

# Hour to hour syllabus for Math21b, Fall 2010

Course Head: Oliver Knill

## Abstract

Here is a brief outline of the lectures for the fall 2010 semester. The section numbers refer to the book of Otto Bretscher, Linear algebra with applications.

## 1. Week: Systems of Linear Equations and Gauss-Jordan

### 1. Lecture: Introduction to linear systems, Section 1.1, September 8, 2010

A central point of this week is **Gauss-Jordan elimination**. While the precise procedure will be introduced in the second lecture, we learn in this first hour what a **system of linear equations** is by looking at examples of systems of linear equations. The aim is to illustrate, where such systems can occur and how one could solve them with 'ad hoc' methods. This involves solving equations by combining equations in a clever way or to eliminate variables until only one variable is left. We see examples with no solution, several solutions or exactly one solution.

### 2. Lecture: Gauss-Jordan elimination, Section 1.2, September 10, 2010

We rewrite systems of linear equations using **matrices** and introduce **Gauss-Jordan elimination steps**: scaling of rows, swapping rows or subtract a multiple of one row to an other row. We also see an example, where one has not only one solution or no solution. Unlike in multi-variable calculus, we distinguish between **column vectors** and **row vectors**. Column vectors are  $n \times 1$  matrices, and row vectors are  $1 \times m$  matrices. A general  $n \times m$  matrix has  $m$  columns and  $n$  rows. The output of Gauss-Jordan elimination is a matrix  $\text{rref}(A)$  which is in row reduced echelon form: the first nonzero entry in each row is 1, called leading 1, every column with a leading 1 has no other nonzero elements and every row above a row with a leading 1 has a leading 1 to the left.

## 2. Week: Linear Transformations and Geometry

### 3. Lecture: On solutions of linear systems, Section 1.3, September 13, 2010

How many solutions does a system of linear equations have? The goal of this lecture is to see that there are three possibilities: exactly one solution, no solution or infinitely many solutions. This can be visualized and explained geometrically in low dimensions. We also learn to determine which case we are in using Gauss-Jordan elimination by looking at the **rank** of the matrix  $A$  as well as the **augmented matrix**  $[A|b]$ . We also mention that one can see a system of linear equations  $Ax = b$  in two different ways: the **column picture** tells that  $b = x_1v_1 + \dots + x_nv_n$  is a sum of column vectors  $v_i$  of the matrix  $A$ , the **row picture** tells that the dot product of the row vectors  $w_j$  with  $x$  are the components  $w_j \cdot x = b_j$  of  $b$ .

### 4. Lecture: Linear transformation, Section 2.1, September 15, 2010

This week provides a link between the geometric and algebraic description of linear transformations. Linear transformations are introduced formally as transformations  $T(x) = Ax$ , where  $A$  is a matrix. We learn how to distinguish between linear and nonlinear, linear and affine transformations. The transformation  $T(x) = x + 5$  for example is not linear because 0 is not mapped to 0. We characterize linear transformations on  $\mathbb{R}^n$  by three properties:  $T(0) = 0$ ,  $T(x+y) = T(x) + T(y)$  and  $T(sx) = sT(x)$ , which means compatibility with the additive structure on  $\mathbb{R}^n$ .

### 5. Lecture: Linear transformations in geometry, Section 2.2, September 17, 2010

We look at examples of rotations, dilations, projections, reflections, rotation-dilations or shears. How are these transformations described algebraically? The main point is to see how to go forth and back between algebraic and geometric description. The key fact is that the column vectors  $v_j$  of a matrix are the images  $v_j = Te_j$  of the basis vectors  $e_j$ . We derive for each of the mentioned geometric transformations the matrix form. Any of them will be important throughout the course.

## 3. Week: Matrix Algebra and Linear Subspaces

### 6. Lecture: Matrix product, Section 2.3, September 20, 2010

The composition of linear transformations leads to the **product of matrices**. The **inverse** of a transformation is described by the inverse of the matrix. Square matrices can be treated in a similar way as numbers: we can add them, multiply them with scalars and many matrices have inverses. There is two things to be careful about: the product of two matrices is not commutative and many nonzero matrices have no inverse. If we take the product of a  $n \times p$  matrix with a  $p \times m$  matrix, we obtain a  $n \times m$  matrix. The dot product as a special case of a matrix product between a  $1 \times n$  matrix and a  $n \times 1$  matrix. It produces a  $1 \times 1$  matrix, a scalar.

### 7. Lecture: The inverse, Section 2.4, September 22, 2010

We first look at invertibility of maps  $f : X \rightarrow X$  in general and then focus on the case of linear maps. If a linear map  $\mathbb{R}^n$  to  $\mathbb{R}^n$  is invertible, how do we find the inverse? We look at examples when this is the case. Finding  $x$  such that  $Ax = y$  is equivalent to solving a system of linear equations. Doing this in parallel gives us an elegant algorithm by row reducing the matrix  $[A|1_n]$  to end up with  $[1_n|A^{-1}]$ . We also might have time to see how upper triangular block matrices  $\left[ \begin{array}{c|c} A & B \\ \hline 0 & C \end{array} \right]$  have the inverse  $\left[ \begin{array}{c|c} A^{-1} & -A^{-1}BC^{-1} \\ \hline 0 & C^{-1} \end{array} \right]$ .

### 8. Lecture: Image and kernel, Section 3.1, September 24, 2010

We define the notion of a **linear subspace** of  $n$ -dimensional space and the **span** of a set of vectors. This is a preparation for the more abstract definition of linear spaces which appear later in the course. The main algorithm is the computation of the **kernel** and the **image** of a linear transformation using row reduction. The image of a matrix  $A$  is spanned by the columns of  $A$  which have a leading 1 in  $\text{rref}(A)$ . The kernel of a matrix  $A$  is parametrized by "**free variables**", the variables for which there is no leading 1 in  $\text{rref}(A)$ . For a  $n \times n$  matrix, the kernel is trivial if and only if the matrix is invertible. The kernel is always nontrivial if the  $n \times m$  matrix satisfies  $m > n$ , that is if there are more variables than equations.

## 4. Week: Basis and Dimension

### 9. Lecture: Basis and linear independence, Section 3.2, September 27, 2010

With the previously defined "span" and the newly introduced linear independence, one can define what a **basis** for a linear space is. It is a **set** of vectors which span the space and which are linear independent. The standard basis in  $R^n$  is an example of a basis. We show that if we have a basis, then every vector can be **uniquely** represented as a linear combination of basis elements. A typical task is to find the basis of the kernel and the basis for the image of a linear transformation.

#### 10. Lecture: Dimension, Section 3.3, September 29, 2010

The concept of abstract linear spaces allows to introduce **linear spaces of functions**. This will be useful for applications in differential equations. We show first that the number of basis elements is independent of the basis. This number is called the **dimension**. The proof uses that if  $p$  vectors are linearly independent and  $q$  vectors span a linear subspace  $V$ , then  $p$  is less or equal to  $q$ . We see the **rank-nullity theorem**:  $\dimker(A) + \dimim(A)$  is the number of columns of  $A$ . Even so the result is not very deep, it is sometimes referred to as the **fundamental theorem of linear algebra**. It will turn out to be quite useful for us, for example, when looking under the hood of data fitting.

#### 11. Lecture: Change of coordinates, Section 3.4, October 1, 2010

Switching to a different basis can be useful for certain problems. For example to find the matrix of the reflection at a line or projection onto a plane, one can first find the matrix  $B$  in a suitable basis  $\mathcal{B} = \{v_1, v_2, v_3\}$ , then use  $B = SAS^{-1}$  to get  $A$ . The matrix  $S$  contains the basis vectors in the columns. We also learn how to express a matrix in a new basis  $Se_i = v_i$ . We derive the formula  $B = SAS^{-1}$ .

### 5. Week: Linear Spaces and Orthogonality

#### 12. Lecture: Linear spaces, Section 4.1, October 4, 2010

In this lecture we generalize the concept of linear subspaces of  $R^n$  and consider **abstract linear spaces**. An abstract linear space is a set  $X$  closed under addition and scalar multiplication and which contains 0. We look at many examples. An important one is the space  $X = C([a, b])$  of continuous functions on the interval  $[a, b]$  or the space  $P_5$  of polynomials of degree smaller or equal to 5, or the linear space of all  $3 \times 3$  matrices.

#### 13. Lecture: Review for the second midterm, October 6, 2010

This is review for the first midterm on October 7th. The plenary review will cover all the material, so that this review can focus on questions or looking at some True/False problems or practice exam problems.

#### 14. Lecture: orthonormal bases and projections, Section 5.1, October 8, 2010

We review orthogonality between vectors  $u, v$  by  $u \cdot v = 0$  and define **orthonormal basis**, a basis which consists of unit vectors which are all orthogonal to each other. The orthogonal complement of a linear space  $V$  in  $R^n$  is defined the set of all vectors perpendicular to all vectors in  $V$ . It can be found as a kernel of the matrix which contains a basis of  $V$  as rows. We then define orthogonal projection onto a linear subspace  $V$ . Given an orthonormal basis  $\{u_1, \dots, u_n\}$  in  $V$ , we have a formula for the orthogonal projection:  $P(x) = (u_1 \cdot x)u_1 + \dots + (u_n \cdot x)u_n$ . This simple formula for a projection only holds if we are given an orthonormal basis in the subspace  $V$ . We mention already that this formula can be written as  $P = AA^T$  where  $A$  is the matrix which contains the orthonormal basis as columns.

### 6. Week: Gram-Schmidt and Projection

Monday is Columbus Day and no lectures take place.

#### 15. Lecture: Gram-Schmidt and QR factorization, Section 5.2, October 13, 2010

The **Gram Schmidt orthogonalization process** lead to the QR factorization of a matrix  $A$ . We will look at this process geometrically as well as algebraically. The geometric process of "straightening out" and "adjusting length" can be illustrated well in 2 and 3 dimensions. Once the formulas for the orthonormal vectors  $w_j$  from a given set of vectors  $v_j$  are derived, one can rewrite it in matrix form. If the  $v_j$  are the  $m$  columns of a  $n \times m$  matrix  $A$  and  $w_j$  the columns of a  $n \times m$  matrix  $Q$ , then  $A = QR$ , where  $R$  is a  $m \times m$  matrix. This is the QR factorization. The QR factorization has its use in numerical methods.

#### 16. Lecture: Orthogonal transformations, Section 5.3, October 15, 2010

We first define the **transpose**  $A^T$  of a matrix  $A$ . **Orthogonal matrices** are defined as matrices for which  $A^T A = 1_n$ . This is equivalent to the fact that the transformation  $T$  defined by  $A$  preserves angles and lengths. Rotations and reflections are examples of orthogonal transformations. We point out the difference between orthogonal projections and orthogonal transformations. The identity matrix is the only orthogonal matrix which is also an orthogonal projection. We also stress that the notion of orthogonal matrix only applies to  $n \times n$  matrices and that the column vectors form an orthonormal basis. A matrix  $A$  for which all columns are orthonormal is not orthogonal if the number of rows is not equal to the number of columns.

### 7. Week: Data fitting and Determinants

#### 17. Lecture: Least squares and data fitting, Section 5.4, October 18, 2010

This is an important lecture from the application point of view. It covers a part of statistics. We learn how to **fit data points** with any finite set of functions. To do so, we write the fitting problem as a in general overdetermined system of linear equations  $Ax = b$  and find from this the **least square solution**  $x_*$  which has geometrically the property that  $Ax_*$  is the projection of  $b$  onto the image of  $A$ . Because this means  $A^T(Ax_* - b) = 0$ , we get the formula

$$x_* = (A^T A)^{-1} A^T b.$$

An example is to fit a set of data  $(x_i, y_i)$  by linear functions  $\{f_1, \dots, f_n\}$ . This is very powerful. We can fit by any type of functions, even functions of several variables.

#### 18. Lecture: Determinants I, Section 6.1, October 20, 2010

We define the determinant of a  $n \times n$  matrix using the **permutation definition**. This immediately implies the Laplace expansion formula and allows comfortably to derive all the properties of determinants from the original definition. In this lecture students learn about permutations in terms of **patterns**. There is no need to talk about permutations and signatures. The equivalent language of "patterns" and "number of **upcrossings**". In this lecture, we see the definition of determinants in all dimensions, see how it fits with 2 and 3 dimensional case. We practice already **Laplace expansion** to compute determinants.

#### 19. Lecture: Determinants II, Section 6.2, October 22, 2010

We learn about the **linearity property** of determinants and how Gauss-Jordan elimination allows a fast computation of determinants. The computation of determinants by Gauss-Jordan elimination is quite efficient. Often we can see the determinant already after a few steps because the matrix has become upper triangular. We also point out how to compute determinants for partitioned matrices. We do lots of examples, also harder examples in which we learn how to decide which of the methods to use: permutation method, Laplace expansion, row reduction to a triangular case or using partitioned matrices.

## 8. Week: Eigenvectors and Diagonalization

### 20. Lecture: Eigenvalues, Section 7.1-2, October 25, 2010

**Eigenvalues and eigenvectors** are introduced in this lecture. It is good to see them first in concrete examples like rotations, reflections, shears. As the book, we can motivate the concept using **discrete dynamical systems**, like the problem to find the growth rate of the Fibonacci sequence. Here it becomes evident, why computing eigenvalues and eigenvectors is useful.

### 21. Lecture: Eigenvectors, Section 7.3, October 27, 2010

This lecture focuses on eigenvectors. Computing eigenvectors relates to the computation of the kernel of a linear transformation. We give also a geometric idea what eigenvectors are and look at lots of examples. A good class of examples are **Markov matrices**, which are important from the application point of view. Markov matrices always have an eigenvalue 1 because the transpose has an eigenvector  $[1, 1, \dots, 1]^T$ . The eigenvector of  $A$  to the eigenvalue 1 has significance. It belongs to a stationary probability distribution.

### 22. Lecture: Diagonalization, Section 7.4, October 29, 2010

A major result of this section is that if all eigenvalues of a matrix are different, one can **diagonalize** the matrix  $A$ . There is an **eigenbasis**. We also see that if the eigenvalues are the same, like for the shear matrix, one can not diagonalize  $A$ . If the eigenvalues are complex like for a rotation, one can not diagonalize over the reals. Since we like to be able to diagonalize in as many situations as possible, we allow complex eigenvalues from now on.

## 9. Week: Stability of Systems and Symmetric Matrices

### 23. Lecture: Complex eigenvalues, Section 7.5, November 1, 2010

We start with a short review on complex numbers. Course assistants will do more to get the class up to speed with complex numbers. The **fundamental theorem of algebra** assures that a polynomial of degree  $n$  has  $n$  solutions, when counted with multiplicities. We express the determinant and trace of a matrix in terms of eigenvalues. Unlike in the real case, these formulas hold for any matrix.

### 24. Lecture: Review for second midterm, November 3, 2010

We review for the second midterm in section. Since there was a plenary review for all students covering the theory, one could focus on student questions and see the big picture or discuss some True/False problems or practice exam problems.

### 25. Lecture: Stability, Section 7.6, November 5, 2010

We study the **stability problem** for discrete dynamical systems. The absolute value of the eigenvalues determines the stability of the transformation. If all eigenvalues are in absolute value smaller than 1, then the origin is **asymptotically stable**. A good example to discuss is the case, when the matrix is not diagonalizable, like for example for a shear dilation  $S = \begin{bmatrix} 0.99 & 1000 \\ 0 & 0.99 \end{bmatrix}$ , where the expansion by the off diagonal shear competes with the contraction in the diagonal.

## 10. Week: Homogeneous Ordinary Differential Equations

### 26. Lecture: Symmetric matrices, Section 8.1, November 8, 2010

The main point of this lecture is to see that **symmetric matrices** can be diagonalized. The key fact is that the eigenvectors of a symmetric matrix are perpendicular to each other. An intuitive proof of the spectral theorem can be given in class: after a small perturbation of the matrix all eigenvalues are different and diagonalization is possible. When making the perturbation smaller and smaller, the eigenspaces stay perpendicular and in particular linearly independent. The shear is the prototype of a matrix, which can not be diagonalized. This lecture also gives plenty of opportunity to practice finding an eigenbasis and possibly for Gram-Schmidt, if an orthonormal eigenbasis needs to be found in a higher dimensional eigenspace.

### 27. Lecture: Differential equations I, Section 9.1, November 10, 2010

We learn to solve **systems of linear differential equations** by diagonalization. We discuss linear stability of the origin. Unlike in the discrete time case, where the absolute value of the eigenvalues mattered, the real part of the eigenvalues is now important. We also keep in mind the one dimensional case, where these facts are obvious. The point is that linear algebra allows to reduce the higher dimensional case to the one-dimensional case.

### 28. Lecture: Differential equations II, Section 9.2, November 12, 2010

A second lecture is necessary for the important topic of applying linear algebra to **solve differential equations**  $x' = Ax$ , where  $A$  is a  $n \times n$  matrix. While the central idea is to diagonalize  $A$  and solve  $y' = By$ , where  $B$  is diagonal, we can do so a bit faster. Write the initial condition  $x(0)$  as a linear combination of eigenvectors  $x(0) = a_1 v_1 + \dots + a_n v_n$  and get  $x(t) = a_1 v_1 e^{\lambda_1 t} + \dots + a_n v_n e^{\lambda_n t}$ . We also look at examples where the eigenvalues  $\lambda_1$  of the matrix  $A$  are complex. An important case for the later is the **harmonic oscillator** with and without damping. There would be many more interesting examples from physics.

## 11. Week: Nonlinear Differential Equations, Function spaces

### 29. Lecture: Nonlinear systems, Section 9.4, November 15, 2010

This section is covered by a separate handout numbered Section 9.4. How can **nonlinear differential equations** in two dimensions  $\dot{x} = f(x, y)$ ,  $\dot{y} = g(x, y)$  be analyzed using linear algebra? The key concepts are finding **null clines**, **equilibria** and their nature using **linearization** of the system near the equilibria by computing the **Jacobian matrix**. Good examples are **competing species systems** like the example of Murray, **predator-pray examples** like the Volterra system or **mechanical systems** like the pendulum.

### 30. Lecture: Linear operators, Section 4.2, November 17, 2010

We study **linear operators** on linear spaces. The main example is the operator  $Df = f'$  as well as polynomials of the operator  $D$  like  $D^2 + D + 1$ . Other examples  $T(f) = xf$ ,  $T(f)(x) = x + 3$  (which is not linear) or  $T(f)(x) = x^2 f(x)$  which is linear. The goal is of this lecture is to get ready to understand that solutions of differential equations are kernels of linear operators or write partial differential equations in the form  $u_t = T(u)$ , where  $T$  is a linear operator.

### 31. Lecture: Linear differential operators, Section 9.3, November 19, 2010

The main goal is to be able to solve linear higher order differential equations  $p(D) = g$  using the **operator method**. The method generalizes the integration process which we use to solve for examples like  $f''' = \sin(x)$  where three fold integration leads to the general solution  $f$ . For a problem  $p(D) = g$  we factor the polynomial  $p(D) = (D - a_1)(D - a_2)\dots(D - a_n)$  into linear parts and invert each linear factor  $(D - a_i)$  using an integration factor. This operator method is very general and always works. It also provides us with a justification for a more convenient way to find solutions.

## 12. Week: Inner product Spaces and Fourier Theory

### 32. Lecture: inhomogeneous differential equations, Handout, November 22, 2010

This operator method to solve differential equations  $p(D)f = g$  works unconditionally. It allows to put together a "cookbook method", which describes, how to find the special solution of the inhomogeneous problem by first finding the general solution to the **homogeneous equation** and then finding a special solution. Very important cases are the situation  $\dot{x} - ax = g(x)$ , the **driven harmonic oscillator**

$$\ddot{x} + c^2x = g(x)$$

or the **driven damped harmonic oscillator**

$$\ddot{x} + b\dot{x} + c^2x = g(x)$$

Special care has to be taken if  $g(x)$  is in the kernel of  $p(D)$  or if the polynomial  $p$  has repeated roots.

### 33. Lecture: Inner product spaces, Section 5.5, November 24, 2010

As a preparation for **Fourier theory** we introduce **inner products** in linear spaces. It generalizes the dot product. For  $2\pi$ -periodic functions, one takes  $\langle f, g \rangle$  as the integral of  $f\bar{g}$  from  $-\pi$  to  $\pi$  and divide by  $2\pi$ . It has all the properties we know from the dot product in finite dimensions. An other example of an inner product on matrices is  $\langle A, B \rangle = \text{tr}(A^T B)$ . Most of the geometry we did before can now be done in a larger context. Examples are Gram-Schmidt orthogonalization, projections, reflections, have the concept of coordinates in a basis, orthogonal transformations etc.

## 13. Week: Fourier Series and Partial differential equations

### 34. Lecture: Fourier series, Handout, November 29, 2010

The expansion of a function with respect to the orthonormal basis  $1/\sqrt{2}, \cos(nx), \sin(nx)$  leads to the **Fourier expansion**

$$f(x) = a_0(1/\sqrt{2}) + \sum_{n=1}^{\infty} a_n \cos(nx) + b_n \sin(nx).$$

A nice example to see how Fourier theory is useful is to derive the **Leibniz series** for  $\pi/4$  by writing

$$x = \sum_{k=1}^{\infty} 2 \frac{(-1)^{k+1}}{k} \sin(kx)$$

and evaluate it at  $\pi/2$ . The main motivation is that the Fourier basis is an orthonormal eigenbasis to the operator  $D^2$ . It diagonalizes this operator because  $D^2 \sin(nx) = -n^2 \sin(nx)$ ,  $D^2 \cos(nx) = -n^2 \cos(nx)$ . We will use this to solve partial differential equations in the same way as we solved ordinary differential equations.

### 35. Lecture: Parseval's identity, Handout, December 1, 2010

**Parseval's identity**

$$\|f\|^2 = a_0^2 + \sum_{n=1}^{\infty} a_n^2 + b_n^2.$$

is the "Pythagorean theorem" for function spaces. It is useful to estimate how fast a finite sum converges. We mention also applications like computations of series by the **Parseval's identity** or by relating them to a Fourier series. Nice examples are computations of  $\zeta(2)$  or  $\zeta(4)$  using the Parseval's identity.

### 36. Lecture: Partial differential equations, Handout, December 3, 2010

**Linear partial differential equations**  $u_t = p(D)u$  or  $u_{tt} = p(D)u$  with a polynomial  $p$  are solved in the same way as ordinary differential equations: by diagonalization. Fourier theory achieves that the "matrix"  $D$  is diagonalized and so the polynomial  $p(D)$ . This is much more powerful than the separation of variable method, which we do **not** do in this course. For example, the partial differential equation

$$u_{tt} = u_{xx} - u_{xxx} + 10u$$

can be solved nicely with Fourier in the same way as we solve the wave equation. The method allows even to solve partial differential equations with a driving force like for example

$$u_{tt} = u_{xx} - u + \sin(t).$$