 13.0.15 Let X_i and $\|\cdot\|_i$ be given in the above definition and consider the norms on $\prod_{i=1}^n X_i$ described there in terms of norms on \mathbb{R}^n . Then any two of these norms on $\prod_{i=1}^n X_i$ obtained in this way are equivalent.

For example, define

$$\|\mathbf{x}\|_1 \equiv \sum_{i=1}^n |x_i|,$$

$$\|\mathbf{x}\|_\infty \equiv \max \{|x_i|, i = 1, \dots, n\},$$

I joined MITAS because
I wanted **real responsibility**

The Graduate Programme
for Engineers and Geoscientists
www.discovermitas.com



Month 16

I was a construction supervisor in the North Sea advising and helping foremen solve problems

Real work
International opportunities
Three work placements



or

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{1/2}$$

and all three are equivalent norms on $\prod_{i=1}^n X_i$.

13.1 The p Norms

In addition to $\|\cdot\|_1$ and $\|\cdot\|_\infty$ mentioned above, it is common to consider the so called p norms for $\mathbf{x} \in \mathbb{C}^n$.

Definition 13.1.1 Let $\mathbf{x} \in \mathbb{C}^n$. Then define for $p \geq 1$,

$$\|\mathbf{x}\|_p \equiv \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$$

The following inequality is called Holder's inequality.

Proposition 13.1.2 For $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$,

$$\sum_{i=1}^n |x_i| |y_i| \leq \left(\sum_{i=1}^n |x_i|^p \right)^{1/p} \left(\sum_{i=1}^n |y_i|^{p'} \right)^{1/p'}$$

The proof will depend on the following lemma.

Lemma 13.1.3 If $a, b \geq 0$ and p' is defined by $\frac{1}{p} + \frac{1}{p'} = 1$, then

$$ab \leq \frac{a^p}{p} + \frac{b^{p'}}{p'}$$

Proof of the Proposition: If \mathbf{x} or \mathbf{y} equals the zero vector there is nothing to prove. Therefore, assume they are both nonzero. Let $A = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$ and $B = \left(\sum_{i=1}^n |y_i|^{p'} \right)^{1/p'}$. Then using Lemma 13.1.3,

$$\begin{aligned} \sum_{i=1}^n \frac{|x_i|}{A} \frac{|y_i|}{B} &\leq \sum_{i=1}^n \left[\frac{1}{p} \left(\frac{|x_i|}{A} \right)^p + \frac{1}{p'} \left(\frac{|y_i|}{B} \right)^{p'} \right] \\ &= \frac{1}{p} \frac{1}{A^p} \sum_{i=1}^n |x_i|^p + \frac{1}{p'} \frac{1}{B^{p'}} \sum_{i=1}^n |y_i|^{p'} = \frac{1}{p} + \frac{1}{p'} = 1 \end{aligned}$$

and so

$$\sum_{i=1}^n |x_i| |y_i| \leq AB = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p} \left(\sum_{i=1}^n |y_i|^{p'} \right)^{1/p'}. \blacksquare$$

Theorem 13.1.4 The p norms do indeed satisfy the axioms of a norm.

Proof: It is obvious that $\|\cdot\|_p$ does indeed satisfy most of the norm axioms. The only one that is not clear is the triangle inequality. To save notation write $\|\cdot\|$ in place of $\|\cdot\|_p$ in what follows. Note also that $\frac{p}{p'} = p - 1$. Then using the Holder inequality,

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^p &= \sum_{i=1}^n |x_i + y_i|^p \\ &\leq \sum_{i=1}^n |x_i + y_i|^{p-1} |x_i| + \sum_{i=1}^n |x_i + y_i|^{p-1} |y_i| \\ &= \sum_{i=1}^n |x_i + y_i|^{\frac{p}{p'}} |x_i| + \sum_{i=1}^n |x_i + y_i|^{\frac{p}{p'}} |y_i| \end{aligned}$$

$$\begin{aligned} &\leq \left(\sum_{i=1}^n |x_i + y_i|^p \right)^{1/p'} \left[\left(\sum_{i=1}^n |x_i|^p \right)^{1/p} + \left(\sum_{i=1}^n |y_i|^p \right)^{1/p} \right] \\ &= \|\mathbf{x} + \mathbf{y}\|^{p/p'} \left(\|\mathbf{x}\|_p + \|\mathbf{y}\|_p \right) \end{aligned}$$

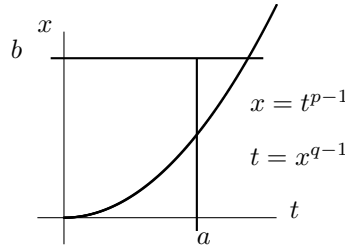
so dividing by $\|\mathbf{x} + \mathbf{y}\|^{p/p'}$, it follows

$$\|\mathbf{x} + \mathbf{y}\| \|\mathbf{x} + \mathbf{y}\|^{-p/p'} = \|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p$$

$\left(p - \frac{p}{p'} = p \left(1 - \frac{1}{p'} \right) = p \frac{1}{p} = 1 \right)$. ■

It only remains to prove Lemma 13.1.3.

Proof of the lemma: Let $p' = q$ to save on notation and consider the following picture:



$$ab \leq \int_0^a t^{p-1} dt + \int_0^b x^{q-1} dx = \frac{a^p}{p} + \frac{b^q}{q}.$$

Note equality occurs when $a^p = b^q$.

Alternate proof of the lemma: For $a, b \geq 0$, let b be fixed and

$$f(a) \equiv \frac{1}{p}a^p + \frac{1}{q}b^q - ab, \quad t > 0$$

If $b = 0$, it is clear that $f(a) \geq 0$ for all a . Then assume $b > 0$. It is clear since $p > 1$ that $\lim_{a \rightarrow \infty} f(a) = \infty$.

$$f'(a) = a^{p-1} - b$$

This is negative for small a and then eventually is positive. Consider the minimum value of f which must occur at $a > 0$ thanks to the observation that the function is initially strictly decreasing. At this point,

$$0 = f'(a) = a^{p-1} - b = a^{(p/q)} - b$$

and so $a^p = b^q$ at the point where this function has a minimum. Thus at this value of a ,

$$f(a) = \frac{1}{p}a^p + \frac{1}{q}a^p - aa^{p-1} = a^p - a^p = 0$$

Hence $f(a) \geq 0$ for all $a \geq 0$ and this proves the inequality. Equality occurs when $a^p = b^q$. ■

Now $\|A\|_p$ may be considered as the operator norm of A taken with respect to $\|\cdot\|_p$. In the case when $p = 2$, this is just the spectral norm. There is an easy estimate for $\|A\|_p$ in terms of the entries of A .

Theorem 13.1.5 *The following holds.*

$$\|A\|_p \leq \left(\sum_k \left(\sum_j |A_{jk}|^p \right)^{q/p} \right)^{1/q}$$

Proof: Let $\|\mathbf{x}\|_p \leq 1$ and let $A = (\mathbf{a}_1, \dots, \mathbf{a}_n)$ where the \mathbf{a}_k are the columns of A . Then

$$A\mathbf{x} = \left(\sum_k x_k \mathbf{a}_k \right)$$

and so by Holder's inequality,

$$\begin{aligned} \|A\mathbf{x}\|_p &\equiv \left\| \sum_k x_k \mathbf{a}_k \right\|_p \leq \sum_k |x_k| \|\mathbf{a}_k\|_p \leq \\ &\leq \left(\sum_k |x_k|^p \right)^{1/p} \left(\sum_k \|\mathbf{a}_k\|_p^q \right)^{1/q} \leq \left(\sum_k \left(\sum_j |A_{jk}|^p \right)^{q/p} \right)^{1/q} \quad \blacksquare \end{aligned}$$

13.2 The Condition Number

Let $A \in \mathcal{L}(X, X)$ be a linear transformation where X is a finite dimensional vector space and consider the problem $Ax = b$ where it is assumed there is a unique solution to this problem. How does the solution change if A is changed a little bit and if b is changed a little bit? This is clearly an interesting question because you often do not know A and b exactly. If a small change in these quantities results in a large change in the solution, x , then it seems clear this would be undesirable. In what follows $\|\cdot\|$ when applied to a linear transformation will always refer to the operator norm. Recall the following property of the operator norm in Theorem 13.0.10.



www.job.oticon.dk

oticon
PEOPLE FIRST



Lemma 13.2.1 Let $A, B \in \mathcal{L}(X, X)$ where X is a normed vector space as above. Then for $\|\cdot\|$ denoting the operator norm,

$$\|AB\| \leq \|A\| \|B\|.$$

Lemma 13.2.2 Let $A, B \in \mathcal{L}(X, X)$, $A^{-1} \in \mathcal{L}(X, X)$, and suppose $\|B\| < 1/\|A^{-1}\|$. Then $(A + B)^{-1}$, $(I + A^{-1}B)^{-1}$ exists and

$$\|(I + A^{-1}B)^{-1}\| \leq (1 - \|A^{-1}B\|)^{-1} \tag{13.6}$$

$$\|(A + B)^{-1}\| \leq \|A^{-1}\| \left| \frac{1}{1 - \|A^{-1}B\|} \right|. \tag{13.7}$$

The above formula makes sense because $\|A^{-1}B\| < 1$.

Proof: By Lemma 13.0.10,

$$\|A^{-1}B\| \leq \|A^{-1}\| \|B\| < \|A^{-1}\| \frac{1}{\|A^{-1}\|} = 1 \tag{13.8}$$

Then from the triangle inequality,

$$\begin{aligned} \|(I + A^{-1}B)x\| &\geq \|x\| - \|A^{-1}Bx\| \\ &\geq \|x\| - \|A^{-1}B\| \|x\| = (1 - \|A^{-1}B\|) \|x\| \end{aligned}$$

It follows that $I + A^{-1}B$ is one to one because from 13.8, $1 - \|A^{-1}B\| > 0$. Thus if $(I + A^{-1}B)x = 0$, then $x = 0$. Thus $I + A^{-1}B$ is also onto, taking a basis to a basis. Then a generic $y \in X$ is of the form $y = (I + A^{-1}B)x$ and the above shows that

$$\|(I + A^{-1}B)^{-1}y\| \leq (1 - \|A^{-1}B\|)^{-1} \|y\|$$

which verifies 13.6. Thus $(A + B) = A(I + A^{-1}B)$ is one to one and this with Lemma 13.0.10 implies 13.7. ■

Proposition 13.2.3 Suppose A is invertible, $b \neq 0$, $Ax = b$, and $(A + B)x_1 = b_1$ where $\|B\| < 1/\|A^{-1}\|$. Then

$$\frac{\|x_1 - x\|}{\|x\|} \leq \frac{\|A^{-1}\| \|A\|}{1 - \|A^{-1}B\|} \left(\frac{\|b_1 - b\|}{\|b\|} + \|B\| \right)$$

Proof: This follows from the above lemma.

$$\begin{aligned} \frac{\|x_1 - x\|}{\|x\|} &= \frac{\|(I + A^{-1}B)^{-1}A^{-1}b_1 - A^{-1}b\|}{\|A^{-1}b\|} \\ &\leq \frac{1}{1 - \|A^{-1}B\|} \frac{\|A^{-1}b_1 - (I + A^{-1}B)A^{-1}b\|}{\|A^{-1}b\|} \\ &\leq \frac{1}{1 - \|A^{-1}B\|} \frac{\|A^{-1}(b_1 - b)\| + \|A^{-1}BA^{-1}b\|}{\|A^{-1}b\|} \\ &\leq \frac{\|A^{-1}\|}{1 - \|A^{-1}B\|} \left(\frac{\|b_1 - b\|}{\|A^{-1}b\|} + \|B\| \right) \end{aligned}$$

because $A^{-1}b/\|A^{-1}b\|$ is a unit vector. Now multiply and divide by $\|A\|$. Then

$$\begin{aligned} &\leq \frac{\|A^{-1}\| \|A\|}{1 - \|A^{-1}B\|} \left(\frac{\|b_1 - b\|}{\|A\| \|A^{-1}b\|} + \frac{\|B\|}{\|A\|} \right) \\ &\leq \frac{\|A^{-1}\| \|A\|}{1 - \|A^{-1}B\|} \left(\frac{\|b_1 - b\|}{\|b\|} + \frac{\|B\|}{\|A\|} \right). \blacksquare \end{aligned}$$

This shows that the number, $\|A^{-1}\| \|A\|$, controls how sensitive the relative change in the solution of $Ax = b$ is to small changes in A and b . This number is called the condition number. It is bad when this number is large because a small relative change in b , for example could yield a large relative change in x .

Recall that for A an $n \times n$ matrix, $\|A\|_2 = \sigma_1$ where σ_1 is the largest singular value. The largest singular value of A^{-1} is therefore, $1/\sigma_n$ where σ_n is the smallest singular value of A . Therefore, the condition number reduces to σ_1/σ_n , the ratio of the largest to the smallest singular value of A provided the norm is the usual Euclidean norm.

13.3 The Spectral Radius

Even though it is in general impractical to compute the Jordan form, its existence is all that is needed in order to prove an important theorem about something which is relatively easy to compute. This is the spectral radius of a matrix.

Definition 13.3.1 Define $\sigma(A)$ to be the eigenvalues of A . Also,

$$\rho(A) \equiv \max(|\lambda| : \lambda \in \sigma(A))$$

The number, $\rho(A)$ is known as the spectral radius of A .

Recall the following symbols and their meaning.

$$\limsup_{n \rightarrow \infty} a_n, \liminf_{n \rightarrow \infty} a_n$$

They are respectively the largest and smallest limit points of the sequence $\{a_n\}$ where $\pm\infty$ is allowed in the case where the sequence is unbounded. They are also defined as

$$\begin{aligned} \limsup_{n \rightarrow \infty} a_n &\equiv \lim_{n \rightarrow \infty} (\sup \{a_k : k \geq n\}), \\ \liminf_{n \rightarrow \infty} a_n &\equiv \lim_{n \rightarrow \infty} (\inf \{a_k : k \geq n\}). \end{aligned}$$

Thus, the limit of the sequence exists if and only if these are both equal to the same real number. Also note that the

Lemma 13.3.2 Let J be a $p \times p$ Jordan matrix

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_s \end{pmatrix}$$

where each J_k is of the form

$$J_k = \lambda_k I + N_k$$

in which N_k is a nilpotent matrix having zeros down the main diagonal and ones down the super diagonal. Then

$$\lim_{n \rightarrow \infty} \|J^n\|^{1/n} = \rho$$

where $\rho = \max\{|\lambda_k|, k = 1, \dots, n\}$. Here the norm is the operator norm.

Proof: Consider one of the blocks, $|\lambda_k| < \rho$. Here J_k is $p \times p$.

$$\frac{1}{\rho^n} J_k^n = \frac{1}{\rho^n} \sum_{i=0}^p \binom{n}{i} N_k^i \lambda_k^{n-i}$$

Then

$$\left\| \frac{1}{\rho^n} J_k^n \right\| \leq \sum_{i=0}^p \binom{n}{i} \|N_k^i\| \frac{|\lambda_k^{n-i}|}{\rho^{n-i}} \frac{1}{\rho^i} \tag{13.9}$$

Now there are p numbers $\|N_k^i\|$ so you could pick the largest, C . Also

$$\frac{|\lambda_k^{n-i}|}{\rho^{n-i}} \leq \frac{|\lambda_k^{n-p}|}{\rho^{n-p}}$$

so 13.9 is dominated by

$$\leq C n^p \frac{|\lambda_k^{n-p}|}{\rho^{n-p}} \sum_{i=0}^p \frac{1}{\rho^i} \equiv \hat{C} \frac{|\lambda_k^{n-p}|}{\rho^{n-p}}$$

The ratio or root test shows that this converges to 0 as $n \rightarrow \infty$.

What happens when $|\lambda_k| = \rho$?

$$\frac{1}{\rho^n} J_k^n = \omega^n I + \sum_{i=1}^p \binom{n}{i} N_k^i \omega^{n-i} \frac{1}{\rho^i}$$

where $|\omega| = 1$.

$$\frac{1}{\rho^n} \|J_k^n\| \leq 1 + n^p C$$

Schlumberger

WHY WAIT FOR PROGRESS?

DARE TO DISCOVER

Discovery means many different things at Schlumberger. But it's the spirit that unites every single one of us. It doesn't matter whether they join our business, engineering or technology teams, our trainees push boundaries, break new ground and deliver the exceptional. If that excites you, then we want to hear from you.

careers.slb.com/recentgraduates





where $C = \max \{ \|N_k^i\|, i = 1, \dots, p, k = 1, \dots, s \} \sum_{i=1}^p \frac{1}{\rho^i}$. Thus

$$\frac{1}{\rho^n} \|J^n\| \leq \frac{1}{\rho^n} \sum_{k=1}^s \|J_k^n\| \leq s(1 + n^p C) = sn^p C \left(\frac{1}{n^p C} + 1 \right)$$

and so

$$\begin{aligned} \frac{1}{\rho} \limsup_{n \rightarrow \infty} \|J^n\|^{1/n} &\leq \limsup_{n \rightarrow \infty} s^{1/n} (n^p C)^{1/n} \left(\frac{1}{n^p C} + 1 \right)^{1/n} = 1 \\ \limsup_{n \rightarrow \infty} \|J^n\|^{1/n} &\leq \rho \end{aligned}$$

Next let \mathbf{x} be an eigenvector for λ , $|\lambda| = \rho$ and let $\|\mathbf{x}\| = 1$. Then

$$\rho^n = \rho^n \|\mathbf{x}\| = \|J^n \mathbf{x}\| \leq \|J^n\|$$

and so

$$\rho \leq \|J^n\|^{1/n}$$

Hence

$$\rho \geq \limsup_{n \rightarrow \infty} \|J^n\|^{1/n} \geq \liminf_{n \rightarrow \infty} \|J^n\|^{1/n} \geq \rho \blacksquare$$

The following theorem is due to Gelfand around 1941.

Theorem 13.3.3 (Gelfand) *Let A be a complex $p \times p$ matrix. Then if ρ is the absolute value of its largest eigenvalue,*

$$\lim_{n \rightarrow \infty} \|A^n\|^{1/n} = \rho.$$

Here $\|\cdot\|$ is any norm on $\mathcal{L}(\mathbb{C}^n, \mathbb{C}^n)$.

Proof: First assume $\|\cdot\|$ is the operator norm with respect to the usual Euclidean metric on \mathbb{C}^n . Then letting J denote the Jordan form of A , $S^{-1}AS = J$, it follows from Lemma 13.3.2

$$\begin{aligned} \limsup_{n \rightarrow \infty} \|A^n\|^{1/n} &= \limsup_{n \rightarrow \infty} \|S J^n S^{-1}\|^{1/n} \leq \limsup_{n \rightarrow \infty} (\|S\| \|S^{-1}\| \|J^n\|)^{1/n} \\ &\leq \limsup_{n \rightarrow \infty} (\|S\| \|S^{-1}\| \|J^n\|)^{1/n} = \rho \end{aligned}$$

Letting λ be the largest eigenvalue of A , $|\lambda| = \rho$, and $A\mathbf{x} = \lambda\mathbf{x}$ where $\|\mathbf{x}\| = 1$,

$$\|A^n\| \geq \|A^n \mathbf{x}\| = \rho^n$$

and so

$$\liminf_{n \rightarrow \infty} \|A^n\|^{1/n} \geq \rho \geq \limsup_{n \rightarrow \infty} \|A^n\|^{1/n}$$

It follows that $\liminf_{n \rightarrow \infty} \|A^n\|^{1/n} = \limsup_{n \rightarrow \infty} \|A^n\|^{1/n} = \lim_{n \rightarrow \infty} \|A^n\|^{1/n} = \rho$.

Now by equivalence of norms, if $\|\cdot\|$ is any other norm for the set of complex $p \times p$ matrices, there exist constants δ, Δ such that

$$\delta \|A^n\| \leq \|\|A^n\|\| \leq \Delta \|A^n\|$$

Then

$$\delta^{1/n} \|A^n\|^{1/n} \leq \|\|A^n\|\|^{1/n} \leq \Delta^{1/n} \|A^n\|^{1/n}$$

The limits exist and equal ρ for the ends of the above inequality. Hence, by the squeezing theorem, $\rho = \lim_{n \rightarrow \infty} \|\|A^n\|\|^{1/n}$. \blacksquare

Example 13.3.4 Consider $\begin{pmatrix} 9 & -1 & 2 \\ -2 & 8 & 4 \\ 1 & 1 & 8 \end{pmatrix}$. Estimate the absolute value of the largest

eigenvalue.

A laborious computation reveals the eigenvalues are 5, and 10. Therefore, the right answer in this case is 10. Consider $\|A^7\|^{1/7}$ where the norm is obtained by taking the maximum of all the absolute values of the entries. Thus

$$\begin{pmatrix} 9 & -1 & 2 \\ -2 & 8 & 4 \\ 1 & 1 & 8 \end{pmatrix}^7 = \begin{pmatrix} 8015\,625 & -1984\,375 & 3968\,750 \\ -3968\,750 & 6031\,250 & 7937\,500 \\ 1984\,375 & 1984\,375 & 6031\,250 \end{pmatrix}$$

and taking the seventh root of the largest entry gives

$$\rho(A) \approx 8015\,625^{1/7} = 9.688\,951\,236\,71.$$

Of course the interest lies primarily in matrices for which the exact roots to the characteristic equation are not known and in the theoretical significance.

13.4 Series And Sequences Of Linear Operators

Before beginning this discussion, it is necessary to define what is meant by convergence in $\mathcal{L}(X, Y)$.

Definition 13.4.1 Let $\{A_k\}_{k=1}^\infty$ be a sequence in $\mathcal{L}(X, Y)$ where X, Y are finite dimensional normed linear spaces. Then $\lim_{n \rightarrow \infty} A_k = A$ if for every $\varepsilon > 0$ there exists N such that if $n > N$, then

$$\|A - A_n\| < \varepsilon.$$

Here the norm refers to any of the norms defined on $\mathcal{L}(X, Y)$. By Corollary 13.0.8 and Theorem 8.2.3 it doesn't matter which one is used. Define the symbol for an infinite sum in the usual way. Thus

$$\sum_{k=1}^\infty A_k \equiv \lim_{n \rightarrow \infty} \sum_{k=1}^n A_k$$

Lemma 13.4.2 Suppose $\{A_k\}_{k=1}^\infty$ is a sequence in $\mathcal{L}(X, Y)$ where X, Y are finite dimensional normed linear spaces. Then if

$$\sum_{k=1}^\infty \|A_k\| < \infty,$$

It follows that

$$\sum_{k=1}^\infty A_k \tag{13.10}$$

exists (converges). In words, absolute convergence implies convergence. Also,

$$\left\| \sum_{k=1}^\infty A_k \right\| \leq \sum_{k=1}^\infty \|A_k\|$$

Proof: For $p \leq m \leq n$,

$$\left\| \sum_{k=1}^n A_k - \sum_{k=1}^m A_k \right\| \leq \sum_{k=p}^\infty \|A_k\|$$

and so for p large enough, this term on the right in the above inequality is less than ε . Since ε is arbitrary, this shows the partial sums of 13.10 are a Cauchy sequence. Therefore by Corollary 13.0.7 it follows that these partial sums converge. As to the last claim,

$$\left\| \sum_{k=1}^n A_k \right\| \leq \sum_{k=1}^n \|A_k\| \leq \sum_{k=1}^{\infty} \|A_k\|$$

Therefore, passing to the limit,

$$\left\| \sum_{k=1}^{\infty} A_k \right\| \leq \sum_{k=1}^{\infty} \|A_k\|. \blacksquare$$

Why is this last step justified? (Recall the triangle inequality $|||A| - |B||| \leq \|A - B\|$.)

Now here is a useful result for differential equations.

Theorem 13.4.3 *Let X be a finite dimensional inner product space and let $A \in \mathcal{L}(X, X)$. Define*

$$\Phi(t) \equiv \sum_{k=0}^{\infty} \frac{t^k A^k}{k!}$$

Then the series converges for each $t \in \mathbb{R}$. Also

$$\Phi'(t) \equiv \lim_{h \rightarrow 0} \frac{\Phi(t+h) - \Phi(t)}{h} = \sum_{k=1}^{\infty} \frac{t^{k-1} A^k}{(k-1)!} = A \sum_{k=0}^{\infty} \frac{t^k A^k}{k!} = A\Phi(t)$$

Also $A\Phi(t) = \Phi(t)A$ and for all $t, \Phi(t)\Phi(-t) = I$ so $\Phi(t)^{-1} = \Phi(-t), \Phi(0) = I$. (It is understood that $A^0 = I$ in the above formula.)

PREPARE FOR A LEADING ROLE.

English-taught MSc programmes in engineering: Aeronautical, Biomedical, Electronics, Mechanical, Communication systems and Transport systems. No tuition fees.

→ liu.se/master

li.u LINKÖPING UNIVERSITY



Proof: First consider the claim about convergence.

$$\sum_{k=0}^{\infty} \left\| \frac{t^k A^k}{k!} \right\| \leq \sum_{k=0}^{\infty} \frac{|t|^k \|A\|^k}{k!} = e^{|t|\|A\|} < \infty$$

so it converges by Lemma 13.4.2.

$$\begin{aligned} \frac{\Phi(t+h) - \Phi(t)}{h} &= \frac{1}{h} \sum_{k=0}^{\infty} \frac{\left((t+h)^k - t^k \right) A^k}{k!} \\ &= \frac{1}{h} \sum_{k=0}^{\infty} \frac{\left(k(t + \theta_k h)^{k-1} h \right) A^k}{k!} = \sum_{k=1}^{\infty} \frac{(t + \theta_k h)^{k-1} A^k}{(k-1)!} \end{aligned}$$

this by the mean value theorem. Note that the series converges thanks to Lemma 13.4.2. Here $\theta_k \in (0, 1)$. Thus

$$\begin{aligned} &\left\| \frac{\Phi(t+h) - \Phi(t)}{h} - \sum_{k=1}^{\infty} \frac{t^{k-1} A^k}{(k-1)!} \right\| = \left\| \sum_{k=1}^{\infty} \frac{\left((t + \theta_k h)^{k-1} - t^{k-1} \right) A^k}{(k-1)!} \right\| \\ &= \left\| \sum_{k=1}^{\infty} \frac{\left((k-1)(t + \tau_k \theta_k h)^{k-2} \theta_k h \right) A^k}{(k-1)!} \right\| = |h| \left\| \sum_{k=2}^{\infty} \frac{\left((t + \tau_k \theta_k h)^{k-2} \theta_k \right) A^k}{(k-2)!} \right\| \\ &\leq |h| \sum_{k=2}^{\infty} \frac{(|t| + |h|)^{k-2} \|A\|^{k-2}}{(k-2)!} \|A\|^2 = |h| e^{(|t|+|h|)\|A\|} \|A\|^2 \end{aligned}$$

so letting $|h| < 1$, this is no larger than $|h| e^{(|t|+1)\|A\|} \|A\|^2$. Hence the desired limit is valid. It is obvious that $A\Phi(t) = \Phi(t)A$. Also the formula shows that

$$\Phi'(t) = A\Phi(t) = \Phi(t)A, \quad \Phi(0) = I.$$

Now consider the claim about $\Phi(-t)$. The above computation shows that $\Phi'(-t) = A\Phi(-t)$ and so $\frac{d}{dt}(\Phi(-t)) = -\Phi'(-t) = -A\Phi(-t)$. Now let x, y be two vectors in X . Consider

$$(\Phi(-t)\Phi(t)x, y)_X$$

Then this equals (x, y) when $t = 0$. Take its derivative.

$$\begin{aligned} &((-\Phi'(-t)\Phi(t) + \Phi(-t)\Phi'(t))x, y)_X \\ &= ((-A\Phi(-t)\Phi(t) + \Phi(-t)A\Phi(t))x, y)_X \\ &= (0, y)_X = 0 \end{aligned}$$

Hence this scalar valued function equals a constant and so the constant must be $(x, y)_X$. Hence for all x, y , $(\Phi(-t)\Phi(t)x - x, y)_X = 0$ for all x, y and this is so in particular for $y = \Phi(-t)\Phi(t)x - x$ which shows that $\Phi(-t)\Phi(t) = I$. ■

As a special case, suppose $\lambda \in \mathbb{C}$ and consider

$$\sum_{k=0}^{\infty} \frac{t^k \lambda^k}{k!}$$

where $t \in \mathbb{R}$. In this case, $A_k = \frac{t^k \lambda^k}{k!}$ and you can think of it as being in $\mathcal{L}(\mathbb{C}, \mathbb{C})$. Then the following corollary is of great interest.

Corollary 13.4.4 *Let*

$$f(t) \equiv \sum_{k=0}^{\infty} \frac{t^k \lambda^k}{k!} \equiv 1 + \sum_{k=1}^{\infty} \frac{t^k \lambda^k}{k!}$$

Then this function is a well defined complex valued function and furthermore, it satisfies the initial value problem,

$$y' = \lambda y, y(0) = 1$$

Furthermore, if $\lambda = a + ib$,

$$|f|(t) = e^{at}.$$

Proof: The first part is a special case of the above theorem. Note that for $f(t) = u(t) + iv(t)$, both u, v are differentiable. This is because

$$u = \frac{f + \bar{f}}{2}, v = \frac{f - \bar{f}}{2i}.$$

Then from the differential equation,

$$(a + ib)(u + iv) = u' + iv'$$

and equating real and imaginary parts,

$$u' = au - bv, v' = av + bu.$$

Then a short computation shows

$$(u^2 + v^2)' = 2uu' + 2vv' = 2u(au - bv) + 2v(av + bu) = 2a(u^2 + v^2)$$

$$(u^2 + v^2)(0) = |f|^2(0) = 1$$

Now in general, if

$$y' = cy, y(0) = 1,$$

with c real it follows $y(t) = e^{ct}$. To see this,

$$y' - cy = 0$$

and so, multiplying both sides by e^{-ct} you get

$$\frac{d}{dt}(ye^{-ct}) = 0$$

and so ye^{-ct} equals a constant which must be 1 because of the initial condition $y(0) = 1$. Thus

$$(u^2 + v^2)(t) = e^{2at}$$

and taking square roots yields the desired conclusion. ■

Definition 13.4.5 The function in Corollary 13.4.4 given by that power series is denoted as

$$\exp(\lambda t) \text{ or } e^{\lambda t}.$$

The next lemma is normally discussed in advanced calculus courses but is proved here for the convenience of the reader. It is known as the root test.

Definition 13.4.6 For $\{a_n\}$ any sequence of real numbers

$$\limsup_{n \rightarrow \infty} a_n \equiv \lim_{n \rightarrow \infty} (\sup \{a_k : k \geq n\})$$

Similarly

$$\liminf_{n \rightarrow \infty} a_n \equiv \lim_{n \rightarrow \infty} (\inf \{a_k : k \geq n\})$$

In case A_n is an increasing (decreasing) sequence which is unbounded above (below) then it is understood that $\lim_{n \rightarrow \infty} A_n = \infty (-\infty)$ respectively. Thus either of \limsup or \liminf can equal $+\infty$ or $-\infty$. However, the important thing about these is that unlike the limit, these always exist.

It is convenient to think of these as the largest point which is the limit of some subsequence of $\{a_n\}$ and the smallest point which is the limit of some subsequence of $\{a_n\}$ respectively. Thus $\lim_{n \rightarrow \infty} a_n$ exists and equals some point of $[-\infty, \infty]$ if and only if the two are equal.

Lemma 13.4.7 Let $\{a_p\}$ be a sequence of nonnegative terms and let

$$r = \limsup_{p \rightarrow \infty} a_p^{1/p}.$$

Then if $r < 1$, it follows the series, $\sum_{k=1}^{\infty} a_k$ converges and if $r > 1$, then a_p fails to converge to 0 so the series diverges. If A is an $n \times n$ matrix and

$$r = \limsup_{p \rightarrow \infty} \|A^p\|^{1/p}, \tag{13.11}$$

then if $r > 1$, then $\sum_{k=0}^{\infty} A^k$ fails to converge and if $r < 1$ then the series converges. Note that the series converges when the spectral radius is less than one and diverges if the spectral radius is larger than one. In fact, $\limsup_{p \rightarrow \infty} \|A^p\|^{1/p} = \lim_{p \rightarrow \infty} \|A^p\|^{1/p}$ from Theorem 13.3.3.

Proof: Suppose $r < 1$. Then there exists N such that if $p > N$,

$$a_p^{1/p} < R$$

where $r < R < 1$. Therefore, for all such p , $a_p < R^p$ and so by comparison with the geometric series, $\sum R^p$, it follows $\sum_{p=1}^{\infty} a_p$ converges.

Next suppose $r > 1$. Then letting $1 < R < r$, it follows there are infinitely many values of p at which

$$R < a_p^{1/p}$$

which implies $R^p < a_p$, showing that a_p cannot converge to 0 and so the series cannot converge either.

To see the last claim, if $r > 1$, then $\|A^p\|$ fails to converge to 0 and so $\{\sum_{k=0}^m A^k\}_{m=0}^{\infty}$ is not a Cauchy sequence. Hence $\sum_{k=0}^{\infty} A^k \equiv \lim_{m \rightarrow \infty} \sum_{k=0}^m A^k$ cannot exist. If $r < 1$, then for all n large enough, $\|A^n\|^{1/n} \leq r < 1$ for some r so $\|A^n\| \leq r^n$. Hence $\sum_n \|A^n\|$ converges and so by Lemma 13.4.2, it follows that $\sum_{k=1}^{\infty} A^k$ also converges. ■

Now denote by $\sigma(A)^p$ the collection of all numbers of the form λ^p where $\lambda \in \sigma(A)$.



Lemma 13.4.8 $\sigma(A^p) = \sigma(A)^p \equiv \{\lambda^p : \lambda \in \sigma(A)\}$.

Proof: In dealing with $\sigma(A^p)$, it suffices to deal with $\sigma(J^p)$ where J is the Jordan form of A because J^p and A^p are similar. Thus if $\lambda \in \sigma(A^p)$, then $\lambda \in \sigma(J^p)$ and so $\lambda = \alpha^p$ where α is one of the entries on the main diagonal of J^p . These entries are of the form λ^p where $\lambda \in \sigma(A)$. Thus $\lambda \in \sigma(A)^p$ and this shows $\sigma(A^p) \subseteq \sigma(A)^p$.
 Now take $\alpha \in \sigma(A)$ and consider α^p .

$$\alpha^p I - A^p = (\alpha^{p-1} I + \dots + \alpha A^{p-2} + A^{p-1})(\alpha I - A)$$

and so $\alpha^p I - A^p$ fails to be one to one which shows that $\alpha^p \in \sigma(A^p)$ which shows that $\sigma(A)^p \subseteq \sigma(A^p)$. ■

13.5 Iterative Methods For Linear Systems

Consider the problem of solving the equation

$$Ax = b \tag{13.12}$$

where A is an $n \times n$ matrix. In many applications, the matrix A is huge and composed mainly of zeros. For such matrices, the method of Gauss elimination (row operations) is not a good way to solve the system because the row operations can destroy the zeros and storing all those zeros takes a lot of room in a computer. These systems are called sparse. To solve them, it is common to use an iterative technique. I am following the treatment given to this subject by Nobel and Daniel [21].

Definition 13.5.1 *The Jacobi iterative technique, also called the method of simultaneous corrections is defined as follows. Let \mathbf{x}^1 be an initial vector, say the zero vector or some other vector. The method generates a succession of vectors, $\mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4, \dots$ and hopefully this sequence of vectors will converge to the solution to 13.12. The vectors in this list are called iterates and they are obtained according to the following procedure. Letting $A = (a_{ij})$,*

$$a_{ii}x_i^{r+1} = - \sum_{j \neq i} a_{ij}x_j^r + b_i. \tag{13.13}$$

In terms of matrices, letting

$$A = \begin{pmatrix} * & \dots & * \\ \vdots & \ddots & \vdots \\ * & \dots & * \end{pmatrix}$$

The iterates are defined as

$$\begin{pmatrix} * & 0 & \dots & 0 \\ 0 & * & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & * \end{pmatrix} \begin{pmatrix} x_1^{r+1} \\ x_2^{r+1} \\ \vdots \\ x_n^{r+1} \end{pmatrix} = - \begin{pmatrix} 0 & * & \dots & * \\ * & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ * & \dots & * & 0 \end{pmatrix} \begin{pmatrix} x_1^r \\ x_2^r \\ \vdots \\ x_n^r \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \tag{13.14}$$

The matrix on the left in 13.14 is obtained by retaining the main diagonal of A and setting every other entry equal to zero. The matrix on the right in 13.14 is obtained from A by setting every diagonal entry equal to zero and retaining all the other entries unchanged.

Example 13.5.2 Use the Jacobi method to solve the system

$$\begin{pmatrix} 3 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 2 & 5 & 1 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

Of course this is solved most easily using row reductions. The Jacobi method is useful when the matrix is very large. This example is just to illustrate how the method works. First lets solve it using row operations. The exact solution from row reduction is $\left(\frac{6}{29} \quad \frac{11}{29} \quad \frac{8}{29} \quad \frac{25}{29} \right)$, which in terms of decimals is approximately equal to

$$\left(0.207 \quad 0.379 \quad 0.276 \quad 0.862 \right)^T.$$

In terms of the matrices, the Jacobi iteration is of the form

$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} x_1^{r+1} \\ x_2^{r+1} \\ x_3^{r+1} \\ x_4^{r+1} \end{pmatrix} = - \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{pmatrix} \begin{pmatrix} x_1^r \\ x_2^r \\ x_3^r \\ x_4^r \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

Multiplying by the inverse of the matrix on the left,¹this iteration reduces to

$$\begin{pmatrix} x_1^{r+1} \\ x_2^{r+1} \\ x_3^{r+1} \\ x_4^{r+1} \end{pmatrix} = - \begin{pmatrix} 0 & \frac{1}{3} & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{2}{5} & 0 & \frac{1}{5} \\ 0 & 0 & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} x_1^r \\ x_2^r \\ x_3^r \\ x_4^r \end{pmatrix} + \begin{pmatrix} \frac{1}{3} \\ \frac{1}{2} \\ \frac{3}{5} \\ 1 \end{pmatrix}. \tag{13.15}$$

Now iterate this starting with $\mathbf{x}^1 \equiv \left(0 \quad 0 \quad 0 \quad 0 \right)^T$.

Thus

$$\mathbf{x}^2 = - \begin{pmatrix} 0 & \frac{1}{3} & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{2}{5} & 0 & \frac{1}{5} \\ 0 & 0 & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{1}{3} \\ \frac{1}{2} \\ \frac{3}{5} \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{2} \\ \frac{3}{5} \\ 1 \end{pmatrix}$$

Then

$$\mathbf{x}^3 = - \begin{pmatrix} 0 & \frac{1}{3} & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{2}{5} & 0 & \frac{1}{5} \\ 0 & 0 & \frac{1}{2} & 0 \end{pmatrix} \overbrace{\begin{pmatrix} \frac{1}{3} \\ \frac{1}{2} \\ \frac{3}{5} \\ 1 \end{pmatrix}}^{\mathbf{x}_2} + \begin{pmatrix} \frac{1}{3} \\ \frac{1}{2} \\ \frac{3}{5} \\ 1 \end{pmatrix} = \begin{pmatrix} .166 \\ .26 \\ .2 \\ .7 \end{pmatrix}$$

Continuing this way one finally gets

$$\mathbf{x}^6 = - \begin{pmatrix} 0 & \frac{1}{3} & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{2}{5} & 0 & \frac{1}{5} \\ 0 & 0 & \frac{1}{2} & 0 \end{pmatrix} \overbrace{\begin{pmatrix} .197 \\ .351 \\ .2566 \\ .822 \end{pmatrix}}^{\mathbf{x}_5} + \begin{pmatrix} \frac{1}{3} \\ \frac{1}{2} \\ \frac{3}{5} \\ 1 \end{pmatrix} = \begin{pmatrix} .216 \\ .386 \\ .295 \\ .871 \end{pmatrix}.$$

You can keep going like this. Recall the solution is approximately equal to

$$\left(0.206 \quad 0.379 \quad 0.275 \quad 0.862 \right)^T$$

¹You certainly would not compute the inverse in solving a large system. This is just to show you how the method works for this simple example. You would use the first description in terms of indices.

so you see that with no care at all and only 6 iterations, an approximate solution has been obtained which is not too far off from the actual solution.

Definition 13.5.3 *The Gauss Seidel method, also called the method of successive corrections is given as follows. For $A = (a_{ij})$, the iterates for the problem $A\mathbf{x} = \mathbf{b}$ are obtained according to the formula*

$$\sum_{j=1}^i a_{ij}x_j^{r+1} = - \sum_{j=i+1}^n a_{ij}x_j^r + b_i. \tag{13.16}$$

In terms of matrices, letting


$$A = \begin{pmatrix} * & \cdots & * \\ \vdots & \ddots & \vdots \\ * & \cdots & * \end{pmatrix}$$

The iterates are defined as

$$\begin{pmatrix} * & 0 & \cdots & 0 \\ * & * & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ * & \cdots & * & * \end{pmatrix} \begin{pmatrix} x_1^{r+1} \\ x_2^{r+1} \\ \vdots \\ x_n^{r+1} \end{pmatrix} = - \begin{pmatrix} 0 & * & \cdots & * \\ 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1^r \\ x_2^r \\ \vdots \\ x_n^r \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \tag{13.17}$$

In words, you set every entry in the original matrix which is strictly above the main diagonal equal to zero to obtain the matrix on the left. To get the matrix on the right, you set every entry of A which is on or below the main diagonal equal to zero. Using the iteration procedure of 13.16 directly, the Gauss Seidel method makes use of the very latest information which is available at that stage of the computation.

The following example is the same as the example used to illustrate the Jacobi method.



WE ARE SHAPING MOBILITY FOR TOMORROW

How will people travel in the future, and how will goods be transported? What resources will we use, and how many will we need? The passenger and freight traffic sector is developing rapidly, and we provide the impetus for innovation and movement. We develop components and systems for internal combustion engines that operate more cleanly and more efficiently than ever before. We are also pushing forward technologies that are bringing hybrid vehicles and alternative drives into a new dimension – for private, corporate, and public use. The challenges are great. We deliver the solutions and offer challenging jobs.

www.schaeffler.com/careers

SCHAEFFLER



Example 13.5.4 Use the Gauss Seidel method to solve the system

$$\begin{pmatrix} 3 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 2 & 5 & 1 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

In terms of matrices, this procedure is

$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 \\ 0 & 2 & 5 & 0 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1^{r+1} \\ x_2^{r+1} \\ x_3^{r+1} \\ x_4^{r+1} \end{pmatrix} = - \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1^r \\ x_2^r \\ x_3^r \\ x_4^r \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

Multiplying by the inverse of the matrix on the left² this yields

$$\begin{pmatrix} x_1^{r+1} \\ x_2^{r+1} \\ x_3^{r+1} \\ x_4^{r+1} \end{pmatrix} = - \begin{pmatrix} 0 & \frac{1}{3} & 0 & 0 \\ 0 & -\frac{1}{12} & \frac{1}{4} & 0 \\ 0 & \frac{1}{30} & -\frac{1}{10} & \frac{1}{5} \\ 0 & -\frac{1}{60} & \frac{1}{20} & -\frac{1}{10} \end{pmatrix} \begin{pmatrix} x_1^r \\ x_2^r \\ x_3^r \\ x_4^r \end{pmatrix} + \begin{pmatrix} \frac{1}{3} \\ \frac{5}{12} \\ \frac{13}{30} \\ \frac{47}{60} \end{pmatrix}$$

As before, I will be totally unoriginal in the choice of \mathbf{x}^1 . Let it equal the zero vector.

Therefore, $\mathbf{x}^2 = \left(\frac{1}{3} \quad \frac{5}{12} \quad \frac{13}{30} \quad \frac{47}{60} \right)^T$. Now

$$\mathbf{x}^3 = - \begin{pmatrix} 0 & \frac{1}{3} & 0 & 0 \\ 0 & -\frac{1}{12} & \frac{1}{4} & 0 \\ 0 & \frac{1}{30} & -\frac{1}{10} & \frac{1}{5} \\ 0 & -\frac{1}{60} & \frac{1}{20} & -\frac{1}{10} \end{pmatrix} \overbrace{\begin{pmatrix} \frac{1}{3} \\ \frac{5}{12} \\ \frac{13}{30} \\ \frac{47}{60} \end{pmatrix}}^{\mathbf{x}^2} + \begin{pmatrix} \frac{1}{3} \\ \frac{5}{12} \\ \frac{13}{30} \\ \frac{47}{60} \end{pmatrix} = \begin{pmatrix} .194 \\ .343 \\ .306 \\ .846 \end{pmatrix}.$$

Continuing this way,

$$\mathbf{x}^4 = - \begin{pmatrix} 0 & \frac{1}{3} & 0 & 0 \\ 0 & -\frac{1}{12} & \frac{1}{4} & 0 \\ 0 & \frac{1}{30} & -\frac{1}{10} & \frac{1}{5} \\ 0 & -\frac{1}{60} & \frac{1}{20} & -\frac{1}{10} \end{pmatrix} \begin{pmatrix} .194 \\ .343 \\ .306 \\ .846 \end{pmatrix} + \begin{pmatrix} \frac{1}{3} \\ \frac{5}{12} \\ \frac{13}{30} \\ \frac{47}{60} \end{pmatrix} = \begin{pmatrix} .219 \\ .36875 \\ .2833 \\ .85835 \end{pmatrix}$$

and so

$$\mathbf{x}^5 = - \begin{pmatrix} 0 & \frac{1}{3} & 0 & 0 \\ 0 & -\frac{1}{12} & \frac{1}{4} & 0 \\ 0 & \frac{1}{30} & -\frac{1}{10} & \frac{1}{5} \\ 0 & -\frac{1}{60} & \frac{1}{20} & -\frac{1}{10} \end{pmatrix} \begin{pmatrix} .219 \\ .36875 \\ .2833 \\ .85835 \end{pmatrix} + \begin{pmatrix} \frac{1}{3} \\ \frac{5}{12} \\ \frac{13}{30} \\ \frac{47}{60} \end{pmatrix} = \begin{pmatrix} .21042 \\ .37657 \\ .2777 \\ .86115 \end{pmatrix}.$$

This is fairly close to the answer. You could continue doing these iterates and it appears they converge to the solution. Now consider the following example.

Example 13.5.5 Use the Gauss Seidel method to solve the system

$$\begin{pmatrix} 1 & 4 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 2 & 5 & 1 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

²As in the case of the Jacobi iteration, the computer would not do this. It would use the iteration procedure in terms of the entries of the matrix directly. Otherwise all benefit to using this method is lost.

The exact solution is given by doing row operations on the augmented matrix. When this is done the solution is seen to be $\begin{pmatrix} 6.0 & -1.25 & 1.0 & 0.5 \end{pmatrix}$. The Gauss Seidel iterations are of the form

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 \\ 0 & 2 & 5 & 0 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1^{r+1} \\ x_2^{r+1} \\ x_3^{r+1} \\ x_4^{r+1} \end{pmatrix} = - \begin{pmatrix} 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1^r \\ x_2^r \\ x_3^r \\ x_4^r \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

and so, multiplying by the inverse of the matrix on the left, the iteration reduces to the following in terms of matrix multiplication.

$$\mathbf{x}^{r+1} = - \begin{pmatrix} 0 & 4 & 0 & 0 \\ 0 & -1 & \frac{1}{4} & 0 \\ 0 & \frac{2}{5} & -\frac{1}{10} & \frac{1}{5} \\ 0 & -\frac{1}{5} & \frac{1}{20} & -\frac{1}{10} \end{pmatrix} \mathbf{x}^r + \begin{pmatrix} 1 \\ \frac{1}{4} \\ \frac{1}{2} \\ \frac{3}{4} \end{pmatrix}.$$

This time, I will pick an initial vector close to the answer. Let $\mathbf{x}^1 = \begin{pmatrix} 6 & -1 & 1 & \frac{1}{2} \end{pmatrix}^T$. This is very close to the answer. Now lets see what the Gauss Seidel iteration does to it.

$$\mathbf{x}^2 = - \begin{pmatrix} 0 & 4 & 0 & 0 \\ 0 & -1 & \frac{1}{4} & 0 \\ 0 & \frac{2}{5} & -\frac{1}{10} & \frac{1}{5} \\ 0 & -\frac{1}{5} & \frac{1}{20} & -\frac{1}{10} \end{pmatrix} \begin{pmatrix} 6 \\ -1 \\ 1 \\ \frac{1}{2} \end{pmatrix} + \begin{pmatrix} 1 \\ \frac{1}{4} \\ \frac{1}{2} \\ \frac{3}{4} \end{pmatrix} = \begin{pmatrix} 5.0 \\ -1.0 \\ .9 \\ .55 \end{pmatrix}$$

It appears that it moved the initial guess far from the solution even though you started with one which was initially close to the solution. This is discouraging. However, you can't expect the method to work well after only one iteration. Unfortunately, if you do multiple iterations, the iterates never seem to get close to the actual solution. Why is the process which worked so well in the other examples not working here? A better question might be: Why does either process ever work at all?

Both iterative procedures for solving

$$A\mathbf{x} = \mathbf{b} \tag{13.18}$$

are of the form

$$B\mathbf{x}^{r+1} = -C\mathbf{x}^r + \mathbf{b}$$

where $A = B + C$. In the Jacobi procedure, the matrix C was obtained by setting the diagonal of A equal to zero and leaving all other entries the same while the matrix B was obtained by making every entry of A equal to zero other than the diagonal entries which are left unchanged. In the Gauss Seidel procedure, the matrix B was obtained from A by making every entry strictly above the main diagonal equal to zero and leaving the others unchanged, and C was obtained from A by making every entry on or below the main diagonal equal to zero and leaving the others unchanged. Thus in the Jacobi procedure, B is a diagonal matrix while in the Gauss Seidel procedure, B is lower triangular. Using matrices to explicitly solve for the iterates, yields

$$\mathbf{x}^{r+1} = -B^{-1}C\mathbf{x}^r + B^{-1}\mathbf{b}. \tag{13.19}$$

This is what you would never have the computer do but this is what will allow the statement of a theorem which gives the condition for convergence of these and all other similar methods. Recall the definition of the spectral radius of $M, \rho(M)$, in Definition 13.3.1 on Page 337.

Theorem 13.5.6 Suppose $\rho(B^{-1}C) < 1$. Then the iterates in 13.19 converge to the unique solution of 13.18.

I will prove this theorem in the next section. The proof depends on analysis which should not be surprising because it involves a statement about convergence of sequences.

What is an easy to verify sufficient condition which will imply the above holds? It is easy to give one in the case of the Jacobi method. Suppose the matrix A is diagonally dominant. That is $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$. Then B would be the diagonal matrix consisting of the entries a_{ii} . You need to find the size of λ where

$$B^{-1}C\mathbf{x} = \lambda\mathbf{x}$$

Thus you need

$$(\lambda B - C)\mathbf{x} = \mathbf{0}$$

Now if $|\lambda| \geq 1$, then the matrix $\lambda B - C$ is diagonally dominant and so this matrix will be invertible so λ is not an eigenvalue. Hence the only eigenvalues have absolute value less than 1.

You might try a similar argument in the case of the Gauss Seidel method.

13.6 Theory Of Convergence

Definition 13.6.1 A normed vector space, E with norm $\|\cdot\|$ is called a Banach space if it is also complete. This means that every Cauchy sequence converges. Recall that a sequence $\{x_n\}_{n=1}^\infty$ is a Cauchy sequence if for every $\varepsilon > 0$ there exists N such that whenever $m, n > N$,

$$\|x_n - x_m\| < \varepsilon.$$

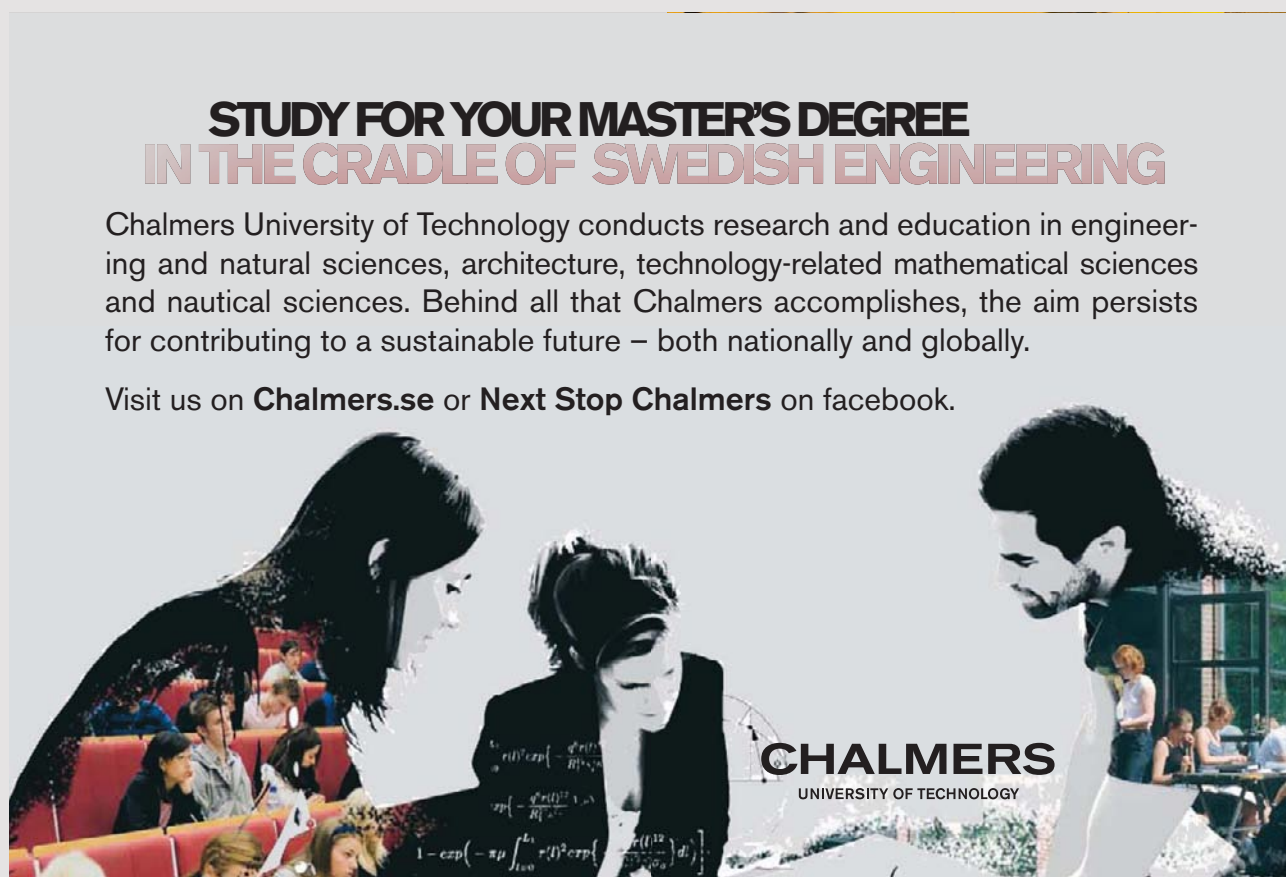
Thus whenever $\{x_n\}$ is a Cauchy sequence, there exists x such that

$$\lim_{n \rightarrow \infty} \|x - x_n\| = 0.$$

STUDY FOR YOUR MASTER'S DEGREE IN THE CRADLE OF SWEDISH ENGINEERING

Chalmers University of Technology conducts research and education in engineering and natural sciences, architecture, technology-related mathematical sciences and nautical sciences. Behind all that Chalmers accomplishes, the aim persists for contributing to a sustainable future – both nationally and globally.

Visit us on Chalmers.se or [Next Stop Chalmers](#) on facebook.



CHALMERS
UNIVERSITY OF TECHNOLOGY



Example 13.6.2 Let E be a Banach space and let Ω be a nonempty subset of a normed linear space F . Let $B(\Omega; E)$ denote those functions f for which

$$\|f\| \equiv \sup \{ \|f(x)\|_E : x \in \Omega \} < \infty$$

Denote by $BC(\Omega; E)$ the set of functions in $B(\Omega; E)$ which are also continuous.

Lemma 13.6.3 The above $\|\cdot\|$ is a norm on $B(\Omega; E)$. The subspace $BC(\Omega; E)$ with the given norm is a Banach space.

Proof: It is obvious $\|\cdot\|$ is a norm. It only remains to verify $BC(\Omega; E)$ is complete. Let $\{f_n\}$ be a Cauchy sequence. Since $\|f_n - f_m\| \rightarrow 0$ as $m, n \rightarrow \infty$, it follows that $\{f_n(x)\}$ is a Cauchy sequence in E for each x . Let $f(x) \equiv \lim_{n \rightarrow \infty} f_n(x)$. Then for any $x \in \Omega$,

$$\|f_n(x) - f_m(x)\|_E \leq \|f_n - f_m\| < \varepsilon$$

whenever m, n are large enough, say as large as N . For $n \geq N$, let $m \rightarrow \infty$. Then passing to the limit, it follows that for all x ,

$$\|f_n(x) - f(x)\|_E \leq \varepsilon$$

and so for all x ,

$$\|f(x)\|_E \leq \varepsilon + \|f_n(x)\|_E \leq \varepsilon + \|f_n\|.$$

It follows that $\|f\| \leq \|f_n\| + \varepsilon$ and $\|f - f_n\| \leq \varepsilon$.

It remains to verify that f is continuous.

$$\begin{aligned} \|f(x) - f(y)\|_E &\leq \|f(x) - f_n(x)\|_E + \|f_n(x) - f_n(y)\|_E + \|f_n(y) - f(y)\|_E \\ &\leq 2\|f - f_n\| + \|f_n(x) - f_n(y)\|_E < \frac{2\varepsilon}{3} + \|f_n(x) - f_n(y)\|_E \end{aligned}$$

for all n large enough. Now pick such an n . By continuity, the last term is less than $\frac{\varepsilon}{3}$ if $\|x - y\|$ is small enough. Hence f is continuous as well. ■

The most familiar example of a Banach space is \mathbb{F}^n . The following lemma is of great importance so it is stated in general.

Lemma 13.6.4 Suppose $T : E \rightarrow E$ where E is a Banach space with norm $|\cdot|$. Also suppose

$$|T\mathbf{x} - T\mathbf{y}| \leq r|\mathbf{x} - \mathbf{y}| \tag{13.20}$$

for some $r \in (0, 1)$. Then there exists a unique fixed point, $\mathbf{x} \in E$ such that

$$T\mathbf{x} = \mathbf{x}. \tag{13.21}$$

Letting $\mathbf{x}^1 \in E$, this fixed point \mathbf{x} , is the limit of the sequence of iterates,

$$\mathbf{x}^1, T\mathbf{x}^1, T^2\mathbf{x}^1, \dots \tag{13.22}$$

In addition to this, there is a nice estimate which tells how close \mathbf{x}^1 is to \mathbf{x} in terms of things which can be computed.

$$|\mathbf{x}^1 - \mathbf{x}| \leq \frac{1}{1-r} |\mathbf{x}^1 - T\mathbf{x}^1|. \tag{13.23}$$

Proof: This follows easily when it is shown that the above sequence, $\{T^k\mathbf{x}^1\}_{k=1}^\infty$ is a Cauchy sequence. Note that

$$|T^2\mathbf{x}^1 - T\mathbf{x}^1| \leq r|T\mathbf{x}^1 - \mathbf{x}^1|.$$

Suppose

$$|T^k\mathbf{x}^1 - T^{k-1}\mathbf{x}^1| \leq r^{k-1}|T\mathbf{x}^1 - \mathbf{x}^1|. \tag{13.24}$$

Then

$$\begin{aligned} |T^{k+1}\mathbf{x}^1 - T^k\mathbf{x}^1| &\leq r|T^k\mathbf{x}^1 - T^{k-1}\mathbf{x}^1| \\ &\leq rr^{k-1}|T\mathbf{x}^1 - \mathbf{x}^1| = r^k|T\mathbf{x}^1 - \mathbf{x}^1|. \end{aligned}$$

By induction, this shows that for all $k \geq 2$, 13.24 is valid. Now let $k > l \geq N$.

$$\begin{aligned} |T^k \mathbf{x}^1 - T^l \mathbf{x}^1| &= \left| \sum_{j=l}^{k-1} (T^{j+1} \mathbf{x}^1 - T^j \mathbf{x}^1) \right| \leq \sum_{j=l}^{k-1} |T^{j+1} \mathbf{x}^1 - T^j \mathbf{x}^1| \\ &\leq \sum_{j=N}^{k-1} r^j |T \mathbf{x}^1 - \mathbf{x}^1| \leq |T \mathbf{x}^1 - \mathbf{x}^1| \frac{r^N}{1-r} \end{aligned}$$

which converges to 0 as $N \rightarrow \infty$. Therefore, this is a Cauchy sequence so it must converge to $\mathbf{x} \in E$. Then

$$\mathbf{x} = \lim_{k \rightarrow \infty} T^k \mathbf{x}^1 = \lim_{k \rightarrow \infty} T^{k+1} \mathbf{x}^1 = T \lim_{k \rightarrow \infty} T^k \mathbf{x}^1 = T \mathbf{x}.$$

This shows the existence of the fixed point. To show it is unique, suppose there were another one, \mathbf{y} . Then

$$|\mathbf{x} - \mathbf{y}| = |T \mathbf{x} - T \mathbf{y}| \leq r |\mathbf{x} - \mathbf{y}|$$

and so $\mathbf{x} = \mathbf{y}$.

It remains to verify the estimate.

$$\begin{aligned} |\mathbf{x}^1 - \mathbf{x}| &\leq |\mathbf{x}^1 - T \mathbf{x}^1| + |T \mathbf{x}^1 - \mathbf{x}| = |\mathbf{x}^1 - T \mathbf{x}^1| + |T \mathbf{x}^1 - T \mathbf{x}| \\ &\leq |\mathbf{x}^1 - T \mathbf{x}^1| + r |\mathbf{x}^1 - \mathbf{x}| \end{aligned}$$

and solving the inequality for $|\mathbf{x}^1 - \mathbf{x}|$ gives the estimate desired. ■

The following corollary is what will be used to prove the convergence condition for the various iterative procedures.

Corollary 13.6.5 *Suppose $T : E \rightarrow E$, for some constant C*

$$|T \mathbf{x} - T \mathbf{y}| \leq C |\mathbf{x} - \mathbf{y}|,$$

for all $\mathbf{x}, \mathbf{y} \in E$, and for some $N \in \mathbb{N}$,

$$|T^N \mathbf{x} - T^N \mathbf{y}| \leq r |\mathbf{x} - \mathbf{y}|,$$

for all $\mathbf{x}, \mathbf{y} \in E$ where $r \in (0, 1)$. Then there exists a unique fixed point for T and it is still the limit of the sequence, $\{T^k \mathbf{x}^1\}$ for any choice of \mathbf{x}^1 .

Proof: From Lemma 13.6.4 there exists a unique fixed point for T^N denoted here as \mathbf{x} . Therefore, $T^N \mathbf{x} = \mathbf{x}$. Now doing T to both sides,

$$T^N T \mathbf{x} = T \mathbf{x}.$$

By uniqueness, $T \mathbf{x} = \mathbf{x}$ because the above equation shows $T \mathbf{x}$ is a fixed point of T^N and there is only one fixed point of T^N . In fact, there is only one fixed point of T because a fixed point of T is automatically a fixed point of T^N .

It remains to show $T^k \mathbf{x}^1 \rightarrow \mathbf{x}$, the unique fixed point of T^N . If this does not happen, there exists $\varepsilon > 0$ and a subsequence, still denoted by T^k such that

$$|T^k \mathbf{x}^1 - \mathbf{x}| \geq \varepsilon$$

Now $k = j_k N + r_k$ where $r_k \in \{0, \dots, N-1\}$ and j_k is a positive integer such that $\lim_{k \rightarrow \infty} j_k = \infty$. Then there exists a single $r \in \{0, \dots, N-1\}$ such that for infinitely many $k, r_k = r$. Taking a further subsequence, still denoted by T^k it follows

$$|T^{j_k N + r} \mathbf{x}^1 - \mathbf{x}| \geq \varepsilon \tag{13.25}$$

However,

$$T^{j_k N + r} \mathbf{x}^1 = T^r T^{j_k N} \mathbf{x}^1 \rightarrow T^r \mathbf{x} = \mathbf{x}$$

and this contradicts 13.25. ■

Theorem 13.6.6 *Suppose $\rho(B^{-1}C) < 1$. Then the iterates in 13.19 converge to the unique solution of 13.18.*

Proof: Consider the iterates in 13.19. Let $T\mathbf{x} = B^{-1}C\mathbf{x} + B^{-1}\mathbf{b}$. Then

$$|T^k\mathbf{x} - T^k\mathbf{y}| = |(B^{-1}C)^k\mathbf{x} - (B^{-1}C)^k\mathbf{y}| \leq \left\| (B^{-1}C)^k \right\| |\mathbf{x} - \mathbf{y}|.$$

Here $\|\cdot\|$ refers to any of the operator norms. It doesn't matter which one you pick because they are all equivalent. I am writing the proof to indicate the operator norm taken with respect to the usual norm on E . Since $\rho(B^{-1}C) < 1$, it follows from Gelfand's theorem, Theorem 13.3.3 on Page 338, there exists N such that if $k \geq N$, then for some $r^{1/k} < 1$,

$$\left\| (B^{-1}C)^k \right\|^{1/k} < r^{1/k} < 1.$$

Consequently,

$$|T^N\mathbf{x} - T^N\mathbf{y}| \leq r |\mathbf{x} - \mathbf{y}|.$$

Also $|T\mathbf{x} - T\mathbf{y}| \leq \|B^{-1}C\| |\mathbf{x} - \mathbf{y}|$ and so Corollary 13.6.5 applies and gives the conclusion of this theorem. ■



Scholarships



Lnu.se

Open your mind to new opportunities

With 31,000 students, Linnaeus University is one of the larger universities in Sweden. We are a modern university, known for our strong international profile. Every year more than 1,600 international students from all over the world choose to enjoy the friendly atmosphere and active student life at Linnaeus University. Welcome to join us!

Linnaeus University

Sweden

Bachelor programmes in
Business & Economics | Computer Science/IT | Design | Mathematics

Master programmes in
Business & Economics | Behavioural Sciences | Computer Science/IT | Cultural Studies & Social Sciences | Design | Mathematics | Natural Sciences | Technology & Engineering

Summer Academy courses



13.7 Exercises

1. Solve the system

$$\begin{pmatrix} 4 & 1 & 1 \\ 1 & 5 & 2 \\ 0 & 2 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

using the Gauss Seidel method and the Jacobi method. Check your answer by also solving it using row operations.

2. Solve the system

$$\begin{pmatrix} 4 & 1 & 1 \\ 1 & 7 & 2 \\ 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

using the Gauss Seidel method and the Jacobi method. Check your answer by also solving it using row operations.

3. Solve the system

$$\begin{pmatrix} 5 & 1 & 1 \\ 1 & 7 & 2 \\ 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

using the Gauss Seidel method and the Jacobi method. Check your answer by also solving it using row operations.

4. If you are considering a system of the form $A\mathbf{x} = \mathbf{b}$ and A^{-1} does not exist, will either the Gauss Seidel or Jacobi methods work? Explain. What does this indicate about finding eigenvectors for a given eigenvalue?
5. For $\|\mathbf{x}\|_\infty \equiv \max\{|x_j| : j = 1, 2, \dots, n\}$, the parallelogram identity does not hold. Explain.
6. A norm $\|\cdot\|$ is said to be strictly convex if whenever $\|x\| = \|y\|, x \neq y$, it follows

$$\left\| \frac{x+y}{2} \right\| < \|x\| = \|y\|.$$

Show the norm $|\cdot|$ which comes from an inner product is strictly convex.

7. A norm $\|\cdot\|$ is said to be uniformly convex if whenever $\|x_n\|, \|y_n\|$ are equal to 1 for all $n \in \mathbb{N}$ and $\lim_{n \rightarrow \infty} \|x_n + y_n\| = 2$, it follows $\lim_{n \rightarrow \infty} \|x_n - y_n\| = 0$. Show the norm $|\cdot|$ coming from an inner product is always uniformly convex. Also show that uniform convexity implies strict convexity which is defined in Problem 6.
8. Suppose $A : \mathbb{C}^n \rightarrow \mathbb{C}^n$ is a one to one and onto matrix. Define

$$\|\mathbf{x}\| \equiv |A\mathbf{x}|.$$

Show this is a norm.

9. If X is a finite dimensional normed vector space and $A, B \in \mathcal{L}(X, X)$ such that $\|B\| < \|A\|$, can it be concluded that $\|A^{-1}B\| < 1$?
10. Let X be a vector space with a norm $\|\cdot\|$ and let $V = \text{span}(v_1, \dots, v_m)$ be a finite dimensional subspace of X such that $\{v_1, \dots, v_m\}$ is a basis for V . Show V is a closed subspace of X . This means that if $w_n \rightarrow w$ and each $w_n \in V$, then so is w . Next show that if $w \notin V$,

$$\text{dist}(w, V) \equiv \inf\{\|w - v\| : v \in V\} > 0$$

is a continuous function of w and

$$|\text{dist}(w, V) - \text{dist}(w_1, V)| \leq \|w_1 - w\|$$

Next show that if $w \notin V$, there exists z such that $\|z\| = 1$ and $\text{dist}(z, V) > 1/2$. For those who know some advanced calculus, show that if X is an infinite dimensional vector space having norm $\|\cdot\|$, then the closed unit ball in X cannot be compact. Thus closed and bounded is never compact in an infinite dimensional normed vector space.

11. Suppose $\rho(A) < 1$ for $A \in \mathcal{L}(V, V)$ where V is a p dimensional vector space having a norm $\|\cdot\|$. You can use \mathbb{R}^p or \mathbb{C}^p if you like. Show there exists a new norm $\|\|\cdot\|\|$ such that with respect to this new norm, $\|\|A\|\| < 1$ where $\|\|A\|\|$ denotes the operator norm of A taken with respect to this new norm on V ,

$$\|\|A\|\| \equiv \sup \{ \|\|A\mathbf{x}\|\| : \|\|\mathbf{x}\|\| \leq 1 \}$$

Hint: You know from Gelfand's theorem that

$$\|A^n\|^{1/n} < r < 1$$

provided n is large enough, this operator norm taken with respect to $\|\cdot\|$. Show there exists $0 < \lambda < 1$ such that

$$\rho\left(\frac{A}{\lambda}\right) < 1.$$

You can do this by arguing the eigenvalues of A/λ are the scalars μ/λ where $\mu \in \sigma(A)$. Now let \mathbb{Z}_+ denote the nonnegative integers.

$$\|\|\mathbf{x}\|\| \equiv \sup_{n \in \mathbb{Z}_+} \left\| \frac{A^n}{\lambda^n} \mathbf{x} \right\|$$

First show this is actually a norm. Next explain why

$$\|\|A\mathbf{x}\|\| \equiv \lambda \sup_{n \in \mathbb{Z}_+} \left\| \frac{A^{n+1}}{\lambda^{n+1}} \mathbf{x} \right\| \leq \lambda \|\|\mathbf{x}\|\|.$$

12. Establish a similar result to Problem 11 without using Gelfand's theorem. Use an argument which depends directly on the Jordan form or a modification of it.
13. Using Problem 11 give an easier proof of Theorem 13.6.6 without having to use Corollary 13.6.5. It would suffice to use a different norm of this problem and the contraction mapping principle of Lemma 13.6.4.
14. A matrix A is diagonally dominant if $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$. Show that the Gauss Seidel method converges if A is diagonally dominant.
15. Suppose $f(\lambda) = \sum_{n=0}^{\infty} a_n \lambda^n$ converges if $|\lambda| < R$. Show that if $\rho(A) < R$ where A is an $n \times n$ matrix, then

$$f(A) \equiv \sum_{n=0}^{\infty} a_n A^n$$

converges in $\mathcal{L}(\mathbb{F}^n, \mathbb{F}^n)$. **Hint:** Use Gelfand's theorem and the root test.

16. Referring to Corollary 13.4.4, for $\lambda = a + ib$ show

$$\exp(\lambda t) = e^{at} (\cos(bt) + i \sin(bt)).$$

Hint: Let $y(t) = \exp(\lambda t)$ and let $z(t) = e^{-at} y(t)$. Show

$$z'' + b^2 z = 0, \quad z(0) = 1, \quad z'(0) = ib.$$

Now letting $z = u + iv$ where u, v are real valued, show

$$\begin{aligned} u'' + b^2 u &= 0, & u(0) &= 1, & u'(0) &= 0 \\ v'' + b^2 v &= 0, & v(0) &= 0, & v'(0) &= b. \end{aligned}$$

Next show $u(t) = \cos(bt)$ and $v(t) = \sin(bt)$ work in the above and that there is at most one solution to

$$w'' + b^2w = 0 \quad w(0) = \alpha, w'(0) = \beta.$$

Thus $z(t) = \cos(bt) + i \sin(bt)$ and so $y(t) = e^{at}(\cos(bt) + i \sin(bt))$. To show there is at most one solution to the above problem, suppose you have two, w_1, w_2 . Subtract them. Let $f = w_1 - w_2$. Thus

$$f'' + b^2f = 0$$

and f is real valued. Multiply both sides by f' and conclude

$$\frac{d}{dt} \left(\frac{(f')^2}{2} + b^2 \frac{f^2}{2} \right) = 0$$

Thus the expression in parenthesis is constant. Explain why this constant must equal 0.

17. Let $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$. Show the following power series converges in $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$.

$$\Psi(t) \equiv \sum_{k=0}^{\infty} \frac{t^k A^k}{k!}$$

This was done in the chapter. Go over it and be sure you understand it. This is how you can define $\exp(tA)$. Next show that $\Psi'(t) = A\Psi(t), \Psi(0) = I$. Next let $\Phi(t) = \sum_{k=0}^{\infty} \frac{t^k (-A)^k}{k!}$. Show each $\Phi(t), \Psi(t)$ each commute with A . Next show that $\Phi(t)\Psi(t) = I$ for all t . Finally, solve the initial value problem

$$\mathbf{x}' = A\mathbf{x} + \mathbf{f}, \quad \mathbf{x}(0) = \mathbf{x}_0$$

e-learning for kids

- The number 1 MOOC for Primary Education
- Free Digital Learning for Children 5-12
- 15 Million Children Reached

About e-Learning for Kids Established in 2004, e-Learning for Kids is a global nonprofit foundation dedicated to fun and free learning on the Internet for children ages 5 - 12 with courses in math, science, language arts, computers, health and environmental skills. Since 2005, more than 15 million children in over 190 countries have benefitted from eLessons provided by EFKI. An all-volunteer staff consists of education and e-learning experts and business professionals from around the world committed to making difference. eLearning for Kids is actively seeking funding, volunteers, sponsors and courseware developers; get involved! For more information, please visit www.e-learningforkids.org.



in terms of Φ and Ψ . This yields most of the substance of a typical differential equations course.

18. In Problem 17 $\Psi(t)$ is defined by the given series. Denote by $\exp(t\sigma(A))$ the numbers $\exp(t\lambda)$ where $\lambda \in \sigma(A)$. Show $\exp(t\sigma(A)) = \sigma(\Psi(t))$. This is like Lemma 13.4.8. Letting J be the Jordan canonical form for A , explain why

$$\Psi(t) \equiv \sum_{k=0}^{\infty} \frac{t^k A^k}{k!} = S \sum_{k=0}^{\infty} \frac{t^k J^k}{k!} S^{-1}$$

and you note that in J^k , the diagonal entries are of the form λ^k for λ an eigenvalue of A . Also $J = D + N$ where N is nilpotent and commutes with D . Argue then that

$$\sum_{k=0}^{\infty} \frac{t^k J^k}{k!}$$

is an upper triangular matrix which has on the diagonal the expressions $e^{\lambda t}$ where $\lambda \in \sigma(A)$. Thus conclude

$$\sigma(\Psi(t)) \subseteq \exp(t\sigma(A))$$

Next take $e^{t\lambda} \in \exp(t\sigma(A))$ and argue it must be in $\sigma(\Psi(t))$. You can do this as follows:

$$\begin{aligned} \Psi(t) - e^{t\lambda} I &= \sum_{k=0}^{\infty} \frac{t^k A^k}{k!} - \sum_{k=0}^{\infty} \frac{t^k \lambda^k}{k!} I = \sum_{k=0}^{\infty} \frac{t^k}{k!} (A^k - \lambda^k I) \\ &= \left(\sum_{k=0}^{\infty} \frac{t^k}{k!} \sum_{j=1}^{k-1} A^{k-j} \lambda^j \right) (A - \lambda I) \end{aligned}$$

Now you need to argue

$$\sum_{k=0}^{\infty} \frac{t^k}{k!} \sum_{j=1}^{k-1} A^{k-j} \lambda^j$$

converges to something in $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$. To do this, use the ratio test and Lemma 13.4.2 after first using the triangle inequality. Since $\lambda \in \sigma(A)$, $\Psi(t) - e^{t\lambda} I$ is not one to one and so this establishes the other inclusion. You fill in the details. This theorem is a special case of theorems which go by the name ‘‘spectral mapping theorem’’.

19. Suppose $\Psi(t) \in \mathcal{L}(V, W)$ where V, W are finite dimensional inner product spaces and $t \rightarrow \Psi(t)$ is continuous for $t \in [a, b]$: For every $\varepsilon > 0$ there exists $\delta > 0$ such that if $|s - t| < \delta$ then $\|\Psi(t) - \Psi(s)\| < \varepsilon$. Show $t \rightarrow (\Psi(t)v, w)$ is continuous. Here it is the inner product in W . Also define what it means for $t \rightarrow \Psi(t)v$ to be continuous and show this is continuous. Do it all for differentiable in place of continuous. Next show $t \rightarrow \|\Psi(t)\|$ is continuous.
20. If $z(t) \in W$, a finite dimensional inner product space, what does it mean for $t \rightarrow z(t)$ to be continuous or differentiable? If z is continuous, define

$$\int_a^b z(t) dt \in W$$

as follows.

$$\left(w, \int_a^b z(t) dt \right) \equiv \int_a^b (w, z(t)) dt.$$

Show that this definition is well defined and furthermore the triangle inequality,

$$\left| \int_a^b z(t) dt \right| \leq \int_a^b |z(t)| dt,$$

and fundamental theorem of calculus,

$$\frac{d}{dt} \left(\int_a^t z(s) ds \right) = z(t)$$

hold along with any other interesting properties of integrals which are true.

21. For V, W two inner product spaces, define

$$\int_a^b \Psi(t) dt \in \mathcal{L}(V, W)$$

as follows.

$$\left(w, \int_a^b \Psi(t) dt (v) \right) \equiv \int_a^b (w, \Psi(t)v) dt.$$

Show this is well defined and does indeed give $\int_a^b \Psi(t) dt \in \mathcal{L}(V, W)$. Also show the triangle inequality

$$\left\| \int_a^b \Psi(t) dt \right\| \leq \int_a^b \|\Psi(t)\| dt$$

where $\|\cdot\|$ is the operator norm and verify the fundamental theorem of calculus holds.

$$\left(\int_a^t \Psi(s) ds \right)' = \Psi(t).$$

Also verify the usual properties of integrals continue to hold such as the fact the integral is linear and

$$\int_a^b \Psi(t) dt + \int_b^c \Psi(t) dt = \int_a^c \Psi(t) dt$$

and similar things. **Hint:** On showing the triangle inequality, it will help if you use the fact that

$$|w|_W = \sup_{|v| \leq 1} |(w, v)|.$$

You should show this also.

22. Prove Gronwall's inequality. Suppose $u(t) \geq 0$ and for all $t \in [0, T]$,

$$u(t) \leq u_0 + \int_0^t Ku(s) ds.$$

where K is some nonnegative constant. Then

$$u(t) \leq u_0 e^{Kt}.$$

Hint: $w(t) = \int_0^t u(s) ds$. Then using the fundamental theorem of calculus, $w(t)$ satisfies the following.

$$u(t) - Kw(t) = w'(t) - Kw(t) \leq u_0, \quad w(0) = 0.$$

Now use the usual techniques you saw in an introductory differential equations class. Multiply both sides of the above inequality by e^{-Kt} and note the resulting left side is now a total derivative. Integrate both sides from 0 to t and see what you have got. If you have problems, look ahead in the book. This inequality is proved later in Theorem D.4.3.

23. With Gronwall's inequality and the integral defined in Problem 21 with its properties listed there, prove there is at most one solution to the initial value problem

$$\mathbf{y}' = A\mathbf{y}, \quad \mathbf{y}(0) = \mathbf{y}_0.$$

Hint: If there are two solutions, subtract them and call the result \mathbf{z} . Then

$$\mathbf{z}' = A\mathbf{z}, \mathbf{z}(0) = \mathbf{0}.$$

It follows

$$\mathbf{z}(t) = \mathbf{0} + \int_0^t A\mathbf{z}(s) ds$$

and so

$$\|\mathbf{z}(t)\| \leq \int_0^t \|A\| \|\mathbf{z}(s)\| ds$$

Now consider Gronwall's inequality of Problem 22.

24. Suppose A is a matrix which has the property that whenever $\mu \in \sigma(A)$, $\text{Re } \mu < 0$. Consider the initial value problem

$$\mathbf{y}' = A\mathbf{y}, \mathbf{y}(0) = \mathbf{y}_0.$$

The existence and uniqueness of a solution to this equation has been established above in preceding problems, Problem 17 to 23. Show that in this case where the real parts of the eigenvalues are all negative, the solution to the initial value problem satisfies

$$\lim_{t \rightarrow \infty} \mathbf{y}(t) = \mathbf{0}.$$

Hint: A nice way to approach this problem is to show you can reduce it to the consideration of the initial value problem

$$\mathbf{z}' = J_\varepsilon \mathbf{z}, \mathbf{z}(0) = \mathbf{z}_0$$

.....Alcatel-Lucent 

www.alcatel-lucent.com/careers

What if you could build your future and create the future?

One generation's transformation is the next's status quo. In the near future, people may soon think it's strange that devices ever had to be "plugged in." To obtain that status, there needs to be "The Shift".



where J_ε is the modified Jordan canonical form where instead of ones down the main diagonal, there are ε down the main diagonal (Problem 19). Then

$$\mathbf{z}' = D\mathbf{z} + N_\varepsilon\mathbf{z}$$

where D is the diagonal matrix obtained from the eigenvalues of A and N_ε is a nilpotent matrix commuting with D which is very small provided ε is chosen very small. Now let $\Psi(t)$ be the solution of

$$\Psi' = -D\Psi, \Psi(0) = I$$

described earlier as

$$\sum_{k=0}^{\infty} \frac{(-1)^k t^k D^k}{k!}.$$

Thus $\Psi(t)$ commutes with D and N_ε . Tell why. Next argue

$$(\Psi(t)\mathbf{z})' = \Psi(t)N_\varepsilon\mathbf{z}(t)$$

and integrate from 0 to t . Then

$$\Psi(t)\mathbf{z}(t) - \mathbf{z}_0 = \int_0^t \Psi(s)N_\varepsilon\mathbf{z}(s) ds.$$

It follows

$$\|\Psi(t)\mathbf{z}(t)\| \leq \|z_0\| + \int_0^t \|N_\varepsilon\| \|\Psi(s)\mathbf{z}(s)\| ds.$$

It follows from Gronwall's inequality

$$\|\Psi(t)\mathbf{z}(t)\| \leq \|z_0\| e^{\|N_\varepsilon\|t}$$

Now look closely at the form of $\Psi(t)$ to get an estimate which is interesting. Explain why

$$\Psi(t) = \begin{pmatrix} e^{\mu_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{\mu_n t} \end{pmatrix}$$

and now observe that if ε is chosen small enough, $\|N_\varepsilon\|$ is so small that each component of $\mathbf{z}(t)$ converges to 0.

25. Using Problem 24 show that if A is a matrix having the real parts of all eigenvalues less than 0 then if

$$\Psi'(t) = A\Psi(t), \Psi(0) = I$$

it follows

$$\lim_{t \rightarrow \infty} \Psi(t) = 0.$$

Hint: Consider the columns of $\Psi(t)$?

26. Let $\Psi(t)$ be a fundamental matrix satisfying

$$\Psi'(t) = A\Psi(t), \Psi(0) = I.$$

Show $\Psi(t)^n = \Psi(nt)$. **Hint:** Subtract and show the difference satisfies $\Phi' = A\Phi$, $\Phi(0) = 0$. Use uniqueness.

27. If the real parts of the eigenvalues of A are all negative, show that for every positive t ,

$$\lim_{n \rightarrow \infty} \Psi(nt) = 0.$$

Hint: Pick $\text{Re}(\sigma(A)) < -\lambda < 0$ and use Problem 18 about the spectrum of $\Psi(t)$ and Gelfand's theorem for the spectral radius along with Problem 26 to argue that $\|\Psi(nt)/e^{-\lambda nt}\| < 1$ for all n large enough.

28. Let H be a Hermitian matrix. ($H = H^*$). Show that $e^{iH} \equiv \sum_{n=0}^{\infty} \frac{(iH)^n}{n!}$ is unitary.
29. Show the converse of the above exercise. If V is unitary, then $V = e^{iH}$ for some H Hermitian.
30. If U is unitary and does not have -1 as an eigenvalue so that $(I + U)^{-1}$ exists, show that

$$H = i(I - U)(I + U)^{-1}$$

is Hermitian. Then, verify that

$$U = (I + iH)(I - iH)^{-1}.$$

31. Suppose that $A \in \mathcal{L}(V, V)$ where V is a normed linear space. Also suppose that $\|A\| < 1$ where this refers to the operator norm on A . Verify that

$$(I - A)^{-1} = \sum_{i=0}^{\infty} A^i$$

This is called the Neumann series. Suppose now that you only know the algebraic condition $\rho(A) < 1$. Is it still the case that the Neumann series converges to $(I - A)^{-1}$?



Nido

Luxurious accommodation

Central zone 1 & 2 locations

Meet hundreds of international students

BOOK NOW and get a £100 voucher from voucherexpress

Nido Student Living - London

Visit www.NidoStudentLiving.com/Bookboon for more info.

+44 (0)20 3102 1060

Chapter 14

Numerical Methods, Eigenvalues

14.1 The Power Method For Eigenvalues

This chapter discusses numerical methods for finding eigenvalues. However, to do this correctly, you must include numerical analysis considerations which are distinct from linear algebra. The purpose of this chapter is to give an introduction to some numerical methods without leaving the context of linear algebra. In addition, some examples are given which make use of computer algebra systems. For a more thorough discussion, you should see books on numerical methods in linear algebra like some listed in the references.

Let A be a complex $p \times p$ matrix and suppose that it has distinct eigenvalues

$$\{\lambda_1, \dots, \lambda_m\}$$

and that $|\lambda_1| > |\lambda_k|$ for all k . Also let the Jordan form of A be

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_m \end{pmatrix}$$

with J_1 an $m_1 \times m_1$ matrix.

$$J_k = \lambda_k I_k + N_k$$

where $N_k^{r_k} \neq 0$ but $N_k^{r_k+1} = 0$. Also let

$$P^{-1}AP = J, \quad A = PJP^{-1}.$$

Now fix $\mathbf{x} \in \mathbb{F}^p$. Take $A\mathbf{x}$ and let s_1 be the entry of the vector $A\mathbf{x}$ which has largest absolute value. Thus $A\mathbf{x}/s_1$ is a vector \mathbf{y}_1 which has a component of 1 and every other entry of this vector has magnitude no larger than 1. If the scalars $\{s_1, \dots, s_{n-1}\}$ and vectors $\{\mathbf{y}_1, \dots, \mathbf{y}_{n-1}\}$ have been obtained, let $\mathbf{y}_n \equiv A\mathbf{y}_{n-1}/s_n$ where s_n is the entry of $A\mathbf{y}_{n-1}$ which has largest absolute value. Thus

$$\mathbf{y}_n = \frac{AA\mathbf{y}_{n-2}}{s_n s_{n-1}} \dots = \frac{A^n \mathbf{x}}{s_n s_{n-1} \dots s_1} \tag{14.1}$$

$$\begin{aligned} &= \frac{1}{s_n s_{n-1} \dots s_1} P \begin{pmatrix} J_1^n & & \\ & \ddots & \\ & & J_m^n \end{pmatrix} P^{-1} \mathbf{x} \\ &= \frac{\lambda_1^n}{s_n s_{n-1} \dots s_1} P \begin{pmatrix} \lambda_1^{-n} J_1^n & & \\ & \ddots & \\ & & \lambda_1^{-n} J_m^n \end{pmatrix} P^{-1} \mathbf{x} \end{aligned} \tag{14.2}$$

Consider one of the blocks in the Jordan form. First consider the k^{th} of these blocks, $k > 1$. It equals

$$\lambda_1^{-n} J_k^n = \sum_{i=0}^{r_k} \binom{n}{i} \lambda_1^{-n} \lambda_k^{n-i} N_k^i$$

which clearly converges to 0 as $n \rightarrow \infty$ since $|\lambda_1| > |\lambda_k|$. An application of the ratio test or root test for each term in the sum will show this. When $k = 1$, this block is

$$\lambda_1^{-n} J_1^n = \lambda_1^{-n} J_k^n = \sum_{i=0}^{r_1} \binom{n}{i} \lambda_1^{-n} \lambda_1^{n-i} N_1^i = \binom{n}{r_1} [\lambda_1^{-r_1} N_1^{r_1} + e_n]$$

where $\lim_{n \rightarrow \infty} e_n = 0$ because it is a sum of bounded matrices which are multiplied by $\binom{n}{i} / \binom{n}{r_1}$. This quotient converges to 0 as $n \rightarrow \infty$ because $i < r_1$. It follows that 14.2 is of the form

$$\mathbf{y}_n = \frac{\lambda_1^n}{s_n s_{n-1} \cdots s_1} \binom{n}{r_1} P \begin{pmatrix} \lambda_1^{-r_1} N_1^{r_1} + e_n & 0 \\ 0 & E_n \end{pmatrix} P^{-1} \mathbf{x} \equiv \frac{\lambda_1^n}{s_n s_{n-1} \cdots s_1} \binom{n}{r_1} \mathbf{w}_n$$

where $E_n \rightarrow 0, e_n \rightarrow 0$. Let $(P^{-1} \mathbf{x})_{m_1}$ denote the first m_1 entries of the vector $P^{-1} \mathbf{x}$. Unless a very unlucky choice for \mathbf{x} was picked, it will follow that $(P^{-1} \mathbf{x})_{m_1} \notin \ker(N_1^{r_1})$. Then for large n , \mathbf{y}_n is close to the vector

$$\frac{\lambda_1^n}{s_n s_{n-1} \cdots s_1} \binom{n}{r_1} P \begin{pmatrix} \lambda_1^{-r_1} N_1^{r_1} & 0 \\ 0 & 0 \end{pmatrix} P^{-1} \mathbf{x} \equiv \frac{\lambda_1^n}{s_n s_{n-1} \cdots s_1} \binom{n}{r_1} \mathbf{w} \equiv \mathbf{z} \neq \mathbf{0}$$

However, this is an eigenvector because

$$\begin{aligned} (A - \lambda_1 I) \mathbf{w} &= \overbrace{P(J - \lambda_1 I)P^{-1}P}^{A - \lambda_1 I} \begin{pmatrix} \lambda_1^{-r_1} N_1^{r_1} & 0 \\ 0 & 0 \end{pmatrix} P^{-1} \mathbf{x} = \\ &= P \begin{pmatrix} N_1 & & & \\ & \ddots & & \\ & & J_m - \lambda_1 I & \\ & & & 0 \end{pmatrix} P^{-1} P \begin{pmatrix} \lambda_1^{-r_1} N_1^{r_1} & & & \\ & \ddots & & \\ & & & 0 \end{pmatrix} P^{-1} \mathbf{x} \\ &= P \begin{pmatrix} N_1 \lambda_1^{-r_1} N_1^{r_1} & 0 \\ 0 & 0 \end{pmatrix} P^{-1} \mathbf{x} = \mathbf{0} \end{aligned}$$

Recall $N_1^{r_1+1} = 0$. Now you could recover an approximation to the eigenvalue as follows.

$$\frac{(A \mathbf{y}_n, \mathbf{y}_n)}{(\mathbf{y}_n, \mathbf{y}_n)} \approx \frac{(A \mathbf{z}, \mathbf{z})}{(\mathbf{z}, \mathbf{z})} = \lambda_1$$

Here \approx means ‘‘approximately equal’’. However, there is a more convenient way to identify the eigenvalue in terms of the scaling factors s_k .

$$\left\| \frac{\lambda_1^n}{s_n \cdots s_1} \binom{n}{r_1} (\mathbf{w}_n - \mathbf{w}) \right\|_\infty \approx 0$$

Pick the largest nonzero entry of \mathbf{w} , w_l . Then for large n , it is also likely the case that the largest entry of \mathbf{w}_n will be in the l^{th} position because \mathbf{w}_n is close to \mathbf{w} . From the construction,

$$\frac{\lambda_1^n}{s_n \cdots s_1} \binom{n}{r_1} w_{nl} = 1 \approx \frac{\lambda_1^n}{s_n \cdots s_1} \binom{n}{r_1} w_l$$

In other words, for large n

$$\frac{\lambda_1^n}{s_n \cdots s_1} \binom{n}{r_1} \approx 1/w_l$$

Therefore, for large n ,

$$\frac{\lambda_1^n}{s_n \cdots s_1} \binom{n}{r_1} \approx \frac{\lambda_1^{n+1}}{s_{n+1} s_n \cdots s_1} \binom{n+1}{r_1}$$

and so

$$\binom{n}{r_1} / \binom{n+1}{r_1} \approx \frac{\lambda_1}{s_{n+1}}$$

But $\lim_{n \rightarrow \infty} \binom{n}{r_1} / \binom{n+1}{r_1} = 1$ and so, for large n it must be the case that $\lambda_1 \approx s_{n+1}$.

This has proved the following theorem which justifies the power method.

Theorem 14.1.1 *Let A be a complex $p \times p$ matrix such that the eigenvalues are*

$$\{\lambda_1, \lambda_2, \dots, \lambda_r\}$$


with $|\lambda_1| > |\lambda_j|$ for all $j \neq 1$. Then for \mathbf{x} a given vector, let

$$\mathbf{y}_1 = \frac{A\mathbf{x}}{s_1}$$


where s_1 is an entry of $A\mathbf{x}$ which has the largest absolute value. If the scalars $\{s_1, \dots, s_{n-1}\}$ and vectors $\{\mathbf{y}_1, \dots, \mathbf{y}_{n-1}\}$ have been obtained, let

$$\mathbf{y}_n \equiv \frac{A\mathbf{y}_{n-1}}{s_n}$$

where s_n is the entry of $A\mathbf{y}_{n-1}$ which has largest absolute value. Then it is probably the case that $\{s_n\}$ will converge to λ_1 and $\{\mathbf{y}_n\}$ will converge to an eigenvector associated with λ_1 . If it doesn't, you picked an incredibly inauspicious initial vector \mathbf{x} .

SIMPLY CLEVER


WE WILL TURN YOUR CV INTO AN OPPORTUNITY OF A LIFETIME



Do you like cars? Would you like to be a part of a successful brand? As a constructor at ŠKODA AUTO you will put great things in motion. Things that will ease everyday lives of people all around Send us your CV. We will give it an entirely new new dimension.

Read more about this and our other international masters degree programmes at www.uu.se/master

Send us your CV on www.employerforlife.com



In summary, here is the procedure.

Finding the largest eigenvalue with its eigenvector.

1. Start with a vector, \mathbf{u}_1 which you hope is not unlucky.
2. If \mathbf{u}_k is known,

$$\mathbf{u}_{k+1} = \frac{A\mathbf{u}_k}{s_{k+1}}$$

where s_{k+1} is the entry of $A\mathbf{u}_k$ which has largest absolute value.

3. When the scaling factors s_k are not changing much, s_{k+1} will be close to the eigenvalue and \mathbf{u}_{k+1} will be close to an eigenvector.
4. Check your answer to see if it worked well. If things don't work well, try another \mathbf{u}_1 . You were miraculously unlucky in your choice.

Example 14.1.2 Find the largest eigenvalue of $A = \begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix}$.

You can begin with $\mathbf{u}_1 = (1, \dots, 1)^T$ and apply the above procedure. However, you can accelerate the process if you begin with $A^n \mathbf{u}_1$ and then divide by the largest entry to get the first approximate eigenvector. Thus

$$\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix}^{20} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2.5558 \times 10^{21} \\ -1.2779 \times 10^{21} \\ -3.6562 \times 10^{15} \end{pmatrix}$$

Divide by the largest entry to obtain a good approximation.

$$\begin{pmatrix} 2.5558 \times 10^{21} \\ -1.2779 \times 10^{21} \\ -3.6562 \times 10^{15} \end{pmatrix} \frac{1}{2.5558 \times 10^{21}} = \begin{pmatrix} 1.0 \\ -0.5 \\ -1.4306 \times 10^{-6} \end{pmatrix}$$

Now begin with this one.

$$\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix} \begin{pmatrix} 1.0 \\ -0.5 \\ -1.4306 \times 10^{-6} \end{pmatrix} = \begin{pmatrix} 12.000 \\ -6.0000 \\ 4.2918 \times 10^{-6} \end{pmatrix}$$

Divide by 12 to get the next iterate.

$$\begin{pmatrix} 12.000 \\ -6.0000 \\ 4.2918 \times 10^{-6} \end{pmatrix} \frac{1}{12} = \begin{pmatrix} 1.0 \\ -0.5 \\ 3.5765 \times 10^{-7} \end{pmatrix}$$

Another iteration will reveal that the scaling factor is still 12. Thus this is an approximate eigenvalue. In fact, it is **the** largest eigenvalue and the corresponding eigenvector is $\begin{pmatrix} 1.0 & -0.5 & 0 \end{pmatrix}$. The process has worked very well.

14.1.1 The Shifted Inverse Power Method

This method can find various eigenvalues and eigenvectors. It is a significant generalization of the above simple procedure and yields very good results. One can find complex eigenvalues using this method. The situation is this: You have a number α which is close to λ , some eigenvalue of an $n \times n$ matrix A . You don't know λ but you know that α is closer to λ than to any other eigenvalue. Your problem is to find both λ and an eigenvector which goes

with λ . Another way to look at this is to start with α and seek the eigenvalue λ , which is closest to α along with an eigenvector associated with λ . If α is an eigenvalue of A , then you have what you want. Therefore, I will always assume α is not an eigenvalue of A and so $(A - \alpha I)^{-1}$ exists. The method is based on the following lemma.

Lemma 14.1.3 *Let $\{\lambda_k\}_{k=1}^n$ be the eigenvalues of A . If \mathbf{x}_k is an eigenvector of A for the eigenvalue λ_k , then \mathbf{x}_k is an eigenvector for $(A - \alpha I)^{-1}$ corresponding to the eigenvalue $\frac{1}{\lambda_k - \alpha}$. Conversely, if*

$$(A - \alpha I)^{-1} \mathbf{y} = \frac{1}{\lambda - \alpha} \mathbf{y} \tag{14.3}$$

and $\mathbf{y} \neq \mathbf{0}$, then $A\mathbf{y} = \lambda\mathbf{y}$.

Proof: Let λ_k and \mathbf{x}_k be as described in the statement of the lemma. Then

$$(A - \alpha I) \mathbf{x}_k = (\lambda_k - \alpha) \mathbf{x}_k$$

and so

$$\frac{1}{\lambda_k - \alpha} \mathbf{x}_k = (A - \alpha I)^{-1} \mathbf{x}_k.$$

Suppose 14.3. Then $\mathbf{y} = \frac{1}{\lambda - \alpha} [A\mathbf{y} - \alpha\mathbf{y}]$. Solving for $A\mathbf{y}$ leads to $A\mathbf{y} = \lambda\mathbf{y}$. ■

Now assume α is closer to λ than to any other eigenvalue. Then the magnitude of $\frac{1}{\lambda - \alpha}$ is greater than the magnitude of all the other eigenvalues of $(A - \alpha I)^{-1}$. Therefore, the power method applied to $(A - \alpha I)^{-1}$ will yield $\frac{1}{\lambda - \alpha}$. You end up with $s_{n+1} \approx \frac{1}{\lambda - \alpha}$ and solve for λ .

14.1.2 The Explicit Description Of The Method

Here is how you use this method to find the eigenvalue closest to α and the corresponding eigenvector.

1. Find $(A - \alpha I)^{-1}$.
2. Pick \mathbf{u}_1 . If you are not phenomenally unlucky, the iterations will converge.
3. If \mathbf{u}_k has been obtained,

$$\mathbf{u}_{k+1} = \frac{(A - \alpha I)^{-1} \mathbf{u}_k}{s_{k+1}}$$

where s_{k+1} is the entry of $(A - \alpha I)^{-1} \mathbf{u}_k$ which has largest absolute value.

4. When the scaling factors, s_k are not changing much and the \mathbf{u}_k are not changing much, find the approximation to the eigenvalue by solving

$$s_{k+1} = \frac{1}{\lambda - \alpha}$$

for λ . The eigenvector is approximated by \mathbf{u}_{k+1} .

5. Check your work by multiplying by the original matrix to see how well what you have found works.

Thus this amounts to the power method for the matrix $(A - \alpha I)^{-1}$ but you are free to pick α .

Example 14.1.4 Find the eigenvalue of $A = \begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix}$ which is closest to -7 .

Also find an eigenvector which goes with this eigenvalue.

In this case the eigenvalues are $-6, 0$, and 12 so the correct answer is -6 for the eigenvalue. Then from the above procedure, I will start with an initial vector, $\mathbf{u}_1 = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}^T$. Then I must solve the following equation.

$$\left(\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix} + 7 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Simplifying the matrix on the left, I must solve

$$\begin{pmatrix} 12 & -14 & 11 \\ -4 & 11 & -4 \\ 3 & 6 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

and then divide by the entry which has largest absolute value to obtain

$$\mathbf{u}_2 = \begin{pmatrix} 1.0 \\ .184 \\ -.76 \end{pmatrix}$$

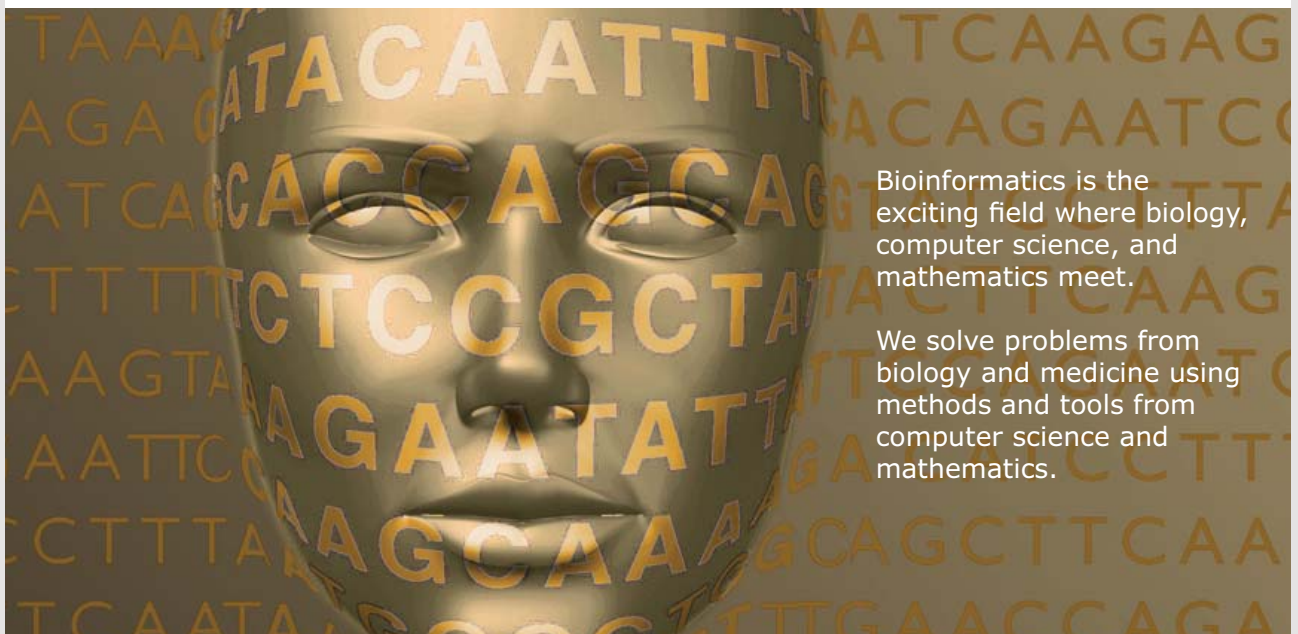
Now solve

$$\begin{pmatrix} 12 & -14 & 11 \\ -4 & 11 & -4 \\ 3 & 6 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1.0 \\ .184 \\ -.76 \end{pmatrix}$$



UPPSALA
UNIVERSITET

Develop the tools we need for Life Science Masters Degree in Bioinformatics



Bioinformatics is the exciting field where biology, computer science, and mathematics meet.

We solve problems from biology and medicine using methods and tools from computer science and mathematics.

Read more about this and our other international masters degree programmes at www.uu.se/master

and divide by the largest entry, 1.0515 to get

$$\mathbf{u}_3 = \begin{pmatrix} 1.0 \\ .0266 \\ -.97061 \end{pmatrix}$$

Solve

$$\begin{pmatrix} 12 & -14 & 11 \\ -4 & 11 & -4 \\ 3 & 6 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1.0 \\ .0266 \\ -.97061 \end{pmatrix}$$

and divide by the largest entry, 1.01 to get

$$\mathbf{u}_4 = \begin{pmatrix} 1.0 \\ 3.8454 \times 10^{-3} \\ -.99604 \end{pmatrix}.$$

These scaling factors are pretty close after these few iterations. Therefore, the predicted eigenvalue is obtained by solving the following for λ .

$$\frac{1}{\lambda + 7} = 1.01$$

which gives $\lambda = -6.01$. You see this is pretty close. In this case the eigenvalue closest to -7 was -6 .

How would you know what to start with for an initial guess? You might apply Gerschgorin's theorem. However, sometimes you can begin with a better estimate.

Example 14.1.5 Consider the symmetric matrix $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 2 \end{pmatrix}$. Find the middle eigenvalue and an eigenvector which goes with it.

Since A is symmetric, it follows it has three real eigenvalues which are solutions to

$$\begin{aligned} p(\lambda) &= \det \left(\lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 2 \end{pmatrix} \right) \\ &= \lambda^3 - 4\lambda^2 - 24\lambda - 17 = 0 \end{aligned}$$

If you use your graphing calculator to graph this polynomial, you find there is an eigenvalue somewhere between $-.9$ and $-.8$ and that this is the middle eigenvalue. Of course you could zoom in and find it very accurately without much trouble but what about the eigenvector which goes with it? If you try to solve

$$\begin{pmatrix} (-.8) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 2 \end{pmatrix} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

there will be only the zero solution because the matrix on the left will be invertible and the same will be true if you replace $-.8$ with a better approximation like $-.86$ or $-.855$. This is because all these are only approximations to the eigenvalue and so the matrix in the above is nonsingular for all of these. Therefore, you will only get the zero solution and

Eigenvectors are never equal to zero!

However, there exists such an eigenvector and you can find it using the shifted inverse power method. Pick $\alpha = -.855$. Then you solve

$$\left(\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 2 \end{pmatrix} + .855 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

or in other words,

$$\begin{pmatrix} 1.855 & 2.0 & 3.0 \\ 2.0 & 1.855 & 4.0 \\ 3.0 & 4.0 & 2.855 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

and after finding the solution, divide by the largest entry -67.944 , to obtain

$$\mathbf{u}_2 = \begin{pmatrix} 1.0 \\ -.58921 \\ -.23044 \end{pmatrix}$$

After a couple more iterations, you obtain

$$\mathbf{u}_3 = \begin{pmatrix} 1.0 \\ -.58777 \\ -.22714 \end{pmatrix} \tag{14.4}$$

Then doing it again, the scaling factor is -513.42 and the next iterate is

$$\mathbf{u}_4 = \begin{pmatrix} 1.0 \\ -.58778 \\ -.22714 \end{pmatrix}$$

Clearly the \mathbf{u}_k are not changing much. This suggests an approximate eigenvector for this eigenvalue which is close to $-.855$ is the above \mathbf{u}_3 and an eigenvalue is obtained by solving

$$\frac{1}{\lambda + .855} = -513.42,$$

which yields $\lambda = -0.85695$ Lets check this.

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 2 \end{pmatrix} \begin{pmatrix} 1.0 \\ -.58778 \\ -.22714 \end{pmatrix} = \begin{pmatrix} -0.85698 \\ 0.50366 \\ 0.1946 \end{pmatrix}$$

$$-0.85695 \begin{pmatrix} 1.0 \\ -.58777 \\ -.22714 \end{pmatrix} = \begin{pmatrix} -0.85695 \\ 0.50369 \\ 0.19465 \end{pmatrix}$$

Thus the vector of 14.4 is very close to the desired eigenvector, just as $-.8569$ is very close to the desired eigenvalue. For practical purposes, I have found both the eigenvector and the eigenvalue.

Example 14.1.6 Find the eigenvalues and eigenvectors of the matrix $A = \begin{pmatrix} 2 & 1 & 3 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{pmatrix}$.

This is only a 3×3 matrix and so it is not hard to estimate the eigenvalues. Just get the characteristic equation, graph it using a calculator and zoom in to find the eigenvalues. If you do this, you find there is an eigenvalue near -1.2 , one near $-.4$, and one near 5.5 . (The characteristic equation is $2 + 8\lambda + 4\lambda^2 - \lambda^3 = 0$.) Of course I have no idea what the eigenvectors are.

Lets first try to find the eigenvector and a better approximation for the eigenvalue near -1.2 . In this case, let $\alpha = -1.2$. Then

$$(A - \alpha I)^{-1} = \begin{pmatrix} -25.357143 & -33.928571 & 50.0 \\ 12.5 & 17.5 & -25.0 \\ 23.214286 & 30.357143 & -45.0 \end{pmatrix}.$$

As before, it helps to get things started if you raise to a power and then go from the approximate eigenvector obtained.

$$\begin{pmatrix} -25.357143 & -33.928571 & 50.0 \\ 12.5 & 17.5 & -25.0 \\ 23.214286 & 30.357143 & -45.0 \end{pmatrix}^7 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2.2956 \times 10^{11} \\ 1.1291 \times 10^{11} \\ 2.0865 \times 10^{11} \end{pmatrix}$$

Then the next iterate will be

$$\begin{pmatrix} -2.2956 \times 10^{11} \\ 1.1291 \times 10^{11} \\ 2.0865 \times 10^{11} \end{pmatrix} \frac{1}{-2.2956 \times 10^{11}} = \begin{pmatrix} 1.0 \\ -0.49185 \\ -0.90891 \end{pmatrix}$$

Next iterate:

$$\begin{pmatrix} -25.357143 & -33.928571 & 50.0 \\ 12.5 & 17.5 & -25.0 \\ 23.214286 & 30.357143 & -45.0 \end{pmatrix} \begin{pmatrix} 1.0 \\ -0.49185 \\ -0.90891 \end{pmatrix} = \begin{pmatrix} -54.115 \\ 26.615 \\ 49.184 \end{pmatrix}$$

UNIVERSITY OF COPENHAGEN

Brain power

By 2020, wind could provide one-tenth of our electricity needs. Already today, SKF's innovative know-how is crucial to running a large proportion of the world's wind turbines.

Up to 25 % of the generation capacity of modern power systems can be reduced by on-line condition monitoring and lubrication. We help you become a greener, cleaner, cheaper energy producer.

By sharing our knowledge, industries can become more efficient and reduce their carbon footprint. Therefore we want to meet this challenge.

The Power of Knowledge Engineering.

Copenhagen Master of Excellence are two-year master degrees taught in English at one of Europe's leading universities

Come to Copenhagen – and see for yourself.

Plug into The Power of Knowledge Engineering. Visit us at www.skf.com/knowledge

www.come.ku.dk

cultural studies

religious studies

science



Divide by largest entry

$$\begin{pmatrix} -54.115 \\ 26.615 \\ 49.184 \end{pmatrix} \frac{1}{-54.115} = \begin{pmatrix} 1.0 \\ -0.49182 \\ -0.90888 \end{pmatrix}$$

You can see the vector didn't change much and so the next scaling factor will not be much different than this one. Hence you need to solve for λ

$$\frac{1}{\lambda + 1.2} = -54.115$$

Then $\lambda = -1.2185$ is an approximate eigenvalue and

$$\begin{pmatrix} 1.0 \\ -0.49182 \\ -0.90888 \end{pmatrix}$$

is an approximate eigenvector. How well does it work?

$$\begin{pmatrix} 2 & 1 & 3 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1.0 \\ -0.49182 \\ -0.90888 \end{pmatrix} = \begin{pmatrix} -1.2185 \\ 0.5993 \\ 1.1075 \end{pmatrix}$$

$$(-1.2185) \begin{pmatrix} 1.0 \\ -0.49182 \\ -0.90888 \end{pmatrix} = \begin{pmatrix} -1.2185 \\ 0.59928 \\ 1.1075 \end{pmatrix}$$

You can see that for practical purposes, this has found the eigenvalue closest to -1.2185 and the corresponding eigenvector.

The other eigenvectors and eigenvalues can be found similarly. In the case of -4 , you could let $\alpha = -4$ and then

$$(A - \alpha I)^{-1} = \begin{pmatrix} 8.0645161 \times 10^{-2} & -9.2741935 & 6.4516129 \\ -.40322581 & 11.370968 & -7.2580645 \\ .40322581 & 3.6290323 & -2.7419355 \end{pmatrix}.$$

Following the procedure of the power method, you find that after about 5 iterations, the scaling factor is 9.7573139, they are not changing much, and

$$\mathbf{u}_5 = \begin{pmatrix} -.7812248 \\ 1.0 \\ .26493688 \end{pmatrix}.$$

Thus the approximate eigenvalue is

$$\frac{1}{\lambda + .4} = 9.7573139$$

which shows $\lambda = -.29751278$ is an approximation to the eigenvalue near $.4$. How well does it work?

$$\begin{pmatrix} 2 & 1 & 3 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} -.7812248 \\ 1.0 \\ .26493688 \end{pmatrix} = \begin{pmatrix} .23236104 \\ -.29751272 \\ -.07873752 \end{pmatrix}.$$

$$-.29751278 \begin{pmatrix} -.7812248 \\ 1.0 \\ .26493688 \end{pmatrix} = \begin{pmatrix} .23242436 \\ -.29751278 \\ -7.8822108 \times 10^{-2} \end{pmatrix}.$$

It works pretty well. For practical purposes, the eigenvalue and eigenvector have now been found. If you want better accuracy, you could just continue iterating. One can find the eigenvector corresponding to the eigenvalue nearest 5.5 the same way.

14.1.3 Complex Eigenvalues

What about complex eigenvalues? If your matrix is real, you won't see these by graphing the characteristic equation on your calculator. Will the shifted inverse power method find these eigenvalues and their associated eigenvectors? The answer is yes. However, for a real matrix, you must pick α to be complex. This is because the eigenvalues occur in conjugate pairs so if you don't pick it complex, it will be the same distance between any conjugate pair of complex numbers and so nothing in the above argument for convergence implies you will get convergence to a complex number. Also, the process of iteration will yield only real vectors and scalars.

Example 14.1.7 Find the complex eigenvalues and corresponding eigenvectors for the matrix

$$\begin{pmatrix} 5 & -8 & 6 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Here the characteristic equation is $\lambda^3 - 5\lambda^2 + 8\lambda - 6 = 0$. One solution is $\lambda = 3$. The other two are $1 + i$ and $1 - i$. I will apply the process to $\alpha = i$ to find the eigenvalue closest to i .

$$(A - \alpha I)^{-1} = \begin{pmatrix} -.02 - .14i & 1.24 + .68i & -.84 + .12i \\ -.14 + .02i & .68 - .24i & .12 + .84i \\ .02 + .14i & -.24 - .68i & .84 + .88i \end{pmatrix}$$

Then let $\mathbf{u}_1 = (1, 1, 1)^T$ for lack of any insight into anything better.

$$\begin{aligned} & \begin{pmatrix} -.02 - .14i & 1.24 + .68i & -.84 + .12i \\ -.14 + .02i & .68 - .24i & .12 + .84i \\ .02 + .14i & -.24 - .68i & .84 + .88i \end{pmatrix}^{20} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} -0.40000 + 0.8i \\ 0.20000 + 0.6i \\ 0.40000 + 0.2i \end{pmatrix} \end{aligned}$$

Now divide by the largest entry to get the next iterate. This yields for an approximate eigenvector approximately

$$\begin{pmatrix} -0.40000 + 0.8i \\ 0.20000 + 0.6i \\ 0.40000 + 0.2i \end{pmatrix} \frac{1}{-0.40000 + 0.8i} = \begin{pmatrix} 1.0 \\ 0.5 - 0.5i \\ -0.5i \end{pmatrix}$$

Now leaving off extremely small terms,

$$\begin{aligned} & \begin{pmatrix} -.02 - .14i & 1.24 + .68i & -.84 + .12i \\ -.14 + .02i & .68 - .24i & .12 + .84i \\ .02 + .14i & -.24 - .68i & .84 + .88i \end{pmatrix} \begin{pmatrix} 1.0 \\ 0.5 - 0.5i \\ -0.5i \end{pmatrix} = \\ & \begin{pmatrix} 1.0 \\ 0.5 - 0.5i \\ -0.5i \end{pmatrix} \end{aligned}$$

so it appears that an eigenvector is the above and an eigenvalue can be obtained by solving

$$\frac{1}{\lambda - i} = 1, \text{ so } \lambda = 1 + i$$

The method has successfully found the complex eigenvalue closest to i as well as the eigenvector. Note that I used essentially 20 iterations of the method.

This illustrates an interesting topic which leads to many related topics. If you have a polynomial, $x^4 + ax^3 + bx^2 + cx + d$, you can consider it as the characteristic polynomial of a certain matrix, called a **companion matrix**. In this case,

$$\begin{pmatrix} -a & -b & -c & -d \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

The above example was just a companion matrix for $\lambda^3 - 5\lambda^2 + 8\lambda - 6$. You can see the pattern which will enable you to obtain a companion matrix for any polynomial of the form $\lambda^n + a_1\lambda^{n-1} + \dots + a_{n-1}\lambda + a_n$. This illustrates that one way to find the complex zeros of a polynomial is to use the shifted inverse power method on a companion matrix for the polynomial. Doubtless there are better ways but this does illustrate how impressive this procedure is. Do you have a better way?

Note that the shifted inverse power method is a way you can begin with something close but not equal to an eigenvalue and end up with something close to an eigenvector.



Brain power

By 2020, wind could provide one-tenth of our planet's electricity needs. Already today, SKF's innovative know-how is crucial to running a large proportion of the world's wind turbines.

Up to 25 % of the generating costs relate to maintenance. These can be reduced dramatically thanks to our systems for on-line condition monitoring and automatic lubrication. We help make it more economical to create cleaner, cheaper energy out of thin air.

By sharing our experience, expertise, and creativity, industries can boost performance beyond expectations. Therefore we need the best employees who can meet this challenge!

The Power of Knowledge Engineering

Plug into The Power of Knowledge Engineering.
Visit us at www.skf.com/knowledge

SKF

14.1.4 Rayleigh Quotients And Estimates for Eigenvalues

There are many specialized results concerning the eigenvalues and eigenvectors for Hermitian matrices. Recall a matrix A is Hermitian if $A = A^*$ where A^* means to take the transpose of the conjugate of A . In the case of a real matrix, Hermitian reduces to symmetric. Recall also that for $\mathbf{x} \in \mathbb{F}^n$,

$$|\mathbf{x}|^2 = \mathbf{x}^* \mathbf{x} = \sum_{j=1}^n |x_j|^2.$$

Recall the following corollary found on Page 168 which is stated here for convenience.

Corollary 14.1.8 *If A is Hermitian, then all the eigenvalues of A are real and there exists an orthonormal basis of eigenvectors.*

Thus for $\{\mathbf{x}_k\}_{k=1}^n$ this orthonormal basis,

$$\mathbf{x}_i^* \mathbf{x}_j = \delta_{ij} \equiv \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

For $\mathbf{x} \in \mathbb{F}^n$, $\mathbf{x} \neq \mathbf{0}$, the Rayleigh quotient is defined by

$$\frac{\mathbf{x}^* A \mathbf{x}}{|\mathbf{x}|^2}.$$

Now let the eigenvalues of A be $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ and $A \mathbf{x}_k = \lambda_k \mathbf{x}_k$ where $\{\mathbf{x}_k\}_{k=1}^n$ is the above orthonormal basis of eigenvectors mentioned in the corollary. Then if \mathbf{x} is an arbitrary vector, there exist constants, a_i such that

$$\mathbf{x} = \sum_{i=1}^n a_i \mathbf{x}_i.$$

Also,

$$|\mathbf{x}|^2 = \sum_{i=1}^n \bar{a}_i \mathbf{x}_i^* \sum_{j=1}^n a_j \mathbf{x}_j = \sum_{ij} \bar{a}_i a_j \mathbf{x}_i^* \mathbf{x}_j = \sum_{ij} \bar{a}_i a_j \delta_{ij} = \sum_{i=1}^n |a_i|^2.$$

Therefore,

$$\begin{aligned} \frac{\mathbf{x}^* A \mathbf{x}}{|\mathbf{x}|^2} &= \frac{(\sum_{i=1}^n \bar{a}_i \mathbf{x}_i^*) (\sum_{j=1}^n a_j \lambda_j \mathbf{x}_j)}{\sum_{i=1}^n |a_i|^2} = \frac{\sum_{ij} \bar{a}_i a_j \lambda_j \mathbf{x}_i^* \mathbf{x}_j}{\sum_{i=1}^n |a_i|^2} \\ &= \frac{\sum_{ij} \bar{a}_i a_j \lambda_j \delta_{ij}}{\sum_{i=1}^n |a_i|^2} = \frac{\sum_{i=1}^n |a_i|^2 \lambda_i}{\sum_{i=1}^n |a_i|^2} \in [\lambda_1, \lambda_n]. \end{aligned}$$

In other words, the Rayleigh quotient is always between the largest and the smallest eigenvalues of A . When $\mathbf{x} = \mathbf{x}_n$, the Rayleigh quotient equals the largest eigenvalue and when $\mathbf{x} = \mathbf{x}_1$ the Rayleigh quotient equals the smallest eigenvalue. Suppose you calculate a Rayleigh quotient. How close is it to some eigenvalue?

Theorem 14.1.9 *Let $\mathbf{x} \neq \mathbf{0}$ and form the Rayleigh quotient,*

$$\frac{\mathbf{x}^* A \mathbf{x}}{|\mathbf{x}|^2} \equiv q.$$

Then there exists an eigenvalue of A , denoted here by λ_q such that

$$|\lambda_q - q| \leq \frac{|A \mathbf{x} - q \mathbf{x}|}{|\mathbf{x}|}. \tag{14.5}$$

Proof: Let $\mathbf{x} = \sum_{k=1}^n a_k \mathbf{x}_k$ where $\{\mathbf{x}_k\}_{k=1}^n$ is the orthonormal basis of eigenvectors.

$$\begin{aligned} |\mathbf{Ax} - q\mathbf{x}|^2 &= (\mathbf{Ax} - q\mathbf{x})^* (\mathbf{Ax} - q\mathbf{x}) \\ &= \left(\sum_{k=1}^n a_k \lambda_k \mathbf{x}_k - q a_k \mathbf{x}_k \right)^* \left(\sum_{k=1}^n a_k \lambda_k \mathbf{x}_k - q a_k \mathbf{x}_k \right) \\ &= \left(\sum_{j=1}^n (\lambda_j - q) \bar{a}_j \mathbf{x}_j^* \right) \left(\sum_{k=1}^n (\lambda_k - q) a_k \mathbf{x}_k \right) \\ &= \sum_{j,k} (\lambda_j - q) \bar{a}_j (\lambda_k - q) a_k \mathbf{x}_j^* \mathbf{x}_k \\ &= \sum_{k=1}^n |a_k|^2 (\lambda_k - q)^2 \end{aligned}$$

Now pick the eigenvalue λ_q which is closest to q . Then

$$|\mathbf{Ax} - q\mathbf{x}|^2 = \sum_{k=1}^n |a_k|^2 (\lambda_k - q)^2 \geq (\lambda_q - q)^2 \sum_{k=1}^n |a_k|^2 = (\lambda_q - q)^2 |\mathbf{x}|^2$$

which implies 14.5. ■

Example 14.1.10 Consider the symmetric matrix $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 1 & 4 \end{pmatrix}$. Let $\mathbf{x} = (1, 1, 1)^T$.

How close is the Rayleigh quotient to some eigenvalue of A ? Find the eigenvector and eigenvalue to several decimal places.

Everything is real and so there is no need to worry about taking conjugates. Therefore, the Rayleigh quotient is

$$\frac{\begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 1 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}{3} = \frac{19}{3}$$

According to the above theorem, there is some eigenvalue of this matrix λ_q such that

$$\begin{aligned} \left| \lambda_q - \frac{19}{3} \right| &\leq \frac{\left| \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 1 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \frac{19}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right|}{\sqrt{3}} = \frac{1}{\sqrt{3}} \begin{pmatrix} -\frac{1}{3} \\ -\frac{4}{3} \\ \frac{5}{3} \end{pmatrix} \\ &= \frac{\sqrt{\frac{1}{9} + \left(\frac{4}{3}\right)^2 + \left(\frac{5}{3}\right)^2}}{\sqrt{3}} = 1.2472 \end{aligned}$$

Could you find this eigenvalue and associated eigenvector? Of course you could. This is what the shifted inverse power method is all about.

Solve

$$\left(\begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 1 & 4 \end{pmatrix} - \frac{19}{3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

In other words solve

$$\begin{pmatrix} -\frac{16}{3} & 2 & 3 \\ 2 & -\frac{13}{3} & 1 \\ 3 & 1 & -\frac{7}{3} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

and divide by the entry which is largest, 3.8707, to get

$$\mathbf{u}_2 = \begin{pmatrix} .69925 \\ .49389 \\ 1.0 \end{pmatrix}$$

Now solve

$$\begin{pmatrix} -\frac{16}{3} & 2 & 3 \\ 2 & -\frac{13}{3} & 1 \\ 3 & 1 & -\frac{7}{3} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} .69925 \\ .49389 \\ 1.0 \end{pmatrix}$$

and divide by the largest entry, 2.9979 to get

$$\mathbf{u}_3 = \begin{pmatrix} .71473 \\ .52263 \\ 1.0 \end{pmatrix}$$

Now solve

$$\begin{pmatrix} -\frac{16}{3} & 2 & 3 \\ 2 & -\frac{13}{3} & 1 \\ 3 & 1 & -\frac{7}{3} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} .71473 \\ .52263 \\ 1.0 \end{pmatrix}$$

and divide by the largest entry, 3.0454, to get

$$\mathbf{u}_4 = \begin{pmatrix} .7137 \\ .52056 \\ 1.0 \end{pmatrix}$$

Trust and responsibility

NNE and Pharmaplan have joined forces to create NNE Pharmaplan, the world's leading engineering and consultancy company focused entirely on the pharma and biotech industries.

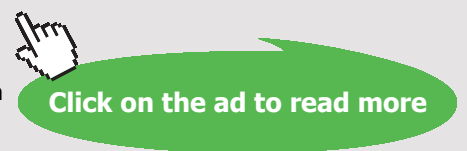
Inés Aréizaga Esteve (Spain), 25 years old
Education: Chemical Engineer

– You have to be proactive and open-minded as a newcomer and make it clear to your colleagues what you are able to cope. The pharmaceutical field is new to me. But busy as they are, most of my colleagues find the time to teach me, and they also trust me. Even though it was a bit hard at first, I can feel over time that I am beginning to be taken seriously and that my contribution is appreciated.



NNE Pharmaplan is the world's leading engineering and consultancy company focused entirely on the pharma and biotech industries. We employ more than 1500 people worldwide and offer global reach and local knowledge along with our all-encompassing list of services.
nnepharmaplan.com

nne pharmaplan®



Solve

$$\begin{pmatrix} -\frac{16}{3} & 2 & 3 \\ 2 & -\frac{13}{3} & 1 \\ 3 & 1 & -\frac{7}{3} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} .7137 \\ .52056 \\ 1.0 \end{pmatrix}$$

and divide by the largest entry, 3.0421 to get

$$\mathbf{u}_5 = \begin{pmatrix} .71378 \\ .52073 \\ 1.0 \end{pmatrix}$$

You can see these scaling factors are not changing much. The predicted eigenvalue is then about

$$\frac{1}{3.0421} + \frac{19}{3} = 6.6621.$$

How close is this?

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 1 & 4 \end{pmatrix} \begin{pmatrix} .71378 \\ .52073 \\ 1.0 \end{pmatrix} = \begin{pmatrix} 4.7552 \\ 3.469 \\ 6.6621 \end{pmatrix}$$

while

$$6.6621 \begin{pmatrix} .71378 \\ .52073 \\ 1.0 \end{pmatrix} = \begin{pmatrix} 4.7553 \\ 3.4692 \\ 6.6621 \end{pmatrix}.$$

You see that for practical purposes, this has found the eigenvalue and an eigenvector.

14.2 The QR Algorithm

14.2.1 Basic Properties And Definition

Recall the theorem about the QR factorization in Theorem 5.7.5. It says that given an $n \times n$ real matrix A , there exists a real orthogonal matrix Q and an upper triangular matrix R such that $A = QR$ and that this factorization can be accomplished by a systematic procedure. One such procedure was given in proving this theorem.

Theorem 14.2.1 *Let A be an $m \times n$ complex matrix. Then there exists a unitary Q and R , where R is all zero below the main diagonal ($R_{ij} = 0$ if $i > j$) such that $A = QR$.*

Proof: This is obvious if $m = 1$.

$$\begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} = (1) \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix}$$

Suppose true for $m - 1$ and let

$$A = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{pmatrix}, \quad A \text{ is } m \times n$$

There exists Q_1 a unitary matrix such that $Q_1(\mathbf{a}_1/|\mathbf{a}_1|) = \mathbf{e}_1$ in case $\mathbf{a}_1 \neq \mathbf{0}$. Thus $Q_1\mathbf{a}_1 = |\mathbf{a}_1|\mathbf{e}_1$. If $\mathbf{a}_1 = \mathbf{0}$, let $Q_1 = I$. Thus

$$Q_1A = \begin{pmatrix} a & \mathbf{b} \\ \mathbf{0} & A_1 \end{pmatrix}$$

where A_1 is $(m - 1) \times (n - 1)$. If $n = 1$, this obtains

$$Q_1A = \begin{pmatrix} a \\ \mathbf{0} \end{pmatrix}, \quad A = Q_1^* \begin{pmatrix} a \\ \mathbf{0} \end{pmatrix}, \quad \text{let } Q = Q_1^*.$$

That which is desired is obtained. So assume $n > 1$. By induction, there exists Q'_2 an $(m - 1) \times (n - 1)$ unitary matrix such that $Q'_2 A_1 = R'$, $R'_{ij} = 0$ if $i > j$. Then

$$\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q'_2 \end{pmatrix} Q_1 A = \begin{pmatrix} a & \mathbf{b} \\ \mathbf{0} & R' \end{pmatrix} = R$$

Since the product of unitary matrices is unitary, there exists Q unitary such that $Q^* A = R$ and so $A = QR$. ■ ► ►

The QR algorithm is described in the following definition.

Definition 14.2.2 *The QR algorithm is the following. In the description of this algorithm, Q is unitary and R is upper triangular having nonnegative entries on the main diagonal. Starting with A an $n \times n$ matrix, form*

$$A_0 \equiv A = Q_1 R_1 \tag{14.6}$$

Then

$$A_1 \equiv R_1 Q_1. \tag{14.7}$$

In general given

$$A_k = R_k Q_k, \tag{14.8}$$

obtain A_{k+1} by

$$A_k = Q_{k+1} R_{k+1}, A_{k+1} = R_{k+1} Q_{k+1} \tag{14.9}$$

This algorithm was proposed by Francis in 1961. The sequence $\{A_k\}$ is the desired sequence of iterates. Now with the above definition of the algorithm, here are its properties. The next lemma shows each of the A_k is unitarily similar to A and the amazing thing about this algorithm is that often it becomes increasingly easy to find the eigenvalues of the A_k .

Lemma 14.2.3 *Let A be an $n \times n$ matrix and let the Q_k and R_k be as described in the algorithm. Then each A_k is unitarily similar to A and denoting by $Q^{(k)}$ the product $Q_1 Q_2 \cdots Q_k$ and $R^{(k)}$ the product $R_k R_{k-1} \cdots R_1$, it follows that*

$$A^k = Q^{(k)} R^{(k)}$$

(The matrix on the left is A raised to the k^{th} power.)

$$A = Q^{(k)} A_k Q^{(k)*}, A_k = Q^{(k)*} A Q^{(k)}.$$

Proof: From the algorithm, $R_{k+1} = A_{k+1} Q_{k+1}^*$ and so

$$A_k = Q_{k+1} R_{k+1} = Q_{k+1} A_{k+1} Q_{k+1}^*$$

Now iterating this, it follows

$$A_{k-1} = Q_k A_k Q_k^* = Q_k Q_{k+1} A_{k+1} Q_{k+1}^* Q_k^*$$

$$A_{k-2} = Q_{k-1} A_{k-1} Q_{k-1}^* = Q_{k-1} Q_k Q_{k+1} A_{k+1} Q_{k+1}^* Q_k^* Q_{k-1}^*$$

etc. Thus, after $k - 2$ more iterations,

$$A = Q^{(k+1)} A_{k+1} Q^{(k+1)*}$$

The product of unitary matrices is unitary and so this proves the first claim of the lemma.

Now consider the part about A^k . From the algorithm, this is clearly true for $k = 1$. ($A^1 = QR$) Suppose then that

$$A^k = Q_1 Q_2 \cdots Q_k R_k R_{k-1} \cdots R_1$$

What was just shown indicated

$$A = Q_1 Q_2 \cdots Q_{k+1} A_{k+1} Q_{k+1}^* Q_k^* \cdots Q_1^*$$

and now from the algorithm, $A_{k+1} = R_{k+1} Q_{k+1}$ and so

$$A = Q_1 Q_2 \cdots Q_{k+1} R_{k+1} Q_{k+1} Q_{k+1}^* Q_k^* \cdots Q_1^*$$

Then

$$\begin{aligned}
 A^{k+1} &= AA^k = \\
 &\overbrace{Q_1 Q_2 \cdots Q_{k+1} R_{k+1} Q_{k+1}^* Q_k^* \cdots Q_1^* Q_1 \cdots Q_k R_k R_{k-1} \cdots R_1}^A \\
 &= Q_1 Q_2 \cdots Q_{k+1} R_{k+1} R_k R_{k-1} \cdots R_1 \equiv Q^{(k+1)} R^{(k+1)} \blacksquare
 \end{aligned}$$

Here is another very interesting lemma.

Lemma 14.2.4 Suppose $Q^{(k)}, Q$ are unitary and R_k is upper triangular such that the diagonal entries on R_k are all positive and

$$Q = \lim_{k \rightarrow \infty} Q^{(k)} R_k$$

Then

$$\lim_{k \rightarrow \infty} Q^{(k)} = Q, \quad \lim_{k \rightarrow \infty} R_k = I.$$

Also the QR factorization of A is unique whenever A^{-1} exists.

Proof: Let

$$Q = (\mathbf{q}_1, \dots, \mathbf{q}_n), \quad Q^{(k)} = (\mathbf{q}_1^k, \dots, \mathbf{q}_n^k)$$

where the \mathbf{q} are the columns. Also denote by r_{ij}^k the ij^{th} entry of R_k . Thus

$$Q^{(k)} R_k = (\mathbf{q}_1^k, \dots, \mathbf{q}_n^k) \begin{pmatrix} r_{11}^k & & * \\ & \ddots & \\ 0 & & r_{nn}^k \end{pmatrix}$$

This e-book
is made with
SetaPDF



PDF components for PHP developers

www.setasign.com



It follows

$$r_{11}^k \mathbf{q}_1^k \rightarrow \mathbf{q}_1$$

and so

$$r_{11}^k = |r_{11}^k \mathbf{q}_1^k| \rightarrow 1$$

Therefore,

$$\mathbf{q}_1^k \rightarrow \mathbf{q}_1.$$

Next consider the second column.

$$r_{12}^k \mathbf{q}_1^k + r_{22}^k \mathbf{q}_2^k \rightarrow \mathbf{q}_2$$

Taking the inner product of both sides with \mathbf{q}_1^k it follows

$$\lim_{k \rightarrow \infty} r_{12}^k = \lim_{k \rightarrow \infty} (\mathbf{q}_2 \cdot \mathbf{q}_1^k) = (\mathbf{q}_2 \cdot \mathbf{q}_1) = 0.$$

Therefore,

$$\lim_{k \rightarrow \infty} r_{22}^k \mathbf{q}_2^k = \mathbf{q}_2$$

and since $r_{22}^k > 0$, it follows as in the first part that $r_{22}^k \rightarrow 1$. Hence

$$\lim_{k \rightarrow \infty} \mathbf{q}_2^k = \mathbf{q}_2.$$

Continuing this way, it follows

$$\lim_{k \rightarrow \infty} r_{ij}^k = 0$$

for all $i \neq j$ and

$$\lim_{k \rightarrow \infty} r_{jj}^k = 1, \quad \lim_{k \rightarrow \infty} \mathbf{q}_j^k = \mathbf{q}_j.$$

Thus $R_k \rightarrow I$ and $Q^{(k)} \rightarrow Q$. This proves the first part of the lemma.

The second part follows immediately. If $QR = Q'R' = A$ where A^{-1} exists, then

$$Q^*Q' = R(R')^{-1}$$

and I need to show both sides of the above are equal to I . The left side of the above is unitary and the right side is upper triangular having positive entries on the diagonal. This is because the inverse of such an upper triangular matrix having positive entries on the main diagonal is still upper triangular having positive entries on the main diagonal and the product of two such upper triangular matrices gives another of the same form having positive entries on the main diagonal. Suppose then that $Q = R$ where Q is unitary and R is upper triangular having positive entries on the main diagonal. Let $Q_k = Q$ and $R_k = R$. It follows

$$IR_k \rightarrow R = Q$$

and so from the first part, $R_k \rightarrow I$ but $R_k = R$ and so $R = I$. Thus applying this to $Q^*Q' = R(R')^{-1}$ yields both sides equal I . ■

A case of all this is of great interest. Suppose A has a largest eigenvalue λ which is real. Then A^n is of the form $(A^{n-1}\mathbf{a}_1, \dots, A^{n-1}\mathbf{a}_n)$ and so likely each of these columns will be pointing roughly in the direction of an eigenvector of A which corresponds to this eigenvalue. Then when you do the QR factorization of this, it follows from the fact that R is upper triangular, that the first column of Q will be a multiple of $A^{n-1}\mathbf{a}_1$ and so will end up being roughly parallel to the eigenvector desired. Also this will require the entries below the top in the first column of $A_n = Q^T A Q$ will all be small because they will be of the form $\mathbf{q}_i^T A \mathbf{q}_1 \approx \lambda \mathbf{q}_i^T \mathbf{q}_1 = 0$. Therefore, A_n will be of the form

$$\begin{pmatrix} \lambda' & \mathbf{a} \\ \mathbf{e} & B \end{pmatrix}$$

where ϵ is small. It follows that λ' will be close to λ and \mathbf{q}_1 will be close to an eigenvector for λ . Then if you like, you could do the same thing with the matrix B to obtain approximations for the other eigenvalues. Finally, you could use the shifted inverse power method to get more exact solutions.

14.2.2 The Case Of Real Eigenvalues

With these lemmas, it is possible to prove that for the QR algorithm and certain conditions, the sequence A_k converges pointwise to an upper triangular matrix having the eigenvalues of A down the diagonal. I will assume all the matrices are real here.

This convergence won't always happen. Consider for example the matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. You can verify quickly that the algorithm will return this matrix for each k . The problem here is that, although the matrix has the two eigenvalues $-1, 1$, they have the same absolute value. The QR algorithm works in somewhat the same way as the power method, exploiting differences in the size of the eigenvalues.

If A has all real eigenvalues and you are interested in finding these eigenvalues along with the corresponding eigenvectors, you could always consider $A + \lambda I$ instead where λ is sufficiently large and positive that $A + \lambda I$ has all positive eigenvalues. (Recall Gerschgorin's theorem.) Then if μ is an eigenvalue of $A + \lambda I$ with

$$(A + \lambda I)\mathbf{x} = \mu\mathbf{x}$$

then

$$A\mathbf{x} = (\mu - \lambda)\mathbf{x}$$

so to find the eigenvalues of A you just subtract λ from the eigenvalues of $A + \lambda I$. Thus there is no loss of generality in assuming at the outset that the eigenvalues of A are all positive. Here is the theorem. It involves a technical condition which will often hold. The proof presented here follows [27] and is a special case of that presented in this reference.

Before giving the proof, note that the product of upper triangular matrices is upper triangular. If they both have positive entries on the main diagonal so will the product. Furthermore, the inverse of an upper triangular matrix is upper triangular. I will use these simple facts without much comment whenever convenient.

Theorem 14.2.5 *Let A be a real matrix having eigenvalues*

$$\lambda_1 > \lambda_2 > \dots > \lambda_n > 0$$

and let

$$A = SDS^{-1} \tag{14.10}$$

where

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

and suppose S^{-1} has an LU factorization. Then the matrices A_k in the QR algorithm described above converge to an upper triangular matrix T' having the eigenvalues of A , $\lambda_1, \dots, \lambda_n$ descending on the main diagonal. The matrices $Q^{(k)}$ converge to Q' , an orthogonal matrix which equals Q except for possibly having some columns multiplied by -1 for Q the unitary part of the QR factorization of S ,

$$S = QR,$$

and

$$\lim_{k \rightarrow \infty} A_k = T' = Q'^T A Q'$$

Proof: From Lemma 14.2.3

$$A^k = Q^{(k)} R^{(k)} = S D^k S^{-1} \tag{14.11}$$

Let $S = QR$ where this is just a QR factorization which is known to exist and let $S^{-1} = LU$ which is assumed to exist. Thus

$$Q^{(k)} R^{(k)} = Q R D^k L U \tag{14.12}$$

and so

$$Q^{(k)} R^{(k)} = Q R D^k L U = Q R D^k L D^{-k} D^k U$$

That matrix in the middle, $D^k L D^{-k}$ satisfies

$$(D^k L D^{-k})_{ij} = \lambda_i^k L_{ij} \lambda_j^{-k} \text{ for } j \leq i, \text{ 0 if } j > i.$$

Thus for $j < i$ the expression converges to 0 because $\lambda_j > \lambda_i$ when this happens. When $i = j$ it reduces to 1. Thus the matrix in the middle is of the form $I + E_k$ where $E_k \rightarrow 0$. Then it follows

$$\begin{aligned} A^k &= Q^{(k)} R^{(k)} = Q R (I + E_k) D^k U \\ &= Q (I + R E_k R^{-1}) R D^k U \equiv Q (I + F_k) R D^k U \end{aligned}$$

where $F_k \rightarrow 0$. Then let $I + F_k = Q_k R_k$ where this is another QR factorization. Then it reduces to

$$Q^{(k)} R^{(k)} = Q Q_k R_k R D^k U$$

This looks really interesting because by Lemma 14.2.4 $Q_k \rightarrow I$ and $R_k \rightarrow I$ because $Q_k R_k = (I + F_k) \rightarrow I$. So it follows $Q Q_k$ is an orthogonal matrix converging to Q while

FOSS

"I studied Sharp Minds - Bright Ideas! English for 16 years but... I finally learned to speak it in just six lessons!"
Jane, Chinese architect

Employees at FOSS Analytical AS are living proof of the company value - First - using new inventions to make dedicated solutions for our customers. With sharp minds and cross functional teamwork, we constantly strive to develop new unique products - *What do you like to do in your team?*

FOSS works diligently with innovation and development as basis for its growth. It is reflected in the fact that more than 200 of the 1200 employees in FOSS work with Research & Development in Scandinavia and USA. Engineers at FOSS work in production, development and marketing within a wide range of different fields: Chemistry, Electronics, Mechanics, Software, Optics, Microbiology, Chromometrics.

A challenging job in an international and innovative company that is leading in its field. You will get the opportunity to work with the most advanced technology together with highly skilled colleagues.

Read more about FOSS at www.foss.dk - or go directly to our student site www.foss.dk/sharpminds where you can learn more about a number of working together with us on projects, your thesis etc.

The Family owned FOSS group is the world leader as supplier of dedicated, high-tech analytical solutions which measure and control the quality and production of agricultural, food, pharmaceutical and chemical products. Main activities are initiated from Denmark, Sweden and USA with headquarters domiciled in Hillerød, DK. The products are marketed globally by 23 sales companies and an extensive net of distributors. In line with the core value to be 'First', the company intends to expand before and after

Dedicated Analytical Solutions

FOSS
 Slangerupgade 69
 3400 Hillerød
 Tel. +45 70103370
www.foss.dk

Click to hear me t, before and aft





$$R_k R D^k U \left(R^{(k)} \right)^{-1}$$

is upper triangular, being the product of upper triangular matrices. Unfortunately, it is not known that the diagonal entries of this matrix are nonnegative because of the U . Let Λ be just like the identity matrix but having some of the ones replaced with -1 in such a way that ΛU is an upper triangular matrix having positive diagonal entries. Note $\Lambda^2 = I$ and also Λ commutes with a diagonal matrix. Thus

$$Q^{(k)} R^{(k)} = Q Q_k R_k R D^k \Lambda^2 U = Q Q_k R_k R \Lambda D^k (\Lambda U)$$

At this point, one does some inspired massaging to write the above in the form

$$\begin{aligned} & Q Q_k (\Lambda D^k) \left[(\Lambda D^k)^{-1} R_k R \Lambda D^k \right] (\Lambda U) \\ &= Q (Q_k \Lambda) D^k \left[(\Lambda D^k)^{-1} R_k R \Lambda D^k \right] (\Lambda U) \\ &= Q (Q_k \Lambda) D^k \overbrace{\left[(\Lambda D^k)^{-1} R_k R \Lambda D^k \right]}^{\equiv G_k} (\Lambda U) \end{aligned}$$

Now I claim the middle matrix in $[\cdot]$ is upper triangular and has all positive entries on the diagonal. This is because it is an upper triangular matrix which is similar to the upper triangular matrix $R_k R$ and so it has the same eigenvalues (diagonal entries) as $R_k R$. Thus the matrix $G_k \equiv D^k \left[(\Lambda D^k)^{-1} R_k R \Lambda D^k \right] (\Lambda U)$ is upper triangular and has all positive entries on the diagonal. Multiply on the right by G_k^{-1} to get

$$Q^{(k)} R^{(k)} G_k^{-1} = Q Q_k \Lambda \rightarrow Q'$$

where Q' is essentially equal to Q but might have some of the columns multiplied by -1 . This is because $Q_k \rightarrow I$ and so $Q_k \Lambda \rightarrow \Lambda$. Now by Lemma 14.2.4, it follows

$$Q^{(k)} \rightarrow Q', \quad R^{(k)} G_k^{-1} \rightarrow I.$$

It remains to verify A_k converges to an upper triangular matrix. Recall that from 14.11 and the definition below this ($S = QR$)

$$A = SDS^{-1} = (QR) D (QR)^{-1} = QRDR^{-1}Q^T = QTQ^T$$

Where T is an upper triangular matrix. This is because it is the product of upper triangular matrices R, D, R^{-1} . Thus $Q^T A Q = T$. If you replace Q with Q' in the above, it still results in an upper triangular matrix T' having the same diagonal entries as T . This is because

$$T = Q^T A Q = (Q' \Lambda)^T A (Q' \Lambda) = \Lambda Q'^T A Q' \Lambda$$

and considering the ii^{th} entry yields

$$(Q'^T A Q')_{ii} \equiv \sum_{j,k} \Lambda_{ij} (Q'^T A Q')_{jk} \Lambda_{ki} = \Lambda_{ii} \Lambda_{ii} (Q'^T A Q')_{ii} = (Q'^T A Q')_{ii}$$

Recall from Lemma 14.2.3, $A_k = Q^{(k)T} A Q^{(k)}$. Thus taking a limit and using the first part,

$$A_k = Q^{(k)T} A Q^{(k)} \rightarrow Q'^T A Q' = T'. \blacksquare$$

An easy case is for A symmetric. Recall Corollary 6.4.13. By this corollary, there exists an orthogonal (real unitary) matrix Q such that

$$Q^T A Q = D$$

where D is diagonal having the eigenvalues on the main diagonal decreasing in size from the upper left corner to the lower right.

Corollary 14.2.6 Let A be a real symmetric $n \times n$ matrix having eigenvalues

$$\lambda_1 > \lambda_2 > \dots > \lambda_n > 0$$

and let Q be defined by

$$QDQ^T = A, \quad D = Q^T A Q, \tag{14.13}$$

where Q is orthogonal and D is a diagonal matrix having the eigenvalues on the main diagonal decreasing in size from the upper left corner to the lower right. Let Q^T have an LU factorization. Then in the QR algorithm, the matrices $Q^{(k)}$ converge to Q' where Q' is the same as Q except having some columns multiplied by (-1) . Thus the columns of Q' are eigenvectors of A . The matrices A_k converge to D .

Proof: This follows from Theorem 14.2.5. Here $S = Q, S^{-1} = Q^T$. Thus

$$Q = S = QR$$

and $R = I$. By Theorem 14.2.5 and Lemma 14.2.3,

$$A_k = Q^{(k)T} A Q^{(k)} \rightarrow Q'^T A Q' = Q^T A Q = D.$$

because formula 14.13 is unaffected by replacing Q with Q' . ■

When using the QR algorithm, it is not necessary to check technical condition about S^{-1} having an LU factorization. The algorithm delivers a sequence of matrices which are similar to the original one. If that sequence converges to an upper triangular matrix, then the algorithm worked. Furthermore, the technical condition is sufficient but not necessary. The algorithm will work even without the technical condition.

Example 14.2.7 Find the eigenvalues and eigenvectors of the matrix

$$A = \begin{pmatrix} 5 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 1 \end{pmatrix}$$

It is a symmetric matrix but other than that, I just pulled it out of the air. By Lemma 14.2.3 it follows $A_k = Q^{(k)T} A Q^{(k)}$. And so to get to the answer quickly I could have the computer raise A to a power and then take the QR factorization of what results to get the k^{th} iteration using the above formula. Lets pick $k = 10$.

$$\begin{pmatrix} 5 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 1 \end{pmatrix}^{10} = \begin{pmatrix} 4.2273 \times 10^7 & 2.5959 \times 10^7 & 1.8611 \times 10^7 \\ 2.5959 \times 10^7 & 1.6072 \times 10^7 & 1.1506 \times 10^7 \\ 1.8611 \times 10^7 & 1.1506 \times 10^7 & 8.2396 \times 10^6 \end{pmatrix}$$

Now take QR factorization of this. The computer will do that also.

This yields

$$\begin{pmatrix} .79785 & -.59912 & -6.6943 \times 10^{-2} \\ .48995 & .70912 & -.50706 \\ .35126 & .37176 & .85931 \end{pmatrix} \cdot \begin{pmatrix} 5.2983 \times 10^7 & 3.2627 \times 10^7 & 2.338 \times 10^7 \\ 0 & 1.2172 \times 10^5 & 71946. \\ 0 & 0 & 277.03 \end{pmatrix}$$

Next it follows

$$A_{10} = \begin{pmatrix} .79785 & -.59912 & -6.6943 \times 10^{-2} \\ .48995 & .70912 & -.50706 \\ .35126 & .37176 & .85931 \end{pmatrix}^T$$

$$\begin{pmatrix} 5 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} .79785 & -.59912 & -6.6943 \times 10^{-2} \\ .48995 & .70912 & -.50706 \\ .35126 & .37176 & .85931 \end{pmatrix}$$

and this equals

$$\begin{pmatrix} 6.0571 & 3.698 \times 10^{-3} & 3.4346 \times 10^{-5} \\ 3.698 \times 10^{-3} & 3.2008 & -4.0643 \times 10^{-4} \\ 3.4346 \times 10^{-5} & -4.0643 \times 10^{-4} & -.2579 \end{pmatrix}$$

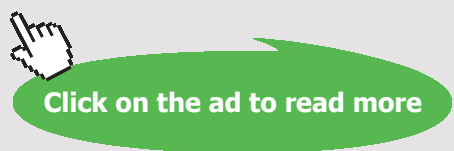
By Gerschgorin’s theorem, the eigenvalues are pretty close to the diagonal entries of the above matrix. Note I didn’t use the theorem, just Lemma 14.2.3 and Gerschgorin’s theorem to verify the eigenvalues are close to the above numbers. The eigenvectors are close to

$$\begin{pmatrix} .79785 \\ .48995 \\ .35126 \end{pmatrix}, \begin{pmatrix} -.59912 \\ .70912 \\ .37176 \end{pmatrix}, \begin{pmatrix} -6.6943 \times 10^{-2} \\ -.50706 \\ .85931 \end{pmatrix}$$

“I studied English for 16 years but...
...I finally learned to speak it in just six lessons”
Jane, Chinese architect

ENGLISH OUT THERE

Click to hear me talking before and after my unique course download



Lets check one of these.

$$\begin{aligned} & \left(\left(\begin{pmatrix} 5 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 1 \end{pmatrix} - 6.0571 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) \begin{pmatrix} .79785 \\ .48995 \\ .35126 \end{pmatrix} \right) \\ &= \begin{pmatrix} -2.1972 \times 10^{-3} \\ 2.5439 \times 10^{-3} \\ 1.3931 \times 10^{-3} \end{pmatrix} \approx \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

Now lets see how well the smallest approximate eigenvalue and eigenvector works.

$$\begin{aligned} & \left(\left(\begin{pmatrix} 5 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 1 \end{pmatrix} - (-.2579) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) \begin{pmatrix} -6.6943 \times 10^{-2} \\ -.50706 \\ .85931 \end{pmatrix} \right) \\ &= \begin{pmatrix} 2.704 \times 10^{-4} \\ -2.7377 \times 10^{-4} \\ -1.3695 \times 10^{-4} \end{pmatrix} \approx \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

For practical purposes, this has found the eigenvalues and eigenvectors.

14.2.3 The QR Algorithm In The General Case

In the case where A has distinct positive eigenvalues it was shown above that under reasonable conditions related to a certain matrix having an LU factorization the QR algorithm produces a sequence of matrices $\{A_k\}$ which converges to an upper triangular matrix. What if A is just an $n \times n$ matrix having possibly complex eigenvalues but A is nondefective? What happens with the QR algorithm in this case? The short answer to this question is that the A_k of the algorithm **typically cannot converge**. However, this does not mean the algorithm is not useful in finding eigenvalues. It turns out the sequence of matrices $\{A_k\}$ have the appearance of a block upper triangular matrix for large k in the sense that the entries below the blocks on the main diagonal are small. Then looking at these blocks gives a way to approximate the eigenvalues. An important example of the concept of a block triangular matrix is the real Schur form for a matrix discussed in Theorem 6.4.7 but the concept as described here allows for any size block centered on the diagonal.

First it is important to note a simple fact about unitary diagonal matrices. In what follows Λ will denote a unitary matrix which is also a diagonal matrix. These matrices are just the identity matrix with some of the ones replaced with a number of the form $e^{i\theta}$ for some θ . The important property of multiplication of any matrix by Λ on either side is that it leaves all the zero entries the same and also preserves the absolute values of the other entries. Thus a block triangular matrix multiplied by Λ on either side is still block triangular. If the matrix is close to being block triangular this property of being close to a block triangular matrix is also preserved by multiplying on either side by Λ . Other patterns depending only on the size of the absolute value occurring in the matrix are also preserved by multiplying on either side by Λ . In other words, in looking for a pattern in a matrix, multiplication by Λ is irrelevant.

Now let A be an $n \times n$ matrix having real or complex entries. By Lemma 14.2.3 and the assumption that A is nondefective, there exists an invertible S ,

$$A^k = Q^{(k)}R^{(k)} = SD^kS^{-1} \tag{14.14}$$

where

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

and by rearranging the columns of S , D can be made such that

$$|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|.$$

Assume S^{-1} has an LU factorization. Then

$$A^k = SD^kLU = SD^kLD^{-k}D^kU.$$

Consider the matrix in the middle, D^kLD^{-k} . The ij^{th} entry is of the form

$$(D^kLD^{-k})_{ij} = \begin{cases} \lambda_i^k L_{ij} \lambda_j^{-k} & \text{if } j < i \\ 1 & \text{if } i = j \\ 0 & \text{if } j > i \end{cases}$$

and these all converge to 0 whenever $|\lambda_i| < |\lambda_j|$. Thus

$$D^kLD^{-k} = (L_k + E_k)$$

where L_k is a lower triangular matrix which has all ones down the diagonal and some subdiagonal terms of the form

$$\lambda_i^k L_{ij} \lambda_j^{-k} \tag{14.15}$$

for which $|\lambda_i| = |\lambda_j|$ while $E_k \rightarrow 0$. (Note the entries of L_k are all bounded independent of k but some may fail to converge.) Then

$$Q^{(k)}R^{(k)} = S(L_k + E_k)D^kU$$

Let

$$SL_k = Q_kR_k \tag{14.16}$$

where this is the QR factorization of SL_k . Then

$$\begin{aligned} Q^{(k)}R^{(k)} &= (Q_kR_k + SE_k)D^kU \\ &= Q_k(I + Q_k^*SE_kR_k^{-1})R_kD^kU \\ &= Q_k(I + F_k)R_kD^kU \end{aligned}$$

where $F_k \rightarrow 0$. Let $I + F_k = Q'_kR'_k$. Then $Q^{(k)}R^{(k)} = Q_kQ'_kR'_kR_kD^kU$. By Lemma 14.2.4

$$Q'_k \rightarrow I \text{ and } R'_k \rightarrow I. \tag{14.17}$$

Now let Λ_k be a diagonal unitary matrix which has the property that $\Lambda_k^*D^kU$ is an upper triangular matrix which has all the diagonal entries positive. Then

$$Q^{(k)}R^{(k)} = Q_kQ'_k\Lambda_k(\Lambda_k^*R'_kR_k\Lambda_k)\Lambda_k^*D^kU$$

That matrix in the middle has all positive diagonal entries because it is itself an upper triangular matrix, being the product of such, and is similar to the matrix R'_kR_k which is upper triangular with positive diagonal entries. By Lemma 14.2.4 again, this time using the uniqueness assertion,

$$Q^{(k)} = Q_kQ'_k\Lambda_k, \quad R^{(k)} = (\Lambda_k^*R'_kR_k\Lambda_k)\Lambda_k^*D^kU$$

Note the term $Q_kQ'_k\Lambda_k$ must be real because the algorithm gives all $Q^{(k)}$ as real matrices. By 14.17 it follows that for k large enough $Q^{(k)} \approx Q_k\Lambda_k$ where \approx means the two matrices are close. Recall $A_k = Q^{(k)T}AQ^{(k)}$ and so for large k ,

$$A_k \approx (Q_k\Lambda_k)^*A(Q_k\Lambda_k) = \Lambda_k^*Q_k^*AQ_k\Lambda_k$$

As noted above, the form of $\Lambda_k^*Q_k^*AQ_k\Lambda_k$ in terms of which entries are large and small is not affected by the presence of Λ_k and Λ_k^* . Thus, in considering what form this is in, it suffices to consider $Q_k^*AQ_k$.

This could get pretty complicated but I will consider the case where

$$\text{if } |\lambda_i| = |\lambda_{i+1}|, \text{ then } |\lambda_{i+2}| < |\lambda_{i+1}|. \tag{14.18}$$

This is typical of the situation where the eigenvalues are all distinct and the matrix A is real so the eigenvalues occur as conjugate pairs. Then in this case, L_k above is lower triangular with some nonzero terms on the diagonal right below the main diagonal but zeros everywhere else. Thus maybe $(L_k)_{s+1,s} \neq 0$ Recall 14.16 which implies

$$Q_k = SL_kR_k^{-1} \tag{14.19}$$

where R_k^{-1} is upper triangular. Also recall from the definition of S in 14.14, it follows that $S^{-1}AS = D$. Thus the columns of S are eigenvectors of A , the i^{th} being an eigenvector for λ_i . Now from the form of L_k , it follows $L_kR_k^{-1}$ is a block upper triangular matrix denoted by T_B and so $Q_k = ST_B$. It follows from the above construction in 14.15 and the given assumption on the sizes of the eigenvalues, there are finitely many 2×2 blocks centered on the main diagonal along with possibly some diagonal entries. Therefore, for large k the matrix $A_k = Q^{(k)T}AQ^{(k)}$ is approximately of the same form as that of

$$Q_k^*AQ_k = T_B^{-1}S^{-1}AST_B = T_B^{-1}DT_B$$

which is a block upper triangular matrix. As explained above, multiplication by the various diagonal unitary matrices does not affect this form. Therefore, for large k , A_k is approximately a block upper triangular matrix.

How would this change if the above assumption on the size of the eigenvalues were relaxed but the matrix was still nondefective with appropriate matrices having an LU factorization as above? It would mean the blocks on the diagonal would be larger. This immediately makes the problem more cumbersome to deal with. However, in the case that the eigenvalues of A are distinct, the above situation really is typical of what occurs and in any case can be quickly reduced to this case.

The Wake
the only emission we want to leave behind

Low-speed Engines Medium-speed Engines Turbochargers Propellers Propulsion Packages PrimeServ

The design of eco-friendly marine power and propulsion solutions is crucial for MAN Diesel & Turbo. Power competencies are offered with the world's largest engine programme – having outputs spanning from 450 to 87,220 kW per engine. Get up front! Find out more at www.mandieselturbo.com

Engineering the Future – since 1758.
MAN Diesel & Turbo



To see this, suppose condition 14.18 is violated and $\lambda_j, \dots, \lambda_{j+p}$ are complex eigenvalues having nonzero imaginary parts such that each has the same absolute value but they are all distinct. Then let $\mu > 0$ and consider the matrix $A + \mu I$. Thus the corresponding eigenvalues of $A + \mu I$ are $\lambda_j + \mu, \dots, \lambda_{j+p} + \mu$. A short computation shows $|\lambda_j + \mu|, \dots, |\lambda_{j+p} + \mu|$ are all distinct and so the above situation of 14.18 is obtained. Of course, if there are repeated eigenvalues, it may not be possible to reduce to the case above and you would end up with large blocks on the main diagonal which could be difficult to deal with.

So how do you identify the eigenvalues? You know A_k and behold that it is close to a block upper triangular matrix T'_B . You know A_k is also similar to A . Therefore, T'_B has eigenvalues which are close to the eigenvalues of A_k and hence those of A provided k is sufficiently large. See Theorem 6.9.2 which depends on complex analysis or the exercise on Page 187 which gives another way to see this. Thus you find the eigenvalues of this block triangular matrix T'_B and assert that these are good approximations of the eigenvalues of A_k and hence to those of A . How do you find the eigenvalues of a block triangular matrix? This is easy from Lemma 6.4.6. Say

$$T'_B = \begin{pmatrix} B_1 & \cdots & * \\ & \ddots & \vdots \\ 0 & & B_m \end{pmatrix}$$

Then forming $\lambda I - T'_B$ and taking the determinant, it follows from Lemma 6.4.6 this equals

$$\prod_{j=1}^m \det(\lambda I_j - B_j)$$

and so all you have to do is take the union of the eigenvalues for each B_j . In the case emphasized here this is very easy because these blocks are just 2×2 matrices.

How do you identify approximate eigenvectors from this? First try to find the approximate eigenvectors for A_k . Pick an approximate eigenvalue λ , an exact eigenvalue for T'_B . Then find \mathbf{v} solving $T'_B \mathbf{v} = \lambda \mathbf{v}$. It follows since T'_B is close to A_k that $A_k \mathbf{v} \approx \lambda \mathbf{v}$ and so

$$Q^{(k)} A Q^{(k)T} \mathbf{v} = A_k \mathbf{v} \approx \lambda \mathbf{v}$$

Hence

$$A Q^{(k)T} \mathbf{v} \approx \lambda Q^{(k)T} \mathbf{v}$$

and so $Q^{(k)T} \mathbf{v}$ is an approximation to the eigenvector which goes with the eigenvalue of A which is close to λ .

Example 14.2.8 Here is a matrix.

$$\begin{pmatrix} 3 & 2 & 1 \\ -2 & 0 & -1 \\ -2 & -2 & 0 \end{pmatrix}$$

It happens that the eigenvalues of this matrix are $1, 1+i, 1-i$. Lets apply the QR algorithm as if the eigenvalues were not known.

Applying the QR algorithm to this matrix yields the following sequence of matrices.

$$A_1 = \begin{pmatrix} 1.2353 & 1.9412 & 4.3657 \\ -.39215 & 1.5425 & 5.3886 \times 10^{-2} \\ -.16169 & -.18864 & .22222 \end{pmatrix}$$

$$\vdots$$

$$A_{12} = \begin{pmatrix} 9.1772 \times 10^{-2} & .63089 & -2.0398 \\ -2.8556 & 1.9082 & -3.1043 \\ 1.0786 \times 10^{-2} & 3.4614 \times 10^{-4} & 1.0 \end{pmatrix}$$

At this point the bottom two terms on the left part of the bottom row are both very small so it appears the real eigenvalue is near 1.0. The complex eigenvalues are obtained from solving

$$\det \left(\lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 9.1772 \times 10^{-2} & .63089 \\ -2.8556 & 1.9082 \end{pmatrix} \right) = 0$$

This yields

$$\lambda = 1.0 - .98828i, 1.0 + .98828i$$

Example 14.2.9 *The equation $x^4 + x^3 + 4x^2 + x - 2 = 0$ has exactly two real solutions. You can see this by graphing it. However, the rational root theorem from algebra shows neither of these solutions are rational. Also, graphing it does not yield any information about the complex solutions. Lets use the QR algorithm to approximate all the solutions, real and complex.*

A matrix whose characteristic polynomial is the given polynomial is

$$\begin{pmatrix} -1 & -4 & -1 & 2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Using the QR algorithm yields the following sequence of iterates for A_k

$$A_1 = \begin{pmatrix} .99999 & -2.5927 & -1.7588 & -1.2978 \\ 2.1213 & -1.7778 & -1.6042 & -.99415 \\ 0 & .34246 & -.32749 & -.91799 \\ 0 & 0 & -.44659 & .10526 \end{pmatrix}$$

⋮

$$A_9 = \begin{pmatrix} -.83412 & -4.1682 & -1.939 & -.7783 \\ 1.05 & .14514 & .2171 & 2.5474 \times 10^{-2} \\ 0 & 4.0264 \times 10^{-4} & -.85029 & -.61608 \\ 0 & 0 & -1.8263 \times 10^{-2} & .53939 \end{pmatrix}$$

Now this is similar to A and the eigenvalues are close to the eigenvalues obtained from the two blocks on the diagonal,

$$\begin{pmatrix} -.83412 & -4.1682 \\ 1.05 & .14514 \end{pmatrix}, \begin{pmatrix} -.85029 & -.61608 \\ -1.8263 \times 10^{-2} & .53939 \end{pmatrix}$$

since 4.0264×10^{-4} is small. After routine computations involving the quadratic formula, these are seen to be

$$-.85834, .54744, -.34449 - 2.0339i, -.34449 + 2.0339i$$

When these are plugged in to the polynomial equation, you see that each is close to being a solution of the equation.

It seems like most of the attention to the QR algorithm has to do with finding ways to get it to “converge” faster. Great and marvelous are the clever tricks which have been proposed to do this but my intent is to present the basic ideas, not to go in to the numerous refinements of this algorithm. However, there is one thing which is usually done. It involves reducing to the case of an upper Hessenberg matrix which is one which is zero below the main sub diagonal. Every matrix is unitarily similar to one of these.

Let A be an invertible $n \times n$ matrix. Let Q'_1 be a unitary matrix

$$Q'_1 \begin{pmatrix} a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} = \begin{pmatrix} \sqrt{\sum_{j=2}^n |a_{j1}|^2} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \equiv \begin{pmatrix} a \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

The vector Q'_1 is multiplying is just the bottom $n - 1$ entries of the first column of A . Then let Q_1 be

$$\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q'_1 \end{pmatrix}$$

It follows

$$\begin{aligned} Q_1 A Q_1^* &= \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q'_1 \end{pmatrix} A Q_1^* = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a & & & \\ \vdots & & A_1 & \\ 0 & & & \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_1'^* \end{pmatrix} \\ &= \begin{pmatrix} * & * & \cdots & * \\ a & & & \\ \vdots & & A_1 & \\ 0 & & & \end{pmatrix} \end{aligned}$$

gaiteye[®]
Challenge the way we run

EXPERIENCE THE POWER OF FULL ENGAGEMENT...

**RUN FASTER.
RUN LONGER..
RUN EASIER...**

READ MORE & PRE-ORDER TODAY
WWW.GAITEYE.COM



Now let Q'_2 be the $n - 2 \times n - 2$ matrix which does to the first column of A_1 the same sort of thing that the $n - 1 \times n - 1$ matrix Q'_1 did to the first column of A . Let

$$Q_2 \equiv \begin{pmatrix} I & 0 \\ 0 & Q'_2 \end{pmatrix}$$

where I is the 2×2 identity. Then applying block multiplication,

$$Q_2 Q_1 A Q_1^* Q_2^* = \begin{pmatrix} * & * & \cdots & * & * \\ * & * & \cdots & * & * \\ 0 & * & & & \\ \vdots & \vdots & & A_2 & \\ 0 & 0 & & & \end{pmatrix}$$

where A_2 is now an $n - 2 \times n - 2$ matrix. Continuing this way you eventually get a unitary matrix Q which is a product of those discussed above such that

$$Q A Q^T = \begin{pmatrix} * & * & \cdots & * & * \\ * & * & \cdots & * & * \\ 0 & * & * & & \vdots \\ \vdots & \vdots & \ddots & \ddots & * \\ 0 & 0 & & * & * \end{pmatrix}$$

This matrix equals zero below the subdiagonal. It is called an upper Hessenberg matrix.

It happens that in the QR algorithm, if A_k is upper Hessenberg, so is A_{k+1} . To see this, note that the matrix is upper Hessenberg means that $A_{ij} = 0$ whenever $i - j \geq 2$.

$$A_{k+1} = R_k Q_k$$

where $A_k = Q_k R_k$. Therefore as shown before,

$$A_{k+1} = R_k A_k R_k^{-1}$$

Let the ij^{th} entry of A_k be a_{ij}^k . Then if $i - j \geq 2$

$$a_{ij}^{k+1} = \sum_{p=i}^n \sum_{q=1}^j r_{ip} a_{pq}^k r_{qj}^{-1}$$

It is given that $a_{pq}^k = 0$ whenever $p - q \geq 2$. However, from the above sum,

$$p - q \geq i - j \geq 2$$

and so the sum equals 0.

Since upper Hessenberg matrices stay that way in the algorithm and it is closer to being upper triangular, it is reasonable to suppose the QR algorithm will yield good results more quickly for this upper Hessenberg matrix than for the original matrix. This would be especially true if the matrix is good sized. The other important thing to observe is that, starting with an upper Hessenberg matrix, the algorithm will restrict the size of the blocks which occur to being 2×2 blocks which are easy to deal with. These blocks allow you to identify the complex roots.

14.3 Exercises

In these exercises which call for a computation, don't waste time on them unless you use a computer or calculator which can raise matrices to powers and take QR factorizations.

1. In Example 14.1.10 an eigenvalue was found correct to several decimal places along with an eigenvector. Find the other eigenvalues along with their eigenvectors.

2. Find the eigenvalues and eigenvectors of the matrix $A = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 1 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ numerically.

In this case the exact eigenvalues are $\pm\sqrt{3}, 6$. Compare with the exact answers.

3. Find the eigenvalues and eigenvectors of the matrix $A = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 5 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ numerically.

The exact eigenvalues are $2, 4 + \sqrt{15}, 4 - \sqrt{15}$. Compare your numerical results with the exact values. Is it much fun to compute the exact eigenvectors?

4. Find the eigenvalues and eigenvectors of the matrix $A = \begin{pmatrix} 0 & 2 & 1 \\ 2 & 5 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ numerically.

I don't know the exact eigenvalues in this case. Check your answers by multiplying your numerically computed eigenvectors by the matrix.

5. Find the eigenvalues and eigenvectors of the matrix $A = \begin{pmatrix} 0 & 2 & 1 \\ 2 & 0 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ numerically.

I don't know the exact eigenvalues in this case. Check your answers by multiplying your numerically computed eigenvectors by the matrix.

6. Consider the matrix $A = \begin{pmatrix} 3 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 0 \end{pmatrix}$ and the vector $(1, 1, 1)^T$. Find the shortest

distance between the Rayleigh quotient determined by this vector and some eigenvalue of A .

7. Consider the matrix $A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 4 \\ 1 & 4 & 5 \end{pmatrix}$ and the vector $(1, 1, 1)^T$. Find the shortest

distance between the Rayleigh quotient determined by this vector and some eigenvalue of A .

8. Consider the matrix $A = \begin{pmatrix} 3 & 2 & 3 \\ 2 & 6 & 4 \\ 3 & 4 & -3 \end{pmatrix}$ and the vector $(1, 1, 1)^T$. Find the shortest

distance between the Rayleigh quotient determined by this vector and some eigenvalue of A .

9. Using Gerschgorin's theorem, find upper and lower bounds for the eigenvalues of $A = \begin{pmatrix} 3 & 2 & 3 \\ 2 & 6 & 4 \\ 3 & 4 & -3 \end{pmatrix}$.

10. Tell how to find a matrix whose characteristic polynomial is a given monic polynomial. This is called a companion matrix. Find the roots of the polynomial $x^3 + 7x^2 + 3x + 7$.

11. Find the roots to $x^4 + 3x^3 + 4x^2 + x + 1$. It has two complex roots.

12. Suppose A is a real symmetric matrix and the technique of reducing to an upper Hessenberg matrix is followed. Show the resulting upper Hessenberg matrix is actually equal to 0 on the top as well as the bottom.

Appendix A

Matrix Calculator On The Web

A.1 Use Of Matrix Calculator On Web

There is a really nice service on the web which will do all of these things very easily. It is www.bluebit.gr/matrix-calculator/ To get to it, you can use the address or google matrix calculator.

When you go to this site, you enter a matrix row by row, placing a space between each number. When you come to the end of a row, you press enter on the keyboard to start the next row. After entering the matrix, you select what you want it to do. You will see that it also solves systems of equations.

**Technical training on
WHAT you need, WHEN you need it**

At IDC Technologies we can tailor our technical and engineering training workshops to suit your needs. We have extensive experience in training technical and engineering staff and have trained people in organisations such as General Motors, Shell, Siemens, BHP and Honeywell to name a few.

Our onsite training is cost effective, convenient and completely customisable to the technical and engineering areas you want covered. Our workshops are all comprehensive hands-on learning experiences with ample time given to practical sessions and demonstrations. We communicate well to ensure that workshop content and timing match the knowledge, skills, and abilities of the participants.

We run onsite training all year round and hold the workshops on your premises or a venue of your choice for your convenience.

For a no obligation proposal, contact us today at training@idc-online.com or visit our website for more information: www.idc-online.com/onsite/

- OIL & GAS ENGINEERING**
- ELECTRONICS**
- AUTOMATION & PROCESS CONTROL**
- MECHANICAL ENGINEERING**
- INDUSTRIAL DATA COMMS**
- ELECTRICAL POWER**

Phone: +61 8 9321 1702
Email: training@idc-online.com
Website: www.idc-online.com

IDC TECHNOLOGIES

Appendix B

Positive Matrices

Earlier theorems about Markov matrices were presented. These were matrices in which all the entries were nonnegative and either the columns or the rows added to 1. It turns out that many of the theorems presented can be generalized to positive matrices. When this is done, the resulting theory is mainly due to Perron and Frobenius. I will give an introduction to this theory here following Karlin and Taylor [19].

Definition B.0.1 For A a matrix or vector, the notation, $A \gg 0$ will mean every entry of A is positive. By $A > 0$ is meant that every entry is nonnegative and at least one is positive. By $A \geq 0$ is meant that every entry is nonnegative. Thus the matrix or vector consisting only of zeros is ≥ 0 . An expression like $A \gg B$ will mean $A - B \gg 0$ with similar modifications for $>$ and \geq .

For the sake of this section only, define the following for $\mathbf{x} = (x_1, \dots, x_n)^T$, a vector.

$$|\mathbf{x}| \equiv (|x_1|, \dots, |x_n|)^T.$$

Thus $|\mathbf{x}|$ is the vector which results by replacing each entry of \mathbf{x} with its absolute value¹. Also define for $\mathbf{x} \in \mathbb{C}^n$,

$$\|\mathbf{x}\|_1 \equiv \sum_k |x_k|.$$

Lemma B.0.2 Let $A \gg 0$ and let $\mathbf{x} > \mathbf{0}$. Then $A\mathbf{x} \gg \mathbf{0}$.

Proof: $(A\mathbf{x})_i = \sum_j A_{ij}x_j > 0$ because all the $A_{ij} > 0$ and at least one $x_j > 0$.

Lemma B.0.3 Let $A \gg 0$. Define

$$S \equiv \{\lambda : A\mathbf{x} > \lambda\mathbf{x} \text{ for some } \mathbf{x} \gg \mathbf{0}\},$$

and let

$$K \equiv \{\mathbf{x} \geq \mathbf{0} \text{ such that } \|\mathbf{x}\|_1 = 1\}.$$

Now define

$$S_1 \equiv \{\lambda : A\mathbf{x} \geq \lambda\mathbf{x} \text{ for some } \mathbf{x} \in K\}.$$

Then

$$\sup(S) = \sup(S_1).$$

Proof: Let $\lambda \in S$. Then there exists $\mathbf{x} \gg \mathbf{0}$ such that $A\mathbf{x} > \lambda\mathbf{x}$. Consider $\mathbf{y} \equiv \mathbf{x} / \|\mathbf{x}\|_1$. Then $\|\mathbf{y}\|_1 = 1$ and $A\mathbf{y} > \lambda\mathbf{y}$. Therefore, $\lambda \in S_1$ and so $S \subseteq S_1$. Therefore, $\sup(S) \leq \sup(S_1)$.

Now let $\lambda \in S_1$. Then there exists $\mathbf{x} \geq \mathbf{0}$ such that $\|\mathbf{x}\|_1 = 1$ so $\mathbf{x} > \mathbf{0}$ and $A\mathbf{x} > \lambda\mathbf{x}$. Letting $\mathbf{y} \equiv A\mathbf{x}$, it follows from Lemma B.0.2 that $A\mathbf{y} \gg \lambda\mathbf{y}$ and $\mathbf{y} \gg \mathbf{0}$. Thus $\lambda \in S$ and so $S_1 \subseteq S$ which shows that $\sup(S_1) \leq \sup(S)$. ■

This lemma is significant because the set, $\{\mathbf{x} \geq \mathbf{0} \text{ such that } \|\mathbf{x}\|_1 = 1\} \equiv K$ is a compact set in \mathbb{R}^n . Define

$$\lambda_0 \equiv \sup(S) = \sup(S_1). \tag{2.1}$$

The following theorem is due to Perron.

¹This notation is just about the most abominable thing imaginable because it is the same notation but entirely different meaning than the norm. However, it saves space in the presentation of this theory of positive matrices and avoids the use of new symbols. Please forget about it when you leave this section.

Theorem B.0.4 Let $A \gg 0$ be an $n \times n$ matrix and let λ_0 be given in 2.1. Then

1. $\lambda_0 > 0$ and there exists $\mathbf{x}_0 \gg \mathbf{0}$ such that $A\mathbf{x}_0 = \lambda_0\mathbf{x}_0$ so λ_0 is an eigenvalue for A .
2. If $A\mathbf{x} = \mu\mathbf{x}$ where $\mathbf{x} \neq \mathbf{0}$, and $\mu \neq \lambda_0$. Then $|\mu| < \lambda_0$.
3. The eigenspace for λ_0 has dimension 1.

Proof: To see $\lambda_0 > 0$, consider the vector, $\mathbf{e} \equiv (1, \dots, 1)^T$. Then

$$(A\mathbf{e})_i = \sum_j A_{ij} > 0$$

and so λ_0 is at least as large as

$$\min_i \sum_j A_{ij}.$$

Let $\{\lambda_k\}$ be an increasing sequence of numbers from S_1 converging to λ_0 . Letting \mathbf{x}_k be the vector from K which occurs in the definition of S_1 , these vectors are in a compact set. Therefore, there exists a subsequence, still denoted by \mathbf{x}_k such that $\mathbf{x}_k \rightarrow \mathbf{x}_0 \in K$ and $\lambda_k \rightarrow \lambda_0$. Then passing to the limit,

$$A\mathbf{x}_0 \geq \lambda_0\mathbf{x}_0, \mathbf{x}_0 > \mathbf{0}.$$

If $A\mathbf{x}_0 > \lambda_0\mathbf{x}_0$, then letting $\mathbf{y} \equiv A\mathbf{x}_0$, it follows from Lemma B.0.2 that $A\mathbf{y} \gg \lambda_0\mathbf{y}$ and $\mathbf{y} \gg \mathbf{0}$. But this contradicts the definition of λ_0 as the supremum of the elements of S because since $A\mathbf{y} \gg \lambda_0\mathbf{y}$, it follows $A\mathbf{y} \gg (\lambda_0 + \varepsilon)\mathbf{y}$ for ε a small positive number. Therefore, $A\mathbf{x}_0 = \lambda_0\mathbf{x}_0$. It remains to verify that $\mathbf{x}_0 \gg \mathbf{0}$. But this follows immediately from

$$0 < \sum_j A_{ij}x_{0j} = (A\mathbf{x}_0)_i = \lambda_0x_{0i}.$$

This proves 1.

Next suppose $A\mathbf{x} = \mu\mathbf{x}$ and $\mathbf{x} \neq \mathbf{0}$ and $\mu \neq \lambda_0$. Then $|A\mathbf{x}| = |\mu||\mathbf{x}|$. But this implies $A|\mathbf{x}| \geq |\mu||\mathbf{x}|$. (See the above abominable definition of $|\mathbf{x}|$.)

Case 1: $|\mathbf{x}| \neq \mathbf{x}$ and $|\mathbf{x}| \neq -\mathbf{x}$.

In this case, $A|\mathbf{x}| > |\mu||\mathbf{x}| = A|\mathbf{x}|$ and letting $\mathbf{y} = A|\mathbf{x}|$, it follows $\mathbf{y} \gg \mathbf{0}$ and $A\mathbf{y} \gg |\mu|\mathbf{y}$ which shows $A\mathbf{y} \gg (|\mu| + \varepsilon)\mathbf{y}$ for sufficiently small positive ε and verifies $|\mu| < \lambda_0$.

Case 2: $|\mathbf{x}| = \mathbf{x}$ or $|\mathbf{x}| = -\mathbf{x}$

In this case, the entries of \mathbf{x} are all real and have the same sign. Therefore, $A|\mathbf{x}| = |A\mathbf{x}| = |\mu||\mathbf{x}|$. Now let $\mathbf{y} \equiv |\mathbf{x}|/|\mathbf{x}|_1$. Then $A\mathbf{y} = |\mu|\mathbf{y}$ and so $|\mu| \in S_1$ showing that $|\mu| \leq \lambda_0$. But also, the fact the entries of \mathbf{x} all have the same sign shows $\mu = |\mu|$ and so $\mu \in S_1$. Since $\mu \neq \lambda_0$, it must be that $\mu = |\mu| < \lambda_0$. This proves 2.

It remains to verify 3. Suppose then that $A\mathbf{y} = \lambda_0\mathbf{y}$ and for all scalars α , $\alpha\mathbf{x}_0 \neq \mathbf{y}$. Then

$$A \operatorname{Re} \mathbf{y} = \lambda_0 \operatorname{Re} \mathbf{y}, A \operatorname{Im} \mathbf{y} = \lambda_0 \operatorname{Im} \mathbf{y}.$$

If $\operatorname{Re} \mathbf{y} = \alpha_1\mathbf{x}_0$ and $\operatorname{Im} \mathbf{y} = \alpha_2\mathbf{x}_0$ for real numbers, α_i , then $\mathbf{y} = (\alpha_1 + i\alpha_2)\mathbf{x}_0$ and it is assumed this does not happen. Therefore, either

$$t \operatorname{Re} \mathbf{y} \neq \mathbf{x}_0 \text{ for all } t \in \mathbb{R}$$

or

$$t \operatorname{Im} \mathbf{y} \neq \mathbf{x}_0 \text{ for all } t \in \mathbb{R}.$$

Assume the first holds. Then varying $t \in \mathbb{R}$, there exists a value of t such that $\mathbf{x}_0 + t \operatorname{Re} \mathbf{y} > \mathbf{0}$ but it is not the case that $\mathbf{x}_0 + t \operatorname{Re} \mathbf{y} \gg \mathbf{0}$. Then $A(\mathbf{x}_0 + t \operatorname{Re} \mathbf{y}) \gg \mathbf{0}$ by Lemma B.0.2. But this implies $\lambda_0(\mathbf{x}_0 + t \operatorname{Re} \mathbf{y}) \gg \mathbf{0}$ which is a contradiction. Hence there exist real numbers, α_1 and α_2 such that $\operatorname{Re} \mathbf{y} = \alpha_1\mathbf{x}_0$ and $\operatorname{Im} \mathbf{y} = \alpha_2\mathbf{x}_0$ showing that $\mathbf{y} = (\alpha_1 + i\alpha_2)\mathbf{x}_0$. This proves 3.

It is possible to obtain a simple corollary to the above theorem.

Corollary B.0.5 *If $A > 0$ and $A^m \gg 0$ for some $m \in \mathbb{N}$, then all the conclusions of the above theorem hold.*

Proof: There exists $\mu_0 > 0$ such that $A^m \mathbf{y}_0 = \mu_0 \mathbf{y}_0$ for $\mathbf{y}_0 \gg 0$ by Theorem B.0.4 and

$$\mu_0 = \sup \{ \mu : A^m \mathbf{x} \geq \mu \mathbf{x} \text{ for some } \mathbf{x} \in K \}.$$

Let $\lambda_0^m = \mu_0$. Then

$$(A - \lambda_0 I) (A^{m-1} + \lambda_0 A^{m-2} + \dots + \lambda_0^{m-1} I) \mathbf{y}_0 = (A^m - \lambda_0^m I) \mathbf{y}_0 = \mathbf{0}$$

and so letting $\mathbf{x}_0 \equiv (A^{m-1} + \lambda_0 A^{m-2} + \dots + \lambda_0^{m-1} I) \mathbf{y}_0$, it follows $\mathbf{x}_0 \gg 0$ and $A \mathbf{x}_0 = \lambda_0 \mathbf{x}_0$.

Suppose now that $A \mathbf{x} = \mu \mathbf{x}$ for $\mathbf{x} \neq \mathbf{0}$ and $\mu \neq \lambda_0$. Suppose $|\mu| \geq \lambda_0$. Multiplying both sides by A , it follows $A^m \mathbf{x} = \mu^m \mathbf{x}$ and $|\mu^m| = |\mu|^m \geq \lambda_0^m = \mu_0$ and so from Theorem B.0.4, since $|\mu^m| \geq \mu_0$, and μ^m is an eigenvalue of A^m , it follows that $\mu^m = \mu_0$. But by Theorem B.0.4 again, this implies $\mathbf{x} = c \mathbf{y}_0$ for some scalar, c and hence $A \mathbf{y}_0 = \mu \mathbf{y}_0$. Since $\mathbf{y}_0 \gg 0$, it follows $\mu \geq 0$ and so $\mu = \lambda_0$, a contradiction. Therefore, $|\mu| < \lambda_0$.

Finally, if $A \mathbf{x} = \lambda_0 \mathbf{x}$, then $A^m \mathbf{x} = \lambda_0^m \mathbf{x}$ and so $\mathbf{x} = c \mathbf{y}_0$ for some scalar, c . Consequently,

$$\begin{aligned} (A^{m-1} + \lambda_0 A^{m-2} + \dots + \lambda_0^{m-1} I) \mathbf{x} &= c (A^{m-1} + \lambda_0 A^{m-2} + \dots + \lambda_0^{m-1} I) \mathbf{y}_0 \\ &= c \mathbf{x}_0. \end{aligned}$$

Hence

$$m \lambda_0^{m-1} \mathbf{x} = c \mathbf{x}_0$$

I joined MITAS because
I wanted **real responsibility**

The Graduate Programme
for Engineers and Geoscientists
www.discovermitas.com



Month 16

I was a construction
supervisor in
the North Sea
advising and
helping foremen
solve problems

Real work
International opportunities
Three work placements



which shows the dimension of the eigenspace for λ_0 is one. ■

The following corollary is an extremely interesting convergence result involving the powers of positive matrices.

Corollary B.0.6 *Let $A > 0$ and $A^m \gg 0$ for some $m \in \mathbb{N}$. Then for λ_0 given in 2.1, there exists a rank one matrix P such that $\lim_{m \rightarrow \infty} \left\| \left(\frac{A}{\lambda_0} \right)^m - P \right\| = 0$.*

Proof: Considering A^T , and the fact that A and A^T have the same eigenvalues, Corollary B.0.5 implies the existence of a vector, $\mathbf{v} \gg \mathbf{0}$ such that

$$A^T \mathbf{v} = \lambda_0 \mathbf{v}.$$

Also let \mathbf{x}_0 denote the vector such that $A\mathbf{x}_0 = \lambda_0\mathbf{x}_0$ with $\mathbf{x}_0 \gg \mathbf{0}$. First note that $\mathbf{x}_0^T \mathbf{v} > 0$ because both these vectors have all entries positive. Therefore, \mathbf{v} may be scaled such that

$$\mathbf{v}^T \mathbf{x}_0 = \mathbf{x}_0^T \mathbf{v} = 1. \tag{2.2}$$

Define

$$P \equiv \mathbf{x}_0 \mathbf{v}^T.$$

Thanks to 2.2,

$$\frac{A}{\lambda_0} P = \mathbf{x}_0 \mathbf{v}^T = P, \quad P \left(\frac{A}{\lambda_0} \right) = \mathbf{x}_0 \mathbf{v}^T \left(\frac{A}{\lambda_0} \right) = \mathbf{x}_0 \mathbf{v}^T = P, \tag{2.3}$$

and

$$P^2 = \mathbf{x}_0 \mathbf{v}^T \mathbf{x}_0 \mathbf{v}^T = \mathbf{v}^T \mathbf{x}_0 = P. \tag{2.4}$$

Therefore,

$$\begin{aligned} \left(\frac{A}{\lambda_0} - P \right)^2 &= \left(\frac{A}{\lambda_0} \right)^2 - 2 \left(\frac{A}{\lambda_0} \right) P + P^2 \\ &= \left(\frac{A}{\lambda_0} \right)^2 - P. \end{aligned}$$

Continuing this way, using 2.3 repeatedly, it follows

$$\left(\left(\frac{A}{\lambda_0} \right) - P \right)^m = \left(\frac{A}{\lambda_0} \right)^m - P. \tag{2.5}$$

The eigenvalues of $\left(\frac{A}{\lambda_0} \right) - P$ are of interest because it is powers of this matrix which determine the convergence of $\left(\frac{A}{\lambda_0} \right)^m$ to P . Therefore, let μ be a nonzero eigenvalue of this matrix. Thus

$$\left(\left(\frac{A}{\lambda_0} \right) - P \right) \mathbf{x} = \mu \mathbf{x} \tag{2.6}$$

for $\mathbf{x} \neq \mathbf{0}$, and $\mu \neq 0$. Applying P to both sides and using the second formula of 2.3 yields

$$\mathbf{0} = (P - P) \mathbf{x} = \left(P \left(\frac{A}{\lambda_0} \right) - P^2 \right) \mathbf{x} = \mu P \mathbf{x}.$$

But since $P\mathbf{x} = \mathbf{0}$, it follows from 2.6 that

$$A\mathbf{x} = \lambda_0 \mu \mathbf{x}$$

which implies $\lambda_0 \mu$ is an eigenvalue of A . Therefore, by Corollary B.0.5 it follows that either $\lambda_0 \mu = \lambda_0$ in which case $\mu = 1$, or $\lambda_0 |\mu| < \lambda_0$ which implies $|\mu| < 1$. But if $\mu = 1$, then \mathbf{x} is a multiple of \mathbf{x}_0 and 2.6 would yield

$$\left(\left(\frac{A}{\lambda_0} \right) - P \right) \mathbf{x}_0 = \mathbf{x}_0$$

which says $\mathbf{x}_0 - \mathbf{x}_0 \mathbf{v}^T \mathbf{x}_0 = \mathbf{x}_0$ and so by 2.2, $\mathbf{x}_0 = \mathbf{0}$ contrary to the property that $\mathbf{x}_0 \gg \mathbf{0}$. Therefore, $|\mu| < 1$ and so this has shown that the absolute values of all eigenvalues of $\left(\frac{A}{\lambda_0}\right) - P$ are less than 1. By Gelfand's theorem, Theorem 13.3.3, it follows

$$\left\| \left(\left(\frac{A}{\lambda_0} \right) - P \right)^m \right\|^{1/m} < r < 1$$

whenever m is large enough. Now by 2.5 this yields

$$\left\| \left(\frac{A}{\lambda_0} \right)^m - P \right\| = \left\| \left(\left(\frac{A}{\lambda_0} \right) - P \right)^m \right\| \leq r^m$$

whenever m is large enough. It follows

$$\lim_{m \rightarrow \infty} \left\| \left(\frac{A}{\lambda_0} \right)^m - P \right\| = 0$$

as claimed.

What about the case when $A > 0$ but maybe it is not the case that $A \gg 0$? As before,

$$K \equiv \{ \mathbf{x} \geq \mathbf{0} \text{ such that } \|\mathbf{x}\|_1 = 1 \}.$$

Now define

$$S_1 \equiv \{ \lambda : A\mathbf{x} \geq \lambda\mathbf{x} \text{ for some } \mathbf{x} \in K \}$$

and

$$\lambda_0 \equiv \sup(S_1) \tag{2.7}$$

Theorem B.0.7 *Let $A > 0$ and let λ_0 be defined in 2.7. Then there exists $\mathbf{x}_0 > \mathbf{0}$ such that $A\mathbf{x}_0 = \lambda_0\mathbf{x}_0$.*

Proof: Let E consist of the matrix which has a one in every entry. Then from Theorem B.0.4 it follows there exists $\mathbf{x}_\delta \gg \mathbf{0}$, $\|\mathbf{x}_\delta\|_1 = 1$, such that $(A + \delta E)\mathbf{x}_\delta = \lambda_{0\delta}\mathbf{x}_\delta$ where

$$\lambda_{0\delta} \equiv \sup \{ \lambda : (A + \delta E)\mathbf{x} \geq \lambda\mathbf{x} \text{ for some } \mathbf{x} \in K \}.$$

Now if $\alpha < \delta$

$$\begin{aligned} \{ \lambda : (A + \alpha E)\mathbf{x} \geq \lambda\mathbf{x} \text{ for some } \mathbf{x} \in K \} &\subseteq \\ \{ \lambda : (A + \delta E)\mathbf{x} \geq \lambda\mathbf{x} \text{ for some } \mathbf{x} \in K \} & \end{aligned}$$

and so $\lambda_{0\delta} \geq \lambda_{0\alpha}$ because $\lambda_{0\delta}$ is the sup of the second set and $\lambda_{0\alpha}$ is the sup of the first. It follows the limit, $\lambda_1 \equiv \lim_{\delta \rightarrow 0+} \lambda_{0\delta}$ exists. Taking a subsequence and using the compactness of K , there exists a subsequence, still denoted by δ such that as $\delta \rightarrow 0$, $\mathbf{x}_\delta \rightarrow \mathbf{x} \in K$. Therefore,

$$A\mathbf{x} = \lambda_1\mathbf{x}$$

and so, in particular, $A\mathbf{x} \geq \lambda_1\mathbf{x}$ and so $\lambda_1 \leq \lambda_0$. But also, if $\lambda \leq \lambda_0$,

$$\lambda\mathbf{x} \leq A\mathbf{x} < (A + \delta E)\mathbf{x}$$

showing that $\lambda_{0\delta} \geq \lambda$ for all such λ . But then $\lambda_{0\delta} \geq \lambda_0$ also. Hence $\lambda_1 \geq \lambda_0$, showing these two numbers are the same. Hence $A\mathbf{x} = \lambda_0\mathbf{x}$. ■

If $A^m \gg 0$ for some m and $A > 0$, it follows that the dimension of the eigenspace for λ_0 is one and that the absolute value of every other eigenvalue of A is less than λ_0 . If it is only assumed that $A > 0$, not necessarily $\gg 0$, this is no longer true. However, there is something which is very interesting which can be said. First here is an interesting lemma.

Lemma B.0.8 Let M be a matrix of the form

$$M = \begin{pmatrix} A & 0 \\ B & C \end{pmatrix}$$

or

$$M = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$$

where A is an $r \times r$ matrix and C is an $(n-r) \times (n-r)$ matrix. Then $\det(M) = \det(A)\det(C)$ and $\sigma(M) = \sigma(A) \cup \sigma(C)$.

Proof: To verify the claim about the determinants, note

$$\begin{pmatrix} A & 0 \\ B & C \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ B & C \end{pmatrix}$$

Therefore,

$$\det \begin{pmatrix} A & 0 \\ B & C \end{pmatrix} = \det \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix} \det \begin{pmatrix} I & 0 \\ B & C \end{pmatrix}.$$

But it is clear from the method of Laplace expansion that

$$\det \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix} = \det A$$

www.job.oticon.dk

oticon
PEOPLE FIRST

and from the multilinear properties of the determinant and row operations that

$$\det \begin{pmatrix} I & 0 \\ B & C \end{pmatrix} = \det \begin{pmatrix} I & 0 \\ 0 & C \end{pmatrix} = \det C.$$

The case where M is upper block triangular is similar.

This immediately implies $\sigma(M) = \sigma(A) \cup \sigma(C)$.

Theorem B.0.9 *Let $A > 0$ and let λ_0 be given in 2.7. If λ is an eigenvalue for A such that $|\lambda| = \lambda_0$, then λ/λ_0 is a root of unity. Thus $(\lambda/\lambda_0)^m = 1$ for some $m \in \mathbb{N}$.*

Proof: Applying Theorem B.0.7 to A^T , there exists $\mathbf{v} > \mathbf{0}$ such that $A^T \mathbf{v} = \lambda_0 \mathbf{v}$. In the first part of the argument it is assumed $\mathbf{v} \gg \mathbf{0}$. Now suppose $A\mathbf{x} = \lambda\mathbf{x}$, $\mathbf{x} \neq \mathbf{0}$ and that $|\lambda| = \lambda_0$. Then

$$A|\mathbf{x}| \geq |\lambda||\mathbf{x}| = \lambda_0|\mathbf{x}|$$

and it follows that if $A|\mathbf{x}| > |\lambda||\mathbf{x}|$, then since $\mathbf{v} \gg \mathbf{0}$,

$$\lambda_0(\mathbf{v}, |\mathbf{x}|) < (\mathbf{v}, A|\mathbf{x}|) = (\mathbf{v}, A^T \mathbf{v}, |\mathbf{x}|) = \lambda_0(\mathbf{v}, |\mathbf{x}|),$$

a contradiction. Therefore,

$$A|\mathbf{x}| = \lambda_0|\mathbf{x}|. \tag{2.8}$$

It follows that

$$\left| \sum_j A_{ij} x_j \right| = \lambda_0 |\mathbf{x}_i| = \sum_j A_{ij} |x_j|$$

and so the complex numbers,

$$A_{ij} x_j, A_{ik} x_k$$

must have the same argument for every k, j because equality holds in the triangle inequality. Therefore, there exists a complex number, μ_i such that

$$A_{ij} x_j = \mu_i A_{ij} |x_j| \tag{2.9}$$

and so, letting $r \in \mathbb{N}$,

$$A_{ij} x_j \mu_j^r = \mu_i A_{ij} |x_j| \mu_j^r.$$

Summing on j yields

$$\sum_j A_{ij} x_j \mu_j^r = \mu_i \sum_j A_{ij} |x_j| \mu_j^r. \tag{2.10}$$

Also, summing 2.9 on j and using that λ is an eigenvalue for \mathbf{x} , it follows from 2.8 that

$$\lambda x_i = \sum_j A_{ij} x_j = \mu_i \sum_j A_{ij} |x_j| = \mu_i \lambda_0 |x_i|. \tag{2.11}$$

From 2.10 and 2.11,

$$\begin{aligned} \sum_j A_{ij} x_j \mu_j^r &= \mu_i \sum_j A_{ij} |x_j| \mu_j^r \\ &= \mu_i \sum_j \overbrace{A_{ij} \mu_j |x_j|}^{\text{see 2.11}} \mu_j^{r-1} \\ &= \mu_i \sum_j A_{ij} \left(\frac{\lambda}{\lambda_0} \right) x_j \mu_j^{r-1} \\ &= \mu_i \left(\frac{\lambda}{\lambda_0} \right) \sum_j A_{ij} x_j \mu_j^{r-1} \end{aligned}$$

Now from 2.10 with r replaced by $r - 1$, this equals

$$\begin{aligned} \mu_i^2 \left(\frac{\lambda}{\lambda_0}\right) \sum_j A_{ij} |x_j| \mu_j^{r-1} &= \mu_i^2 \left(\frac{\lambda}{\lambda_0}\right) \sum_j A_{ij} \mu_j |x_j| \mu_j^{r-2} \\ &= \mu_i^2 \left(\frac{\lambda}{\lambda_0}\right)^2 \sum_j A_{ij} x_j \mu_j^{r-2}. \end{aligned}$$

Continuing this way,

$$\sum_j A_{ij} x_j \mu_j^r = \mu_i^k \left(\frac{\lambda}{\lambda_0}\right)^k \sum_j A_{ij} x_j \mu_j^{r-k}$$

and eventually, this shows

$$\begin{aligned} \sum_j A_{ij} x_j \mu_j^r &= \mu_i^r \left(\frac{\lambda}{\lambda_0}\right)^r \sum_j A_{ij} x_j \\ &= \left(\frac{\lambda}{\lambda_0}\right)^r \lambda(x_i \mu_i^r) \end{aligned}$$

and this says $\left(\frac{\lambda}{\lambda_0}\right)^{r+1}$ is an eigenvalue for $\left(\frac{A}{\lambda_0}\right)$ with the eigenvector being

$$(x_1 \mu_1^r, \dots, x_n \mu_n^r)^T.$$

Now recall that $r \in \mathbb{N}$ was arbitrary and so this has shown that $\left(\frac{\lambda}{\lambda_0}\right)^2, \left(\frac{\lambda}{\lambda_0}\right)^3, \left(\frac{\lambda}{\lambda_0}\right)^4, \dots$ are each eigenvalues of $\left(\frac{A}{\lambda_0}\right)$ which has only finitely many and hence this sequence must repeat. Therefore, $\left(\frac{\lambda}{\lambda_0}\right)$ is a root of unity as claimed. This proves the theorem in the case that $\mathbf{v} \gg \mathbf{0}$.

Now it is necessary to consider the case where $\mathbf{v} > \mathbf{0}$ but it is not the case that $\mathbf{v} \gg \mathbf{0}$. Then in this case, there exists a permutation matrix P such that

$$P\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} \equiv \begin{pmatrix} \mathbf{u} \\ \mathbf{0} \end{pmatrix} \equiv \mathbf{v}_1$$

Then

$$\lambda_0 \mathbf{v} = A^T \mathbf{v} = A^T P \mathbf{v}_1.$$

Therefore,

$$\lambda_0 \mathbf{v}_1 = P A^T P \mathbf{v}_1 = G \mathbf{v}_1$$

Now $P^2 = I$ because it is a permutation matrix. Therefore, the matrix $G \equiv P A^T P$ and A are similar. Consequently, they have the same eigenvalues and it suffices from now on to consider the matrix G rather than A . Then

$$\lambda_0 \begin{pmatrix} \mathbf{u} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} M_1 & M_2 \\ M_3 & M_4 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{0} \end{pmatrix}$$

where M_1 is $r \times r$ and M_4 is $(n - r) \times (n - r)$. It follows from block multiplication and the assumption that A and hence G are > 0 that

$$G = \begin{pmatrix} A' & B \\ 0 & C \end{pmatrix}.$$

Now let λ be an eigenvalue of G such that $|\lambda| = \lambda_0$. Then from Lemma B.0.8, either $\lambda \in \sigma(A')$ or $\lambda \in \sigma(C)$. Suppose without loss of generality that $\lambda \in \sigma(A')$. Since $A' > 0$ it has a largest positive eigenvalue λ'_0 which is obtained from 2.7. Thus $\lambda'_0 \leq \lambda_0$ but λ being an eigenvalue of A' , has its absolute value bounded by λ'_0 and so $\lambda_0 = |\lambda| \leq \lambda'_0 \leq \lambda_0$ showing that $\lambda_0 \in \sigma(A')$. Now if there exists $\mathbf{v} \gg \mathbf{0}$ such that $A'^T \mathbf{v} = \lambda_0 \mathbf{v}$, then the first part of this proof applies to the matrix A and so (λ/λ_0) is a root of unity. If such a vector, \mathbf{v} does not exist, then let A' play the role of A in the above argument and reduce to the consideration of

$$G' \equiv \begin{pmatrix} A'' & B' \\ 0 & C' \end{pmatrix}$$

where G' is similar to A' and $\lambda, \lambda_0 \in \sigma(A'')$. Stop if $A''^T \mathbf{v} = \lambda_0 \mathbf{v}$ for some $\mathbf{v} \gg \mathbf{0}$. Otherwise, decompose A'' similar to the above and add another prime. Continuing this way you must eventually obtain the situation where $(A'^{\dots'})^T \mathbf{v} = \lambda_0 \mathbf{v}$ for some $\mathbf{v} \gg \mathbf{0}$. Indeed, this happens no later than when $A'^{\dots'}$ is a 1×1 matrix. ■

Schlumberger

WHY WAIT FOR PROGRESS?

DARE TO DISCOVER

Discovery means many different things at Schlumberger. But it's the spirit that unites every single one of us. It doesn't matter whether they join our business, engineering or technology teams, our trainees push boundaries, break new ground and deliver the exceptional. If that excites you, then we want to hear from you.

careers.slb.com/recentgraduates



Appendix C

Functions Of Matrices

The existence of the Jordan form also makes it possible to define various functions of matrices. Suppose

$$f(\lambda) = \sum_{n=0}^{\infty} a_n \lambda^n \tag{3.1}$$

for all $|\lambda| < R$. There is a formula for $f(A) \equiv \sum_{n=0}^{\infty} a_n A^n$ which makes sense whenever $\rho(A) < R$. Thus you can speak of $\sin(A)$ or e^A for A an $n \times n$ matrix. To begin with, define

$$f_P(\lambda) \equiv \sum_{n=0}^P a_n \lambda^n$$

so for $k < P$

$$\begin{aligned} f_P^{(k)}(\lambda) &= \sum_{n=k}^P a_n n \cdots (n-k+1) \lambda^{n-k} \\ &= \sum_{n=k}^P a_n \binom{n}{k} k! \lambda^{n-k}. \end{aligned} \tag{3.2}$$

Thus

$$\frac{f_P^{(k)}(\lambda)}{k!} = \sum_{n=k}^P a_n \binom{n}{k} \lambda^{n-k} \tag{3.3}$$

To begin with consider $f(J_m(\lambda))$ where $J_m(\lambda)$ is an $m \times m$ Jordan block. Thus $J_m(\lambda) = D + N$ where $N^m = 0$ and N commutes with D . Therefore, letting $P > m$

$$\begin{aligned} \sum_{n=0}^P a_n J_m(\lambda)^n &= \sum_{n=0}^P a_n \sum_{k=0}^n \binom{n}{k} D^{n-k} N^k \\ &= \sum_{k=0}^P \sum_{n=k}^P a_n \binom{n}{k} D^{n-k} N^k \\ &= \sum_{k=0}^{m-1} N^k \sum_{n=k}^P a_n \binom{n}{k} D^{n-k}. \end{aligned} \tag{3.4}$$

From 3.3 this equals

$$\sum_{k=0}^{m-1} N^k \text{diag} \left(\frac{f_P^{(k)}(\lambda)}{k!}, \dots, \frac{f_P^{(k)}(\lambda)}{k!} \right) \tag{3.5}$$

where for $k = 0, \dots, m-1$, define $\text{diag}_k(a_1, \dots, a_{m-k})$ the $m \times m$ matrix which equals zero everywhere except on the k^{th} super diagonal where this diagonal is filled with the numbers, $\{a_1, \dots, a_{m-k}\}$ from the upper left to the lower right. With no subscript, it is just the diagonal matrices having the indicated entries. Thus in 4×4 matrices, $\text{diag}_2(1, 2)$ would be the matrix

$$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Then from 3.5 and 3.2,

$$\sum_{n=0}^P a_n J_m(\lambda)^n = \sum_{k=0}^{m-1} \text{diag}_k \left(\frac{f_P^{(k)}(\lambda)}{k!}, \dots, \frac{f_P^{(k)}(\lambda)}{k!} \right).$$

Therefore, $\sum_{n=0}^P a_n J_m(\lambda)^n =$

$$\begin{pmatrix} f_P(\lambda) & \frac{f'_P(\lambda)}{1!} & \frac{f_P^{(2)}(\lambda)}{2!} & \dots & \frac{f_P^{(m-1)}(\lambda)}{(m-1)!} \\ & f_P(\lambda) & \frac{f'_P(\lambda)}{1!} & \ddots & \vdots \\ & & f_P(\lambda) & \ddots & \frac{f_P^{(2)}(\lambda)}{2!} \\ & & & \ddots & \frac{f'_P(\lambda)}{1!} \\ 0 & & & & f_P(\lambda) \end{pmatrix} \tag{3.6}$$

Now let A be an $n \times n$ matrix with $\rho(A) < R$ where R is given above. Then the Jordan form of A is of the form

$$J = \begin{pmatrix} J_1 & & 0 \\ & J_2 & \\ & & \ddots \\ 0 & & & J_r \end{pmatrix} \tag{3.7}$$

where $J_k = J_{m_k}(\lambda_k)$ is an $m_k \times m_k$ Jordan block and $A = S^{-1}JS$. Then, letting $P > m_k$ for all k ,

$$\sum_{n=0}^P a_n A^n = S^{-1} \sum_{n=0}^P a_n J^n S,$$

and because of block multiplication of matrices,

$$\sum_{n=0}^P a_n J^n = \begin{pmatrix} \sum_{n=0}^P a_n J_1^n & & 0 \\ & \ddots & \\ & & \ddots \\ 0 & & & \sum_{n=0}^P a_n J_r^n \end{pmatrix}$$

and from 3.6 $\sum_{n=0}^P a_n J_k^n$ converges as $P \rightarrow \infty$ to the $m_k \times m_k$ matrix

$$\begin{pmatrix} f(\lambda_k) & \frac{f'(\lambda_k)}{1!} & \frac{f^{(2)}(\lambda_k)}{2!} & \dots & \frac{f^{(m_k-1)}(\lambda_k)}{(m_k-1)!} \\ 0 & f(\lambda_k) & \frac{f'(\lambda_k)}{1!} & \ddots & \vdots \\ 0 & 0 & f(\lambda_k) & \ddots & \frac{f^{(2)}(\lambda_k)}{2!} \\ \vdots & & \ddots & \ddots & \frac{f'(\lambda_k)}{1!} \\ 0 & 0 & \dots & 0 & f(\lambda_k) \end{pmatrix} \tag{3.8}$$

There is no convergence problem because $|\lambda| < R$ for all $\lambda \in \sigma(A)$. This has proved the following theorem.

Theorem C.0.1 *Let f be given by 3.1 and suppose $\rho(A) < R$ where R is the radius of convergence of the power series in 3.1. Then the series,*

$$\sum_{k=0}^{\infty} a_n A^n \tag{3.9}$$

converges in the space $\mathcal{L}(\mathbb{F}^n, \mathbb{F}^n)$ with respect to any of the norms on this space and furthermore,

$$\sum_{k=0}^{\infty} a_n A^n = S^{-1} \begin{pmatrix} \sum_{n=0}^{\infty} a_n J_1^n & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \sum_{n=0}^{\infty} a_n J_r^n \end{pmatrix} S$$

where $\sum_{n=0}^{\infty} a_n J_k^n$ is an $m_k \times m_k$ matrix of the form given in 3.8 where $A = S^{-1}JS$ and the Jordan form of A , J is given by 3.7. Therefore, you can define $f(A)$ by the series in 3.9.

Here is a simple example.

Example C.0.2 Find $\sin(A)$ where $A = \begin{pmatrix} 4 & 1 & -1 & 1 \\ 1 & 1 & 0 & -1 \\ 0 & -1 & 1 & -1 \\ -1 & 2 & 1 & 4 \end{pmatrix}$.

In this case, the Jordan canonical form of the matrix is not too hard to find.

$$\begin{pmatrix} 4 & 1 & -1 & 1 \\ 1 & 1 & 0 & -1 \\ 0 & -1 & 1 & -1 \\ -1 & 2 & 1 & 4 \end{pmatrix} = \begin{pmatrix} 2 & 0 & -2 & -1 \\ 1 & -4 & -2 & -1 \\ 0 & 0 & -2 & 1 \\ -1 & 4 & 4 & 2 \end{pmatrix} \\ \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{8} & -\frac{3}{8} & 0 & -\frac{1}{8} \\ 0 & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Then from the above theorem $\sin(J)$ is given by

$$\sin \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix} = \begin{pmatrix} \sin 4 & 0 & 0 & 0 \\ 0 & \sin 2 & \cos 2 & -\frac{\sin 2}{2} \\ 0 & 0 & \sin 2 & \cos 2 \\ 0 & 0 & 0 & \sin 2 \end{pmatrix}$$

Therefore, $\sin(A) =$

$$\begin{pmatrix} 2 & 0 & -2 & -1 \\ 1 & -4 & -2 & -1 \\ 0 & 0 & -2 & 1 \\ -1 & 4 & 4 & 2 \end{pmatrix} \begin{pmatrix} \sin 4 & 0 & 0 & 0 \\ 0 & \sin 2 & \cos 2 & -\frac{\sin 2}{2} \\ 0 & 0 & \sin 2 & \cos 2 \\ 0 & 0 & 0 & \sin 2 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{8} & -\frac{3}{8} & 0 & -\frac{1}{8} \\ 0 & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix} = M$$

where the columns of M are as follows from left to right,

$$\begin{pmatrix} \sin 4 \\ \frac{1}{2} \sin 4 - \frac{1}{2} \sin 2 \\ 0 \\ -\frac{1}{2} \sin 4 + \frac{1}{2} \sin 2 \end{pmatrix}, \begin{pmatrix} \sin 4 - \sin 2 - \cos 2 \\ \frac{1}{2} \sin 4 + \frac{3}{2} \sin 2 - 2 \cos 2 \\ -\cos 2 \\ -\frac{1}{2} \sin 4 - \frac{1}{2} \sin 2 + 3 \cos 2 \end{pmatrix}, \begin{pmatrix} -\cos 2 \\ \sin 2 \\ \sin 2 - \cos 2 \\ \cos 2 - \sin 2 \end{pmatrix} \\ \begin{pmatrix} \sin 4 - \sin 2 - \cos 2 \\ \frac{1}{2} \sin 4 + \frac{1}{2} \sin 2 - 2 \cos 2 \\ -\cos 2 \\ -\frac{1}{2} \sin 4 + \frac{1}{2} \sin 2 + 3 \cos 2 \end{pmatrix}$$

Perhaps this isn't the first thing you would think of. Of course the ability to get this nice closed form description of $\sin(A)$ was dependent on being able to find the Jordan form along with a similarity transformation which will yield the Jordan form.

The following corollary is known as the spectral mapping theorem.

Corollary C.0.3 *Let A be an $n \times n$ matrix and let $\rho(A) < R$ where for $|\lambda| < R$,*

$$f(\lambda) = \sum_{n=0}^{\infty} a_n \lambda^n.$$

Then $f(A)$ is also an $n \times n$ matrix and furthermore, $\sigma(f(A)) = f(\sigma(A))$. Thus the eigenvalues of $f(A)$ are exactly the numbers $f(\lambda)$ where λ is an eigenvalue of A . Furthermore, the algebraic multiplicity of $f(\lambda)$ coincides with the algebraic multiplicity of λ .

All of these things can be generalized to linear transformations defined on infinite dimensional spaces and when this is done the main tool is the Dunford integral along with the methods of complex analysis. It is good to see it done for finite dimensional situations first because it gives an idea of what is possible. Actually, some of the most interesting functions in applications do not come in the above form as a power series expanded about 0. One example of this situation has already been encountered in the proof of the right polar decomposition with the square root of an Hermitian transformation which had all nonnegative eigenvalues. Another example is that of taking the positive part of an Hermitian matrix. This is important in some physical models where something may depend on the positive part of the strain which is a symmetric real matrix. Obviously there is no way to consider this as a power series expanded about 0 because the function $f(r) = r^+$ is not even differentiable at 0. Therefore, a totally different approach must be considered. First the notion of a positive part is defined.



PREPARE FOR A LEADING ROLE.

English-taught MSc programmes in engineering: Aeronautical, Biomedical, Electronics, Mechanical, Communication systems and Transport systems. No tuition fees.

→ liu.se/master

li.u LINKÖPING UNIVERSITY

Definition C.0.4 Let A be an Hermitian matrix. Thus it suffices to consider A as an element of $\mathcal{L}(\mathbb{F}^n, \mathbb{F}^n)$ according to the usual notion of matrix multiplication. Then there exists an orthonormal basis of eigenvectors, $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ such that

$$A = \sum_{j=1}^n \lambda_j \mathbf{u}_j \otimes \mathbf{u}_j,$$

for λ_j the eigenvalues of A , all real. Define

$$A^+ \equiv \sum_{j=1}^n \lambda_j^+ \mathbf{u}_j \otimes \mathbf{u}_j$$

where $\lambda^+ \equiv \frac{|\lambda| + \lambda}{2}$.

This gives us a nice definition of what is meant but it turns out to be very important in the applications to determine how this function depends on the choice of symmetric matrix A . The following addresses this question.

Theorem C.0.5 If A, B be Hermitian matrices, then for $|\cdot|$ the Frobenius norm,

$$|A^+ - B^+| \leq |A - B|.$$

Proof: Let $A = \sum_i \lambda_i \mathbf{v}_i \otimes \mathbf{v}_i$ and let $B = \sum_j \mu_j \mathbf{w}_j \otimes \mathbf{w}_j$ where $\{\mathbf{v}_i\}$ and $\{\mathbf{w}_j\}$ are orthonormal bases of eigenvectors.

$$\begin{aligned} |A^+ - B^+|^2 &= \text{trace} \left(\sum_i \lambda_i^+ \mathbf{v}_i \otimes \mathbf{v}_i - \sum_j \mu_j^+ \mathbf{w}_j \otimes \mathbf{w}_j \right)^2 = \\ &\text{trace} \left[\sum_i (\lambda_i^+)^2 \mathbf{v}_i \otimes \mathbf{v}_i + \sum_j (\mu_j^+)^2 \mathbf{w}_j \otimes \mathbf{w}_j \right. \\ &\left. - \sum_{i,j} \lambda_i^+ \mu_j^+ (\mathbf{w}_j, \mathbf{v}_i) \mathbf{v}_i \otimes \mathbf{w}_j - \sum_{i,j} \lambda_i^+ \mu_j^+ (\mathbf{v}_i, \mathbf{w}_j) \mathbf{w}_j \otimes \mathbf{v}_i \right] \end{aligned}$$

Since the trace of $\mathbf{v}_i \otimes \mathbf{w}_j$ is $(\mathbf{v}_i, \mathbf{w}_j)$, a fact which follows from $(\mathbf{v}_i, \mathbf{w}_j)$ being the only possibly nonzero eigenvalue,

$$= \sum_i (\lambda_i^+)^2 + \sum_j (\mu_j^+)^2 - 2 \sum_{i,j} \lambda_i^+ \mu_j^+ |(\mathbf{v}_i, \mathbf{w}_j)|^2. \tag{3.10}$$

Since these are orthonormal bases,

$$\sum_i |(\mathbf{v}_i, \mathbf{w}_j)|^2 = 1 = \sum_j |(\mathbf{v}_i, \mathbf{w}_j)|^2$$

and so 3.10 equals

$$= \sum_i \sum_j \left((\lambda_i^+)^2 + (\mu_j^+)^2 - 2\lambda_i^+ \mu_j^+ \right) |(\mathbf{v}_i, \mathbf{w}_j)|^2.$$

Similarly,

$$|A - B|^2 = \sum_i \sum_j \left((\lambda_i)^2 + (\mu_j)^2 - 2\lambda_i \mu_j \right) |(\mathbf{v}_i, \mathbf{w}_j)|^2.$$

Now it is easy to check that $(\lambda_i)^2 + (\mu_j)^2 - 2\lambda_i \mu_j \geq (\lambda_i^+)^2 + (\mu_j^+)^2 - 2\lambda_i^+ \mu_j^+$. ■

Appendix D

Differential Equations

D.1 Theory Of Ordinary Differential Equations

First are some analytical preliminaries which are easy consequences of some of the considerations presented earlier. These things have to do with existence and uniqueness of the initial value problem

$$\mathbf{x}' = \mathbf{f}(t, \mathbf{x}), \mathbf{x}(c) = \mathbf{x}_0$$

Suppose that $\mathbf{f} : [a, b] \times \mathbb{F}^n \rightarrow \mathbb{F}^n$ satisfies the following two conditions.

$$|\mathbf{f}(t, \mathbf{x}) - \mathbf{f}(t, \mathbf{x}_1)| \leq K |\mathbf{x} - \mathbf{x}_1|, \tag{4.1}$$

$$\mathbf{f} \text{ is continuous.} \tag{4.2}$$

The first of these conditions is known as a Lipschitz condition.

Lemma D.1.1 *Suppose $\mathbf{x} : [a, b] \rightarrow \mathbb{F}^n$ is a continuous function and $c \in [a, b]$. Then \mathbf{x} is a solution to the initial value problem,*

$$\mathbf{x}' = \mathbf{f}(t, \mathbf{x}), \mathbf{x}(c) = \mathbf{x}_0 \tag{4.3}$$

if and only if \mathbf{x} is a solution to the integral equation,

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_c^t \mathbf{f}(s, \mathbf{x}(s)) ds. \tag{4.4}$$

Proof: If \mathbf{x} solves 4.4, then since \mathbf{f} is continuous, we may apply the fundamental theorem of calculus to differentiate both sides and obtain $\mathbf{x}'(t) = \mathbf{f}(t, \mathbf{x}(t))$. Also, letting $t = c$ on both sides, gives $\mathbf{x}(c) = \mathbf{x}_0$. Conversely, if \mathbf{x} is a solution of the initial value problem, we may integrate both sides from c to t to see that \mathbf{x} solves 4.4. ■

Theorem D.1.2 *Let \mathbf{f} satisfy 4.1 and 4.2. Then there exists a unique solution to the initial value problem, 4.3 on the interval $[a, b]$.*

Proof: Let $\|\mathbf{x}\|_\lambda \equiv \sup \{e^{\lambda t} |\mathbf{x}(t)| : t \in [a, b]\}$. Then this norm is equivalent to the usual norm on $BC([a, b], \mathbb{F}^n)$ described in Example 13.6.2. This means that for $\|\cdot\|$ the norm given there, there exist constants δ and Δ such that

$$\|\mathbf{x}\|_\lambda \delta \leq \|\mathbf{x}\| \leq \Delta \|\mathbf{x}\|_\lambda$$

for all $\mathbf{x} \in BC([a, b], \mathbb{F}^n)$. In fact, you can take $\delta \equiv e^{\lambda a}$ and $\Delta \equiv e^{\lambda b}$ in case $\lambda > 0$ with the two reversed in case $\lambda < 0$. Thus $BC([a, b], \mathbb{F}^n)$ is a Banach space with this norm, $\|\cdot\|_\lambda$. Then let $F : BC([a, b], \mathbb{F}^n) \rightarrow BC([a, b], \mathbb{F}^n)$ be defined by

$$F\mathbf{x}(t) \equiv \mathbf{x}_0 + \int_c^t \mathbf{f}(s, \mathbf{x}(s)) ds.$$

Let $\lambda < 0$. It follows

$$\begin{aligned} e^{\lambda t} |F\mathbf{x}(t) - F\mathbf{y}(t)| &\leq \left| e^{\lambda t} \int_c^t |\mathbf{f}(s, \mathbf{x}(s)) - \mathbf{f}(s, \mathbf{y}(s))| ds \right| \\ &\leq \left| \int_c^t K e^{\lambda(t-s)} |\mathbf{x}(s) - \mathbf{y}(s)| e^{\lambda s} ds \right| \end{aligned}$$

If $t \geq c$, this is no larger than

$$\begin{aligned} \|\mathbf{x} - \mathbf{y}\|_\lambda \int_c^t K e^{\lambda(t-s)} ds &\leq \|\mathbf{x} - \mathbf{y}\|_\lambda \frac{1}{|\lambda|} \\ &\leq \|\mathbf{x} - \mathbf{y}\|_\lambda \int_a^t K e^{\lambda(t-s)} ds \leq \|\mathbf{x} - \mathbf{y}\|_\lambda \frac{K}{|\lambda|} \end{aligned}$$

If $t < c$, this equals

$$\int_t^c K e^{\lambda(t-s)} |\mathbf{x}(s) - \mathbf{y}(s)| e^{\lambda s} ds$$

which is no larger than

$$\|\mathbf{x} - \mathbf{y}\|_\lambda e^{|\lambda|(b-a)} \frac{K}{|\lambda|}$$

Therefore, it is always the case that

$$\|F\mathbf{x} - F\mathbf{y}\|_\lambda \leq \|\mathbf{x} - \mathbf{y}\|_\lambda e^{|\lambda|(b-a)} \frac{K}{|\lambda|}.$$

If $|\lambda|$ is chosen larger than $Ke^{|\lambda|(b-a)}$, this implies F is a contraction mapping on $BC([a, b], \mathbb{F}^n)$. Therefore, there exists a unique fixed point. With Lemma D.1.1 this proves the theorem. ■

D.2 Linear Systems

As an example of the above theorem, consider for $t \in [a, b]$ the system

$$\mathbf{x}' = A(t)\mathbf{x}(t) + \mathbf{g}(t), \quad \mathbf{x}(c) = \mathbf{x}_0 \tag{4.5}$$

where $A(t)$ is an $n \times n$ matrix whose entries are continuous functions of t , $(a_{ij}(t))$ and $\mathbf{g}(t)$ is a vector whose components are continuous functions of t satisfies the conditions of Theorem D.1.2 with $\mathbf{f}(t, \mathbf{x}) = A(t)\mathbf{x} + \mathbf{g}(t)$. To see this, let $\mathbf{x} = (x_1, \dots, x_n)^T$ and $\mathbf{x}_1 = (x_{11}, \dots, x_{1n})^T$. Then letting $M = \max\{|a_{ij}(t)| : t \in [a, b], i, j \leq n\}$,

$$\begin{aligned} |\mathbf{f}(t, \mathbf{x}) - \mathbf{f}(t, \mathbf{x}_1)| &= |A(t)(\mathbf{x} - \mathbf{x}_1)| \\ &= \left| \left(\sum_{i=1}^n \left| \sum_{j=1}^n a_{ij}(t)(x_j - x_{1j}) \right|^2 \right)^{1/2} \right| \leq M \left| \left(\sum_{i=1}^n \left(\sum_{j=1}^n |x_j - x_{1j}| \right)^2 \right)^{1/2} \right| \\ &\leq M \left| \left(\sum_{i=1}^n n \sum_{j=1}^n |x_j - x_{1j}|^2 \right)^{1/2} \right| = Mn \left(\sum_{j=1}^n |x_j - x_{1j}|^2 \right)^{1/2} = Mn |\mathbf{x} - \mathbf{x}_1|. \end{aligned}$$

Therefore, let $K = Mn$. This proves

Theorem D.2.1 *Let $A(t)$ be a continuous $n \times n$ matrix and let $\mathbf{g}(t)$ be a continuous vector for $t \in [a, b]$ and let $c \in [a, b]$ and $\mathbf{x}_0 \in \mathbb{F}^n$. Then there exists a unique solution to 4.5 valid for $t \in [a, b]$.*

This includes more examples of linear equations than are typically encountered in an entire differential equations course.

D.3 Local Solutions

Lemma D.3.1 *Let $D(\mathbf{x}_0, r) \equiv \{\mathbf{x} \in \mathbb{F}^n : |\mathbf{x} - \mathbf{x}_0| \leq r\}$ and suppose U is an open set containing $D(\mathbf{x}_0, r)$ such that $\mathbf{f} : U \rightarrow \mathbb{F}^n$ is $C^1(U)$. (Recall this means all partial derivatives of \mathbf{f} exist and are continuous.) Then for $K = Mn$, where M denotes the maximum of $\left| \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{z}) \right|$ for $\mathbf{z} \in D(\mathbf{x}_0, r)$, it follows that for all $\mathbf{x}, \mathbf{y} \in D(\mathbf{x}_0, r)$,*

$$|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| \leq K |\mathbf{x} - \mathbf{y}|.$$

Proof: Let $\mathbf{x}, \mathbf{y} \in D(\mathbf{x}_0, r)$ and consider the line segment joining these two points, $\mathbf{x} + t(\mathbf{y} - \mathbf{x})$ for $t \in [0, 1]$. Letting $\mathbf{h}(t) = \mathbf{f}(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))$ for $t \in [0, 1]$, then

$$\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x}) = \mathbf{h}(1) - \mathbf{h}(0) = \int_0^1 \mathbf{h}'(t) dt.$$

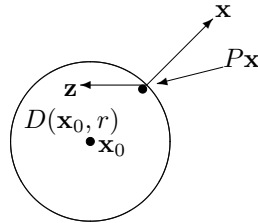
Also, by the chain rule,

$$\mathbf{h}'(t) = \sum_{i=1}^n \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) (y_i - x_i).$$

Therefore,

$$\begin{aligned} |\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x})| &= \\ &\left| \int_0^1 \sum_{i=1}^n \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) (y_i - x_i) dt \right| \\ &\leq \int_0^1 \sum_{i=1}^n \left| \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) \right| |y_i - x_i| dt \\ &\leq M \sum_{i=1}^n |y_i - x_i| \leq Mn |\mathbf{x} - \mathbf{y}|. \blacksquare \end{aligned}$$

Now consider the map, P which maps all of \mathbb{R}^n to $D(\mathbf{x}_0, r)$ given as follows. For $\mathbf{x} \in D(\mathbf{x}_0, r)$, $P\mathbf{x} = \mathbf{x}$. For $\mathbf{x} \notin D(\mathbf{x}_0, r)$, $P\mathbf{x}$ will be the closest point in $D(\mathbf{x}_0, r)$ to \mathbf{x} . Such a closest point exists because $D(\mathbf{x}_0, r)$ is a closed and bounded set. Taking $f(\mathbf{y}) \equiv |\mathbf{y} - \mathbf{x}|$, it follows f is a continuous function defined on $D(\mathbf{x}_0, r)$ which must achieve its minimum value by the extreme value theorem from calculus.



Lemma D.3.2 For any pair of points, $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$, $|P\mathbf{x} - P\mathbf{y}| \leq |\mathbf{x} - \mathbf{y}|$.

Proof: The above picture suggests the geometry of what is going on. Letting $\mathbf{z} \in D(\mathbf{x}_0, r)$, it follows that for all $t \in [0, 1]$,

$$\begin{aligned} |\mathbf{x} - P\mathbf{x}|^2 &\leq |\mathbf{x} - (P\mathbf{x} + t(\mathbf{z} - P\mathbf{x}))|^2 \\ &= |\mathbf{x} - P\mathbf{x}|^2 + 2t \operatorname{Re}((\mathbf{x} - P\mathbf{x}) \cdot (P\mathbf{x} - \mathbf{z})) + t^2 |\mathbf{z} - P\mathbf{x}|^2 \end{aligned}$$

Hence

$$2t \operatorname{Re}((\mathbf{x} - P\mathbf{x}) \cdot (P\mathbf{x} - \mathbf{z})) + t^2 |\mathbf{z} - P\mathbf{x}|^2 \geq 0$$

and this can only happen if

$$\operatorname{Re}((\mathbf{x} - P\mathbf{x}) \cdot (P\mathbf{x} - \mathbf{z})) \geq 0.$$

Therefore,

$$\begin{aligned} \operatorname{Re}((\mathbf{x} - P\mathbf{x}) \cdot (P\mathbf{x} - P\mathbf{y})) &\geq 0 \\ \operatorname{Re}((\mathbf{y} - P\mathbf{y}) \cdot (P\mathbf{y} - P\mathbf{x})) &\geq 0 \end{aligned}$$

and so

$$\operatorname{Re}(\mathbf{x} - P\mathbf{x} - (\mathbf{y} - P\mathbf{y})) \cdot (P\mathbf{x} - P\mathbf{y}) \geq 0$$

which implies

$$\operatorname{Re}(\mathbf{x} - \mathbf{y}) \cdot (P\mathbf{x} - P\mathbf{y}) \geq |P\mathbf{x} - P\mathbf{y}|^2$$

Then using the Cauchy Schwarz inequality it follows

$$|\mathbf{x} - \mathbf{y}| \geq |P\mathbf{x} - P\mathbf{y}|.$$

■

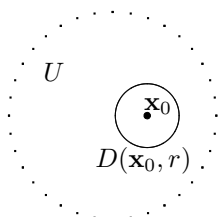
With this here is the local existence and uniqueness theorem.

Theorem D.3.3 Let $[a, b]$ be a closed interval and let U be an open subset of \mathbb{F}^n . Let $\mathbf{f} : [a, b] \times U \rightarrow \mathbb{F}^n$ be continuous and suppose that for each $t \in [a, b]$, the map $\mathbf{x} \rightarrow \frac{\partial \mathbf{f}}{\partial x_i}(t, \mathbf{x})$ is continuous. Also let $\mathbf{x}_0 \in U$ and $c \in [a, b]$. Then there exists an interval, $I \subseteq [a, b]$ such that $c \in I$ and there exists a unique solution to the initial value problem,

$$\mathbf{x}' = \mathbf{f}(t, \mathbf{x}), \mathbf{x}(c) = \mathbf{x}_0 \tag{4.6}$$

valid for $t \in I$.

Proof: Consider the following picture.



The large dotted circle represents U and the little solid circle represents $D(\mathbf{x}_0, r)$ as indicated. Here r is so small that $D(\mathbf{x}_0, r)$ is contained in U as shown. Now let P denote the projection map defined above. Consider the initial value problem

$$\mathbf{x}' = \mathbf{f}(t, P\mathbf{x}), \quad \mathbf{x}(c) = \mathbf{x}_0. \tag{4.7}$$

From Lemma D.3.1 and the continuity of $\mathbf{x} \rightarrow \frac{\partial \mathbf{f}}{\partial x_i}(t, \mathbf{x})$, there exists a constant, K such that if $\mathbf{x}, \mathbf{y} \in D(\mathbf{x}_0, r)$, then $|\mathbf{f}(t, \mathbf{x}) - \mathbf{f}(t, \mathbf{y})| \leq K|\mathbf{x} - \mathbf{y}|$ for all $t \in [a, b]$. Therefore, by Lemma D.3.2

$$|\mathbf{f}(t, P\mathbf{x}) - \mathbf{f}(t, P\mathbf{y})| \leq K|P\mathbf{x} - P\mathbf{y}| \leq K|\mathbf{x} - \mathbf{y}|.$$

It follows from Theorem D.1.2 that 4.7 has a unique solution valid for $t \in [a, b]$. Since \mathbf{x} is continuous, it follows that there exists an interval, I containing c such that for $t \in I$, $\mathbf{x}(t) \in D(\mathbf{x}_0, r)$. Therefore, for these values of t , $\mathbf{f}(t, P\mathbf{x}) = \mathbf{f}(t, \mathbf{x})$ and so there is a unique solution to 4.6 on I . ■

Now suppose \mathbf{f} has the property that for every $R > 0$ there exists a constant, K_R such that for all $\mathbf{x}, \mathbf{x}_1 \in B(0, R)$,

$$|\mathbf{f}(t, \mathbf{x}) - \mathbf{f}(t, \mathbf{x}_1)| \leq K_R|\mathbf{x} - \mathbf{x}_1|. \tag{4.8}$$



**WE ARE SHAPING
MOBILITY FOR
TOMORROW**

How will people travel in the future, and how will goods be transported? What resources will we use, and how many will we need? The passenger and freight traffic sector is developing rapidly, and we provide the impetus for innovation and movement. We develop components and systems for internal combustion engines that operate more cleanly and more efficiently than ever before. We are also pushing forward technologies that are bringing hybrid vehicles and alternative drives into a new dimension – for private, corporate, and public use. The challenges are great. We deliver the solutions and offer challenging jobs.

www.schaeffler.com/careers

SCHAEFFLER



Corollary D.3.4 Let \mathbf{f} satisfy 4.8 and suppose also that $(t, \mathbf{x}) \rightarrow \mathbf{f}(t, \mathbf{x})$ is continuous. Suppose now that \mathbf{x}_0 is given and there exists an estimate of the form $|\mathbf{x}(t)| < R$ for all $t \in [0, T)$ where $T \leq \infty$ on the local solution to

$$\mathbf{x}' = \mathbf{f}(t, \mathbf{x}), \quad \mathbf{x}(0) = \mathbf{x}_0. \tag{4.9}$$

Then there exists a unique solution to the initial value problem, 4.9 valid on $[0, T)$.

Proof: Replace $\mathbf{f}(t, \mathbf{x})$ with $\mathbf{f}(t, P\mathbf{x})$ where P is the projection onto $\overline{B(0, R)}$. Then by Theorem D.1.2 there exists a unique solution to the system

$$\mathbf{x}' = \mathbf{f}(t, P\mathbf{x}), \quad \mathbf{x}(0) = \mathbf{x}_0$$

valid on $[0, T_1]$ for every $T_1 < T$. Therefore, the above system has a unique solution on $[0, T)$ and from the estimate, $P\mathbf{x} = \mathbf{x}$. ■

D.4 First Order Linear Systems

Here is a discussion of linear systems of the form

$$\mathbf{x}' = A\mathbf{x} + \mathbf{f}(t)$$

where A is a constant $n \times n$ matrix and \mathbf{f} is a vector valued function having all entries continuous. Of course the existence theory is a very special case of the general considerations above but I will give a self contained presentation based on elementary first order scalar differential equations and linear algebra.

Definition D.4.1 Suppose $t \rightarrow M(t)$ is a matrix valued function of t . Thus $M(t) = (m_{ij}(t))$. Then define

$$M'(t) \equiv (m'_{ij}(t)).$$

In words, the derivative of $M(t)$ is the matrix whose entries consist of the derivatives of the entries of $M(t)$. Integrals of matrices are defined the same way. Thus

$$\int_a^b M(t) dt \equiv \left(\int_a^b m_{ij}(t) dt \right).$$

In words, the integral of $M(t)$ is the matrix obtained by replacing each entry of $M(t)$ by the integral of that entry.

With this definition, it is easy to prove the following theorem.

Theorem D.4.2 Suppose $M(t)$ and $N(t)$ are matrices for which $M(t)N(t)$ makes sense. Then if $M'(t)$ and $N'(t)$ both exist, it follows that

$$(M(t)N(t))' = M'(t)N(t) + M(t)N'(t).$$

Proof:

$$\begin{aligned} ((M(t)N(t))')_{ij} &\equiv \left((M(t)N(t))_{ij} \right)' = \left(\sum_k M(t)_{ik} N(t)_{kj} \right)' \\ &= \sum_k (M(t)_{ik})' N(t)_{kj} + M(t)_{ik} (N(t)_{kj})' \\ &\equiv \sum_k (M'(t)_{ik}) N(t)_{kj} + M(t)_{ik} (N'(t)_{kj}) \\ &\equiv (M'(t)N(t) + M(t)N'(t))_{ij} \quad \blacksquare \end{aligned}$$

In the study of differential equations, one of the most important theorems is Gronwall's inequality which is next.

Theorem D.4.3 Suppose $u(t) \geq 0$ and for all $t \in [0, T]$,

$$u(t) \leq u_0 + \int_0^t Ku(s) ds. \tag{4.10}$$

where K is some nonnegative constant. Then

$$u(t) \leq u_0 e^{Kt}. \tag{4.11}$$

Proof: Let $w(t) = \int_0^t u(s) ds$. Then using the fundamental theorem of calculus, 4.10 $w(t)$ satisfies the following.

$$u(t) - Kw(t) = w'(t) - Kw(t) \leq u_0, \quad w(0) = 0. \tag{4.12}$$

Multiply both sides of this inequality by e^{-Kt} and using the product rule and the chain rule,

$$e^{-Kt}(w'(t) - Kw(t)) = \frac{d}{dt}(e^{-Kt}w(t)) \leq u_0 e^{-Kt}.$$

Integrating this from 0 to t ,

$$e^{-Kt}w(t) \leq u_0 \int_0^t e^{-Ks} ds = u_0 \left(-\frac{e^{-tK} - 1}{K} \right).$$

Now multiply through by e^{Kt} to obtain

$$w(t) \leq u_0 \left(-\frac{e^{-tK} - 1}{K} \right) e^{Kt} = -\frac{u_0}{K} + \frac{u_0}{K} e^{tK}.$$

Therefore, 4.12 implies

$$u(t) \leq u_0 + K \left(-\frac{u_0}{K} + \frac{u_0}{K} e^{tK} \right) = u_0 e^{Kt}.$$

■

With Gronwall's inequality, here is a theorem on uniqueness of solutions to the initial value problem,

$$\mathbf{x}' = A\mathbf{x} + \mathbf{f}(t), \quad \mathbf{x}(a) = \mathbf{x}_a, \tag{4.13}$$

in which A is an $n \times n$ matrix and \mathbf{f} is a continuous function having values in \mathbb{C}^n .

Theorem D.4.4 Suppose \mathbf{x} and \mathbf{y} satisfy 4.13. Then $\mathbf{x}(t) = \mathbf{y}(t)$ for all t .

Proof: Let $\mathbf{z}(t) = \mathbf{x}(t+a) - \mathbf{y}(t+a)$. Then for $t \geq 0$,

$$\mathbf{z}' = A\mathbf{z}, \quad \mathbf{z}(0) = \mathbf{0}. \tag{4.14}$$

Note that for $K = \max\{|a_{ij}|\}$, where $A = (a_{ij})$,

$$|(A\mathbf{z}, \mathbf{z})| = \left| \sum_{ij} a_{ij} z_j \bar{z}_i \right| \leq K \sum_{ij} |z_i| |z_j| \leq K \sum_{ij} \left(\frac{|z_i|^2}{2} + \frac{|z_j|^2}{2} \right) = nK |\mathbf{z}|^2.$$

(For x and y real numbers, $xy \leq \frac{x^2}{2} + \frac{y^2}{2}$ because this is equivalent to saying $(x - y)^2 \geq 0$.) Similarly, $|(z, Az)| \leq nK |\mathbf{z}|^2$. Thus,

$$|(z, Az)|, |(Az, z)| \leq nK |\mathbf{z}|^2. \tag{4.15}$$

Now multiplying 4.14 by \mathbf{z} and observing that

$$\frac{d}{dt} (|\mathbf{z}|^2) = (\mathbf{z}', \mathbf{z}) + (\mathbf{z}, \mathbf{z}') = (A\mathbf{z}, \mathbf{z}) + (\mathbf{z}, A\mathbf{z}),$$

it follows from 4.15 and the observation that $\mathbf{z}(0) = 0$,

$$|\mathbf{z}(t)|^2 \leq \int_0^t 2nK |\mathbf{z}(s)|^2 ds$$

and so by Gronwall's inequality, $|\mathbf{z}(t)|^2 = 0$ for all $t \geq 0$. Thus,

$$\mathbf{x}(t) = \mathbf{y}(t)$$

for all $t \geq a$.

Now let $\mathbf{w}(t) = \mathbf{x}(a-t) - \mathbf{y}(a-t)$ for $t \geq 0$. Then $\mathbf{w}'(t) = (-A)\mathbf{w}(t)$ and you can repeat the argument which was just given to conclude that $\mathbf{x}(t) = \mathbf{y}(t)$ for all $t \leq a$. ■

Definition D.4.5 Let A be an $n \times n$ matrix. We say $\Phi(t)$ is a fundamental matrix for A if

$$\Phi'(t) = A\Phi(t), \quad \Phi(0) = I, \tag{4.16}$$

and $\Phi(t)^{-1}$ exists for all $t \in \mathbb{R}$.

Why should anyone care about a fundamental matrix? The reason is that such a matrix valued function makes possible a convenient description of the solution of the initial value problem,

$$\mathbf{x}' = A\mathbf{x} + \mathbf{f}(t), \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{4.17}$$

on the interval, $[0, T]$. First consider the special case where $n = 1$. This is the first order linear differential equation,

$$r' = \lambda r + g, \quad r(0) = r_0, \tag{4.18}$$

where g is a continuous scalar valued function. First consider the case where $g = 0$.

**STUDY FOR YOUR MASTER'S DEGREE
IN THE CRADLE OF SWEDISH ENGINEERING**

Chalmers University of Technology conducts research and education in engineering and natural sciences, architecture, technology-related mathematical sciences and nautical sciences. Behind all that Chalmers accomplishes, the aim persists for contributing to a sustainable future – both nationally and globally.

Visit us on **Chalmers.se** or **Next Stop Chalmers** on facebook.

CHALMERS
UNIVERSITY OF TECHNOLOGY



Lemma D.4.6 *There exists a unique solution to the initial value problem,*

$$r' = \lambda r, r(0) = 1, \tag{4.19}$$

and the solution for $\lambda = a + ib$ is given by

$$r(t) = e^{at} (\cos bt + i \sin bt). \tag{4.20}$$

This solution to the initial value problem is denoted as $e^{\lambda t}$. (If λ is real, $e^{\lambda t}$ as defined here reduces to the usual exponential function so there is no contradiction between this and earlier notation seen in calculus.)

Proof: From the uniqueness theorem presented above, Theorem D.4.4, applied to the case where $n = 1$, there can be no more than one solution to the initial value problem, 4.19. Therefore, it only remains to verify 4.20 is a solution to 4.19. However, this is an easy calculus exercise. ■

Note the differential equation in 4.19 says

$$\frac{d}{dt} (e^{\lambda t}) = \lambda e^{\lambda t}. \tag{4.21}$$

With this lemma, it becomes possible to easily solve the case in which $g \neq 0$.

Theorem D.4.7 *There exists a unique solution to 4.18 and this solution is given by the formula,*

$$r(t) = e^{\lambda t} r_0 + e^{\lambda t} \int_0^t e^{-\lambda s} g(s) ds. \tag{4.22}$$

Proof: By the uniqueness theorem, Theorem D.4.4, there is no more than one solution. It only remains to verify that 4.22 is a solution. But $r(0) = e^{\lambda 0} r_0 + \int_0^0 e^{-\lambda s} g(s) ds = r_0$ and so the initial condition is satisfied. Next differentiate this expression to verify the differential equation is also satisfied. Using 4.21, the product rule and the fundamental theorem of calculus,

$$r'(t) = \lambda e^{\lambda t} r_0 + \lambda e^{\lambda t} \int_0^t e^{-\lambda s} g(s) ds + e^{\lambda t} e^{-\lambda t} g(t) = \lambda r(t) + g(t). \blacksquare$$

Now consider the question of finding a fundamental matrix for A . When this is done, it will be easy to give a formula for the general solution to 4.17 known as the variation of constants formula, arguably the most important result in differential equations.

The next theorem gives a formula for the fundamental matrix 4.16. It is known as Putzer's method [1],[22].

Theorem D.4.8 *Let A be an $n \times n$ matrix whose eigenvalues are $\{\lambda_1, \dots, \lambda_n\}$ listed according to multiplicity as roots of the characteristic equation. Define*

$$P_k(A) \equiv \prod_{m=1}^k (A - \lambda_m I), P_0(A) \equiv I,$$

and let the scalar valued functions, $r_k(t)$ be defined as the solutions to the following initial value problem

$$\begin{pmatrix} r'_0(t) \\ r'_1(t) \\ r'_2(t) \\ \vdots \\ r'_n(t) \end{pmatrix} = \begin{pmatrix} 0 \\ \lambda_1 r_1(t) + r_0(t) \\ \lambda_2 r_2(t) + r_1(t) \\ \vdots \\ \lambda_n r_n(t) + r_{n-1}(t) \end{pmatrix}, \begin{pmatrix} r_0(0) \\ r_1(0) \\ r_2(0) \\ \vdots \\ r_n(0) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Note the system amounts to a list of single first order linear differential equations. Now define

$$\Phi(t) \equiv \sum_{k=0}^{n-1} r_{k+1}(t) P_k(A).$$

Then

$$\Phi'(t) = A\Phi(t), \quad \Phi(0) = I. \tag{4.23}$$

Furthermore, if $\Phi(t)$ is a solution to 4.23 for all t , then it follows $\Phi(t)^{-1}$ exists for all t and $\Phi(t)$ is the unique fundamental matrix for A .

Proof: The first part of this follows from a computation. First note that by the Cayley Hamilton theorem, $P_n(A) = 0$ and $r_0(t) = 0$. Also from the formula, if we define $\prod_{m=1}^0 (A - \lambda_m I) \equiv I$ to correspond to the above definition, for all $k \geq 1$,

$$P_k(A) = (A - \lambda_k I) \prod_{m=1}^{k-1} (A - \lambda_m I) = (A - \lambda_k I) P_{k-1}(A)$$

Now for the computation:

$$\begin{aligned} \Phi'(t) &= \sum_{k=0}^{n-1} r'_{k+1}(t) P_k(A) = \sum_{k=0}^{n-1} (\lambda_{k+1} r_{k+1}(t) + r_k(t)) P_k(A) = \\ &= \sum_{k=0}^{n-1} \lambda_{k+1} r_{k+1}(t) P_k(A) + \sum_{k=1}^n r_k(t) P_k(A) = \sum_{k=0}^{n-1} \lambda_{k+1} r_{k+1}(t) P_k(A) + \sum_{k=0}^{n-1} r_{k+1}(t) P_{k+1}(A) \\ &= \sum_{k=0}^{n-1} \lambda_{k+1} r_{k+1}(t) P_k(A) + \sum_{k=0}^{n-1} r_{k+1}(t) (A - \lambda_{k+1} I) P_k(A) = A \sum_{k=0}^{n-1} r_{k+1}(t) P_k(A) = A\Phi(t) \end{aligned}$$

That $\Phi(0) = I$ follows from

$$\Phi(0) = \sum_{k=0}^{n-1} r_{k+1}(0) P_k(A) = r_1(0) P_0(A) = I.$$

It remains to verify that if 4.23 holds, then $\Phi(t)^{-1}$ exists for all t . To do so, consider $\mathbf{v} \neq \mathbf{0}$ and suppose for some t_0 , $\Phi(t_0) \mathbf{v} = \mathbf{0}$. Let $\mathbf{x}(t) \equiv \Phi(t_0 + t) \mathbf{v}$. Then

$$\mathbf{x}'(t) = A\Phi(t_0 + t) \mathbf{v} = A\mathbf{x}(t), \quad \mathbf{x}(0) = \Phi(t_0) \mathbf{v} = \mathbf{0}.$$

But also $\mathbf{z}(t) \equiv \mathbf{0}$ also satisfies

$$\mathbf{z}'(t) = A\mathbf{z}(t), \quad \mathbf{z}(0) = \mathbf{0},$$

and so by the theorem on uniqueness, it must be the case that $\mathbf{z}(t) = \mathbf{x}(t)$ for all t , showing that $\Phi(t + t_0) \mathbf{v} = \mathbf{0}$ for all t , and in particular for $t = -t_0$. Therefore,

$$\Phi(-t_0 + t_0) \mathbf{v} = I\mathbf{v} = \mathbf{0}$$

and so $\mathbf{v} = \mathbf{0}$, a contradiction. It follows that $\Phi(t)$ must be one to one for all t and so, $\Phi(t)^{-1}$ exists for all t .

It only remains to verify the solution to 4.23 is unique. Suppose Ψ is another fundamental matrix solving 4.23. Then letting \mathbf{v} be an arbitrary vector,

$$\mathbf{z}(t) \equiv \Phi(t) \mathbf{v}, \quad \mathbf{y}(t) \equiv \Psi(t) \mathbf{v}$$

both solve the initial value problem,

$$\mathbf{x}' = A\mathbf{x}, \quad \mathbf{x}(0) = \mathbf{v},$$

and so by the uniqueness theorem, $\mathbf{z}(t) = \mathbf{y}(t)$ for all t showing that $\Phi(t)\mathbf{v} = \Psi(t)\mathbf{v}$ for all t . Since \mathbf{v} is arbitrary, this shows that $\Phi(t) = \Psi(t)$ for every t . ■

It is useful to consider the differential equations for the r_k for $k \geq 1$. As noted above, $r_0(t) = 0$ and $r_1(t) = e^{\lambda_1 t}$.

$$r'_{k+1} = \lambda_{k+1}r_{k+1} + r_k, \quad r_{k+1}(0) = 0.$$

Thus

$$r_{k+1}(t) = \int_0^t e^{\lambda_{k+1}(t-s)} r_k(s) ds.$$

Therefore,

$$r_2(t) = \int_0^t e^{\lambda_2(t-s)} e^{\lambda_1 s} ds = \frac{e^{\lambda_1 t} - e^{\lambda_2 t}}{-\lambda_2 + \lambda_1}$$

assuming $\lambda_1 \neq \lambda_2$.

Sometimes people define a fundamental matrix to be a matrix $\Phi(t)$ such that $\Phi'(t) = A\Phi(t)$ and $\det(\Phi(t)) \neq 0$ for all t . Thus this avoids the initial condition, $\Phi(0) = I$. The next proposition has to do with this situation.

Proposition D.4.9 *Suppose A is an $n \times n$ matrix and suppose $\Phi(t)$ is an $n \times n$ matrix for each $t \in \mathbb{R}$ with the property that*

$$\Phi'(t) = A\Phi(t). \tag{4.24}$$

Then either $\Phi(t)^{-1}$ exists for all $t \in \mathbb{R}$ or $\Phi(t)^{-1}$ fails to exist for all $t \in \mathbb{R}$.

Proof: Suppose $\Phi(0)^{-1}$ exists and 4.24 holds. Let $\Psi(t) \equiv \Phi(t)\Phi(0)^{-1}$. Then $\Psi(0) = I$ and

$$\Psi'(t) = \Phi'(t)\Phi(0)^{-1} = A\Phi(t)\Phi(0)^{-1} = A\Psi(t)$$

so by Theorem D.4.8, $\Psi(t)^{-1}$ exists for all t . Therefore, $\Phi(t)^{-1}$ also exists for all t .



Scholarships



Lnu.se

Open your mind to new opportunities

With 31,000 students, Linnaeus University is one of the larger universities in Sweden. We are a modern university, known for our strong international profile. Every year more than 1,600 international students from all over the world choose to enjoy the friendly atmosphere and active student life at Linnaeus University. Welcome to join us!

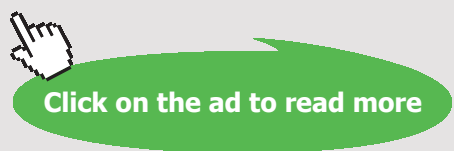
Linnaeus University

Sweden

Bachelor programmes in
Business & Economics | Computer Science/IT | Design | Mathematics

Master programmes in
Business & Economics | Behavioural Sciences | Computer Science/IT | Cultural Studies & Social Sciences | Design | Mathematics | Natural Sciences | Technology & Engineering

Summer Academy courses



Next suppose $\Phi(0)^{-1}$ does not exist. I need to show $\Phi(t)^{-1}$ does not exist for any t . Suppose then that $\Phi(t_0)^{-1}$ does exist. Then let $\Psi(t) \equiv \Phi(t_0 + t)\Phi(t_0)^{-1}$. Then $\Psi(0) = I$ and $\Psi' = A\Psi$ so by Theorem D.4.8 it follows $\Psi(t)^{-1}$ exists for all t and so for all t , $\Phi(t + t_0)^{-1}$ must also exist, even for $t = -t_0$ which implies $\Phi(0)^{-1}$ exists after all. ■

The conclusion of this proposition is usually referred to as the Wronskian alternative and another way to say it is that if 4.24 holds, then either $\det(\Phi(t)) = 0$ for all t or $\det(\Phi(t))$ is never equal to 0. The Wronskian is the usual name of the function, $t \rightarrow \det(\Phi(t))$.

The following theorem gives the variation of constants formula,.

Theorem D.4.10 *Let \mathbf{f} be continuous on $[0, T]$ and let A be an $n \times n$ matrix and \mathbf{x}_0 a vector in \mathbb{C}^n . Then there exists a unique solution to 4.17, \mathbf{x} , given by the variation of constants formula,*

$$\mathbf{x}(t) = \Phi(t)\mathbf{x}_0 + \Phi(t) \int_0^t \Phi(s)^{-1} \mathbf{f}(s) ds \tag{4.25}$$

for $\Phi(t)$ the fundamental matrix for A . Also, $\Phi(t)^{-1} = \Phi(-t)$ and $\Phi(t + s) = \Phi(t)\Phi(s)$ for all t, s and the above variation of constants formula can also be written as

$$\mathbf{x}(t) = \Phi(t)\mathbf{x}_0 + \int_0^t \Phi(t - s) \mathbf{f}(s) ds \tag{4.26}$$

$$= \Phi(t)\mathbf{x}_0 + \int_0^t \Phi(s) \mathbf{f}(t - s) ds \tag{4.27}$$

Proof: From the uniqueness theorem there is at most one solution to 4.17. Therefore, if 4.25 solves 4.17, the theorem is proved. The verification that the given formula works is identical with the verification that the scalar formula given in Theorem D.4.7 solves the initial value problem given there. $\Phi(s)^{-1}$ is continuous because of the formula for the inverse of a matrix in terms of the transpose of the cofactor matrix. Therefore, the integrand in 4.25 is continuous and the fundamental theorem of calculus applies. To verify the formula for the inverse, fix s and consider $\mathbf{x}(t) = \Phi(s + t)\mathbf{v}$, and $\mathbf{y}(t) = \Phi(t)\Phi(s)\mathbf{v}$. Then

$$\mathbf{x}'(t) = A\Phi(t + s)\mathbf{v} = A\mathbf{x}(t), \quad \mathbf{x}(0) = \Phi(s)\mathbf{v}$$

$$\mathbf{y}'(t) = A\Phi(t)\Phi(s)\mathbf{v} = A\mathbf{y}(t), \quad \mathbf{y}(0) = \Phi(s)\mathbf{v}.$$

By the uniqueness theorem, $\mathbf{x}(t) = \mathbf{y}(t)$ for all t . Since s and \mathbf{v} are arbitrary, this shows $\Phi(t + s) = \Phi(t)\Phi(s)$ for all t, s . Letting $s = -t$ and using $\Phi(0) = I$ verifies $\Phi(t)^{-1} = \Phi(-t)$.

Next, note that this also implies $\Phi(t - s)\Phi(s) = \Phi(t)$ and so $\Phi(t - s) = \Phi(t)\Phi(s)^{-1}$. Therefore, this yields 4.26 and then 4.27 follows from changing the variable. ■

If $\Phi' = A\Phi$ and $\Phi(t)^{-1}$ exists for all t , you should verify that the solution to the initial value problem

$$\mathbf{x}' = A\mathbf{x} + \mathbf{f}, \quad \mathbf{x}(t_0) = \mathbf{x}_0$$

is given by

$$\mathbf{x}(t) = \Phi(t - t_0)\mathbf{x}_0 + \int_{t_0}^t \Phi(t - s) \mathbf{f}(s) ds.$$

Theorem D.4.10 is general enough to include all constant coefficient linear differential equations or any order. Thus it includes as a special case the main topics of an entire elementary differential equations class. This is illustrated in the following example. One can reduce an arbitrary linear differential equation to a first order system and then apply the above theory to solve the problem. The next example is a differential equation of damped vibration.

Example D.4.11 The differential equation is $y'' + 2y' + 2y = \cos t$ and initial conditions, $y(0) = 1$ and $y'(0) = 0$.

To solve this equation, let $x_1 = y$ and $x_2 = x_1' = y'$. Then, writing this in terms of these new variables, yields the following system.

$$\begin{aligned} x_2' + 2x_2 + 2x_1 &= \cos t \\ x_1' &= x_2 \end{aligned}$$

This system can be written in the above form as

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}' = \begin{pmatrix} x_2 \\ -2x_2 - 2x_1 \end{pmatrix} + \begin{pmatrix} 0 \\ \cos t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -2 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ \cos t \end{pmatrix}.$$

and the initial condition is of the form

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} (0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Now $P_0(A) \equiv I$. The eigenvalues are $-1 + i, -1 - i$ and so

$$P_1(A) = \left(\begin{pmatrix} 0 & 1 \\ -2 & -2 \end{pmatrix} - (-1 + i) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) = \begin{pmatrix} 1 - i & 1 \\ -2 & -1 - i \end{pmatrix}.$$

Recall $r_0(t) \equiv 0$ and $r_1(t) = e^{(-1+i)t}$. Then

$$r_2' = (-1 - i)r_2 + e^{(-1+i)t}, \quad r_2(0) = 0$$

and so

$$r_2(t) = \frac{e^{(-1+i)t} - e^{(-1-i)t}}{2i} = e^{-t} \sin(t)$$

Putzer's method yields the fundamental matrix as

$$\begin{aligned} \Phi(t) &= e^{(-1+i)t} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + e^{-t} \sin(t) \begin{pmatrix} 1 - i & 1 \\ -2 & -1 - i \end{pmatrix} \\ &= \begin{pmatrix} e^{-t} (\cos(t) + \sin(t)) & e^{-t} \sin t \\ -2e^{-t} \sin t & e^{-t} (\cos(t) - \sin(t)) \end{pmatrix} \end{aligned}$$

From variation of constants formula the desired solution is

$$\begin{aligned} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} (t) &= \begin{pmatrix} e^{-t} (\cos(t) + \sin(t)) & e^{-t} \sin t \\ -2e^{-t} \sin t & e^{-t} (\cos(t) - \sin(t)) \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &+ \int_0^t \begin{pmatrix} e^{-s} (\cos(s) + \sin(s)) & e^{-s} \sin s \\ -2e^{-s} \sin s & e^{-s} (\cos(s) - \sin(s)) \end{pmatrix} \begin{pmatrix} 0 \\ \cos(t-s) \end{pmatrix} ds \\ &= \begin{pmatrix} e^{-t} (\cos(t) + \sin(t)) \\ -2e^{-t} \sin t \end{pmatrix} + \int_0^t \begin{pmatrix} e^{-s} \sin(s) \cos(t-s) \\ e^{-s} (\cos s - \sin s) \cos(t-s) \end{pmatrix} ds \\ &= \begin{pmatrix} e^{-t} (\cos(t) + \sin(t)) \\ -2e^{-t} \sin t \end{pmatrix} + \begin{pmatrix} -\frac{1}{5} (\cos t) e^{-t} - \frac{3}{5} e^{-t} \sin t + \frac{1}{5} \cos t + \frac{2}{5} \sin t \\ -\frac{2}{5} (\cos t) e^{-t} + \frac{4}{5} e^{-t} \sin t + \frac{2}{5} \cos t - \frac{1}{5} \sin t \end{pmatrix} \\ &= \begin{pmatrix} \frac{4}{5} (\cos t) e^{-t} + \frac{2}{5} e^{-t} \sin t + \frac{1}{5} \cos t + \frac{2}{5} \sin t \\ -\frac{6}{5} e^{-t} \sin t - \frac{2}{5} (\cos t) e^{-t} + \frac{2}{5} \cos t - \frac{1}{5} \sin t \end{pmatrix} \end{aligned}$$

Thus $y(t) = x_1(t) = \frac{4}{5} (\cos t) e^{-t} + \frac{2}{5} e^{-t} \sin t + \frac{1}{5} \cos t + \frac{2}{5} \sin t$.

D.5 Geometric Theory Of Autonomous Systems

Here a sufficient condition is given for stability of a first order system. First of all, here is a fundamental estimate for the entries of a fundamental matrix.

Lemma D.5.1 *Let the functions, r_k be given in the statement of Theorem D.4.8 and suppose that A is an $n \times n$ matrix whose eigenvalues are $\{\lambda_1, \dots, \lambda_n\}$. Suppose that these eigenvalues are ordered such that*

$$\operatorname{Re}(\lambda_1) \leq \operatorname{Re}(\lambda_2) \leq \dots \leq \operatorname{Re}(\lambda_n) < 0.$$

Then if $0 > -\delta > \operatorname{Re}(\lambda_n)$ is given, there exists a constant, C such that for each $k = 0, 1, \dots, n$,

$$|r_k(t)| \leq Ce^{-\delta t} \tag{4.28}$$

for all $t > 0$.

Proof: This is obvious for $r_0(t)$ because it is identically equal to 0. From the definition of the $r_k, r'_1 = \lambda_1 r_1, r_1(0) = 1$ and so $r_1(t) = e^{\lambda_1 t}$ which implies

$$|r_1(t)| \leq e^{\operatorname{Re}(\lambda_1)t}.$$

Suppose for some $m \geq 1$ there exists a constant, C_m such that

$$|r_k(t)| \leq C_m t^m e^{\operatorname{Re}(\lambda_m)t}$$

for all $k \leq m$ for all $t > 0$. Then

$$r'_{m+1}(t) = \lambda_{m+1} r_{m+1}(t) + r_m(t), \quad r_{m+1}(0) = 0$$

e-learning for kids

- The number 1 MOOC for Primary Education
- Free Digital Learning for Children 5-12
- 15 Million Children Reached

About e-Learning for Kids Established in 2004, e-Learning for Kids is a global nonprofit foundation dedicated to fun and free learning on the Internet for children ages 5 - 12 with courses in math, science, language arts, computers, health and environmental skills. Since 2005, more than 15 million children in over 190 countries have benefitted from eLessons provided by EFKI. An all-volunteer staff consists of education and e-learning experts and business professionals from around the world committed to making difference. eLearning for Kids is actively seeking funding, volunteers, sponsors and courseware developers; get involved! For more information, please visit www.e-learningforkids.org.



and so

$$r_{m+1}(t) = e^{\lambda_{m+1}t} \int_0^t e^{-\lambda_{m+1}s} r_m(s) ds.$$

Then by the induction hypothesis,

$$\begin{aligned} |r_{m+1}(t)| &\leq e^{\operatorname{Re}(\lambda_{m+1})t} \int_0^t |e^{-\lambda_{m+1}s}| C_m s^m e^{\operatorname{Re}(\lambda_m)s} ds \\ &\leq e^{\operatorname{Re}(\lambda_{m+1})t} \int_0^t s^m C_m e^{-\operatorname{Re}(\lambda_{m+1})s} e^{\operatorname{Re}(\lambda_m)s} ds \\ &\leq e^{\operatorname{Re}(\lambda_{m+1})t} \int_0^t s^m C_m ds = \frac{C_m}{m+1} t^{m+1} e^{\operatorname{Re}(\lambda_{m+1})t} \end{aligned}$$

It follows by induction there exists a constant, C such that for all $k \leq n$,

$$|r_k(t)| \leq Ct^n e^{\operatorname{Re}(\lambda_n)t}$$

and this obviously implies the conclusion of the lemma.

The proof of the above lemma yields the following corollary.

Corollary D.5.2 *Let the functions, r_k be given in the statement of Theorem D.4.8 and suppose that A is an $n \times n$ matrix whose eigenvalues are $\{\lambda_1, \dots, \lambda_n\}$. Suppose that these eigenvalues are ordered such that*

$$\operatorname{Re}(\lambda_1) \leq \operatorname{Re}(\lambda_2) \leq \dots \leq \operatorname{Re}(\lambda_n).$$

Then there exists a constant C such that for all $k \leq m$

$$|r_k(t)| \leq Ct^m e^{\operatorname{Re}(\lambda_m)t}.$$

With the lemma, the following sloppy estimate is available for a fundamental matrix.

Theorem D.5.3 *Let A be an $n \times n$ matrix and let $\Phi(t)$ be the fundamental matrix for A . That is,*

$$\Phi'(t) = A\Phi(t), \quad \Phi(0) = I.$$

Suppose also the eigenvalues of A are $\{\lambda_1, \dots, \lambda_n\}$ where these eigenvalues are ordered such that

$$\operatorname{Re}(\lambda_1) \leq \operatorname{Re}(\lambda_2) \leq \dots \leq \operatorname{Re}(\lambda_n) < 0.$$

Then if $0 > -\delta > \operatorname{Re}(\lambda_n)$, is given, there exists a constant, C such that $|\Phi(t)_{ij}| \leq Ce^{-\delta t}$ for all $t > 0$. Also

$$|\Phi(t)\mathbf{x}| \leq Cn^{3/2}e^{-\delta t} |\mathbf{x}|. \tag{4.29}$$

Proof: Let

$$M \equiv \max \left\{ |P_k(A)_{ij}| \text{ for all } i, j, k \right\}.$$

Then from Putzer's formula for $\Phi(t)$ and Lemma D.5.1, there exists a constant, C such that

$$|\Phi(t)_{ij}| \leq \sum_{k=0}^{n-1} Ce^{-\delta t} M.$$

Let the new C be given by nCM . ■

Next,

$$\begin{aligned} |\Phi(t)\mathbf{x}|^2 &\equiv \sum_{i=1}^n \left(\sum_{j=1}^n \Phi_{ij}(t) x_j \right)^2 \leq \sum_{i=1}^n \left(\sum_{j=1}^n |\Phi_{ij}(t)| |x_j| \right)^2 \\ &\leq \sum_{i=1}^n \left(\sum_{j=1}^n Ce^{-\delta t} |\mathbf{x}| \right)^2 = C^2 e^{-2\delta t} \sum_{i=1}^n (n|\mathbf{x}|)^2 = C^2 e^{-2\delta t} n^3 |\mathbf{x}|^2 \end{aligned}$$

This proves 4.29 and completes the proof.

Definition D.5.4 Let $\mathbf{f} : U \rightarrow \mathbb{R}^n$ where U is an open subset of \mathbb{R}^n such that $\mathbf{a} \in U$ and $\mathbf{f}(\mathbf{a}) = \mathbf{0}$. A point, \mathbf{a} where $\mathbf{f}(\mathbf{a}) = \mathbf{0}$ is called an equilibrium point. Then \mathbf{a} is asymptotically stable if for any $\varepsilon > 0$ there exists $r > 0$ such that whenever $|\mathbf{x}_0 - \mathbf{a}| < r$ and $\mathbf{x}(t)$ the solution to the initial value problem,

$$\mathbf{x}' = \mathbf{f}(\mathbf{x}), \mathbf{x}(0) = \mathbf{x}_0,$$

it follows

$$\lim_{t \rightarrow \infty} \mathbf{x}(t) = \mathbf{a}, |\mathbf{x}(t) - \mathbf{a}| < \varepsilon$$

A differential equation of the form $\mathbf{x}' = \mathbf{f}(\mathbf{x})$ is called autonomous as opposed to a nonautonomous equation of the form $\mathbf{x}' = \mathbf{f}(t, \mathbf{x})$. The equilibrium point \mathbf{a} is stable if for every $\varepsilon > 0$ there exists $\delta > 0$ such that if $|\mathbf{x}_0 - \mathbf{a}| < \delta$, then if \mathbf{x} is the solution of

$$\mathbf{x}' = \mathbf{f}(\mathbf{x}), \mathbf{x}(0) = \mathbf{x}_0, \tag{4.30}$$

then $|\mathbf{x}(t) - \mathbf{a}| < \varepsilon$ for all $t > 0$.

Obviously asymptotic stability implies stability.

An ordinary differential equation is called almost linear if it is of the form

$$\mathbf{x}' = A\mathbf{x} + \mathbf{g}(\mathbf{x})$$

where A is an $n \times n$ matrix and

$$\lim_{\mathbf{x} \rightarrow \mathbf{0}} \frac{\mathbf{g}(\mathbf{x})}{|\mathbf{x}|} = \mathbf{0}.$$

Now the stability of an equilibrium point of an autonomous system, $\mathbf{x}' = \mathbf{f}(\mathbf{x})$ can always be reduced to the consideration of the stability of $\mathbf{0}$ for an almost linear system. Here is why. If you are considering the equilibrium point, \mathbf{a} for $\mathbf{x}' = \mathbf{f}(\mathbf{x})$, you could define a new variable, \mathbf{y} by $\mathbf{a} + \mathbf{y} = \mathbf{x}$. Then asymptotic stability would involve $|\mathbf{y}(t)| < \varepsilon$ and $\lim_{t \rightarrow \infty} \mathbf{y}(t) = \mathbf{0}$ while stability would only require $|\mathbf{y}(t)| < \varepsilon$. Then since \mathbf{a} is an equilibrium point, \mathbf{y} solves the following initial value problem.

$$\mathbf{y}' = \mathbf{f}(\mathbf{a} + \mathbf{y}) - \mathbf{f}(\mathbf{a}), \mathbf{y}(0) = \mathbf{y}_0,$$

where $\mathbf{y}_0 = \mathbf{x}_0 - \mathbf{a}$.

Let $A = D\mathbf{f}(\mathbf{a})$. Then from the definition of the derivative of a function,

$$\mathbf{y}' = A\mathbf{y} + \mathbf{g}(\mathbf{y}), \mathbf{y}(0) = \mathbf{y}_0 \tag{4.31}$$

where

$$\lim_{\mathbf{y} \rightarrow \mathbf{0}} \frac{\mathbf{g}(\mathbf{y})}{|\mathbf{y}|} = \mathbf{0}.$$

Thus there is never any loss of generality in considering only the equilibrium point $\mathbf{0}$ for an almost linear system.¹ Therefore, from now on I will only consider the case of almost linear systems and the equilibrium point $\mathbf{0}$.

Theorem D.5.5 Consider the almost linear system of equations,

$$\mathbf{x}' = A\mathbf{x} + \mathbf{g}(\mathbf{x}) \tag{4.32}$$

where

$$\lim_{\mathbf{x} \rightarrow \mathbf{0}} \frac{\mathbf{g}(\mathbf{x})}{|\mathbf{x}|} = \mathbf{0}$$

and \mathbf{g} is a C^1 function. Suppose that for all λ an eigenvalue of A , $\text{Re } \lambda < 0$. Then $\mathbf{0}$ is asymptotically stable.

¹This is no longer true when you study partial differential equations as ordinary differential equations in infinite dimensional spaces.

Proof: By Theorem D.5.3 there exist constants $\delta > 0$ and K such that for $\Phi(t)$ the fundamental matrix for A ,

$$|\Phi(t)\mathbf{x}| \leq Ke^{-\delta t}|\mathbf{x}|.$$

Let $\varepsilon > 0$ be given and let r be small enough that $Kr < \varepsilon$ and for $|\mathbf{x}| < (K+1)r$, $|\mathbf{g}(\mathbf{x})| < \eta|\mathbf{x}|$ where η is so small that $K\eta < \delta$, and let $|\mathbf{y}_0| < r$. Then by the variation of constants formula, the solution to 4.32, at least for small t satisfies

$$\mathbf{y}(t) = \Phi(t)\mathbf{y}_0 + \int_0^t \Phi(t-s)\mathbf{g}(\mathbf{y}(s))ds.$$

The following estimate holds.

$$|\mathbf{y}(t)| \leq Ke^{-\delta t}|\mathbf{y}_0| + \int_0^t Ke^{-\delta(t-s)}\eta|\mathbf{y}(s)|ds < Ke^{-\delta t}r + \int_0^t Ke^{-\delta(t-s)}\eta|\mathbf{y}(s)|ds.$$

Therefore,

$$e^{\delta t}|\mathbf{y}(t)| < Kr + \int_0^t K\eta e^{\delta s}|\mathbf{y}(s)|ds.$$

By Gronwall's inequality,

$$e^{\delta t}|\mathbf{y}(t)| < Kre^{K\eta t}$$

and so

$$|\mathbf{y}(t)| < Kre^{(K\eta-\delta)t} < \varepsilon e^{(K\eta-\delta)t}$$

.....Alcatel-Lucent 

www.alcatel-lucent.com/careers

What if you could build your future and create the future?

One generation's transformation is the next's status quo. In the near future, people may soon think it's strange that devices ever had to be "plugged in." To obtain that status, there needs to be "The Shift".



Therefore, $|\mathbf{y}(t)| < Kr < \varepsilon$ for all t and so from Corollary D.3.4, the solution to 4.32 exists for all $t \geq 0$ and since $K\eta - \delta < 0$,

$$\lim_{t \rightarrow \infty} |\mathbf{y}(t)| = 0. \blacksquare$$

D.6 General Geometric Theory

Here I will consider the case where the matrix A has both positive and negative eigenvalues. First here is a useful lemma.

Lemma D.6.1 *Suppose A is an $n \times n$ matrix and there exists $\delta > 0$ such that*

$$0 < \delta < \operatorname{Re}(\lambda_1) \leq \dots \leq \operatorname{Re}(\lambda_n)$$

where $\{\lambda_1, \dots, \lambda_n\}$ are the eigenvalues of A , with possibly some repeated. Then there exists a constant, C such that for all $t < 0$,

$$|\Phi(t)\mathbf{x}| \leq Ce^{\delta t} |\mathbf{x}|$$

Proof: I want an estimate on the solutions to the system

$$\Phi'(t) = A\Phi(t), \quad \Phi(0) = I.$$

for $t < 0$. Let $s = -t$ and let $\Psi(s) = \Phi(t)$. Then writing this in terms of Ψ ,

$$\Psi'(s) = -A\Psi(s), \quad \Psi(0) = I.$$

Now the eigenvalues of $-A$ have real parts less than $-\delta$ because these eigenvalues are obtained from the eigenvalues of A by multiplying by -1 . Then by Theorem D.5.3 there exists a constant, C such that for any \mathbf{x} ,

$$|\Psi(s)\mathbf{x}| \leq Ce^{-\delta s} |\mathbf{x}|.$$

Therefore, from the definition of Ψ ,

$$|\Phi(t)\mathbf{x}| \leq Ce^{\delta t} |\mathbf{x}|. \blacksquare$$

Here is another essential lemma which is found in Coddington and Levinson [6]

Lemma D.6.2 *Let $p_j(t)$ be polynomials with complex coefficients and let*

$$f(t) = \sum_{j=1}^m p_j(t) e^{\lambda_j t}$$

where $m \geq 1, \lambda_j \neq \lambda_k$ for $j \neq k$, and none of the $p_j(t)$ vanish identically. Let

$$\sigma = \max(\operatorname{Re}(\lambda_1), \dots, \operatorname{Re}(\lambda_m)).$$

Then there exists a positive number, r and arbitrarily large positive values of t such that

$$e^{-\sigma t} |f(t)| > r.$$

In particular, $|f(t)|$ is unbounded.

Proof: Suppose the largest exponent of any of the p_j is M and let $\lambda_j = a_j + ib_j$. First assume each $a_j = 0$. This is convenient because $\sigma = 0$ in this case and the largest of the $\operatorname{Re}(\lambda_j)$ occurs in every λ_j .

Then arranging the above sum as a sum of decreasing powers of t ,

$$f(t) = t^M f_M(t) + \dots + t f_1(t) + f_0(t).$$

Then

$$t^{-M} f(t) = f_M(t) + O\left(\frac{1}{t}\right)$$

where the last term means that $tO\left(\frac{1}{t}\right)$ is bounded. Then

$$f_M(t) = \sum_{j=1}^m c_j e^{ib_j t}$$

It can't be the case that all the c_j are equal to 0 because then M would not be the highest power exponent. Suppose $c_k \neq 0$. Then

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T t^{-M} f(t) e^{-ib_k t} dt = \sum_{j=1}^m c_j \frac{1}{T} \int_0^T e^{i(b_j - b_k)t} dt = c_k \neq 0.$$

Letting $r = |c_k/2|$, it follows $|t^{-M} f(t) e^{-ib_k t}| > r$ for arbitrarily large values of t . Thus it is also true that $|f(t)| > r$ for arbitrarily large values of t .

Next consider the general case in which σ is given above. Thus

$$e^{-\sigma t} f(t) = \sum_{j:a_j=\sigma} p_j(t) e^{b_j t} + g(t)$$

where $\lim_{t \rightarrow \infty} g(t) = 0$, $g(t)$ being of the form $\sum_s p_s(t) e^{(a_s - \sigma + ib_s)t}$ where $a_s - \sigma < 0$. Then this reduces to the case above in which $\sigma = 0$. Therefore, there exists $r > 0$ such that

$$|e^{-\sigma t} f(t)| > r$$

for arbitrarily large values of t . ■

Next here is a Banach space which will be useful.

Lemma D.6.3 For $\gamma > 0$, let

$$E_\gamma = \{ \mathbf{x} \in BC([0, \infty), \mathbb{F}^n) : t \rightarrow e^{\gamma t} \mathbf{x}(t) \text{ is also in } BC([0, \infty), \mathbb{F}^n) \}$$

and let the norm be given by

$$\|\mathbf{x}\|_\gamma \equiv \sup \{ |e^{\gamma t} \mathbf{x}(t)| : t \in [0, \infty) \}$$

Then E_γ is a Banach space.

Proof: Let $\{\mathbf{x}_k\}$ be a Cauchy sequence in E_γ . Then since $BC([0, \infty), \mathbb{F}^n)$ is a Banach space, there exists $\mathbf{y} \in BC([0, \infty), \mathbb{F}^n)$ such that $e^{\gamma t} \mathbf{x}_k(t)$ converges uniformly on $[0, \infty)$ to $\mathbf{y}(t)$. Therefore $e^{-\gamma t} e^{\gamma t} \mathbf{x}_k(t) = \mathbf{x}_k(t)$ converges uniformly to $e^{-\gamma t} \mathbf{y}(t)$ on $[0, \infty)$. Define $\mathbf{x}(t) \equiv e^{-\gamma t} \mathbf{y}(t)$. Then $\mathbf{y}(t) = e^{\gamma t} \mathbf{x}(t)$ and by definition,

$$\|\mathbf{x}_k - \mathbf{x}\|_\gamma \rightarrow 0. \quad \blacksquare$$

D.7 The Stable Manifold

Here assume

$$A = \begin{pmatrix} A_- & 0 \\ 0 & A_+ \end{pmatrix} \tag{4.33}$$

where A_- and A_+ are square matrices of size $k \times k$ and $(n - k) \times (n - k)$ respectively. Also assume A_- has eigenvalues whose real parts are all less than $-\alpha$ while A_+ has eigenvalues whose real parts are all larger than α . Assume also that each of A_- and A_+ is upper triangular.

Also, I will use the following convention. For $\mathbf{v} \in \mathbb{F}^n$,

$$\mathbf{v} = \begin{pmatrix} \mathbf{v}_- \\ \mathbf{v}_+ \end{pmatrix}$$

where \mathbf{v}_- consists of the first k entries of \mathbf{v} .

Then from Theorem D.5.3 and Lemma D.6.1 the following lemma is obtained.

Lemma D.7.1 *Let A be of the form given in 4.33 as explained above and let $\Phi_+(t)$ and $\Phi_-(t)$ be the fundamental matrices corresponding to A_+ and A_- respectively. Then there exist positive constants, α and γ such that*

$$|\Phi_+(t) \mathbf{y}| \leq C e^{\alpha t} \text{ for all } t < 0 \tag{4.34}$$

$$|\Phi_-(t) \mathbf{y}| \leq C e^{-(\alpha+\gamma)t} \text{ for all } t > 0. \tag{4.35}$$

Also for any nonzero $\mathbf{x} \in \mathbb{C}^{n-k}$,

$$|\Phi_+(t) \mathbf{x}| \text{ is unbounded.} \tag{4.36}$$

Proof: The first two claims have been established already. It suffices to pick α and γ such that $-(\alpha + \gamma)$ is larger than all eigenvalues of A_- and α is smaller than all eigenvalues of A_+ . It remains to verify 4.36. From the Putzer formula for $\Phi_+(t)$,

$$\Phi_+(t) \mathbf{x} = \sum_{k=0}^{n-1} r_{k+1}(t) P_k(A) \mathbf{x}$$

Nido

Luxurious accommodation

Central zone 1 & 2 locations

Meet hundreds of international students

BOOK NOW and get a £100 voucher from voucherexpress

Nido Student Living - London

Visit www.NidoStudentLiving.com/Bookboon for more info.

+44 (0)20 3102 1060



where $P_0(A) \equiv I$. Now each r_k is a polynomial (possibly a constant) times an exponential. This follows easily from the definition of the r_k as solutions of the differential equations

$$r'_{k+1} = \lambda_{k+1}r_{k+1} + r_k.$$

Now by assumption the eigenvalues have positive real parts so

$$\sigma \equiv \max(\operatorname{Re}(\lambda_1), \dots, \operatorname{Re}(\lambda_{n-k})) > 0.$$

It can also be assumed

$$\operatorname{Re}(\lambda_1) \geq \dots \geq \operatorname{Re}(\lambda_{n-k})$$

By Lemma D.6.2 it follows $|\Phi_+(t)\mathbf{x}|$ is unbounded. This follows because

$$\Phi_+(t)\mathbf{x} = r_1(t)\mathbf{x} + \sum_{k=1}^{n-1} r_{k+1}(t)\mathbf{y}_k, \quad r_1(t) = e^{\lambda_1 t}.$$

Since $\mathbf{x} \neq \mathbf{0}$, it has a nonzero entry, say $x_m \neq 0$. Consider the m^{th} entry of the vector $\Phi_+(t)\mathbf{x}$. By this Lemma the m^{th} entry is unbounded and this is all it takes for $\mathbf{x}(t)$ to be unbounded. ■

Lemma D.7.2 Consider the initial value problem for the almost linear system

$$\mathbf{x}' = A\mathbf{x} + \mathbf{g}(\mathbf{x}), \quad \mathbf{x}(0) = \mathbf{x}_0,$$

where \mathbf{g} is C^1 and A is of the special form

$$A = \begin{pmatrix} A_- & 0 \\ 0 & A_+ \end{pmatrix}$$

in which A_- is a $k \times k$ matrix which has eigenvalues for which the real parts are all negative and A_+ is a $(n - k) \times (n - k)$ matrix for which the real parts of all the eigenvalues are positive. Then $\mathbf{0}$ is not stable. More precisely, there exists a set of points $(\mathbf{a}_-, \psi(\mathbf{a}_-))$ for \mathbf{a}_- small such that for \mathbf{x}_0 on this set,

$$\lim_{t \rightarrow \infty} \mathbf{x}(t, \mathbf{x}_0) = \mathbf{0}$$

and for \mathbf{x}_0 not on this set, there exists a $\delta > 0$ such that $|\mathbf{x}(t, \mathbf{x}_0)|$ cannot remain less than δ for all positive t .

Proof: Consider the initial value problem for the almost linear equation,

$$\mathbf{x}' = A\mathbf{x} + \mathbf{g}(\mathbf{x}), \quad \mathbf{x}(0) = \mathbf{a} = \begin{pmatrix} \mathbf{a}_- \\ \mathbf{a}_+ \end{pmatrix}.$$

Then by the variation of constants formula, a local solution has the form

$$\begin{aligned} \mathbf{x}(t, \mathbf{a}) &= \begin{pmatrix} \Phi_-(t) & 0 \\ 0 & \Phi_+(t) \end{pmatrix} \begin{pmatrix} \mathbf{a}_- \\ \mathbf{a}_+ \end{pmatrix} \\ &+ \int_0^t \begin{pmatrix} \Phi_-(t-s) & 0 \\ 0 & \Phi_+(t-s) \end{pmatrix} \mathbf{g}(\mathbf{x}(s, \mathbf{a})) ds \end{aligned} \tag{4.37}$$

Write $\mathbf{x}(t)$ for $\mathbf{x}(t, \mathbf{a})$ for short. Let $\varepsilon > 0$ be given and suppose δ is such that if $|\mathbf{x}| < \delta$, then $|\mathbf{g}_{\pm}(\mathbf{x})| < \varepsilon|\mathbf{x}|$. Assume from now on that $|\mathbf{a}| < \delta$. Then suppose $|\mathbf{x}(t)| < \delta$ for all $t > 0$. Writing 4.37 differently yields

$$\begin{aligned} \mathbf{x}(t, \mathbf{a}) &= \begin{pmatrix} \Phi_{-}(t) & 0 \\ 0 & \Phi_{+}(t) \end{pmatrix} \begin{pmatrix} \mathbf{a}_{-} \\ \mathbf{a}_{+} \end{pmatrix} + \begin{pmatrix} \int_0^t \Phi_{-}(t-s) \mathbf{g}_{-}(\mathbf{x}(s, \mathbf{a})) ds \\ 0 \end{pmatrix} \\ &\quad + \begin{pmatrix} 0 \\ \int_0^t \Phi_{+}(t-s) \mathbf{g}_{+}(\mathbf{x}(s, \mathbf{a})) ds \end{pmatrix} \\ &= \begin{pmatrix} \Phi_{-}(t) & 0 \\ 0 & \Phi_{+}(t) \end{pmatrix} \begin{pmatrix} \mathbf{a}_{-} \\ \mathbf{a}_{+} \end{pmatrix} + \begin{pmatrix} \int_0^t \Phi_{-}(t-s) \mathbf{g}_{-}(\mathbf{x}(s, \mathbf{a})) ds \\ 0 \end{pmatrix} \\ &\quad + \begin{pmatrix} 0 \\ \int_0^{\infty} \Phi_{+}(t-s) \mathbf{g}_{+}(\mathbf{x}(s, \mathbf{a})) ds - \int_t^{\infty} \Phi_{+}(t-s) \mathbf{g}_{+}(\mathbf{x}(s, \mathbf{a})) ds \end{pmatrix}. \end{aligned}$$

These improper integrals converge thanks to the assumption that \mathbf{x} is bounded and the estimates 4.34 and 4.35. Continuing the rewriting,

$$\begin{pmatrix} \mathbf{x}_{-}(t) \\ \mathbf{x}_{+}(t) \end{pmatrix} = \begin{pmatrix} \Phi_{-}(t) \mathbf{a}_{-} + \int_0^t \Phi_{-}(t-s) \mathbf{g}_{-}(\mathbf{x}(s, \mathbf{a})) ds \\ \Phi_{+}(t) (\mathbf{a}_{+} + \int_0^{\infty} \Phi_{+}(-s) \mathbf{g}_{+}(\mathbf{x}(s, \mathbf{a})) ds) \end{pmatrix} + \begin{pmatrix} 0 \\ -\int_t^{\infty} \Phi_{+}(t-s) \mathbf{g}_{+}(\mathbf{x}(s, \mathbf{a})) ds \end{pmatrix}.$$

It follows from Lemma D.7.1 that if $|\mathbf{x}(t, \mathbf{a})|$ is bounded by δ as asserted, then it must be the case that $\mathbf{a}_{+} + \int_0^{\infty} \Phi_{+}(-s) \mathbf{g}_{+}(\mathbf{x}(s, \mathbf{a})) ds = \mathbf{0}$. Consequently, it must be the case that

$$\mathbf{x}(t) = \Phi(t) \begin{pmatrix} \mathbf{a}_{-} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \int_0^t \Phi_{-}(t-s) \mathbf{g}_{-}(\mathbf{x}(s, \mathbf{a})) ds \\ -\int_t^{\infty} \Phi_{+}(t-s) \mathbf{g}_{+}(\mathbf{x}(s, \mathbf{a})) ds \end{pmatrix} \tag{4.38}$$

Letting $t \rightarrow 0$, this requires that for a solution to the initial value problem to exist and also satisfy $|\mathbf{x}(t)| < \delta$ for all $t > 0$ it must be the case that

$$\mathbf{x}(0) = \begin{pmatrix} \mathbf{a}_{-} \\ -\int_0^{\infty} \Phi_{+}(-s) \mathbf{g}_{+}(\mathbf{x}(s, \mathbf{a})) ds \end{pmatrix}$$

where $\mathbf{x}(t, \mathbf{a})$ is the solution of

$$\mathbf{x}' = A\mathbf{x} + \mathbf{g}(\mathbf{x}), \quad \mathbf{x}(0) = \begin{pmatrix} \mathbf{a}_{-} \\ -\int_0^{\infty} \Phi_{+}(-s) \mathbf{g}_{+}(\mathbf{x}(s, \mathbf{a})) ds \end{pmatrix}$$

This is because in 4.38, if \mathbf{x} is bounded by δ then the reverse steps show \mathbf{x} is a solution of the above differential equation and initial condition.

It follows if I can show that for all \mathbf{a}_{-} sufficiently small and $\mathbf{a} = (\mathbf{a}_{-}, \mathbf{0})^T$, there exists a solution to 4.38 $\mathbf{x}(s, \mathbf{a})$ on $(0, \infty)$ for which $|\mathbf{x}(s, \mathbf{a})| < \delta$, then I can define

$$\boldsymbol{\psi}(\mathbf{a}) \equiv -\int_0^{\infty} \Phi_{+}(-s) \mathbf{g}_{+}(\mathbf{x}(s, \mathbf{a})) ds$$

and conclude that $|\mathbf{x}(t, \mathbf{x}_0)| < \delta$ for all $t > 0$ if and only if $\mathbf{x}_0 = (\mathbf{a}_{-}, \boldsymbol{\psi}(\mathbf{a}_{-}))^T$ for some sufficiently small \mathbf{a}_{-} .

Let C, α, γ be the constants of Lemma D.7.1. Let η be a small positive number such that

$$\frac{C\eta}{\alpha} < \frac{1}{6}$$

Note that $\frac{\partial \mathbf{g}}{\partial x_i}(0) = \mathbf{0}$. Therefore, by Lemma D.3.1, there exists $\delta > 0$ such that if $|\mathbf{x}|, |\mathbf{y}| \leq \delta$, then

$$|\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{y})| < \eta |\mathbf{x} - \mathbf{y}|$$

and in particular,

$$|\mathbf{g}_{\pm}(\mathbf{x}) - \mathbf{g}_{\pm}(\mathbf{y})| < \eta |\mathbf{x} - \mathbf{y}| \tag{4.39}$$

because each $\frac{\partial \mathbf{g}}{\partial x_i}(\mathbf{x})$ is very small. In particular, this implies


$$|\mathbf{g}_{-}(\mathbf{x})| < \eta |\mathbf{x}|, |\mathbf{g}_{+}(\mathbf{x})| < \eta |\mathbf{x}|.$$


For $\mathbf{x} \in E_{\gamma}$ defined in Lemma D.6.3 and $|\mathbf{a}_{-}| < \frac{\delta}{2C}$,

$$F\mathbf{x}(t) \equiv \begin{pmatrix} \Phi_{-}(t)\mathbf{a}_{-} + \int_0^t \Phi_{-}(t-s)\mathbf{g}_{-}(\mathbf{x}(s))ds \\ - \int_t^{\infty} \Phi_{+}(t-s)\mathbf{g}_{+}(\mathbf{x}(s))ds \end{pmatrix}.$$

I need to find a fixed point of F . Letting $\|\mathbf{x}\|_{\gamma} < \delta$, and using the estimates of Lemma D.7.1,

$$\begin{aligned} e^{\gamma t} |F\mathbf{x}(t)| &\leq e^{\gamma t} |\Phi_{-}(t)\mathbf{a}_{-}| + e^{\gamma t} \int_0^t C e^{-(\alpha+\gamma)(t-s)} \eta |\mathbf{x}(s)| ds \\ &\quad + e^{\gamma t} \int_t^{\infty} C e^{\alpha(t-s)} \eta |\mathbf{x}(s)| ds \\ &\leq e^{\gamma t} C \frac{\delta}{2C} e^{-(\alpha+\gamma)t} + e^{\gamma t} \|\mathbf{x}\|_{\gamma} C \eta \int_0^t e^{-(\alpha+\gamma)(t-s)} e^{-\gamma s} ds \\ &\quad + e^{\gamma t} C \eta \int_t^{\infty} e^{\alpha(t-s)} e^{-\gamma s} ds \|\mathbf{x}\|_{\gamma} \end{aligned}$$

SIMPLY CLEVER




WE WILL TURN YOUR CV INTO AN OPPORTUNITY OF A LIFETIME

Do you like cars? Would you like to be a part of a successful brand? As a constructor at ŠKODA AUTO you will put great things in motion. Things that will ease everyday lives of people all around Send us your CV. We will give it an entirely new new dimension.
Read more about this and our other international masters degree programmes at www.uu.se/master

Send us your CV on www.employerforlife.com



$$\begin{aligned} &< \frac{\delta}{2} + \delta C\eta \int_0^t e^{-\alpha(t-s)} ds + C\eta\delta \int_t^\infty e^{(\alpha+\gamma)(t-s)} ds \\ &< \frac{\delta}{2} + \delta C\eta \frac{1}{\alpha} + \frac{\delta C\eta}{\alpha + \gamma} \leq \delta \left(\frac{1}{2} + \frac{C\eta}{\alpha} \right) < \frac{2\delta}{3}. \end{aligned}$$

Thus F maps every $\mathbf{x} \in E_\gamma$ having $\|\mathbf{x}\|_\gamma < \delta$ to $F\mathbf{x}$ where $\|F\mathbf{x}\|_\gamma \leq \frac{2\delta}{3}$.

Now let $\mathbf{x}, \mathbf{y} \in E_\gamma$ where $\|\mathbf{x}\|_\gamma, \|\mathbf{y}\|_\gamma < \delta$. Then

$$\begin{aligned} e^{\gamma t} |F\mathbf{x}(t) - F\mathbf{y}(t)| &\leq e^{\gamma t} \int_0^t |\Phi_-(t-s)| \eta e^{-\gamma s} e^{\gamma s} |\mathbf{x}(s) - \mathbf{y}(s)| ds \\ &\quad + e^{\gamma t} \int_t^\infty |\Phi_+(t-s)| e^{-\gamma s} e^{\gamma s} \eta |\mathbf{x}(s) - \mathbf{y}(s)| ds \\ &\leq C\eta \|\mathbf{x} - \mathbf{y}\|_\gamma \left(\int_0^t e^{-\alpha(t-s)} ds + \int_t^\infty e^{(\alpha+\gamma)(t-s)} ds \right) \\ &\leq C\eta \left(\frac{1}{\alpha} + \frac{1}{\alpha + \gamma} \right) \|\mathbf{x} - \mathbf{y}\|_\gamma < \frac{2C\eta}{\alpha} \|\mathbf{x} - \mathbf{y}\|_\gamma < \frac{1}{3} \|\mathbf{x} - \mathbf{y}\|_\gamma. \end{aligned}$$

It follows from Lemma 13.6.4, for each \mathbf{a}_- such that $|\mathbf{a}_-| < \frac{\delta}{2C}$, there exists a unique solution to 4.38 in E_γ .

As pointed out earlier, if

$$\boldsymbol{\psi}(\mathbf{a}) \equiv - \int_0^\infty \Phi_+(-s) \mathbf{g}_+(\mathbf{x}(s, \mathbf{a})) ds$$

then for $\mathbf{x}(t, \mathbf{x}_0)$ the solution to the initial value problem

$$\mathbf{x}' = A\mathbf{x} + \mathbf{g}(\mathbf{x}), \quad \mathbf{x}(0) = \mathbf{x}_0$$

has the property that if \mathbf{x}_0 is not of the form $\begin{pmatrix} \mathbf{a}_- \\ \boldsymbol{\psi}(\mathbf{a}_-) \end{pmatrix}$, then $|\mathbf{x}(t, \mathbf{x}_0)|$ cannot be less than δ for all $t > 0$.

On the other hand, if $\mathbf{x}_0 = \begin{pmatrix} \mathbf{a}_- \\ \boldsymbol{\psi}(\mathbf{a}_-) \end{pmatrix}$ for $|\mathbf{a}_-| < \frac{\delta}{2C}$, then $\mathbf{x}(t, \mathbf{x}_0)$, the solution to 4.38 is the unique solution to the initial value problem

$$\mathbf{x}' = A\mathbf{x} + \mathbf{g}(\mathbf{x}), \quad \mathbf{x}(0) = \mathbf{x}_0.$$

and it was shown that $\|\mathbf{x}(\cdot, \mathbf{x}_0)\|_\gamma < \delta$ and so in fact,

$$|\mathbf{x}(t, \mathbf{x}_0)| \leq \delta e^{-\gamma t}$$

showing that

$$\lim_{t \rightarrow \infty} \mathbf{x}(t, \mathbf{x}_0) = \mathbf{0}.$$

■

The following theorem is the main result. It involves a use of linear algebra and the above lemma.

Theorem D.7.3 Consider the initial value problem for the almost linear system

$$\mathbf{x}' = A\mathbf{x} + \mathbf{g}(\mathbf{x}), \quad \mathbf{x}(0) = \mathbf{x}_0$$

in which \mathbf{g} is C^1 and where at there are $k < n$ eigenvalues of A which have negative real parts and $n - k$ eigenvalues of A which have positive real parts. Then $\mathbf{0}$ is not stable. More precisely, there exists a set of points $(\mathbf{a}, \boldsymbol{\psi}(\mathbf{a}))$ for \mathbf{a} small and in a k dimensional subspace such that for \mathbf{x}_0 on this set,

$$\lim_{t \rightarrow \infty} \mathbf{x}(t, \mathbf{x}_0) = \mathbf{0}$$

and for \mathbf{x}_0 not on this set, there exists a $\delta > 0$ such that $|\mathbf{x}(t, \mathbf{x}_0)|$ cannot remain less than δ for all positive t .

Proof: This involves nothing more than a reduction to the situation of Lemma D.7.2. From Theorem 9.5.2 on Page 9.5.2 A is similar to a matrix of the form described in Lemma

D.7.2. Thus $A = S^{-1} \begin{pmatrix} A_- & 0 \\ 0 & A_+ \end{pmatrix} S$. Letting $\mathbf{y} = S\mathbf{x}$, it follows

$$\mathbf{y}' = \begin{pmatrix} A_- & 0 \\ 0 & A_+ \end{pmatrix} \mathbf{y} + \mathbf{g}(S^{-1}\mathbf{y})$$

Now $|\mathbf{x}| = |S^{-1}S\mathbf{x}| \leq \|S^{-1}\| |\mathbf{y}|$ and $|\mathbf{y}| = |SS^{-1}\mathbf{y}| \leq \|S\| |\mathbf{x}|$. Therefore,

$$\frac{1}{\|S\|} |\mathbf{y}| \leq |\mathbf{x}| \leq \|S^{-1}\| |\mathbf{y}|.$$

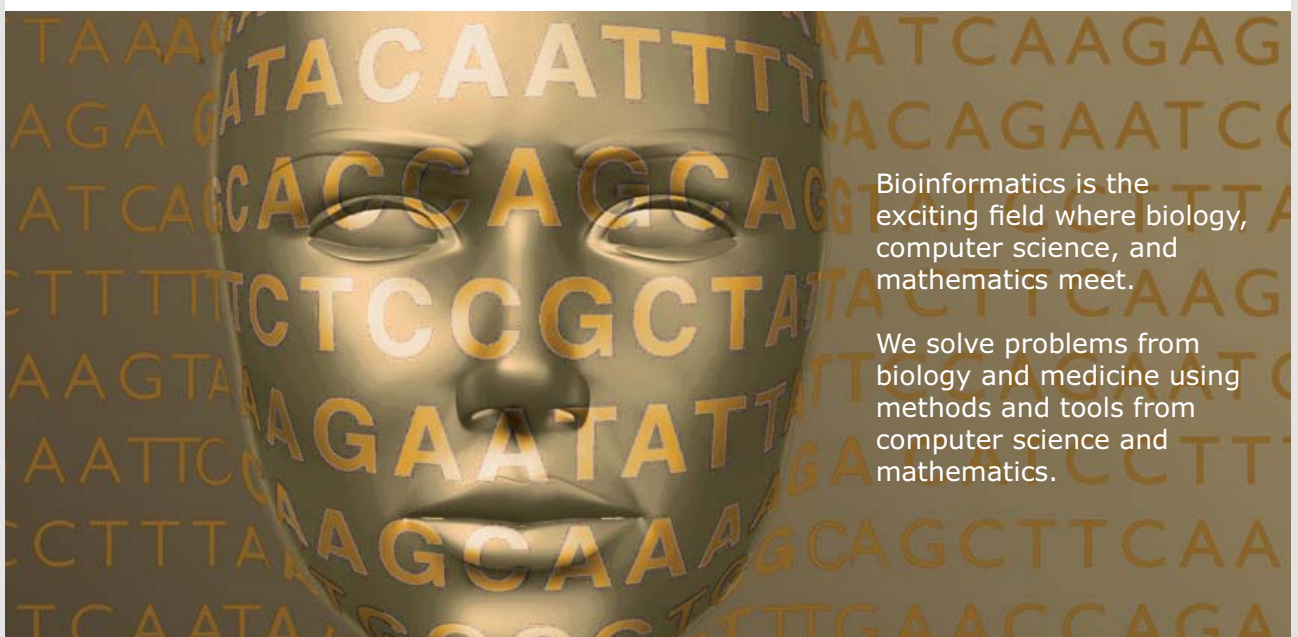
It follows all conclusions of Lemma D.7.2 are valid for this theorem. ■

The set of points $(\mathbf{a}, \psi(\mathbf{a}))$ for \mathbf{a} small is called the stable manifold. Much more can be said about the stable manifold and you should look at a good differential equations book for this.



UPPSALA
UNIVERSITET

Develop the tools we need for Life Science Masters Degree in Bioinformatics



Bioinformatics is the exciting field where biology, computer science, and mathematics meet.

We solve problems from biology and medicine using methods and tools from computer science and mathematics.

Read more about this and our other international masters degree programmes at www.uu.se/master



Appendix E

Compactness And Completeness

E.1 The Nested Interval Lemma

First, here is the one dimensional nested interval lemma.

Lemma E.1.1 *Let $I_k = [a_k, b_k]$ be closed intervals, $a_k \leq b_k$, such that $I_k \supseteq I_{k+1}$ for all k . Then there exists a point c which is contained in all these intervals. If $\lim_{k \rightarrow \infty} (b_k - a_k) = 0$, then there is exactly one such point.*

Proof: Note that the $\{a_k\}$ are an increasing sequence and that $\{b_k\}$ is a decreasing sequence. Now note that if $m < n$, then

$$a_m \leq a_n \leq b_n$$

while if $m > n$,

$$b_n \geq b_m \geq a_m.$$

It follows that $a_m \leq b_n$ for any pair m, n . Therefore, each b_n is an upper bound for all the a_m and so if $c \equiv \sup \{a_k\}$, then for each n , it follows that $c \leq b_n$ and so for all, $a_n \leq c \leq b_n$ which shows that c is in all of these intervals.

If the condition on the lengths of the intervals holds, then if c, c' are in all the intervals, then if they are not equal, then eventually, for large enough k , they cannot both be contained in $[a_k, b_k]$ since eventually $b_k - a_k < |c - c'|$. This would be a contradiction. Hence $c = c'$. ■

Definition E.1.2 *The **diameter** of a set S , is defined as*

$$\text{diam}(S) \equiv \sup \{|\mathbf{x} - \mathbf{y}| : \mathbf{x}, \mathbf{y} \in S\}.$$

Thus $\text{diam}(S)$ is just a careful description of what you would think of as the diameter. It measures how stretched out the set is.

Here is a multidimensional version of the nested interval lemma.

Lemma E.1.3 *Let $I_k = \prod_{i=1}^p [a_i^k, b_i^k] \equiv \{\mathbf{x} \in \mathbb{R}^p : x_i \in [a_i^k, b_i^k]\}$ and suppose that for all $k = 1, 2, \dots$,*

$$I_k \supseteq I_{k+1}.$$

Then there exists a point $\mathbf{c} \in \mathbb{R}^p$ which is an element of every I_k . If $\lim_{k \rightarrow \infty} \text{diam}(I_k) = 0$, then the point \mathbf{c} is unique.

Proof: For each $i = 1, \dots, p$, $[a_i^k, b_i^k] \supseteq [a_i^{k+1}, b_i^{k+1}]$ and so, by Lemma E.1.1, there exists a point $c_i \in [a_i^k, b_i^k]$ for all k . Then letting $\mathbf{c} \equiv (c_1, \dots, c_p)$ it follows $\mathbf{c} \in I_k$ for all k . If the condition on the diameters holds, then the lengths of the intervals $\lim_{k \rightarrow \infty} [a_i^k, b_i^k] = 0$ and so by the same lemma, each c_i is unique. Hence \mathbf{c} is unique. ■

E.2 Convergent Sequences, Sequential Compactness

A mapping $\mathbf{f} : \{k, k + 1, k + 2, \dots\} \rightarrow \mathbb{R}^p$ is called a sequence. We usually write it in the form $\{\mathbf{a}_j\}$ where it is understood that $\mathbf{a}_j \equiv \mathbf{f}(j)$.

Definition E.2.1 A sequence, $\{\mathbf{a}_k\}$ is said to **converge** to \mathbf{a} if for every $\varepsilon > 0$ there exists n_ε such that if $n > n_\varepsilon$, then $|\mathbf{a} - \mathbf{a}_n| < \varepsilon$. The usual notation for this is $\lim_{n \rightarrow \infty} \mathbf{a}_n = \mathbf{a}$ although it is often written as $\mathbf{a}_n \rightarrow \mathbf{a}$. A closed set $K \subseteq \mathbb{R}^n$ is one which has the property that if $\{\mathbf{k}_j\}_{j=1}^\infty$ is a sequence of points of K which converges to \mathbf{x} , then $\mathbf{x} \in K$.

One can also define a subsequence.

Definition E.2.2 $\{\mathbf{a}_{n_k}\}$ is a **subsequence** of $\{\mathbf{a}_n\}$ if $n_1 < n_2 < \dots$.

The following theorem says the limit, if it exists, is unique.

Theorem E.2.3 If a sequence, $\{\mathbf{a}_n\}$ converges to \mathbf{a} and to \mathbf{b} then $\mathbf{a} = \mathbf{b}$.

Proof: There exists n_ε such that if $n > n_\varepsilon$ then $|\mathbf{a}_n - \mathbf{a}| < \frac{\varepsilon}{2}$ and if $n > n_\varepsilon$, then $|\mathbf{a}_n - \mathbf{b}| < \frac{\varepsilon}{2}$. Then pick such an n .

$$|\mathbf{a} - \mathbf{b}| < |\mathbf{a} - \mathbf{a}_n| + |\mathbf{a}_n - \mathbf{b}| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Since ε is arbitrary, this proves the theorem. ■

The following is the definition of a Cauchy sequence in \mathbb{R}^p .

Definition E.2.4 $\{\mathbf{a}_n\}$ is a **Cauchy sequence** if for all $\varepsilon > 0$, there exists n_ε such that whenever $n, m \geq n_\varepsilon$, it follows that $|\mathbf{a}_n - \mathbf{a}_m| < \varepsilon$.

A sequence is Cauchy, means the terms are “bunching up to each other” as m, n get large.

Theorem E.2.5 The set of terms in a Cauchy sequence in \mathbb{R}^p is bounded in the sense that for all n , $|\mathbf{a}_n| < M$ for some $M < \infty$.

Proof: Let $\varepsilon = 1$ in the definition of a Cauchy sequence and let $n > n_1$. Then from the definition, $|\mathbf{a}_n - \mathbf{a}_{n_1}| < 1$. It follows that for all $n > n_1$, $|\mathbf{a}_n| < 1 + |\mathbf{a}_{n_1}|$. Therefore, for all n ,

$$|\mathbf{a}_n| \leq 1 + |\mathbf{a}_{n_1}| + \sum_{k=1}^{n_1} |\mathbf{a}_k|. \quad \blacksquare$$

Theorem E.2.6 If a sequence $\{\mathbf{a}_n\}$ in \mathbb{R}^p converges, then the sequence is a Cauchy sequence. Also, if some subsequence of a Cauchy sequence converges, then the original sequence converges.

Proof: Let $\varepsilon > 0$ be given and suppose $\mathbf{a}_n \rightarrow \mathbf{a}$. Then from the definition of convergence, there exists n_ε such that if $n > n_\varepsilon$, it follows that $|\mathbf{a}_n - \mathbf{a}| < \frac{\varepsilon}{2}$. Therefore, if $m, n \geq n_\varepsilon + 1$, it follows that

$$|\mathbf{a}_n - \mathbf{a}_m| \leq |\mathbf{a}_n - \mathbf{a}| + |\mathbf{a} - \mathbf{a}_m| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

showing that, since $\varepsilon > 0$ is arbitrary, $\{\mathbf{a}_n\}$ is a Cauchy sequence. It remains to that the last claim.

Suppose then that $\{\mathbf{a}_n\}$ is a Cauchy sequence and $\mathbf{a} = \lim_{k \rightarrow \infty} \mathbf{a}_{n_k}$ where $\{\mathbf{a}_{n_k}\}_{k=1}^\infty$ is a subsequence. Let $\varepsilon > 0$ be given. Then there exists K such that if $k, l \geq K$, then $|\mathbf{a}_k - \mathbf{a}_l| < \frac{\varepsilon}{2}$. Then if $k > K$, it follows $n_k > K$ because n_1, n_2, n_3, \dots is strictly increasing as the subscript increases. Also, there exists K_1 such that if $k > K_1$, $|\mathbf{a}_{n_k} - \mathbf{a}| < \frac{\varepsilon}{2}$. Then letting $n > \max(K, K_1)$, pick $k > \max(K, K_1)$. Then

$$|\mathbf{a} - \mathbf{a}_n| \leq |\mathbf{a} - \mathbf{a}_{n_k}| + |\mathbf{a}_{n_k} - \mathbf{a}_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Therefore, the sequence converges. ■

Definition E.2.7 A set K in \mathbb{R}^p is said to be **sequentially compact** if every sequence in K has a subsequence which converges to a point of K .

Theorem E.2.8 *If $I_0 = \prod_{i=1}^p [a_i, b_i]$ where $a_i \leq b_i$, then I_0 is sequentially compact.*

Proof: Let $\{\mathbf{a}_k\}_{k=1}^\infty \subseteq I_0$ and consider all sets of the form $\prod_{i=1}^p [c_i, d_i]$ where $[c_i, d_i]$ equals either $[a_i, \frac{a_i+b_i}{2}]$ or $[c_i, d_i] = [\frac{a_i+b_i}{2}, b_i]$. Thus there are 2^p of these sets because there are two choices for the i^{th} slot for $i = 1, \dots, p$. Also, if \mathbf{x} and \mathbf{y} are two points in one of these sets, $|x_i - y_i| \leq 2^{-1} |b_i - a_i|$ where $\text{diam}(I_0) = \left(\sum_{i=1}^p |b_i - a_i|^2\right)^{1/2}$,

$$|\mathbf{x} - \mathbf{y}| = \left(\sum_{i=1}^p |x_i - y_i|^2\right)^{1/2} \leq 2^{-1} \left(\sum_{i=1}^p |b_i - a_i|^2\right)^{1/2} \equiv 2^{-1} \text{diam}(I_0).$$

In particular, since $\mathbf{d} \equiv (d_1, \dots, d_p)$ and $\mathbf{c} \equiv (c_1, \dots, c_p)$ are two such points,

$$D_1 \equiv \left(\sum_{i=1}^p |d_i - c_i|^2\right)^{1/2} \leq 2^{-1} \text{diam}(I_0)$$

Denote by $\{J_1, \dots, J_{2^p}\}$ these sets determined above. Since the union of these sets equals all of $I_0 \equiv I$, it follows that for some J_k , the sequence, $\{\mathbf{a}_i\}$ is contained in J_k for infinitely many k . Let that one be called I_1 . Next do for I_1 what was done for I_0 to get $I_2 \subseteq I_1$ such that the diameter is half that of I_1 and I_2 contains $\{\mathbf{a}_k\}$ for infinitely many values of k . Continue in this way obtaining a nested sequence $\{I_k\}$ such that $I_k \supseteq I_{k+1}$, and if $\mathbf{x}, \mathbf{y} \in I_k$, then $|\mathbf{x} - \mathbf{y}| \leq 2^{-k} \text{diam}(I_0)$, and I_n contains $\{\mathbf{a}_k\}$ for infinitely many values of k for each n . Then by the nested interval lemma, there exists \mathbf{c} such that \mathbf{c} is contained in each I_k . Pick $\mathbf{a}_{n_1} \in I_1$. Next pick $n_2 > n_1$ such that $\mathbf{a}_{n_2} \in I_2$. If $\mathbf{a}_{n_1}, \dots, \mathbf{a}_{n_k}$ have been chosen, let $\mathbf{a}_{n_{k+1}} \in I_{k+1}$ and $n_{k+1} > n_k$. This can be done because in the construction, I_n contains $\{\mathbf{a}_k\}$ for infinitely many k . Thus the distance between \mathbf{a}_{n_k} and \mathbf{c} is no larger than $2^{-k} \text{diam}(I_0)$, and so $\lim_{k \rightarrow \infty} \mathbf{a}_{n_k} = \mathbf{c} \in I_0$. ■

Corollary E.2.9 *Let K be a closed and bounded set of points in \mathbb{R}^p . Then K is sequentially compact.*

Proof: Since K is closed and bounded, there exists a closed rectangle, $\prod_{k=1}^p [a_k, b_k]$ which contains K . Now let $\{\mathbf{x}_k\}$ be a sequence of points in K . By Theorem E.2.8, there exists a subsequence $\{\mathbf{x}_{n_k}\}$ such that $\mathbf{x}_{n_k} \rightarrow \mathbf{x} \in \prod_{k=1}^p [a_k, b_k]$. However, K is closed and each \mathbf{x}_{n_k} is in K so $\mathbf{x} \in K$. ■

Theorem E.2.10 *Every Cauchy sequence in \mathbb{R}^p converges.*

Proof: Let $\{\mathbf{a}_k\}$ be a Cauchy sequence. By Theorem E.2.5, there is some box $\prod_{i=1}^p [a_i, b_i]$ containing all the terms of $\{\mathbf{a}_k\}$. Therefore, by Theorem E.2.8, a subsequence converges to a point of $\prod_{i=1}^p [a_i, b_i]$. By Theorem E.2.6, the original sequence converges. ■

Appendix F

Some Topics Flavored With Linear Algebra

F.1 The Symmetric Polynomial Theorem

First here is a definition of polynomials in many variables which have coefficients in a commutative ring. A commutative ring would be a field except you don't know that every nonzero element has a multiplicative inverse. If you like, let these coefficients be in a field. It is still interesting. A good example of a commutative ring is the integers. In particular, every field is a commutative ring.

Definition F.1.1 Let $\mathbf{k} \equiv (k_1, k_2, \dots, k_n)$ where each k_i is a nonnegative integer. Let

$$|\mathbf{k}| \equiv \sum_i k_i$$

Polynomials of degree p in the variables x_1, x_2, \dots, x_n are expressions of the form

$$g(x_1, x_2, \dots, x_n) = \sum_{|\mathbf{k}| \leq p} a_{\mathbf{k}} x_1^{k_1} \dots x_n^{k_n}$$

UNIVERSITY OF COPENHAGEN

Brain power

By 2020, wind could provide one-tenth of our electricity needs. Already today, SKF's innovative know-how is crucial to running a large proportion of the world's wind turbines.

Up to 25 % of the generated energy can be reduced by using SKF's advanced on-line condition monitoring systems for on-line condition monitoring and lubrication. We help you reduce your energy consumption by using our innovative products and solutions. By sharing our knowledge and expertise, industries can benefit in many ways. Therefore, we have an obligation to help and support you. This is our brain power.

Copenhagen Master of Excellence

Copenhagen Master of Excellence are two-year master degrees taught in English at one of Europe's leading universities

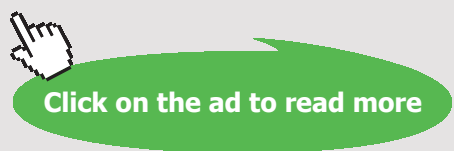
Come to Copenhagen – and see for yourself

Plug into The Power of Knowledge Engineering. Visit us at www.skf.com/knowledge

www.come.ku.dk

The Power of Knowledge Engineering

- cultural studies
- religious studies
- science



where each $a_{\mathbf{k}}$ is in a commutative ring. If all $a_{\mathbf{k}} = 0$, the polynomial has no degree. Such a polynomial is said to be symmetric if whenever σ is a permutation of $\{1, 2, \dots, n\}$,

$$g(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}) = g(x_1, x_2, \dots, x_n)$$

An example of a symmetric polynomial is

$$s_1(x_1, x_2, \dots, x_n) \equiv \sum_{i=1}^n x_i$$

Another one is

$$s_n(x_1, x_2, \dots, x_n) \equiv x_1 x_2 \cdots x_n$$

Definition F.1.2 The elementary symmetric polynomial $s_k(x_1, x_2, \dots, x_n)$, $k = 1, \dots, n$ is the coefficient of $(-1)^k x^{n-k}$ in the following polynomial.

$$\begin{aligned} & (x - x_1)(x - x_2) \cdots (x - x_n) \\ &= x^n - s_1 x^{n-1} + s_2 x^{n-2} - \cdots \pm s_n \end{aligned}$$

Thus

$$\begin{aligned} s_1 &= x_1 + x_2 + \cdots + x_n \\ s_2 &= \sum_{i < j} x_i x_j, \quad s_3 = \sum_{i < j < k} x_i x_j x_k, \dots, \quad s_n = x_1 x_2 \cdots x_n \end{aligned}$$

Then the following result is the fundamental theorem in the subject. It is the symmetric polynomial theorem. It says that these elementary symmetric polynomials are a lot like a basis for the symmetric polynomials.

Theorem F.1.3 Let $g(x_1, x_2, \dots, x_n)$ be a symmetric polynomial. Then $g(x_1, x_2, \dots, x_n)$ equals a polynomial in the elementary symmetric functions.

$$g(x_1, x_2, \dots, x_n) = \sum_{\mathbf{k}} a_{\mathbf{k}} s_1^{k_1} \cdots s_n^{k_n}$$

and the $a_{\mathbf{k}}$ are unique.

Proof: If $n = 1$, it is obviously true because $s_1 = x_1$. Suppose the theorem is true for $n - 1$ and $g(x_1, x_2, \dots, x_n)$ has degree d . Let

$$g'(x_1, x_2, \dots, x_{n-1}) \equiv g(x_1, x_2, \dots, x_{n-1}, 0)$$

By induction, there are unique $a_{\mathbf{k}}$ such that

$$g'(x_1, x_2, \dots, x_{n-1}) = \sum_{\mathbf{k}} a_{\mathbf{k}} s_1'^{k_1} \cdots s_{n-1}'^{k_{n-1}}$$

where s_i' is the corresponding symmetric polynomial which pertains to x_1, x_2, \dots, x_{n-1} . Note that

$$s_k(x_1, x_2, \dots, x_{n-1}, 0) = s_k'(x_1, x_2, \dots, x_{n-1})$$

Now consider

$$g(x_1, x_2, \dots, x_n) - \sum_{\mathbf{k}} a_{\mathbf{k}} s_1^{k_1} \cdots s_{n-1}^{k_{n-1}} \equiv q(x_1, x_2, \dots, x_n)$$

is a symmetric polynomial and it equals 0 when x_n equals 0. Since it is symmetric, it is also 0 whenever $x_i = 0$. Therefore,

$$q(x_1, x_2, \dots, x_n) = s_n h(x_1, x_2, \dots, x_n)$$

and it follows that $h(x_1, x_2, \dots, x_n)$ is symmetric of degree no more than $d - n$ and is uniquely determined. Thus, if $g(x_1, x_2, \dots, x_n)$ is symmetric of degree d ,

$$g(x_1, x_2, \dots, x_n) = \sum_{\mathbf{k}} a_{\mathbf{k}} s_1^{k_1} \cdots s_{n-1}^{k_{n-1}} + s_n h(x_1, x_2, \dots, x_n)$$

where h has degree no more than $d - n$. Now apply the same argument to $h(x_1, x_2, \dots, x_n)$ and continue, repeatedly obtaining a sequence of symmetric polynomials h_i , of strictly decreasing degree, obtaining expressions of the form

$$g(x_1, x_2, \dots, x_n) = \sum_{\mathbf{k}} b_{\mathbf{k}} s_1^{k_1} \cdots s_{n-1}^{k_{n-1}} s_n^{k_n} + s_n h_m(x_1, x_2, \dots, x_n)$$

Eventually h_m must be a constant or zero. By induction, each step in the argument yields uniqueness and so, the final sum of combinations of elementary symmetric functions is uniquely determined. ■

Here is a very interesting result which I saw claimed in a paper by Steinberg and Redheffer on Lindemann's theorem which follows from the above theorem.

Theorem F.1.4 *Let $\alpha_1, \dots, \alpha_n$ be roots of the polynomial equation*

$$p(x) \equiv a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0$$

where each a_i is an integer. Then any symmetric polynomial in the quantities $a_n \alpha_1, \dots, a_n \alpha_n$ having integer coefficients is also an integer. Also any symmetric polynomial in the quantities $\alpha_1, \dots, \alpha_n$ having rational coefficients is a rational number.

Proof: Let $f(x_1, \dots, x_n)$ be the symmetric polynomial. Thus

$$f(x_1, \dots, x_n) \in \mathbb{Z}[x_1 \cdots x_n], \text{ the polynomials having integer coefficients}$$

From Theorem F.1.3 it follows there are integers $a_{k_1 \dots k_n}$ such that

$$f(x_1, \dots, x_n) = \sum_{k_1 + \dots + k_n \leq m} a_{k_1 \dots k_n} p_1^{k_1} \cdots p_n^{k_n}$$

where the p_i are the elementary symmetric polynomials defined as the coefficients of

$$\prod_{j=1}^n (x - x_j)$$

Earlier we had them \pm these coefficients. Thus

$$\begin{aligned} & f(a_n \alpha_1, \dots, a_n \alpha_n) \\ &= \sum_{k_1 + \dots + k_n = d} a_{k_1 \dots k_n} p_1^{k_1}(a_n \alpha_1, \dots, a_n \alpha_n) \cdots p_n^{k_n}(a_n \alpha_1, \dots, a_n \alpha_n) \end{aligned}$$

Now the given polynomial $p(x)$ is of the form

$$\begin{aligned} a_n \prod_{j=1}^n (x - \alpha_j) &\equiv a_n \left(\sum_{k=0}^n p_k(\alpha_1, \dots, \alpha_n) x^{n-k} \right) \\ &= a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \end{aligned}$$

Thus, equating coefficients, $a_n p_k(\alpha_1, \dots, \alpha_n) = a_{n-k}$. Multiply both sides by a_n^{k-1} . Thus

$$p_k(a_n \alpha_1, \dots, a_n \alpha_n) = a_n^{k-1} a_{n-k}$$

an integer. Therefore,

$$\begin{aligned} & f(a_n \alpha_1, \dots, a_n \alpha_n) \\ &= \sum_{k_1 + \dots + k_n = d} a_{k_1 \dots k_n} p_1^{k_1}(a_n \alpha_1, \dots, a_n \alpha_n) \cdots p_n^{k_n}(a_n \alpha_1, \dots, a_n \alpha_n) \end{aligned}$$

and each $p_k(a_n \alpha_1, \dots, a_n \alpha_n)$ is an integer. Thus $f(a_n \alpha_1, \dots, a_n \alpha_n)$ is an integer as claimed. From this, it is obvious that $f(\alpha_1, \dots, \alpha_n)$ is rational. Indeed,

$$f(\alpha_1, \dots, \alpha_n) = \sum_{k_1 + \dots + k_n = d} a_{k_1 \dots k_n} p_1^{k_1}(\alpha_1, \dots, \alpha_n) \cdots p_n^{k_n}(\alpha_1, \dots, \alpha_n)$$

Now multiply both sides by $a_n a_n^2 a_n^3 \cdots a_n^n$, an integer. Then

$$a_n a_n^2 a_n^3 \cdots a_n^n f(\alpha_1, \dots, \alpha_n) = \sum_{k_1 + \dots + k_n = d} a_{k_1 \dots k_n} p_1^{k_1}(a_n \alpha_1, \dots, a_n \alpha_n) \cdots p_n^{k_n}(a_n \alpha_1, \dots, a_n \alpha_n)$$

with the right side an integer. Thus $f(\alpha_1, \dots, \alpha_n)$ is rational. If the f had rational coefficients, then mf would have integer coefficients for a suitable m and so $mf(\alpha_1, \dots, \alpha_n)$ would be rational which yields $f(\alpha_1, \dots, \alpha_n)$ is rational. ■

F.2 The Fundamental Theorem Of Algebra

This is devoted to a mostly algebraic proof of the fundamental theorem of algebra. It depends on the interesting results about symmetric polynomials which are presented above. I found it on the Wikipedia article about the fundamental theorem of algebra. You google “fundamental theorem of algebra” and go to the Wikipedia article. It gives several other proofs in addition to this one. According to this article, the first completely correct proof of this major theorem is due to Argand in 1806. Gauss and others did it earlier but their arguments had gaps in them.

You can’t completely escape analysis when you prove this theorem. The necessary analysis is in the following lemma.

Lemma F.2.1 *Suppose $p(x) = x^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0$ where n is odd and the coefficients are real. Then $p(x)$ has a real root.*

Proof: This follows from the intermediate value theorem from calculus.

Next is an algebraic consideration. First recall some notation.

$$\prod_{i=1}^m a_i \equiv a_1 a_2 \cdots a_m$$

Brain power

By 2020, wind could provide one-tenth of our planet's electricity needs. Already today, SKF's innovative know-how is crucial to running a large proportion of the world's wind turbines.

Up to 25 % of the generating costs relate to maintenance. These can be reduced dramatically thanks to our systems for on-line condition monitoring and automatic lubrication. We help make it more economical to create cleaner, cheaper energy out of thin air.

By sharing our experience, expertise, and creativity, industries can boost performance beyond expectations. Therefore we need the best employees who can meet this challenge!

The Power of Knowledge Engineering

Plug into The Power of Knowledge Engineering. Visit us at www.skf.com/knowledge

SKF



Recall a polynomial in $\{z_1, \dots, z_n\}$ is symmetric only if it can be written as a sum of elementary symmetric polynomials raised to various powers multiplied by constants.

The following is the main part of the theorem. In fact this is one version of the fundamental theorem of algebra which people studied earlier in the 1700's.

Lemma F.2.2 *Let $p(x) = x^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0$ be a polynomial with real coefficients. Then it has a complex root.*

Proof: It is possible to write

$$n = 2^k m$$

where m is odd. If n is odd, $k = 0$. If n is even, keep dividing by 2 until you are left with an odd number. If $k = 0$ so that n is odd, it follows from Lemma F.2.1 that $p(x)$ has a real, hence complex root. The proof will be by induction on k , the case $k = 0$ being done. Suppose then that it works for $n = 2^l m$ where m is odd and $l \leq k - 1$ and let $n = 2^k m$ where m is odd. Let $\{z_1, \dots, z_n\}$ be the roots of the polynomial in a splitting field, the existence of this field being given by the above proposition. Then

$$p(x) = \prod_{j=1}^n (x - z_j) = \sum_{k=0}^n (-1)^k p_k(z_1, \dots, z_n) x^k \tag{6.1}$$

where $p_k(z_1, \dots, z_n)$ is the k^{th} elementary symmetric polynomial. Note this shows

$$a_{n-k} = p_k(z_1, \dots, z_n) (-1)^k, \text{ a real number.} \tag{6.2}$$

There is another polynomial which has coefficients which are sums of real numbers times the p_k raised to various powers and it is

$$q_t(x) \equiv \prod_{1 \leq i < j \leq n} (x - (z_i + z_j + tz_i z_j)), \quad t \in \mathbb{R}$$

I need to verify this is really the case for $q_t(x)$. When you switch any two of the z_i in $q_t(x)$ the polynomial does not change. Thus the coefficients of $q_t(x)$ must be symmetric polynomials in the z_i with real coefficients. Hence by Proposition F.1.3 these coefficients are real polynomials in terms of the elementary symmetric polynomials p_k . Thus by 6.2 the coefficients of $q_t(x)$ are real polynomials in terms of the a_k of the original polynomial. Recall these were all real. It follows, and this is what was wanted, that $q_t(x)$ has all real coefficients.

Note that the degree of $q_t(x)$ is $\binom{n}{2}$ because there are this number of ways to pick $i < j$ out of $\{1, \dots, n\}$. Now

$$\begin{aligned} \binom{n}{2} &= \frac{n(n-1)}{2} = 2^{k-1} m (2^k m - 1) \\ &= 2^{k-1} (\text{odd}) \end{aligned}$$

and so by induction, for each $t \in \mathbb{R}$, $q_t(x)$ has a complex root.

There must exist $s \neq t$ such that for a single pair of indices i, j , with $i < j$,

$$(z_i + z_j + tz_i z_j), (z_i + z_j + sz_i z_j)$$

are both complex. Here is why. Let $A(i, j)$ denote those $t \in \mathbb{R}$ such that $(z_i + z_j + tz_i z_j)$ is complex. It was just shown that every $t \in \mathbb{R}$ must be in some $A(i, j)$. There are infinitely many $t \in \mathbb{R}$ and so some $A(i, j)$ contains two of them.

Now for that t, s ,

$$\begin{aligned} z_i + z_j + tz_i z_j &= a \\ z_i + z_j + sz_i z_j &= b \end{aligned}$$

where $t \neq s$ and so by Cramer's rule,

$$z_i + z_j = \frac{\begin{vmatrix} a & t \\ b & s \end{vmatrix}}{\begin{vmatrix} 1 & t \\ 1 & s \end{vmatrix}} \in \mathbb{C}$$

and also

$$z_i z_j = \frac{\begin{vmatrix} 1 & a \\ 1 & b \end{vmatrix}}{\begin{vmatrix} 1 & t \\ 1 & s \end{vmatrix}} \in \mathbb{C}$$

At this point, note that z_i, z_j are both solutions to the equation

$$x^2 - (z_1 + z_2)x + z_1 z_2 = 0,$$

which from the above has complex coefficients. By the quadratic formula the z_i, z_j are both complex. Thus the original polynomial has a complex root. ■

With this lemma, it is easy to prove the fundamental theorem of algebra. The difference between the lemma and this theorem is that in the theorem, the coefficients are only assumed to be complex. What this means is that if you have any polynomial with complex coefficients it has a complex root and so it is not irreducible. Hence the field extension is the same field. Another way to say this is that for **every** complex polynomial there exists a factorization into linear factors or in other words a splitting field for a complex polynomial is the field of complex numbers.

Theorem F.2.3 *Let $p(x) \equiv a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$ be any complex polynomial, $n \geq 1, a_n \neq 0$. Then it has a complex root. Furthermore, there exist complex numbers z_1, \dots, z_n such that*

$$p(x) = a_n \prod_{k=1}^n (x - z_k)$$

Proof: First suppose $a_n = 1$. Consider the polynomial $q(x) \equiv p(x) \overline{p(\bar{x})}$

$$\begin{aligned} &(x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0) \cdot \\ &(x^n + \overline{a_{n-1}} x^{n-1} + \dots + \overline{a_1} x + \overline{a_0}) \end{aligned}$$

This polynomial has real coefficients because the coefficient of x^m is of the form

$$\sum_{k=0}^m a^{m-k} \overline{a_k}$$

and the sum involves adding terms of the form

$$a_k \overline{a_j} + \overline{a_k} a_j = a_k \overline{a_j} + \overline{a_k a_j}$$

so it is of the form of a complex number added to its conjugate. Hence $q(x)$ has real coefficients as claimed. Therefore, by Lemma F.2.2 it has a complex root z . Hence either $p(z) = 0$ or $p(\bar{z}) = 0$. Thus $p(x)$ has a complex root.

Next suppose $a_n \neq 1$. Then simply divide by it and get a polynomial in which $a_n = 1$. Denote this modified polynomial as $q(x)$. Then by what was just shown and the Euclidean algorithm, there exists $z_1 \in \mathbb{C}$ such that

$$q(x) = (x - z_1) q_1(x)$$

where $q_1(x)$ has complex coefficients. Now do the same thing for $q_1(x)$ to obtain

$$q(x) = (x - z_1)(x - z_2)q_2(x)$$

and continue this way. Thus

$$\frac{p(x)}{a_n} = \prod_{j=1}^n (x - z_j) \blacksquare$$

F.3 Transcendental Numbers

Most numbers are like this. Here the algebraic numbers are those which are roots of a polynomial equation having rational numbers as coefficients. By the fundamental theorem of algebra, all these numbers are in \mathbb{C} . There are only countably many of these algebraic numbers, (Problem 41 on Page 215). Therefore, most numbers are transcendental. Nevertheless, it is very hard to prove that this or that number is transcendental. Probably the most famous theorem about this is the Lindemann Weierstrass theorem.

Theorem F.3.1 *Let the α_i be distinct nonzero algebraic numbers and let the a_i be nonzero algebraic numbers. Then*

$$\sum_{i=1}^n a_i e^{\alpha_i} \neq 0$$


I am following the interesting Wikipedia article on this subject. You can also look at the book by Baker [4], Transcendental Number Theory, Cambridge University Press. There are also many other treatments which you can find on the web including an interesting article by Steinberg and Redheffer which appeared in about 1950 part of which I am following here.

Trust and responsibility

NNE and Pharmaplan have joined forces to create NNE Pharmaplan, the world's leading engineering and consultancy company focused entirely on the pharma and biotech industries.

Inés Aréizaga Esteve (Spain), 25 years old
Education: Chemical Engineer

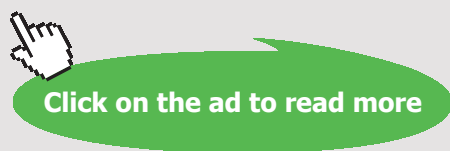
– You have to be proactive and open-minded as a newcomer and make it clear to your colleagues what you are able to cope. The pharmaceutical field is new to me. But busy as they are, most of my colleagues find the time to teach me, and they also trust me. Even though it was a bit hard at first, I can feel over time that I am beginning to be taken seriously and that my contribution is appreciated.



NNE Pharmaplan is the world's leading engineering and consultancy company focused entirely on the pharma and biotech industries. We employ more than 1500 people worldwide and offer global reach and local knowledge along with our all-encompassing list of services.

nnepharmaplan.com

nne pharmaplan®



The proof makes use of the following identity. For $f(x)$ a polynomial,

$$I(s) \equiv \int_0^s e^{s-x} f(x) dx = e^s \sum_{j=0}^{\deg(f)} f^{(j)}(0) - \sum_{j=0}^{\deg(f)} f^{(j)}(s). \tag{6.3}$$

where $f^{(j)}$ denotes the j^{th} derivative. In this formula, $s \in \mathbb{C}$ and the integral is defined in the natural way as

$$\int_0^1 sf(ts) e^{s-ts} dt \tag{6.4}$$

The identity follows from integration by parts.

$$\begin{aligned} \int_0^1 sf(ts) e^{s-ts} dt &= se^s \int_0^1 f(ts) e^{-ts} dt \\ &= se^s \left[-\frac{e^{-ts}}{s} f(ts) \Big|_0^1 + \int_0^1 \frac{e^{-ts}}{s} sf'(st) dt \right] \\ &= se^s \left[-\frac{e^{-s}}{s} f(s) + \frac{1}{s} f(0) + \int_0^1 e^{-ts} f'(st) dt \right] \\ &= e^s f(0) - f(s) + \int_0^1 se^{s-ts} f'(st) dt \\ &\equiv e^s f(0) - f(s) + \int_0^s e^{s-x} f'(x) dx \end{aligned}$$

Continuing this way establishes the identity since at the right end looks just like what we started with except with a derivative on the f .

Lemma F.3.2 *If K and c are nonzero integers, and β_1, \dots, β_m are the roots of a single polynomial with integer coefficients,*

$$Q(x) = vx^m + \dots + u$$

where $v, u \neq 0$, then

$$K + c(e^{\beta_1} + \dots + e^{\beta_m}) \neq 0.$$

Letting

$$f(x) = \frac{v^{(m+1)p} Q^p(x) x^{p-1}}{(p-1)!}$$

and $I(s)$ be defined in terms of $f(x)$ as above, it follows,

$$\lim_{p \rightarrow \infty} \sum_{i=1}^m I(\beta_i) = 0$$

and

$$\begin{aligned} \sum_{j=0}^n f^{(j)}(0) &= v^{p(m+1)} u^p + m_1(p)p \\ \sum_{i=1}^m \sum_{j=0}^n f^{(j)}(\beta_i) &= m_2(p)p \end{aligned}$$

where $m_i(p)$ is some integer.

Proof: Let p be a large prime number. Then consider the polynomial $f(x)$ of degree $n \equiv pm + p - 1$,

$$f(x) = \frac{v^{(m+1)p} Q^p(x) x^{p-1}}{(p-1)!}$$

From 6.3,

$$c \sum_{i=1}^m I(\beta_i) = c \sum_{i=1}^m \left(e^{\beta_i} \sum_{j=0}^n f^{(j)}(0) - \sum_{j=0}^n f^{(j)}(\beta_i) \right)$$

$$= \left(K + c \sum_{i=1}^m e^{\beta_i} \right) \sum_{j=0}^n f^{(j)}(0) - K \sum_{j=0}^n f^{(j)}(0) - c \sum_{i=1}^m \sum_{j=0}^n f^{(j)}(\beta_i) \tag{6.5}$$

Claim 1: $\lim_{p \rightarrow \infty} c \sum_{i=1}^m I(\beta_i) = 0$.

Proof: This follows right away from the definition of $I(\beta_j)$ and the definition of $f(x)$.

$$\begin{aligned} |I(\beta_j)| &\leq \int_0^1 |\beta_j f(t\beta_j) e^{\beta_j - t\beta_j}| dt \\ &\leq \int_0^1 \left| \frac{|v|^{(m-1)p} |Q(t\beta_j)|^p t^{p-1} |\beta_j|^{p-1}}{(p-1)!} dt \right| \end{aligned}$$

which clearly converges to 0. This proves the claim.

The next thing to consider is the term on the end in 6.5,

$$K \sum_{j=0}^n f^{(j)}(0) + c \sum_{i=1}^m \sum_{j=0}^n f^{(j)}(\beta_i) \tag{6.6}$$

The idea is to show that for large enough p it is always a nonzero integer. When this is done, it can't happen that $K + c \sum_{i=1}^m e^{\beta_i} = 0$ because if this were so, you would have a very small number equal to an integer.

$$f(x) = \frac{v^{(m+1)p} (vx^m + \dots + u)^p x^{p-1}}{(p-1)!}$$

Then $f^j(0) = 0$ unless $j \geq p - 1$ because otherwise, that x^{p-1} term will result in some $x^r, r > 0$ and everything is zero when you plug in $x = 0$. Now say $j = p - 1$. Then it is clear that

$$f^{(p-1)}(0) = u^p v^{(m+1)p}$$

So what if $j > p - 1$? Then by Leibniz formula,

$$f^j(x) = \binom{j}{p-1} v^{(m+1)p} \frac{d}{dx^{j-(p-1)}} [(vx^m + \dots + u)^p + Stuf f$$

where the *Stuf f* equals 0 when $x = 0$. Thus $f^j(0) = pm_j$ where m_j is some integer depending on the integer coefficients of the polynomial $Q(x)$. Therefore,

$$\sum_{j=0}^n f^{(j)}(0) = v^{(m+1)p} u^p + m(p)p \tag{6.7}$$

where $m(p)$ is some integer.

Now consider the other sum in 6.6,

$$c \sum_{i=1}^m \sum_{j=0}^n f^{(j)}(\beta_i)$$

Also it follows that

$$f(x) = \frac{v^{(m+1)p} ((x - \beta_1)(x - \beta_2) \dots (x - \beta_m))^p x^{p-1}}{(p-1)!}$$

it follows that for $j < p$, $f^{(j)}(\beta_i) = 0$. This is because for such derivatives, each term will have that product of the $(x - \beta_i)$ in it.

To get something non zero, the nonzero terms must involve at least p derivatives of the expression

$$((x - \beta_1)(x - \beta_2) \dots (x - \beta_m))^p$$

since otherwise, when evaluated at any β_k the result would be 0.

Now say $j \geq p$. Then by Liebnez formula, $f^j(x)$ is of the form

$$\begin{aligned} & \frac{v^{(m+1)p}}{(p-1)!} \sum_{r=0}^j \binom{j}{r} \frac{d}{dx^r} (((x - \beta_1)(x - \beta_2) \cdots (x - \beta_m))^p) \frac{d}{dx^{j-r}} x^{p-1} \\ = & \frac{v^{p(m+1)-2p+1}}{(p-1)!} \sum_{r=0}^j \binom{j}{r} \frac{d}{dx^r} (((vx - v\beta_1)(vx - v\beta_2) \cdots (vx - v\beta_m))^p) \frac{d}{dx^{j-r}} (vx)^{p-1} \end{aligned}$$

Note that for r too small, the term will be zero when evaluated at any of the β_i . You only get something nonzero if $r \geq p$ and so there will be a $p!$ produced which will cancel with the $(p-1)!$ to yield an extra p .

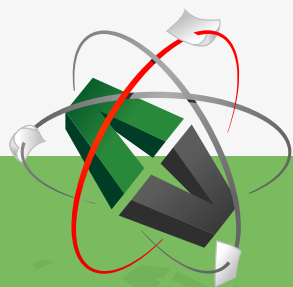
Now if you do the computations using the product rule and then replace x with β_i and sum these over all $v\beta_i$, you will get a symmetric polynomial in the quantities $\{v\beta_1, \dots, v\beta_m\}$ and by Theorem F.1.4 this is an integer. To see this is symmetric note that switching $v\beta_a, v\beta_b$ in

$$\frac{d}{dx^r} ((vx - v\beta_1)(vx - v\beta_2) \cdots (vx - v\beta_m))$$

does not change anything. The other term is just $v^{p-1} (p-1)(p-2) \cdots (p-j+r) x^{p-j+r-1}$ or zero if $j-r > p-1$. It follows that when adding these over i ,

$$c \sum_{i=1}^m \sum_{j=0}^n f^{(j)}(\beta_i) = L(p)p$$

This e-book
is made with
SetaPDF



PDF components for PHP developers

www.setasign.com



where $L(p)$ is some integer. Therefore, 6.6 is of the form

$$Kv^{p(m+1)}u^p + m(p)p + L(p)p \equiv Kv^{p(m+1)}u^p + M(p)p$$

for some integer $M(p)$. Summarizing, it follows

$$c \sum_{i=1}^m I(\beta_i) = \left(K + c \sum_{i=1}^m e^{\beta_i} \right) \overbrace{\sum_{j=0}^n f^{(j)}(0)}^{\neq 0} + Kv^{p(m+1)}u^p + M(p)p$$

where the left side is very small whenever p is large enough. Let p be larger than $\max(K, v, u)$. Since p is prime, it follows that it cannot divide $Kv^{p(m+1)}u^p$ and so the last two terms must sum to a nonzero integer and so the equation 6.5 cannot hold unless

$$K + c \sum_{i=1}^m e^{\beta_i} \neq 0 \quad \blacksquare$$

Note that this shows π is irrational. If $\pi = k/m$ where k, m are integers, then both $i\pi$ and $-i\pi$ are roots of the polynomial with integer coefficients,

$$m^2x^2 + k^2$$

which would require, from what was just shown that

$$0 \neq 2 + e^{i\pi} + e^{-i\pi}$$

which is not the case since the sum on the right equals 0.

The following corollary follows from this.

Corollary F.3.3 *Let K and c_i for $i = 1, \dots, n$ be nonzero integers. For each k between 1 and n let $\{\beta(k)_i\}_{i=1}^{m_k}$ be the roots of a polynomial with integer coefficients,*

$$Q_k(x) \equiv v_kx^{m_k} + \dots + u_k$$

where $v_k, u_k \neq 0$. Then

$$K + c_1 \left(\sum_{j=1}^{m_1} e^{\beta(1)_j} \right) + c_2 \left(\sum_{j=1}^{m_2} e^{\beta(2)_j} \right) + \dots + c_n \left(\sum_{j=1}^{m_n} e^{\beta(n)_j} \right) \neq 0.$$

Proof: Defining $f_k(x)$ and $I_k(s)$ as in Lemma F.3.2, it follows from Lemma F.3.2 that for each $k = 1, \dots, n$,

$$\begin{aligned} c_k \sum_{i=1}^{m_k} I_k(\beta(k)_i) &= \left(K_k + c_k \sum_{i=1}^{m_k} e^{\beta(k)_i} \right) \sum_{j=0}^{\deg(f_k)} f_k^{(j)}(0) \\ &\quad - K_k \sum_{j=0}^{\deg(f_k)} f_k^{(j)}(0) - c_k \sum_{i=1}^{m_k} \sum_{j=0}^{\deg(f_k)} f_k^{(j)}(\beta(k)_i) \end{aligned}$$

This is exactly the same computation as in the beginning of that lemma except one adds and subtracts $K_k \sum_{j=0}^{\deg(f_k)} f_k^{(j)}(0)$ rather than $K \sum_{j=0}^{\deg(f_k)} f_k^{(j)}(0)$ where the K_k are chosen such that their sum equals K . By Lemma F.3.2,

$$\begin{aligned} c_k \sum_{i=1}^{m_k} I_k(\beta(k)_i) &= \left(K_k + c_k \sum_{i=1}^{m_k} e^{\beta(k)_i} \right) \left(v_k^{(m_k+1)p} u_k^p + N_k p \right) \\ &\quad - K_k \left(v_k^{(m_k+1)p} u_k^p + N_k p \right) - c_k N_k' p \end{aligned}$$

and so

$$c_k \sum_{i=1}^{m_k} I_k(\beta(k)_i) = \left(K_k + c_k \sum_{i=1}^{m_k} e^{\beta(k)_i} \right) \left(v_k^{(m_k+1)p} u_k^p + N_k p \right) - K_k v_k^{(m_k+1)p} u_k^p + M_k p$$

for some integer M_k . By multiplying each $Q_k(x)$ by a suitable constant depending on k , it can be assumed without loss of generality that all the $v_k^{m_k+1} u_k$ are equal to a constant integer U . Then the above equals

$$c_k \sum_{i=1}^{m_k} I_k(\beta(k)_i) = \left(K_k + c_k \sum_{i=1}^{m_k} e^{\beta(k)_i} \right) (U^p + N_k p) - K_k U^p + M_k p$$

Adding these for all k gives

$$\begin{aligned} \sum_{k=1}^n c_k \sum_{i=1}^{m_k} I_k(\beta(k)_i) &= U^p \left(K + \sum_{k=1}^n c_k \sum_{i=1}^{m_k} e^{\beta(k)_i} \right) - K U^p + M p \\ &+ \sum_{k=1}^n N_k p \left(K_k + c_k \sum_{i=1}^{m_k} e^{\beta(k)_i} \right) \end{aligned} \tag{6.8}$$

For large p it follows from Lemma F.3.2 that the left side is very small. If

$$K + \sum_{k=1}^n c_k \sum_{i=1}^{m_k} e^{\beta(k)_i} = 0$$

then $\sum_{k=1}^n c_k \sum_{i=1}^{m_k} e^{\beta(k)_i}$ is an integer and so the last term in 6.8 is an integer times p . Thus for large p it reduces to

$$\text{small number} = -K U^p + I p$$

where I is an integer. Picking prime $p > \max(U, K)$ it follows $-K U^p + I p$ is a nonzero integer and this contradicts the left side being a small number less than 1 in absolute value. ■

Next is an even more interesting Lemma which follows from the above corollary.

Lemma F.3.4 *If b_0, b_1, \dots, b_n are non zero integers, and $\gamma_1, \dots, \gamma_n$ are distinct algebraic numbers, then*

$$b_0 e^{\gamma_0} + b_1 e^{\gamma_1} + \dots + b_n e^{\gamma_n} \neq 0$$

Proof: Assume

$$b_0 e^{\gamma_0} + b_1 e^{\gamma_1} + \dots + b_n e^{\gamma_n} = 0 \tag{6.9}$$

Divide by e^{γ_0} and letting $K = b_0$,

$$K + b_1 e^{\alpha(1)} + \dots + b_n e^{\alpha(n)} = 0 \tag{6.10}$$

where $\alpha(k) = \gamma_k - \gamma_0$. These are still distinct algebraic numbers none of which is 0 thanks to Theorem 7.3.32. Therefore, $\alpha(k)$ is a root of a polynomial

$$v_k x^{m_k} + \dots + u_k \tag{6.11}$$

having integer coefficients, $v_k, u_k \neq 0$. Recall algebraic numbers were defined as roots of polynomial equations having rational coefficients. Just multiply by the denominators to get one with integer coefficients. Let the roots of this polynomial equation be

$$\{\alpha(k)_1, \dots, \alpha(k)_{m_k}\}$$

and suppose they are listed in such a way that $\alpha(k)_1 = \alpha(k)$. Letting i_k be an integer in $\{1, \dots, m_k\}$ it follows from the assumption 6.9 that

$$\prod_{\substack{(i_1, \dots, i_n) \\ i_k \in \{1, \dots, m_k\}}} \left(K + b_1 e^{\alpha(1)_{i_1}} + b_2 e^{\alpha(2)_{i_2}} + \dots + b_n e^{\alpha(n)_{i_n}} \right) = 0 \quad (6.12)$$

This is because one of the factors is the one occurring in 6.10 when $i_k = 1$ for every k . The product is taken over all distinct ordered lists (i_1, \dots, i_n) where i_k is as indicated. Expand this possibly huge product. This will yield something like the following.

$$K^r + c_1 \left(e^{\beta(1)_1} + \dots + e^{\beta(1)_{\mu(1)}} \right) + c_2 \left(e^{\beta(2)_1} + \dots + e^{\beta(2)_{\mu(2)}} \right) + \dots + c_N \left(e^{\beta(N)_1} + \dots + e^{\beta(N)_{\mu(N)}} \right) = 0 \quad (6.13)$$

These integers c_j come from products of the b_i and K . The $\beta(i)_j$ are the distinct exponents which result. Note that a typical term in this product 6.12 would be something like

$$\underbrace{K^p b_{k_1} \dots b_{k_{n-p}}}_{\text{integer}} e^{\overbrace{\alpha(k_1)_{i_1} + \alpha(k_2)_{i_2} + \dots + \alpha(k_{n-p})_{i_{n-p}}}^{\beta(j)_r}}$$

the k_r possibly not distinct and each $i_k \in \{1, \dots, m_{i_k}\}$. A given term in the sum of 6.13 corresponds to such a choice of $\{b_{k_1}, \dots, b_{k_{n-p}}\}$ which leads to $K^p b_{k_1} \dots b_{k_{n-p}}$ times a sum of exponentials like those just described. Since the product in 6.12 is taken over all choices $i_k \in \{1, \dots, m_k\}$, it follows that if you switch $\alpha(r)_i$ and $\alpha(r)_j$, two of the roots of the polynomial

$$v_r x^{m_r} + \dots + u_r$$

FOSS

"I studied Sharp Minds - Bright Ideas! English for 16 years but... I finally learned to speak it in just six lessons!"
Jane, Chinese architect

Employees at FOSS Analytical AS are living proof of the company value - First - using new inventions to make dedicated solutions for our customers. With sharp minds and cross functional teamwork, we constantly strive to develop new unique products - **What do you like to join our team?**

FOSS works diligently with innovation and development as basis for its growth. It is reflected in the fact that more than 200 of the 1200 employees in FOSS work with Research & Development in Scandinavia and USA. Engineers at FOSS work in production, development and marketing within a wide range of different fields: Chemistry, Electronics, Mechanics, Software, Optics, Microbiology, Chromometrics.

A challenging job in an international and innovative company that is leading in its field. You will get the opportunity to work with the most advanced technology together with highly skilled colleagues.

Read more about FOSS at www.foss.dk - or go directly to our student site www.foss.dk/sharpminds where you can learn more about a variety of working together with us on projects, your thesis etc.

Dedicated Analytical Solutions
FOSS
Slangerupgade 69
3400 Hillerød
Tel. +45 70103370
www.foss.dk

The Family owned FOSS group is the world leader as supplier of dedicated, high-tech analytical solutions which measure and control the quality and production of agricultural, food, pharmaceutical and chemical products. Main activities are initiated from Denmark, Sweden and USA with headquarters domiciled in Hillerød, DK. The products are marketed globally by 23 sales companies and an extensive net of distributors. In line with the core value to be 'First', the company intends to expand before and after





Click to hear me before and after



mentioned above, the result in 6.13 would be the same except for permuting the

$$\beta(s)_1, \beta(s)_2, \dots, \beta(s)_{\mu(s)}.$$

Thus a symmetric polynomial in

$$\beta(s)_1, \beta(s)_2, \dots, \beta(s)_{\mu(s)}$$

is also a symmetric polynomial in the $\alpha(k)_1, \alpha(k)_2, \dots, \alpha(k)_{m_k}$ for each k . Thus for a given $r, \beta(r)_1, \dots, \beta(r)_{\mu(r)}$ are roots of the polynomial

$$(x - \beta(r)_1)(x - \beta(r)_2) \cdots (x - \beta(r)_{\mu(r)})$$

whose coefficients are symmetric polynomials in the $\beta(r)_j$ which is a symmetric polynomial in the $\alpha(k)_j, j = 1, \dots, m_k$ for each k . Letting g be one of these symmetric polynomials, a coefficient of the above polynomial, and writing it in terms of the $\alpha(k)_i$ you would have

$$\sum_{l_1, \dots, l_n} A_{l_1 \dots l_n} \alpha(n)_1^{l_1} \alpha(n)_2^{l_2} \cdots \alpha(n)_{m_n}^{l_n}$$

where $A_{l_1 \dots l_n}$ is a symmetric polynomial in $\alpha(k)_j, j = 1, \dots, m_k$ for each $k \leq n - 1$. (It is desired to show g is rational.) These coefficients are in the field (Proposition 7.3.31) $\mathbb{Q}[A(1), \dots, A(n - 1)]$ where $A(k)$ denotes

$$\{\alpha(k)_1, \dots, \alpha(k)_{m_k}\}$$

and so from Theorem F.1.3, the above symmetric polynomial is of the form

$$\sum_{(k_1 \dots k_{m_n})} B_{k_1 \dots k_{m_n}} p_1^{k_1}(\alpha(n)_1, \dots, \alpha(n)_{m_n}) \cdots p_{m_n}^{k_{m_n}}(\alpha(n)_1, \dots, \alpha(n)_{m_n})$$

where $B_{k_1 \dots k_{m_n}}$ is a symmetric polynomial in $\alpha(k)_j, j = 1, \dots, m_k$ for each $k \leq n - 1$. Now do for each $B_{k_1 \dots k_{m_n}}$ what was just done for g featuring this time

$$\{\alpha(n - 1)_1, \dots, \alpha(n - 1)_{m_{n-1}}\}$$

and continuing this way, it must be the case that eventually you have a sum of integer multiples of products of elementary symmetric polynomials in $\alpha(k)_j, j = 1, \dots, m_k$ for each $k \leq n$. By Theorem F.1.4, these are rational numbers. Therefore, each such g is a rational number and so the $\beta(r)_j$ are algebraic roots of a polynomial having rational coefficients, hence also roots of one which has integer coefficients. Now 6.13 contradicts Corollary F.3.3. ■

Note this lemma is sufficient to prove Lindemann's theorem that π is transcendental. Here is why. If π is algebraic, then so is $i\pi$ and so from this lemma, $e^0 + e^{i\pi} \neq 0$ but this is not the case because $e^{i\pi} = -1$.

The next theorem is the main result, the Lindemann Weierstrass theorem.

Theorem F.3.5 *Suppose $a(1), \dots, a(n)$ are nonzero algebraic numbers and suppose*

$$\alpha(1), \dots, \alpha(n)$$

are distinct algebraic numbers. Then

$$a(1)e^{\alpha(1)} + a(2)e^{\alpha(2)} + \dots + a(n)e^{\alpha(n)} \neq 0$$

Proof: Suppose $a(j) \equiv a(j)_1$ is a root of the polynomial

$$v_j x^{m_j} + \dots + u_j$$

where $v_j, u_j \neq 0$. Let the roots of this polynomial be $a(j)_1, \dots, a(j)_{m_j}$. Suppose to the contrary that

$$a(1)_1 e^{\alpha(1)} + a(2)_1 e^{\alpha(2)} + \dots + a(n)_1 e^{\alpha(n)} = 0$$

Then consider the big product

$$\prod_{\substack{(i_1, \dots, i_n) \\ i_k \in \{1, \dots, m_k\}}} \left(a(1)_{i_1} e^{\alpha(1)} + a(2)_{i_2} e^{\alpha(2)} + \dots + a(n)_{i_n} e^{\alpha(n)} \right) \tag{6.14}$$

the product taken over all ordered lists (i_1, \dots, i_n) . This product equals

$$0 = b_1 e^{\beta(1)} + b_2 e^{\beta(2)} + \dots + b_N e^{\beta(N)} \tag{6.15}$$

where the $\beta(j)$ are the distinct exponents which result. The $\beta(i)$ are clearly algebraic because they are the sum of the $\alpha(i)$. Since the product in 6.14 is taken for all ordered lists as described above, it follows that for a given k , if $a(k)_i$ is switched with $a(k)_j$, that is, two of the roots of $v_k x^{m_k} + \dots + u_k$ are switched, then the product is unchanged and so 6.15 is also unchanged. Thus each b_k is a symmetric polynomial in the $a(k)_j, j = 1, \dots, m_k$ for each k . It follows

$$b_k = \sum_{(j_1, \dots, j_{m_n})} A_{j_1, \dots, j_{m_n}} a(n)_{j_1}^{j_1} \dots a(n)_{j_{m_n}}^{j_{m_n}}$$

and this is symmetric in the $\{a(n)_1, \dots, a(n)_{m_n}\}$ the coefficients $A_{j_1, \dots, j_{m_n}}$ being in the field (Proposition 7.3.31) $\mathbb{Q}[A(1), \dots, A(n-1)]$ where $A(k)$ denotes

$$a(k)_1, \dots, a(k)_{m_k}$$

and so from Theorem F.1.3,

$$b_k = \sum_{(j_1, \dots, j_{m_n})} B_{j_1, \dots, j_{m_n}} p_1^{j_1} (a(n)_1 \dots a(n)_{m_n}) \dots p_{m_n}^{j_{m_n}} (a(n)_1 \dots a(n)_{m_n})$$

where the $B_{j_1, \dots, j_{m_n}}$ are symmetric in $\{a(k)_j\}_{j=1}^{m_k}$ for each $k \leq n-1$. Now doing to $B_{j_1, \dots, j_{m_n}}$ what was just done to b_k and continuing this way, it follows b_k is a finite sum of integers times elementary polynomials in the various $\{a(k)_j\}_{j=1}^{m_k}$ for $k \leq n$. By Theorem F.1.4 this is a rational number. Thus b_k is a rational number. Multiplying by the product of all the denominators, it follows there exist integers c_i such that

$$0 = c_1 e^{\beta(1)} + c_2 e^{\beta(2)} + \dots + c_N e^{\beta(N)}$$

which contradicts Lemma F.3.4. ■

This theorem is sufficient to show e is transcendental. If it were algebraic, then

$$e e^{-1} + (-1) e^0 \neq 0$$

but this is not the case. If $a \neq 1$ is algebraic, then $\ln(a)$ is transcendental. To see this, note that

$$1 e^{\ln(a)} + (-1) a e^0 = 0$$

which cannot happen if $\ln(a)$ is algebraic according to the above theorem. If a is algebraic and $\sin(a) \neq 0$, then $\sin(a)$ is transcendental because

$$\frac{1}{2i} e^{ia} - \frac{1}{2i} e^{-ia} + (-1) \sin(a) e^0 = 0$$

which cannot occur if $\sin(a)$ is algebraic. There are doubtless other examples of numbers which are transcendental by this amazing theorem.

F.4 More On Algebraic Field Extensions

The next few sections have to do with fields and field extensions. There are many linear algebra techniques which are used in this discussion and it seems to me to be very interesting. However, this is definitely far removed from my own expertise so there may be some parts of this which are not too good. I am following various algebra books in putting this together.

Consider the notion of splitting fields. It is desired to show that any two are isomorphic, meaning that there exists a one to one and onto mapping from one to the other which preserves all the algebraic structure. To begin with, here is a theorem about extending homomorphisms. [18]

Definition F.4.1 Suppose $\mathbb{F}, \bar{\mathbb{F}}$ are two fields and that $f : \mathbb{F} \rightarrow \bar{\mathbb{F}}$ is a homomorphism. This means that

$$f(xy) = f(x)f(y), \quad f(x+y) = f(x) + f(y)$$

An isomorphism is a homomorphism which is one to one and onto. A monomorphism is a homomorphism which is one to one. An automorphism is an isomorphism of a single field. Sometimes people use the symbol \simeq to indicate something is an isomorphism. Then if $p(x) \in \mathbb{F}[x]$, say

$$p(x) = \sum_{k=0}^n a_k x^k,$$

$\bar{p}(x)$ will be the polynomial in $\bar{\mathbb{F}}[x]$ defined as

$$\bar{p}(x) \equiv \sum_{k=0}^n f(a_k) x^k.$$

"I studied English for 16 years but...
...I finally learned to speak it in just six lessons"

Jane, Chinese architect

ENGLISH OUT THERE

Click to hear me talking before and after my unique course download

Also consider f as a homomorphism of $\mathbb{F}[x]$ and $\bar{\mathbb{F}}[x]$ in the obvious way.

$$f(p(x)) = \bar{p}(x)$$

It is clear that if f is an isomorphism of the two fields $\mathbb{F}, \bar{\mathbb{F}}$, then it is also an isomorphism of the commutative rings $\mathbb{F}[x], \bar{\mathbb{F}}[x]$ meaning that it is one to one and onto and preserves the two operations of addition and multiplication.

The following is a nice theorem which will be useful.

Theorem F.4.2 Let \mathbb{F} be a field and let r be algebraic over \mathbb{F} . Let $p(x)$ be the minimal polynomial of r . Thus $p(r) = 0$ and $p(x)$ is monic and no nonzero polynomial having coefficients in \mathbb{F} of smaller degree has r as a root. In particular, $p(x)$ is irreducible over \mathbb{F} . Then define $f : \mathbb{F}[x] \rightarrow \mathbb{F}[r]$, the polynomials in r by

$$f\left(\sum_{i=0}^m a_i x^i\right) \equiv \sum_{i=0}^m a_i r^i$$

Then f is a homomorphism. Also, defining $g : \mathbb{F}[x]/(p(x))$ by

$$g([q(x)]) \equiv f(q(x)) \equiv q(r)$$

it follows that g is an isomorphism from the field $\mathbb{F}[x]/(p(x))$ to $\mathbb{F}[r]$.

Proof: First of all, consider why f is a homomorphism. The preservation of sums is obvious. Consider products.

$$\begin{aligned} f\left(\sum_i a_i x^i \sum_j b_j x^j\right) &= f\left(\sum_{i,j} a_i b_j x^{i+j}\right) = \sum_{ij} a_i b_j r^{i+j} \\ &= \sum_i a_i r^i \sum_j b_j r^j = f\left(\sum_i a_i x^i\right) f\left(\sum_j b_j x^j\right) \end{aligned}$$

Thus it is clear that f is a homomorphism.

First consider why g is even well defined. If $[q(x)] = [q_1(x)]$, this means that

$$q_1(x) - q(x) = p(x)l(x)$$

for some $l(x) \in \mathbb{F}[x]$. Therefore,

$$\begin{aligned} f(q_1(x)) &= f(q(x)) + f(p(x)l(x)) \\ &= f(q(x)) + f(p(x))f(l(x)) \\ &\equiv q(r) + p(r)l(r) = q(r) = f(q(x)) \end{aligned}$$

Now from this, it is obvious that g is a homomorphism.

$$\begin{aligned} g([q(x)][q_1(x)]) &= g([q(x)q_1(x)]) = f(q(x)q_1(x)) = q(r)q_1(r) \\ g([q(x)])g([q_1(x)]) &\equiv q(r)q_1(r) \end{aligned}$$

Similarly, g preserves sums. Now why is g one to one? It suffices to show that if $g([q(x)]) = 0$, then $[q(x)] = 0$. Suppose then that

$$g([q(x)]) \equiv q(r) = 0$$

Then

$$q(x) = p(x)l(x) + \rho(x)$$

where the degree of $\rho(x)$ is less than the degree of $p(x)$ or else $\rho(x) = 0$. If $\rho(x) \neq 0$, then it follows that

$$\rho(r) = 0$$

and $\rho(x)$ has smaller degree than that of $p(x)$ which contradicts the definition of $p(x)$ as the minimal polynomial of r . Thus $q(x) = p(x)l(x)$ and so $[q(x)] = 0$. Since $p(x)$ is irreducible, $\mathbb{F}[x]/(p(x))$ is a field. It is clear that g is onto. Therefore, $\mathbb{F}[r]$ is a field also. (This was shown earlier by different reasoning.) ■

Here is a diagram of what the following theorem says.

Extending f to g

$$\begin{array}{ccc}
 \mathbb{F} & \xrightarrow{f} & \bar{\mathbb{F}} \\
 \downarrow \cong & & \downarrow \cong \\
 p(x) \in \mathbb{F}[x] & \xrightarrow{f} & \bar{p}(x) \in \bar{\mathbb{F}}[x] \\
 p(x) = \sum_{k=0}^n a_k x^k & \rightarrow & \sum_{k=0}^n f(a_k) x^k = \bar{p}(x) \\
 p(r) = 0 & & \bar{p}(\bar{r}) = 0 \\
 \mathbb{F}[r] & \xrightarrow{g} & \bar{\mathbb{F}}[\bar{r}] \\
 \downarrow \cong & & \downarrow \cong \\
 r & \xrightarrow{g} & \bar{r}
 \end{array}$$

The idea illustrated is the following question: For r algebraic over \mathbb{F} and f an isomorphism of \mathbb{F} and $\bar{\mathbb{F}}$, when does there exist \bar{r} algebraic over $\bar{\mathbb{F}}$ and an isomorphism of $\mathbb{F}[r]$ and $\bar{\mathbb{F}}[\bar{r}]$ which extends f ? This is the content of the following theorem.

Theorem F.4.3 *Let $f : \mathbb{F} \rightarrow \bar{\mathbb{F}}$ be an isomorphism and let r be algebraic over \mathbb{F} with minimal polynomial $p(x)$. Then the following are equivalent.*

1. *There exists \bar{r} algebraic over $\bar{\mathbb{F}}$ such that $\bar{p}(\bar{r}) = 0$ in which case $\bar{p}(x)$ is the minimal polynomial of \bar{r} .*
2. *There exists $g : \mathbb{F}[r] \rightarrow \bar{\mathbb{F}}[\bar{r}]$ an isomorphism which extends f such that $g(r) = \bar{r}$. In this case, there is only one such isomorphism.*

Proof: 2.) \Rightarrow 1.) Let $g(r) = \bar{r}$ with g an isomorphism extending $f, g(r) = \bar{g}(\bar{r})$. Then since it is an isomorphism,

$$0 = g(p(r)) = \bar{p}(g(r)) = \bar{p}(\bar{r}) \tag{*}$$

Define β as $\beta([k(x)]) \equiv \bar{k}(\bar{r})$ relative to this $\bar{r} \equiv g(r)$ and let $\alpha : \mathbb{F}[x]/(p(x)) \rightarrow \mathbb{F}[r]$ be the isomorphism mentioned in Theorem F.4.2 called g there, given by $\alpha([k(x)]) \equiv k(r)$. Thus

$$\mathbb{F}[r] \xleftarrow{\alpha} \mathbb{F}[x]/(p(x)) \xrightarrow{\beta} \bar{\mathbb{F}}[\bar{r}]$$

Then if β is a well defined homomorphism, it follows that g must equal $\beta \circ \alpha^{-1}$ because

$$\beta \circ \alpha^{-1}(k(r)) \equiv \beta([k(x)]) \equiv \bar{k}(\bar{r}) \equiv \bar{k}(g(r)) = g(k(r)).$$

This is because g is a homomorphism which takes r to \bar{r} . It only remains to verify that β is well defined.

Why is β well defined? Suppose $[k(x)] = [k'(x)]$ so that $k(x) - k'(x) = l(x)p(x)$. Then since f is a homomorphism, it follows from * that

$$\bar{k}(x) - \bar{k}'(x) = \bar{l}(x)\bar{p}(x) \Rightarrow \bar{k}(\bar{r}) - \bar{k}'(\bar{r}) = \bar{l}(\bar{r})\bar{p}(\bar{r}) = 0$$

so β is indeed well defined. It is clear from the definition that β is a homomorphism.

1.) \Rightarrow 2.) Next suppose there exists \bar{r} algebraic over $\bar{\mathbb{F}}$ such that $\bar{p}(\bar{r}) = 0$. Why is $\bar{p}(x)$ the minimal polynomial of \bar{r} ? Call it $\bar{q}(x)$. There is no loss of generality because f is an isomorphism so the minimal polynomial can be written this way. Then $\beta([q(x)]) \equiv \bar{q}(\bar{r}) = 0 = \bar{p}(\bar{r})$. Then $\bar{p}(x) = \bar{q}(x)\bar{m}(x) + R(x)$ where the degree of $R(x)$ is less than the degree of $\bar{q}(x)$ or equal to zero and so $R(\bar{r}) = 0$ which is contrary to $\bar{q}(x)$ being minimal polynomial for \bar{r} unless $R(x) = 0$. Therefore, $R(x) = 0$. It follows that, since f is a isomorphism, we have $p(x) = q(x)m(x)$ contrary to $p(x)$ being the minimal polynomial for r . Indeed, if the degree of $q(x)$ is less than that of $p(x)$, we have $0 = p(r) = q(r)m(r)$ and so one of $q(r), m(r)$ equals 0 contrary to $p(x)$ having smallest possible degree for sending r to 0. Thus the degree of $q(x)$ is the same as the degree of $p(x)$ and since both are monic by definition, $m(x) = 1$. Hence $p(x) = q(x)$ and so $\bar{p}(x) = \bar{q}(x)$.

Now let α, β be defined as above. It was shown above that β is a well defined homomorphism. It is also clear that β is onto. It only remains to verify that β is one to one and when this is done, the isomorphism will be $\beta \circ \alpha^{-1}$. Suppose $\beta([k(x)]) \equiv \bar{k}(\bar{r}) = 0$. Does it follow that $[k(x)] = 0$? By assumption, $\bar{p}(\bar{r}) = 0$ and also,

$$\bar{k}(x) = \bar{p}(x)\bar{l}(x) + \bar{\rho}(x) \tag{*}$$

where the degree of $\bar{\rho}(x)$ is less than the degree of $\bar{p}(x)$ which is the same as the degree of $p(x)$ or else it equals 0. It follows that $\bar{\rho}(\bar{r}) = 0$ and this is a contradiction because $\bar{p}(x)$ is the minimal polynomial for \bar{r} which was shown above. Hence $\bar{k}(x) = \bar{p}(x)\bar{l}(x)$ and since f is an isomorphism, this says that $k(x) = p(x)l(x)$ and so $[k(x)] = 0$. Hence β is indeed one to one and so an example of g would be $\beta \circ \alpha^{-1}$. Also $\beta \circ \alpha^{-1}(r) = \beta([x]) = \bar{r}$. ■

What is the meaning of the above in simple terms? It says that the monomorphisms from $\mathbb{F}[r]$ to a field $\bar{\mathbb{K}}$ containing $\bar{\mathbb{F}}$ correspond to the roots of $\bar{p}(x)$ in $\bar{\mathbb{K}}$. That is, for each root of $\bar{p}(x)$, there is a monomorphism and for each monomorphism, there is a root. Also, for each root \bar{r} of $\bar{p}(x)$ in $\bar{\mathbb{K}}$, there is an isomorphism from $\mathbb{F}[r]$ to $\bar{\mathbb{F}}[\bar{r}]$. Here $p(x)$ is the minimal polynomial for r .

Note that if $p(x)$ is a monic **irreducible** polynomial, then it is the minimal polynomial for each of its roots. Consider why this is. If r is a root of $p(x)$, then let $q(x)$ be the minimal polynomial for r . Then

$$p(x) = q(x)k(x) + R(x)$$

where $R(x)$ is 0 or else has smaller degree than $q(x)$. However, $R(r) = 0$ and this contradicts $q(x)$ being the minimal polynomial of r . Hence $q(x)$ divides $p(x)$ or else $k(x) = 1$. The latter possibility must be the case because $p(x)$ is irreducible.

This is the situation which is about to be considered. It involves the splitting fields $\mathbb{K}, \bar{\mathbb{K}}$ of $p(x), \bar{p}(x)$ where η is an isomorphism of \mathbb{F} and $\bar{\mathbb{F}}$ as described above. See [18]. Here is a little diagram which describes what this theorem says.

The Wake
the only emission we want to leave behind

Low-speed Engines Medium-speed Engines Turbochargers Propellers Propulsion Packages PrimeServ

The design of eco-friendly marine power and propulsion solutions is crucial for MAN Diesel & Turbo. Power competencies are offered with the world's largest engine programme – having outputs spanning from 450 to 87,220 kW per engine. Get up front! Find out more at www.mandieselturbo.com

Engineering the Future – since 1758.
MAN Diesel & Turbo



Definition F.4.4 The symbol $[\mathbb{K} : \mathbb{F}]$ where \mathbb{K} is a field extension of \mathbb{F} means the dimension of the vector space \mathbb{K} with field of scalars \mathbb{F} .

$$\begin{array}{ccc}
 \mathbb{F} & \xrightarrow[\cong]{\eta} & \bar{\mathbb{F}} \\
 p(x) & \xrightarrow[\cong]{\eta p(x) = \bar{p}(x)} & \bar{p}(x) \\
 \mathbb{F}[r_1, \dots, r_n] & \xrightarrow[\cong]{\zeta_i} & \bar{\mathbb{F}}[\bar{r}_1, \dots, \bar{r}_n] \\
 & \left\{ \begin{array}{l} m \leq [\mathbb{K} : \mathbb{F}] \\ m = [\mathbb{K} : \mathbb{F}], \bar{r}_i \neq \bar{r}_j \end{array} \right. &
 \end{array}$$

Theorem F.4.5 Let η be an isomorphism from \mathbb{F} to $\bar{\mathbb{F}}$ and let $\mathbb{K} = \mathbb{F}[r_1, \dots, r_n], \bar{\mathbb{K}} = \bar{\mathbb{F}}[\bar{r}_1, \dots, \bar{r}_n]$ be splitting fields of $p(x)$ and $\bar{p}(x)$ respectively. Then there exist at most $[\mathbb{K} : \mathbb{F}]$ isomorphisms $\zeta_i : \mathbb{K} \rightarrow \bar{\mathbb{K}}$ which extend η . If $\{\bar{r}_1, \dots, \bar{r}_n\}$ are distinct, then there exist exactly $[\mathbb{K} : \mathbb{F}]$ isomorphisms of the above sort. In either case, the two splitting fields are isomorphic with any of these ζ_i serving as an isomorphism.

Proof: Suppose $[\mathbb{K} : \mathbb{F}] = 1$. Say a basis for \mathbb{K} is $\{r\}$. Then $\{1, r\}$ is dependent and so there exist $a, b \in \mathbb{F}$, not both zero such that $a + br = 0$. Then it follows that $r \in \mathbb{F}$ and so in this case $\mathbb{F} = \mathbb{K}$. Then the isomorphism which extends η is just η itself and there is exactly 1 isomorphism.

Next suppose $[\mathbb{K} : \mathbb{F}] > 1$. Then $p(x)$ has an irreducible factor over \mathbb{F} of degree larger than 1, $q(x)$. If not, you would have

$$p(x) = x^n + a_{n-1}x^{n-1} + \dots + a_n$$

and it would factor as

$$= (x - r_1) \cdots (x - r_n)$$

with each $r_j \in \mathbb{F}$, so $\mathbb{F} = \mathbb{K}$ contrary to $[\mathbb{K} : \mathbb{F}] > 1$. Without loss of generality, let the roots of $q(x)$ in \mathbb{K} be $\{r_1, \dots, r_m\}$. Thus

$$q(x) = \prod_{i=1}^m (x - r_i), \quad p(x) = \prod_{i=1}^n (x - r_i)$$

Now $\bar{q}(x)$ defined analogously to $\bar{p}(x)$, also has degree at least 2. Furthermore, it divides $\bar{p}(x)$ all of whose roots are in $\bar{\mathbb{K}}$. This is obvious because η is an isomorphism. You have

$$l(x)q(x) = p(x) \text{ so } \bar{l}(x)\bar{q}(x) = \bar{p}(x).$$

Denote the roots of $\bar{q}(x)$ in $\bar{\mathbb{K}}$ as $\{\bar{r}_1, \dots, \bar{r}_m\}$ where they are counted according to multiplicity.

Then from Theorem F.4.3, there exist $k \leq m$ one to one homomorphisms (monomorphisms) ζ_i mapping $\mathbb{F}[r_1]$ to $\bar{\mathbb{K}} \equiv \bar{\mathbb{F}}[\bar{r}_1, \dots, \bar{r}_n]$, one for each distinct root of $\bar{q}(x)$ in $\bar{\mathbb{K}}$. If the roots of $\bar{p}(x)$ are distinct, then this is sufficient to imply that the roots of $\bar{q}(x)$ are also distinct, and $k = m$, the dimension of $q(x)$. Otherwise, maybe $k < m$. (It is conceivable that $\bar{q}(x)$ might have repeated roots in $\bar{\mathbb{K}}$.) Then

$$[\mathbb{K} : \mathbb{F}] = [\mathbb{K} : \mathbb{F}[r_1]] [\mathbb{F}[r_1] : \mathbb{F}]$$

and since the degree of $q(x) > 1$ and $q(x)$ is irreducible, this shows that $[\mathbb{F}[r_1] : \mathbb{F}] = m > 1$ and so

$$[\mathbb{K} : \mathbb{F}[r_1]] < [\mathbb{K} : \mathbb{F}]$$

Therefore, by induction, using Theorem F.4.3, each of these $k \leq m = [\mathbb{F}[r_1] : \mathbb{F}]$ one to one homomorphisms extends to an isomorphism from \mathbb{K} to $\bar{\mathbb{K}}$ and for each of these ζ_i , there are no more than $[\mathbb{K} : \mathbb{F}[r_1]]$ of these isomorphisms extending \mathbb{F} . If the roots of $\bar{p}(x)$ are distinct, then there are exactly m of these ζ_i and for each, there are $[\mathbb{K} : \mathbb{F}[r_1]]$ extensions. Therefore, if the roots of $\bar{p}(x)$ are distinct, this has identified

$$[\mathbb{K} : \mathbb{F}[r_1]] m = [\mathbb{K} : \mathbb{F}[r_1]] [\mathbb{F}[r_1] : \mathbb{F}] = [\mathbb{K} : \mathbb{F}]$$

isomorphisms of \mathbb{K} to $\bar{\mathbb{K}}$ which agree with η on \mathbb{F} . If the roots of $\bar{p}(x)$ are not distinct, then maybe there are fewer than $[\mathbb{K} : \mathbb{F}]$ extensions of η .

Is this all of them? Suppose ζ is such an isomorphism of \mathbb{K} and $\bar{\mathbb{K}}$. Then consider its restriction to $\mathbb{F}[r_1]$. By Theorem F.4.3, this restriction must coincide with one of the ζ_i chosen earlier. Then by induction, ζ is one of the extensions of the ζ_i just mentioned. ■

Definition F.4.6 Let \mathbb{K} be a finite dimensional extension of a field \mathbb{F} such that every element of \mathbb{K} is algebraic over \mathbb{F} , that is, each element of \mathbb{K} is a root of some polynomial in $\mathbb{F}[x]$. Then \mathbb{K} is called a normal extension if for every $k \in \mathbb{K}$ all roots of the minimal polynomial of k are contained in \mathbb{K} .

So what are some ways to tell that a field is a normal extension? It turns out that if \mathbb{K} is a splitting field of $f(x) \in \mathbb{F}[x]$, then \mathbb{K} is a normal extension. I found this in [18]. This is an amazing result.

Proposition F.4.7 Let \mathbb{K} be a splitting field of $f(x) \in \mathbb{F}[x]$. Then \mathbb{K} is a normal extension. In fact, if \mathbb{L} is an intermediate field between \mathbb{F} and \mathbb{K} , then \mathbb{L} is also a normal extension of \mathbb{F} .

Proof: Let $r \in \mathbb{K}$ be a root of $g(x)$, an irreducible monic polynomial in $\mathbb{F}[x]$. It is required to show that every other root of $g(x)$ is in \mathbb{K} . Let the roots of $g(x)$ in a splitting field be $\{r_1 = r, r_2, \dots, r_m\}$. Now $g(x)$ is the minimal polynomial of r_j over \mathbb{F} because $g(x)$ is irreducible. Recall why this was. If $p(x)$ is the minimal polynomial of r_j ,

$$g(x) = p(x)l(x) + r(x)$$

where $r(x)$ either is 0 or it has degree less than the degree of $p(x)$. However, $r(r_j) = 0$ and this is impossible if $p(x)$ is the minimal polynomial. Hence $r(x) = 0$ and now it follows that $g(x)$ was not irreducible unless $l(x) = 1$.

By Theorem F.4.3, there exists an isomorphism η of $\mathbb{F}[r_1]$ and $\mathbb{F}[r_j]$ which fixes \mathbb{F} and maps r_1 to r_j . Now $\mathbb{K}[r_1]$ and $\mathbb{K}[r_j]$ are splitting fields of $f(x)$ over $\mathbb{F}[r_1]$ and $\mathbb{F}[r_j]$ respectively. By Theorem F.4.5, the two fields $\mathbb{K}[r_1]$ and $\mathbb{K}[r_j]$ are isomorphic, the isomorphism, ζ extending η . Hence

$$[\mathbb{K}[r_1] : \mathbb{K}] = [\mathbb{K}[r_j] : \mathbb{K}]$$

But $r_1 \in \mathbb{K}$ and so $\mathbb{K}[r_1] = \mathbb{K}$. Therefore, $\mathbb{K} = \mathbb{K}[r_j]$ and so r_j is also in \mathbb{K} . Thus all the roots of $g(x)$ are actually in \mathbb{K} . Consider the last assertion.

Suppose $r = r_1 \in \mathbb{L}$ where the minimal polynomial for r is denoted by $q(x)$. Then letting the roots of $q(x)$ in \mathbb{K} be $\{r_1, \dots, r_m\}$. By Theorem F.4.3 applied to the identity map on \mathbb{L} , there exists an isomorphism $\theta : \mathbb{L}[r_1] \rightarrow \mathbb{L}[r_j]$ which fixes \mathbb{L} and takes r_1 to r_j . But this implies that

$$1 = [\mathbb{L}[r_1] : \mathbb{L}] = [\mathbb{L}[r_j] : \mathbb{L}]$$

Hence $r_j \in \mathbb{L}$ also. Since r was an arbitrary element of \mathbb{L} , this shows that \mathbb{L} is normal. ■

Definition F.4.8 When you have $\mathbb{F}[a_1, \dots, a_m]$ with each a_i algebraic so that $\mathbb{F}[a_1, \dots, a_m]$ is a field, you could consider

$$f(x) \equiv \prod_{i=1}^m f_i(x)$$

where $f_i(x)$ is the minimal polynomial of a_i . Then if \mathbb{K} is a splitting field for $f(x)$, this \mathbb{K} is called the normal closure. It is at least as large as $\mathbb{F}[a_1, \dots, a_m]$ and it has the advantage of being a normal extension.

Bibliography

- [1] **Apostol T.**, *Calculus Volume II Second edition*, Wiley 1969.
- [2] **Artin M.**, *Algebra*, Pearson 2011.
- [3] **Baker, Roger**, *Linear Algebra*, Rinton Press 2001.
- [4] **Baker, A.** *Transcendental Number Theory*, Cambridge University Press 1975.
- [5] **Chahal J.S.**, *Historical Perspective of Mathematics 2000 B.C. - 2000 A.D. Kendrick Press, Inc. (2007)*
- [6] **Coddington and Levinson**, *Theory of Ordinary Differential Equations* McGraw Hill 1955.
- [7] **Davis H. and Snider A.**, *Vector Analysis* Wm. C. Brown 1995.
- [8] **Edwards C.H.**, *Advanced Calculus of several Variables*, Dover 1994.
- [9] **Evans L.C.** *Partial Differential Equations*, Berkeley Mathematics Lecture Notes. 1993.
- [10] **Friedberg S. Insel A. and Spence L.**, *Linear Algebra*, Prentice Hall, 2003.
- [11] **Golub, G. and Van Loan, C.**, *Matrix Computations*, Johns Hopkins University Press, 1996.
- [12] **Gurtin M.**, *An introduction to continuum mechanics*, Academic press 1981.
- [13] **Hardy G.**, *A Course Of Pure Mathematics, Tenth edition*, Cambridge University Press 1992.
- [14] **Herstein I. N.**, *Topics In Algebra*, Xerox, 1964.
- [15] **Hofman K. and Kunze R.**, *Linear Algebra*, Prentice Hall, 1971.
- [16] **Householder A.** *The theory of matrices in numerical analysis* , Dover, 1975.
- [17] **Horn R. and Johnson C.**, *matrix Analysis*, Cambridge University Press, 1985.
- [18] **Jacobsen N.** *Basic Algebra* Freeman 1974.
- [19] **Karlin S. and Taylor H.**, *A First Course in Stochastic Processes*, Academic Press, 1975.
- [20] **Marcus M., and Minc H.**, *A Survey Of Matrix Theory and Matrix Inequalities*, Allyn and Bacon, INC. Boston, 1964
- [21] **Nobel B. and Daniel J.**, *Applied Linear Algebra*, Prentice Hall, 1977.
- [22] **E. J. Putzer**, American Mathematical Monthly, Vol. 73 (1966), pp. 2-7.
- [23] **Rudin W.**, *Principles of Mathematical Analysis*, McGraw Hill, 1976.
- [24] **Rudin W.**, *Functional Analysis*, McGraw Hill, 1991.
- [25] **Salas S. and Hille E.**, *Calculus One and Several Variables*, Wiley 1990.
- [26] **Strang Gilbert**, *Linear Algebra and its Applications*, Harcourt Brace Jovanovich 1980.
- [27] **Wilkinson, J.H.**, *The Algebraic Eigenvalue Problem*, Clarendon Press Oxford 1965.

INDEX

- \cap , 1
- \cup , 1
- A close to B!eigenvalues, 135
- A invariant, 184
- Abel's formula, 83, 197, 387, 462
- absolute convergence!convergence, 268
- adjugate, 63, 75
- algebraic number!minimal polynomial, 159
- algebraic numbers, 158
- algebraic numbers!field, 160
- almost linear, 333
- almost linear system, 333
- analytic function of matrix, 318
- Archimedean property, 10
- asymptotically stable, 333
- augmented matrix, 16
- autonomous, 333
- Banach space, 259
- basis, 45, 146
- Binet Cauchy ! volumes, 230
- Binet Cauchy formula, 71
- block matrix, 79
- block matrix!multiplication, 80
- block multiplication, 79
- bounded linear transformations, 259
- Cauchy Schwarz inequality, 21, 215, 257
- Cauchy sequence, 227, 258, 341, 485
- Cayley Hamilton theorem, 78, 196, 205, 459, 471
- centrifugal acceleration@centrifugal acceleration, 51
- centripetal acceleration@centripetal acceleration, 51
- characteristic and minimal polynomial, 179, 450
- characteristic equation, 109
- characteristic polynomial, 78, 177
- characteristic value, 109
- Cholesky factorization, 256, 499
- codomain, 1
- cofactor, 62, 73
- column rank, 75, 89
- commutative ring, 343
- companion matrix, 199, 293
- complete, 277
- completeness axiom, 9
- complex conjugate, 4
- complex numbers!absolute value, 4
- complex numbers!field, 4
- complex numbers@complex numbers, 4
- complex roots, 5
- composition of linear transformations, 174
- comutator, 144, 440
- condition number, 265
- conformable, 28
- conjugate linear, 220
- converge, 341
- convex combination, 180, 453
- convex hull, 180, 453
- convex hull!compactness, 180, 453
- coordinate axis, 19
- coordinates, 19
- Coriolis acceleration, 51
- Coriolis acceleration@Coriolis acceleration!earth@earth, 53
- Coriolis force, 51
- counting zeros, 135
- Courant Fischer theorem, 238
- Cramer's rule, 64, 75
- cyclic basis, 189
- cyclic set, 187
- damped vibration, 330
- defective, 113
- DeMoivre identity, 5
- dense, 11
- density of rationals, 11
- determinant!block upper triangular matrix, 124, 384
- determinant!definition, 68
- determinant!estimate for Hermitian matrix, 214
- determinant!expansion along a column, 62
- determinant!expansion along a row, 62
- determinant!expansion along row, column, 73
- determinant!Hadamard inequality, 214
- determinant!inverse of matrix, 63

- determinant!matrix inverse, 74
- determinant!partial derivative, cofactor, 83, 388
- determinant!permutation of rows, 69
- determinant!product, 71
- determinant!product of eigenvalues, 129
- determinant!product of eigenvalues, 139, 427
- determinant!row, column operations, 63, 70
- determinant!summary of properties, 77
- determinant!symmetric definition, 69
- determinant!transpose, 69
- diagonalizable, 172, 231
- diagonalizable! minimal polynomial condition, 198, 465
- diagonalizable!basis of eigenvectors, 121, 421
- diagonalization, 235
- diameter, 340
- differentiable matrix, 48
- differential equations!first order systems, 141, 434
- digraph, 29
- dimension of vector space, 147
- direct sum, 60, 182, 378
- directed graph, 29
- discrete Fourier transform, 254, 494
- division of real numbers, 11
- Dolittle's method, 100
- domain, 1
- dot product, 20
- dyadics, 167
- dynamical system, 121, 423
- eigenspace, 110, 184
- eigenvalue, 61, 109, 380
- eigenvalues, 78, 135, 177
- eigenvalues!AB and BA, 81
- eigenvector, 61, 109, 380
- eigenvectors!distinct eigenvalues independence, 113
- elementary matrices, 85
- elementary symmetric polynomials, 343
- empty set, 1
- equality of mixed partial derivatives, 131
- equilibrium point, 333
- equivalence class, 154, 170
- equivalence of norms, 259
- equivalence relation, 154, 170
- Euclidean algorithm, 11
- exchange theorem, 44
- existence of a fixed point, 278
- field axioms, 2
- field extension, 154
- field extension!dimension, 156
- field extension!finite, 156
- field extensions, 156
- field!ordered, 3
- finite dimensional normed linear space!completeness, 259
- finite dimensional normed linear space!equivalence of norms, 259
- Foucault pendulum@Foucault pendulum, 53
- Fourier series, 226, 484
- Fredholm alternative, 95, 224
- free variable, 17
- Frobenius norm, 248
- Frobenius norm!singular value decomposition, 248
- Frobenius! inner product, 143, 438
- Frobenius norm, 253, 493
- functions, 1
- fundamental matrix, 327
- fundamental theorem of algebra, 347
- fundamental theorem of algebra ! plausibility argument, 7
- fundamental theorem of algebra ! rigorous proof, 8
- fundamental theorem of arithmetic, 13
- Gauss Jordan method for inverses, 33
- Gauss Seidel method, 273
- Gelfand, 267
- generalized eigenspace, 61, 380
- generalized eigenspaces, 184, 192
- generalized eigenvectors, 193
- Gerschgorin's theorem, 133
- Gram Schmidt procedure, 108, 123, 216, 403
- Gram Schmidt process, 216
- Gramm Schmidt process, 123
- greatest common divisor, 11, 150
- greatest common divisor!characterization, 12
- greatest lower bound, 9
- Gronwall's inequality, 283, 326, 509
- Hermitian, 126

- Hermitian matrix! factorization, 213, 478
- Hermitian matrix!positive part, 320
- Hermitian matrix!positive part, Lipschitz continuous, 320
- Hermitian operator, 220
- Hermitian operator!largest, smallest, eigenvalues, 238
- Hermitian operator!spectral representation, 235
- Hermitian!orthonormal basis eigenvectors, 236
- Hermitian!positive definite, 239
- Hermitian!real eigenvalues, 127
- Hessian matrix, 132
- Holder's inequality, 262
- Householder matrix, 105
- Householder!reflection, 106
- idempotent, 57, 372
- inconsistent, 17
- initial value problem!existence, 321
- initial value problem!global solutions, 325
- initial value problem!linear system, 323
- initial value problem!local solutions, existence, uniqueness, 324
- initial value problem!uniqueness, 283, 321, 509
- injective, 1
- inner product, 20, 214
- inner product space, 214
- inner product space!adjoint operator, 219
- inner product space!parallelogram identity, 215
- inner product space!triangle inequality, 215
- integers mod a prime, 165, 445
- integral!operator valued function, 282, 508
- integral!vector valued function, 282, 507
- intersection, 1
- intervals!notation, 1
- invariant, 234
- invariant subspaces!direct sum, block diagonal matrix, 186
- invariant!subspace, 184
- inverses and determinants, 74
- invertible, 33
- invertible matrix!product of elementary matrices, 92
- irreducible, 150
- irreducible!relatively prime, 151
- iterative methods!alternate proof of convergence, 280, 503
- iterative methods!convergence criterion, 276
- iterative methods!diagonally dominant, 281, 503
- iterative methods!proof of convergence, 279
- Jacobi method, 272
- Jordan block, 191, 193
- Jordan canonical form!existence and uniqueness, 193
- Jordan canonical form!powers of a matrix, 194
- ker, 93
- kernel, 42
- kernel of a product!direct sum decomposition, 183
- Krylov sequence, 187
- Lagrange form of remainder, 131
- Laplace expansion, 73
- least squares, 98, 223, 398
- least upper bound, 9
- Lindemann Weierstrass theorem, 353
- linear combination, 25, 43, 70
- linear transformation, 38, 166
- linear transformation!defined on a basis, 167
- linear transformation!dimension of vector space, 167
- linear transformation!existence of eigenvector, 178
- linear transformation!kernel, 181
- linear transformation!matrix, 39
- linear transformation!minimal polynomial, 178
- linear transformation!rotation, 40
- linear transformations!a vector space, 167
- linear transformations!commuting, 183
- linear transformations!composition, matrices, 174
- linear transformations!sum, 167, 221
- linearly dependent, 43
- linearly independent, 43, 145
- linearly independent set!extend to basis, 149
- Lipschitz condition, 321
- LU factorization!justification for multiplier method, 102
- LU factorization!multiplier method, 99
- LU factorization!solutions of linear systems, 100
- main diagonal, 62
- Markov matrix, 205
- Markov matrix!limit, 208
- Markov matrix!regular, 208

- Markov matrix!steady state, 205, 208
- mathematical induction, 10
- matrices!commuting, 233
- matrices!notation, 24
- matrices!transpose, 32
- matrix, 23
- matrix ! positive definite, 255, 497
- matrix exponential, 281, 504
- matrix multiplication!definition, 26
- matrix multiplication!entries of the product, 28
- matrix multiplication!not commutative, 27
- matrix multiplication!properties, 31
- matrix multiplication!vectors, 25
- matrix of linear transformation!orthonormal bases, 172
- matrix!differentiation operator, 169
- matrix!injective, 47
- matrix!inverse, 32
- matrix!left inverse, 75
- matrix!lower triangular, 62, 75
- matrix!Markov, 205
- matrix!non defective, 126
- matrix!normal, 126
- matrix!polynomial, 84, 391
- matrix!rank and existence of solutions, 94
- matrix!rank and nullity, 93
- matrix!right and left inverse, 47
- matrix!right inverse, 75
- matrix!right, left inverse, 74
- matrix!row, column, determinant rank, 75
- matrix!self adjoint, 121, 420
- matrix!stochastic, 205
- matrix!surjective, 47
- matrix!symmetric, 119
- matrix!symmetric, 418
- matrix!unitary, 123
- matrix!upper triangular, 62, 75
- migration matrix, 209
- minimal polynomial, 60, 177, 184, 379
- minimal polynomial ! algebraic number, 158
- minimal polynomial!eigenvalues, eigenvectors, 178
- minimal polynomial!finding it, 196, 457
- minimal polynomial!generalized eigenspaces, 184
- minor, 62, 73
- mixed partial derivatives, 130
- Moore Penrose inverse, 251
- Moore Penrose inverse!least squares, 251
- Moore Penrose inverse!uniqueness, 255
- moving coordinate system@moving coordinate system, 49
- moving coordinate system@moving coordinate system!acceleration @acceleration, 51
- negative definite, 239
- Neuman!series, 285, 512
- nilpotent!block diagonal matrix, 191
- nilpotent!Jordan form, uniqueness, 191
- nilpotent!Jordan normal form, 191
- non defective, 198, 465
- nonnegative self adjoint!square root, 241
- norm, 214
- norm!strictly convex, 280, 500
- norm!uniformly convex, 280, 500
- normal, 245
- normal!diagonalizable, 127
- normal!non defective, 126
- normed linear space, 214, 256
- normed vector space, 214
- norms!equivalent, 257
- null and rank, 227, 487
- null space, 42
- nullity, 93
- one to one, 1
- onto, 1
- operator norm, 259
- orthogonal matrix, 61, 66, 105, 124, 380, 385
- orthonormal basis, 215
- orthonormal polynomials, 225, 482
- p norms, 262
- p norms!axioms of a norm, 263
- parallelepiped!volume, 228
- partitioned matrix, 79
- Penrose conditions, 252
- permutation, 68
- permutation matrices, 85
- permutation!even, 86
- permutation!odd, 86

- perp, 94
- Perron's theorem, 311
- pivot column, 91
- PLU factorization, 101
- PLU factorization!existence, 105
- polar decomposition!left, 244
- polar decomposition!right, 243
- polar form complex number, 5
- polynomial, 14, 150
- polynomial ! leading coefficient, 150
- polynomial ! leading term, 14
- polynomial ! matrix coefficients, 84, 391
- polynomial ! monic, 14, 150
- polynomial!addition, 14
- polynomial!degree, 14, 150
- polynomial!divides, 150
- polynomial!division, 14, 150
- polynomial!equal, 150
- polynomial!equality, 14
- polynomial!greatest common divisor, 150
- polynomial!greatest common divisor description, 151
- polynomial!greatest common divisor, uniqueness, 151
- polynomial!irreducible, 150
- polynomial!irreducible factorization, 152
- polynomial!multiplication, 14
- polynomial!relatively prime, 150
- polynomial!root, 150
- polynomials!canceling, 152
- polynomials!factorization, 152
- positive definite matrix, 255, 497
- positive definite!positive eigenvalues, 239
- positive definite!principle minors, 240
- positive definite, 239
- power method, 287
- prime number, 11
- prime numbers!infinity of primes, 164, 445
- principle directions, 115
- principle minors, 240
- product rule!matrices, 48
- projection map!convex set, 227, 486
- Putzer's method, 328
- QR algorithm, 138, 297, 425
- QR algorithm! convergence, 300
- QR algorithm!convergence theorem, 300
- QR algorithm!non convergence, 138, 303
- QR algorithm!nonconvergence, 426
- QR factorization, 106
- QR factorization!existence, 107
- QR factorization!Gram Schmidt procedure, 108, 403
- quadratic form, 129
- quotient space, 165, 446
- quotient vector space, 165
- range, 1
- rank, 90
- rank of a matrix, 75, 89
- rank one transformation, 221
- rank!number of pivot columns, 93
- rational canonical form, 200
- rational canonical form!uniqueness, 202
- Rayleigh quotient, 294
- Rayleigh quotient!how close?, 294
- real numbers, 2
- real Schur form, 124
- regression line, 223
- regular Sturm Liouville problem, 225, 483
- relatively prime, 12
- Riesz representation theorem, 219
- right Cauchy Green strain tensor, 243
- right polar decomposition, 243
- row equivalence!determination, 92
- row equivalent, 91
- row operations, 16, 85
- row operations!inverse, 16
- row operations!linear relations between columns, 89
- row rank, 75, 89
- row reduced echelon form!definition, 91
- row reduced echelon form!examples, 91
- row reduced echelon form!existence, 91
- row reduced echelon form!uniqueness, 92
- scalar product, 20
- scalars, 6, 19, 23
- Schur's theorem, 123, 234
- Schur's theorem!inner product space, 234

- second derivative test, 133
- self adjoint, 126, 220
- self adjoint nonnegative!roots, 242
- sequential compactness, 342
- sequentially compact, 342
- set notation, 0
- sgn, 67
- sgn!uniqueness, 68
- shifted inverse power method, 288
- shifted inverse power method!complex eigenvalues, 292
- sign of a permutation, 68
- similar matrices, 65, 83, 170, 382, 387
- similar!matrix and its transpose, 198, 466
- similarity transformation, 170
- simple field extension, 160
- simultaneous corrections, 272
- simultaneously diagonalizable, 232
- simultaneously diagonalizable!commuting family, 234
- singular value decomposition, 247
- singular values, 247
- skew symmetric, 32, 119, 418
- space of linear transformations!vector space, 221
- span, 43, 70
- spanning set!restricting to a basis, 149
- spectral mapping theorem, 320
- spectral norm, 261
- spectral radius, 266, 267
- spectrum, 109
- splitting field, 157
- stable, 333
- stable manifold, 339
- stochastic matrix, 205
- subsequence, 341
- subspace, 43, 145
- subspace!basis, 46, 149
- subspace!complementary, 231, 490
- subspace!dimension, 46
- subspace!invariant, 184
- subspaces!direct sum, 182
- subspaces!direct sum, basis, 183
- substituting matrix into polynomial identity, 84, 391
- surjective, 1
- Sylvester, 60, 377
- Sylvester! law of inertia, 142, 437
- Sylvester!dimension of kernel of product, 181
- Sylvester's equation, 230, 489
- symmetric, 32, 119, 418
- symmetric polynomial theorem, 343
- symmetric polynomials, 343
- system of linear equations, 17
- tensor product, 221
- the space AU , 231
- trace, 129
- trace! AB and BA , 129
- trace!sum of eigenvalues, 139, 427
- transpose, 32
- transpose!properties, 32
- triangle inequality, 22
- trivial, 43
- union, 1
- Unitary matrix! representation, 285
- upper Hessenberg matrix, 307
- Vandermonde determinant, 84, 390
- variation of constants formula, 142, 329, 435
- variational inequality, 227, 486
- vector space axioms, 20
- vector space!axioms, 25, 144
- vector space!basis, 45
- vector space!dimension, 46
- vector space!examples, 145
- vector!angular velocity, 49
- vectors, 25
- volume!parallelepiped, 228
- well ordered, 10
- Wronskian, 82, 142, 197, 329, 386, 435, 462
- Wronskian alternative, 142, 329, 435