Mathematics 21b
Spring 2005

The subject: This is a course on *linear algebra and differential equations* with a special section that also introduces statistical techniques as used in the sciences. As everyone in the course will be learning linear algebra and learning about differential equations, an introductory digression is in order to explain what these subjects are about and why they should be understood by practicing scientists. These discussions constitute Parts 1 and 2 of the digression. The digression has a third part that says some things about the role of statistics in the sciences.

Part 1: The digression starts with the following definition of science:

The purpose of science is to predict future behavior from present circumstance.

Here is a somewhat simplistic elaboration:

The goal of a scientific program is to predict the outcome of experiments done in the future from some prescribed amount of presently known data.

For example, Watson and Crick's original proposal for genetic inheritance asserts that knowledge of the sequence of bases (adenine, guanine, cytosine, or thymine) along the DNA in a living cell is sufficient to predict the sorts of proteins that the cell can produce. Watson and Crick made this proposal based on their understanding from experiments of the workings of a cell, and then further experimental work subsequently verified that their proposal is fundamental for understanding how cells encode their intrinsic operating instructions.

By the way, note how theory and experiment work hand in hand: Experiment tells us the present state of the world, theory provides the prediction for the future, and future experiments either falsify the theory or are consistent with its predictions. In this regard, a scientific theory must be falsifiable.

To a first approximation, a scientific theory can be viewed in the following abstract manner: The known data constitutes a labeled collection of numbers, $x = \{x_1, . . ., x_n\}$; here and below, integer subscripts play the role of labels. Meanwhile, data that arises from future experiments can always be labeled so as to constitutes a second ordered set of numbers, $\{y_1, . . . , y_N\}$. Of course, the latter are not known until the experiments are carried out.

Granted this notation, a theory in science must give a well defined and reproducible method for predicting the future data, $\{y_1, . . ., y_N\}$, from the initial data, $\{x_1,$

. . ., x_n}. Thus, a theory can be viewed as a *function* that assigns an ordered collection of N numbers (the y's) to the collection of n numbers (the x's).

Of course, these experiments, once performed, provide a labeled set of N real data values, $\{y_1^{real}, \cdots, y_N^{real}\}$; and if each y_k^{real} is close to its predicted value, y_k, then the theory can be said to be an accurate description of reality. Of course, if some y_k^{real} is far from its prediction, y_k, the theory needs some revising.

The simplest non-constant functions are the *linear functions*; these have the schematic form

$$
\begin{aligned}
y_1 &= a_{11} x_1 + a_{12} x_2 + \cdots a_{1n} x_n \\
&\vdots \\
y_N &= a_{N1} x_1 + a_{N2} x_2 + \cdots a_{Nn} x_n
\end{aligned}
$$

where the collection $\{a_{ij}\}_{1\le i\le N, 1\le\le n}$ are numbers. Thus, a scientific theory that is based on such a linear function would have to specify the collection $\{a_{ij}\}_{1\le i\le N, 1\le\le n}$ and then knowledge of the input data $\{x_1, \ldots, x_n\}$ predicts the output, $\{y_1, \ldots, y_N\}$, of the future experiments using the preceding equation.

As it turns out, the linear functions are among the most relevant to the sciences. This is because an appropriate linear function usually provides the mathematical equivalent of a 'first approximation' for a predictive description of any given phenomena. In any event, a good grasp of the mathematics of linear functions is a prerequisite for further explorations because the techniques that are used to study more complicated functions employ most of the mathematics for linear functions.

The subject of *Linear algebra* concerns the mathematics of linear functions.

Part 2: It is often the case that the quantity of interest in an experiment can be viewed as a function of some auxiliary variable. Indeed, consider the case when the quantity of interest changes with time, and so can be viewed as a function of time. Think of time as a variable, t, that can take values on the real line, and then the quantity of interest at time t can be written as a function of t. This is to say that there is a function of one variable, $t \to u(t)$, and the values of this function at time t are defined to be those of the quantity of interest. Here is a hypothetical example: Radioactive iodine is inadvertently dumped in a reservoir at time t = today. The concentration of iodine in the reservoir at any given future time can be called u(t), and so the assignment $t \to u(t)$ defines a function of time.

In any case, a full theoretical understanding of the time varying behavior of what is represented abstractly by one or several functions of time entails predicting their values at future times from present data. In the reservoir example, the present data might consist of the concentration of iodine measured today and, in addition, the average rates of inflow, outflow and evaporation of water from the reservoir.

As it turns out, theoretical predictions for the future behavior of real world quantities are often expressed via a system of equations that relate the *rate of change* of the quantities of interest at any given time to the values of the quantities at that same time. For example, such a theory for predicting the time dependence of some quantity that is modeled by a function $t \to u(t)$ would have the form

$$\frac{du}{dt} = f(u) ,$$

where the function f is specified by the theory. Such an equation is a simple example of a differential equation. In general, a differential equation can be said to be any equation for a function or set of functions that constrains the functions and their derivatives in some specified manner.

The subject of *differential equations* concerns techniques for finding solutions to differential equations. The subject also concerns techniques for estimating properties of interest of hypothetical solutions without knowledge of their explicit form.

Linear algebra enters the differential equation story in the following way: Just as linear functions are good first approximations for modeling many phenomenon, so differential equations that involve linear functions also offer reasonable first approximations in many situations. As is explained in this course, differential equations that involve only the linear functions can almost always be solved by closed form expressions using linear algebra. In this regard, note that most of the known techniques for dealing with non-linear differential equations involve generalizations of those that work for the class of linear differential equations.

Because of this relation between the subject of linear algebra and the subject of differential equations, this course studies these subjects together.

Part 3: As I remarked at the outset, there is a special section in this course that focuses not just on linear algebra and differential equations, but on statistics as well. Below, I lay out a guide as to who should be in the special section. But first comes a discussion of the role of statistics in the sciences.

To start, *statistics* and its partner, *probability*, constitute the mathematics of uncertainty. To elaborate, note that uncertainty arises in the sciences in three ways: First, it is the rare measurement that can be done with absolute precision. This is to say that repeated measurements of a particular item will typically differ. Granted this uncertainty, one role of statistics is to supply an answer to the following question: Are the theoretical predictions consistent with the output data given that both output and input measurements are unavoidably imprecise?

Uncertainty also appears because most theoretical models do not take into account every possible factor. Indeed, theories must discard factors so as to insure tractable calculations. The strategy then is to discard factors with presumably small influence while estimating the error in having done so. This done, statistics provides the mathematics for estimating such errors.

Finally, certain natural phenomena have an intrinsic probabilistic nature. A prime example comes from physics where quantum mechanics dictates that the submicroscopic world is ruled by probabilistic laws rather than laws that are deterministic.

To summarize: *Statistics* provides the tools for incorporating uncertainties in the evaluation of a given scientific theory. Meanwhile, *probability* provides various theoretical models for how the values of repeated measurements should distribute themselves.

Linear algebra and statistics are related for the very simple reason that many results in statistics can be formulated using linear equations. For example, the spread in the suite of values from repeated measurements of a particular system is an important clue as to how the system works. As it turns out, the relative frequencies of appearances of the various possible values from such measurements is often a linear function of the relative frequencies of the various values that can appear by repeatedly measuring the input data. By input data, I mean things such as the temperature at which the experiment is carried out, the ambient pressure, the precise concentration of various chemicals used, etc. Note that this is not to say that the output values are linear functions of the input values. Rather, the assertion is that the frequency of the occurrence of any particular output value is often a linear function of the frequencies of the occurrence of the input values.

Differential equations enter the story here by virtue of the fact that it is often the case that the frequencies for the appearances of various real world quantities are time dependent. This said, the theoretical predictions for such time dependence are most often obtained using a differential equation for the functions whose values at a given time represent these same frequencies.

The sections: This course is taught in relatively small, separately meeting sections rather than in a big and rather impersonal room all together. This means that each of you will be assigned a section based on your request (instructions for how to make this request are given below). You then attend all meetings of your assigned section, hand in your homework to the assigned section, and collect your graded homework and graded exams from the assigned section. Most sections will be of size on the order of 25 students.

All sections are deemed 'regular' sections except for one specially named section, the 'bio/statistics' section. The various regular sections teach the same material on the same week to week schedule, give the same weekly homework assignments, and take the same version of the midterm exam and final exam. If you opt for a regular section, choose the one whose scheduled meeting time best fits with your schedule.

As just noted, there is one very special section of this course, the bio/statistics section. The curriculum as planned for the bio/statistics section differs substantially from that of the regular sections starting from the very beginning. Thus, transfer to and from the bio/statistics section after the first week of lectures is very much discouraged. I will be teaching the bio/statistics section from 11:30-1 on Tuesdays and Thursdays.

What follows tersely summarizes the curricula of the two sorts of sections:

∑ The regular section cover a great deal of linear algebra. Any one completing a regular section of this course will see all of the basics of linear algebra that might be required for higher level courses in the physical sciences, social sciences, and in mathematics, both pure and applied. In particular, the regular sections teach
  > methods to solve systems of linear equations,
  > methods to analyze and solve systems of linear differential equations,
  > methods to solve discrete linear dynamical systems such as Markov processes,
  > least square fitting of data with arbitrary function sets,
  > the basics of Fourier series and its use to solve partial differential equations.
- The bio/statistic section will cover less linear algebra than the regular sections, providing instead basics of statistics and probability. Moreover, the linear algebra and also the differential equations will be taught using examples from statistics and probability, with many coming from applications in the life sciences and in bioinformatics. In this regard, no apriori background in the life sciences is required here; this section makes sense for those with interests in other physical sciences and in the social sciences as well. Those in the biostatistics section of this course will see the core of the linear algebra and differential equations from the regular sections, and also see enough statistics and probability to take various higher level statistics courses (such as Stat 111 and 139).

By the way, successful completion of any section of this course will provide you with a level of sophistication in mathematics that you will not get by learning linear algebra elsewhere, either at Harvard or at any other institution. This sophistication alone will serve you well in your future science courses, and in your future scientific career.

Syllabi for the two kinds of sections are provided below.

Which kind of section should you choose? Here is some advice:

∑ **Those who should take the regular section:**
  a) If you contemplate physics beyond Physics 11a,b, then you are better off in a regular section. In particular, if you plan to take the Physics 15-16 sequence, or

are planning to be a physics concentrator, then you should enroll in a regular section.

b) If you are planning to be a mathematics concentrator, then you should also enroll in a regular section.

∑ **Those who should take the bio/statistics section:**

a) If you are planning to concentrate in BioChemical Sciences, then you are strongly urged by that department to enroll in the bio/statistics section.

b) You will also be better served in the bio/statistics section if you are planning to concentrate in Biology or in Environmental Science and Public Policy.

∑ **Those who can take either section:**

a) If you plan to major in economics or other social sciences, the bio/statistics section might also be your best choice. Likewise, chemistry concentrators who are not planning to take advanced physics courses might find the bio/statistics section useful.

b) If you took Math 19, the bio/statistics section for Math 21b is a perfect follow-up.

c) If you took the BioChem section of Math 21a, the bio/statistics section of this course is a great way to learn more statistics and probability.

d) If you are planning to take the Physics 11a,b sequence, either a regular section or the bio/statistic section is appropriate.

∑ **No worries mate:**

No matter what, you can't go wrong in a regular section so if you aren't sure of your concentration, take a regular section. You can always learn probability and statistics in a later course.

Here are the instructions for making a section request: If you have an email account, log on to the Harvard computer system, and instead of typing "pine," type "ssh section@ulam.fas.harvard.edu" and follow the instructions. Sectioning starts Monday 1/31 and ends at 1pm on Thursday, 2/3. Section assignments are emailed at 5 on Friday, 2/4. If you are having trouble sectioning, if you miss the sectioning period later, or if there is a problem with your section assignment, contact Susan Milano via email at milano@math.harvard.edu.

Course Head: Clifford Taubes, Science Center 504, email chtaubes@math. Drop in office hours on Tuesdays 1-2:30 and Fridays 2-3:30.

Prerequisites: Math 1b with a satisfactory grade, or AB-BC score of at least 4, or scores of at least 20, 8, 4 on the respective three Harvard University Math Placement Tests. Moreover, knowledge of multi-variable calculus as taught in Math 19 or Math 21a is recommended.

Textbooks: All sections require the third edition of Linear Algebra with Applications by Otto Bretscher. The publisher is Prentice Hall. Note that it is crucial that you get the third edition. Those in the bio/statistic section must also obtain the book Probability Theory, a Precise Course, by Y. Rozanov. This book is published by Dover.

Class meetings and problem sessions: The first class meeting, which everyone should attend, is on *Wednesday, February 2 at 8:30 am* in Science Center lecture hall A. This is going to be a roughly 1/2 hour meeting. As short as it is, don't miss it. Except for this one meeting and for the course wide exams, you meet in your assigned section. The section meets for a total of three hours per week, either one hour each on Mondays, Wednesdays and Fridays, or for one and one half hours each on Tuesdays and Thursdays. Each student is also assigned to a 1-hour math problem session, conducted weekly by a course assistant. The meeting time for the problem session will be arranged in your section during the first week of classes. You may attend more than one problem session per week; and the schedule of all problem sessions will be posted on the Math 21b website and on the Calculus Office bulletin board outside of Science Center 308.

The first meeting of MWF sections is on *Monday, February 7* at the posted time. The first meeting for the TTh sections is on *Tuesday, February 8.*

Homework: Problems are assigned each meeting of your section. These are due in the first scheduled section meeting of the subsequent week except if stated explicitly otherwise. However, you are strongly urged to work the problems that are assigned in any given section meeting before the next section meeting, rather than collecting the bunch for the week and working them all at once. Here is the reason: The problems that are assigned at a given section are designed to help you digest the material so as to be ready for its applications in the subsequent section. In particular, you will more easily and efficiently follow the lecture in the subsequent section if you have already done the previous section's homework.

In any event, you are very much encouraged to discuss the homework with your fellow students and to form study groups to work these assignments. However, you must write up the solutions by yourself, and you must note the names of your coworkers somewhere on the homework. (This last point is simply a matter of professional ethics.) The lowest homework score will be disregarded when your average homework grade is computed.

The homework assignments will be posted on the Math 21b web site. The answers to the homework assignments will appear after the due date on the web site as well. Moreover, selected problems from the homework will be discussed in the problem sessions.

Homework assignments that are submitted after their assigned due date will be accepted only at my discretion. In any event, no more than two late homework assignments will be accepted per student over the course of the semester.

In addition to the weekly homework assignment, various problems of a more routine sort will be suggested for the subsequent class meeting. These are not to be turned in as their answers are either in the text book or will be provided otherwise. However, you are strongly urged to work them on your own or with others in the class because their purpose is to supply practice with the techniques and ideas that are presented in the lectures. By the way, you are also encouraged to try on your own other problems from the text to hone your ability with the concepts and techniques. Don't feel that you should limit yourself to the suggested or assigned problems. In this regard, note that the linear algebra text has answers to most of its odd numbered problems. Meanwhile, the probability book supplies answers to many of its problems.

Exams: There will be single course-wide 'midterm' exam and a final exam. The midterm will take place on Tuesday, March 22 from 7-9pm in Science Center lecture halls A and B. Be sure to mark this date on your calendar now, as no make-up will be given. The final exam is scheduled by the University for a date in May. According to the Course Catalogue, the preliminary schedule has the final on Thursday May 26 (Exam Group 1). The University will confirm this date later in the semester.

Grading: Your final grade will be based on your performance on the homework (30%), the midterm (30%), and the final (40%). A small upward adjustment in the grade is possible when the final is dramatically better than the average of the midterms and the homework. The Bio/statistics section will also have an optional extra credit project that can be used to boost the grade.

Computers and calculators: This course is teaching various concepts whose applications are often facilitated by a computer or calculator. Even so, without a strong understanding of the underlying concepts, the computer will be of minimal use to you in your future scientific career. The point here is that there are now many widely available computer programs that will solve linear algebra problems in the blink of an eye. Yet, none of us will live to see the day when a computer can turn processed pig feed into gold. Thus, you should avoid at all costs using computer programs to help with your homework assignments.

In any event, the use of computers and other electronic aids will not be permitted during exams. (Bring only your brain and some pencils.)

Words of Caution and Advice: This course may well be more demanding then your previous mathematics courses at Harvard and elsewhere. In particular, any given assignment might take more time than expected, and you should plan now to set aside regular hours to wrestle with them. It is virtually impossible to do well in this course without working the homework assignments in a timely fashion. Note also that this course is fast paced, and new material builds on old. Thus, do not fall behind. If you find yourself falling behind, *please* contact your section's teacher immediately to discuss options for personal help. Indeed, Harvard provides many services along these lines for its students, and your section teacher can help you find them.

When you are working your assignments, keep in mind that your success in this course will require more than just memorizing formulas and "plugging in values". In this regard, you will consistently be asked to battle with homework and exam problems that differ significantly from material discussed in class.

Here is the key to success: Understand the underlying concepts and then work enough problems so that you can employ them in any example thrown at you.

Math 21b Regular Section Syllabus

The course will cover topics from the following chapters of Linear Algebra with Applications by Bretscher and from the indicated supplementary material.

Chapter 1.1:	Introduction to linear systems.
Chapter 1.2:	Matrices and Gauss-Jordan elimination
Chapter 1.3:	On solutions to linear systems
Chapter 2.1:	Linear transformations and their derivatives
Chapter 2.2:	Linear transformations in geometry
Chapter 2.3:	Inverse of a linear transformation
Chapter 2.4:	Matrix products
Chapter 3.1:	Image and kernel
Chapter 3.2:	Subspaces, bases and linear independence
Chapter 3.3:	Dimension
Chapter 3.4:	Coordinates
Chapter 5.1:	Orthonormal bases and orthogonal projections
Chapter 5.2:	Gram-Schmidt and QR factorization
Chapter 5.3:	Orthogonal transformations
Chapter 5.4:	Least squares and data fitting
Chapter 6.1:	Introduction to determinants
Chapter 6.2:	Determinants
Chapter 7.1:	Introduction to eigenvalues
Chapter 7.2:	Eigenvalues
Chapter 7.3:	Eigenvectors
Chapter 7.4:	Diagonalization
Chapter 7.5:	Complex eigenvalues
Chapter 7.6:	Stability
Chapter 8.1:	Symmetric matrices
Chapter 9.1:	Introduction to differential equations
Chapter 9.2:	Differential equations
Chapter 9.4:	Nonlinear systems
Handout:	Function spaces
Chapter 9.3:	Differential operators
Handout:	Fourier series
Handout:	Introduction to partial differential equations
Handout:	Partial differential equations

- One ‘midterm’ exams: Tue., March 23 from 7-9pm in Science Center A & D
- A final exam scheduled by the University; the preliminary date is Fri., May 21.

Math 21b Bio/statistics section syllabus

The bio/statistics section syllabus will cover the linear algebra topics, and the probability and statistic topics that are listed below.

The linear algebra material will come from the following chapters of Linear Algebra with Applications by Bretscher.

Chapter 1.1: Introduction to linear systems.
Chapter 1.2: Matrices and Gauss-Jordan elimination
Chapter 1.3: On solutions of linear systems
Chapter 2.1: Linear transformations and their derivatives
Chapter 2.2: Linear transformations in geometry
Chapter 2.3: Inverse of a linear transformation
Chapter 2.4: Matrix products
Chapter 3.1: Image and kernel
Chapter 3.2: Subspaces, bases and linear independence
Chapter 3.3: Dimension
Chapter 3.4: Coordinates
Chapter 5.1: Orthonormal bases and orthogonal projections
Chapter 5.2: Gram-Schmidt and QR factorization
Chapter 5.3: Orthogonal transformations
Chapter 5.4: Least squares and data fitting
Chapter 6.1: Introduction to determinants
Chapter 6.2: Determinants
Chapter 7.1: Introduction to eigenvalues
Chapter 7.2: Eigenvalues
Chapter 7.3: Eigenvectors
Chapter 7.4: Diagonalization
Chapter 7.5: Complex eigenvalues
Chapter 7.6: Stability
Chapter 8.1: Symmetric matrices
Chapter 9.1: Introduction to differential equations
Chapter 9.2: Differential equations
Chapter 9.4: Stability in nonlinear systems

The probability and statistics topics will include material of the following sort:

- Introduction to probability: Set theoretic definitions, conditional probability, independence, Bayes' theorem.
- Common discrete and continuous probability distributions, their properties, and their utility.
- Random variables and associated constructions.
- Correlation matrices.

- Central limit theorem.
- Transition matrices and Markov chains.
- Use of data to estimate model parameters (point estimation)
- Use of data to derive probability distributions for model parameters.
- Statistical tests for ‘random’ fluctuations.
- Testing for dependent data.
- The use of time series data to estimate the number of relevant variables.
- Fitting lines and curves to data.
- Identifying functional dependencies in a large dimensional data set.

As noted in the introduction, most of these topics will be introduced in a life science context, but no specific life science background is required.