Course Announcement Spring 2012
M 532 Mathematical Modeling of Large Data Sets

Description: A time-honored approach for the investigation of unexplained phenomena is to attempt to infer laws, or explain processes, from the patterns present in collected data. However, our phenomenal ability to acquire data has outstripped our ability to process and analyze it. Thus, researchers today are confronted with a modern dilemma. Presumably the more information available concerning a phenomenon the better. Yet, a massive data set storing the information, in and of itself, a potentially significant barrier to the investigation. We will present an overview of several mathematical tools for attacking problems associated with analyzing high-dimensional and massive data sets. Our approach is geometric in nature and the main tool is the dimensionality reducing mapping. These mappings are required for the analysis and representation of information (patterns) in large data sets generated by physical or numerical experiments. Topics will include (not necessarily in this order)

- singular value decomposition and the generalized singular value decomposition with applications to signal separation and missing data.
- multidimensional scaling
- subspace analysis, angles between subspaces, canonical correlation analysis
- linear discriminant analysis, support vector machines
- data on manifolds, isomap, Laplacian eigenmaps, local linear embedding, Whitney's theorem and bilipschitz data reduction.
- Fourier and Wavelet analysis

Prerequisites: It is expected that the student will have a mathematics background typical of a graduate student in applied mathematics or engineering including linear algebra, e.g., M369 (M560 preferred). Knowledge of some programming language, e.g. Matlab, Fortran or C++ is also required.

Problem Sets: This will consist of computer assignments as well as the mathematical foundations of the subject. There will be approximately six problem sets all carrying equal weight in computing the final grade.

RamCT: This course will be implemented with the aid of RamCT. All course materials will be posted on RamCT and completed problem sets *must* be uploaded by students to the course RamCT site as single pdf files. Students may discuss projects and lectures on the course blog, chat interactively with other students as well as send high priority email to the instructor via RamCT. If you collaborate with other students at an idea level, indicate the names of these students on your submission; there is no penalty for this. However, each problem set must be completed individually. This means no code sharing and no specific solution sharing.

Required Text: There is no required text.

Background Text: Geometric Data Analysis, Michael Kirby, Wiley & Sons

Time and Place: t.b.a.

Questions? Contact Professor Michael Kirby, Phone: 491-6850, Email: kirby@math . colostate