IT411

Course: Mathematics-II

Objective: To introduce differential equations and other techniques to the students which will be necessary for the academic programme.

Prerequisite: Knowledge of basic calculus.

Course overview:

Ordinary Differential Equations (Linear and Non Linear, Linear stability analysis, Stability criteria)

Higher Order Linear Equations and linear Systems (Wronskian, variation of constants, matrix exponential solution, behaviour of solutions)

Partial Differential Equations (Reaction diffusion, Reaction-advection, reaction-advection-diffusion equation, spatio-temporal patterns, Turing wave and instability)

Stochastic Differential Equations (General SDE, Langevin equation, Master equation, Chemical Langevin equation)

Special Functions (Power Series methods: Bessel functions, Legendre polynomials and Laguerre polynomials, Applications)

Mathematical Modeling: Reaction network, Modeling biological systems, Modeling large networks (biological, social, ecological).

Reference Books:
1. V.I. Arnold, Ordinary Differential Equation, MIT Press.
2. J.D. Murray, Mathematical Biology-I, II, 3rd Edition, Springer.
3. C.H. Taubes, Modeling Differential Equations in Biology, Cambridge University Press, Prentice Hall, Second Edition, 2008.
4. J. Arfken, Weber and Harris, Mathematical methods for Physicists, Elsevier, Seventh Edition, 2013.
5. Bernt Oksendal, Stochastic differential equations, Springer-Verlag Heidelberg New York, Fifth Edition 2003.

Syllabus IT451 Computational Structural Biology M.Sc Third Semester

Instructors : Dr Naidu Subbarao and Arnab Bhattacharjee

Basics of Protein Structure: Physical and Chemical Properties of aminoacids and Polypeptides. Theoretical and experimental methods for determination of sizes of proteins. Physical nature of noncovalent interactions. Motifs of protein structure: Hydrophobic and hydrophilic regions, Alpha-helix, Beta sheets, Loops, Topology diagrams & various structural motifs Conformational properties of proteins, Ramachandran Plot, Secondary, Super Secondary, Tertiary and Quaternary structure of Proteins. Protein folding & Levinthal Paradox, Post translation modification of proteins, Protein sequence, structure and function relationships

Protein Structure Determination by X-ray diffraction: Isolation and Purification of Proteins, Basic principles of X-ray diffraction studies, phase determination, calaculation of Electron Density Map, Interpretation of the electron density map. Refinement of the Structures.. NMR, Electron microscopy Techniques for structure determination, Structure evaluation methods.

Structural Databases: PDB, NDB,, MMDB, CATH, SCOP, FSSP, DALI, protein structural alignment databases, protein-protein interaction database, protein-ligand databases,PubChem, ChEMBL and ZINC databases

Basics of Nucleic Acid Structure, Carbohydrates and Lipids

Protein Structure prediction methods/Molecular Modeling

Domain Assignment methods: Identifying structural domains in protein: first and second generation algorithms for domain assignments, domain assignment based on graph theoretical methods.

Secondary Structure Prediction methods: Secondary structure assignment methods:DSSP/STRIDE for known structures. Statistical methods of Chou and Fasman, Garnier-Osguthorpe-Robson, Stereochemical method of Lim and nearest neighbour, Neural network method etc, Fold Recognition and threading
 methods

Classification of three dimensional structures of proteins, structural alignment methods:

Prediction of structural classes, motifs, folds and domains, classification of three dimensional structures in Brook haven Protein Data Bank (HSSP, SCOP, FSSP, CATH). pairwise/multiple structural alignment methods CE, Dali, SSAP, VAST, concept of rigid body translation and rotation, RMSD, Z score for structural comparison and classification.

Comparative modeling of proteins

Muliple structural alignment of proteins, Template Selection, Backbone modeling , loop building (search / generation), sidechain generation and model evaluation and validation (Ramachandran plot analysis, Procheck, Whatcheck,ERRAT score,Prove, VERIFY3D and PROSA), Energy minimization, Model databases and software for molecular modeling.Molecular visualization and graphics. Solvent accessible surfaces, Identification of Binding site(s) and validation and comparision of binding site(s). protein function assignment at molecular level, protein-protein interaction

Overview of Protein folding process and abnitio Structure prediction methods

Advanced Molecular simulation Techniques:

Empirical force field for biomolecular simulations, Potential Energy Function (bond length potential, aond angle potential, torsional potential, van der wals potential and coulomb potential), classical representations of electrostatics (Poisson-Boltzmann, Generalized Born) Energy minimization – minimization algorithms, simplex method, steepest descent
method, Newton-Raphson, Conjugate gradient, Variable metrics methods. Geometry optimization and energy minimization. Conformational analysis and optimization methods – using systematic searching, simulated annealing, Monte Carlo search,
genetic algorithms, and distance geometry methods. Molecular Dynamics simulations, Monte Carlo Simulations, Calculation of Free energy using simulation techniques. Free energy estimations, thermodynamic cycle, incorporation of solvation effects (implicit and explicit)

Reading :

Introduction to Protein Structure, Branden & Tooze, Garland Publishing, Inc, New York

Structural Bioinformatics, Phil Bourne

Molecular Moelling : Principles & Applications, Andrew R. Leach, Prentice Hall

The art of Molecular Dynamics Simulations. Rapport, Cambridge University Press

IT- 412 Topics and Statistics and Linear Algebra

Credits: 3

Prerequisite: Maths1 and Introduction to Probability and Statistics.

Lectures: 3 hrs a week (minimum)

Objective: This course is meant to provide a solid and rigorous foundation in

a) advanced linear algebra (LA)
b) advanced regression analysis

Course overview: Topics to be covered:

Linear Algebra: The Four Fundamental Vector spaces (row space, column space, null space and left-null space), Eigenvalues and Eigenvectors, Diagonalization, Various forms for decomposition/factorization of matrices (eg. LU-decomposition, QR-decomposition, Singular Value Decomposition)

Statistics: Hypothesis Testing, Analysis of Variance, (Multiple) Linear Regression (LR/MLR), Principal Components Regression (PCR).

The course may be modified based on needs of the class. Make sure you check with the course instructor for sections covered on the exams, HW problems and other course related announcements.

References:

1) Introduction to Linear Algebra (4th edition) by Gilbert Strang

2) Selected video lectures on Linear Algebra by Gilbert Strang downloadable from http://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/video-lectures/

3) An Introduction to Generalized Linear Models (2nd Edition) by Annette J. Dobson

Handouts of relevant articles/book chapters will be given as required.

Grading Policy: The policy for grading subjects is as follows: Quiz/Assignment: 20%; Mid-semester exam: 30%; Final: 50%

Assignments: The assignments will be a mixture of theoretical and programming jobs. You are encouraged to discuss the solutions amongst your fellow students, but the submissions have to be your own and will be graded likewise. Homework submitted late will not be graded. Retake of missed quizzes is not allowed.

Programming assignments are often graded based on your presentation in the class, as such any queries wrt. the grades given should be taken up immediately.

Programming: All assignments must be submitted in R

Remedial Classes: Students in need of remedial classes are requested to contact to contact the instructor at the earliest possible. Students are also encouraged to contact the instructor (preferable via email) for any questions pertaining to this course. These will then be discussed and clarified in the class.

Policy on cheating and plagiarism: Although, cooperation between the students is highly recommended, any act of cheating and plagiarism will not be tolerated and the matter will be handled strictly as per the JNU regulations.

Advanced Physical Sciences
Electrodynamics, Advanced Statistical Mechanics (Ergodic Hypothesis, Ensembles, Distributions, Ising Model, Lattice Model of Proteins), Transport Phenomena

Course Name: Advance Physical Science
Course Code: IT 413
Credit: 3
Course offered to: M.Sc (Sem-III)

Course description: This course aims to provide an elementary introduction to statistical mechanics and connecting it to problems in physical and life science.

Pre-requisites:

Course Outcome(s) (CO): At the end of the course, the students should be able to

1. Explain the basics of statistical mechanics.
2. Can apply to simple discrete systems.

Tentative plan:

Week number	Lecture topic	CO met
1-2	A test will be conducted to check students' knowledge of basic physics and maths (to decide the level of teaching). Review of basic thermodynamics, application. Need for statistical mechanics. Examples	CO-1
3-6	Concept of ensemble, distribution. Ideal gas example. Non-interacting system – harmonic oscillator. Connecting harmonic oscillator to molecular vibration. Interacting system. Examples in physical and biological sciences. Concept of transport phenomena. Evolution of PDF. Derivation of Einstein-Stokes equation for diffusion. Examples.	CO-1 and CO-2.

Text Books and References

(1) Statistical Physics (Berkeley lecture series – volume 5) by F. Reif. McGraw-Hill

(2) Some parts from D. A. McQuarrie, Statistical Mechanics (Harper Collins).

Related articles from journals and Related topics from authentic sources

Course: Graphs, Networks and Concepts of System Sciences

Objective:

Topics to be covered: Graphs, Networks and concepts of system sciences, Introduction to graph algorithms, Spanning Trees, DAGs, Shortest paths, Depth-and-Breadth First search, Random networks, Power law, Graph Matrix implementations, Network, flows

Reference Books:
Narsingh Deo, Graph Theory *with applications to engineering and computer science,* Prentice Hall (India).

Syllabus for Numerical Techniques and Programming Lab – IT421

(Each section contains several methods. Some of them will be explicitly taught in the class and some will be given as task to solve by the students)

1) The Solution of Nonlinear Equations $f(x) = 0$

1. Fixed Point Iteration
2. Bisection Method
3. False Position or Regula Falsi Method
4. Newton-Raphson Method
5. Secant Method
6. Muller's Method
7. Aitken's Method & Steffensen's Acceleration
8. Accelerated & Modified Newton-Raphson
9. Improved Newton Method
10. Halley's Method
11. Horner's Method
12. Lin-Bairstow Method
13. Brent's Method
14. Graeffe's Method
15. Nonlinear Systems
16. Broyden's Method

2) The Solution of Linear Systems $AX = B$

1. Triangular Systems and Back Substitution
2. Gauss-Jordan Elimination and Pivoting
3. Tri-Diagonal Matrices
4. Inverse Matrix
5. LU Factorization
6. Cholesky, Doolittle and Crout Factorizations
7. Jacobi and Gauss-Seidel Iteration
8. Successive Over Relaxation - SOR
9. Pivoting Methods
10. Iterative Refinement
11. Row Reduced Echelon Form
12. Homogeneous Linear Systems
13. Kirchoff's Law
14. Leontief Model
15. Linear Programming-Simplex Method

3) Interpolation and Polynomial Approximation

1. Maclaurin and Taylor Series
2. Lagrange Polynomial Interpolation and Approximation
3. Newton Interpolation Polynomial
4. Hermite Polynomial Interpolation
5. Cubic Splines
6. B-Splines
7. Bézier Curves Bézier Curves
8. Chebyshev Approximation Polynomial
9. Pade Approximation
10. Rational Approximation
11. Aitken's and Neville's Interpolation

12. Legendre Polynomials
13. The Tangent Parabola
14. Catenary

4) Curve Fitting

1. Least Squares Lines
2. Least Squares Polynomials
3. Nonlinear Curve Fitting
4. Logistic Curve
5. FFT and Trigonometric Polynomials
6. Conic Fit
7. Circle of Curvature

5) Numerical Differentiation

1. Numerical Differentiation
2. Richardson Extrapolation
3. Derive Numerical Differentiation Formulae

6) Numerical Integration

1. Riemann Sums
2. Midpoint Rule
3. Newton-Cotes Integration
4. Trapezoidal Rule for Numerical Integration
5. Simpson's Rule for Numerical Integration
6. Simpson's 3/8 Rule for Numerical Integration
7. Boole's Rule
8. Romberg Integration
9. Adaptive Simpson's Rule
10. Gauss-Legendre Quadrature
11. Cubic Spline Quadrature
12. Monte Carlo Pi
13. Monte Carlo Integration
14. 2D Trapezoidal and Simpson Rules

7) Solution of Differential Equations

1. Euler's Method for ODE's
2. Taylor Series Method for ODE's
3. Runge-Kutta Method
4. Runge-Kutta-Fehlberg Method
5. Adams-Bashforth-Moulton Method
6. Milne-Simpson's Method
7. Predictor-Corrector Methods
8. Shooting Methods for ODE's
9. Finite Difference Method for ODE's
10. Galerkin's Method
11. Painleve Property
12. Lotka-Volterra Model
13. Pendulum
14. Projectile Motion
15. Lorenz Attractor

16. van der Pol System
17. Harvesting Model
18. Frobenius Series Solution
19. Picard Iteration
20. Spring-Mass Systems

8) Solution of Partial Differential Equations

1. Finite Difference Method
2. Crank-Nicolson Method
3. Elliptic PDE's

9) Eigenvalues and Eigenvectors

1. Eigenvalues and Eigenvectors
2. Power method
3. Jacobi method
4. Householder Transformations
5. QR method
6. Compartment Model
7. Earthquake Model
8. Matrix Exponential
9. Faddeev-Leverrier Method
10. Hessenberg Factorization

10) Numerical Optimization

1. Golden Ratio Search
2. Fibonacci Search
3. Quadratic Interpolative Search
4. Nelder Mead Method
5. Powell's Method
6. Steepest Descent - Gradient Search
7. Newton's Search Method

Course name : Biological Sequence Analysis and Algorithms
Course Code : IT453
Credit : 4 credits
Course Offeered to : M.Sc III Semester

Course Description:

This course will provide the necessary introduction about the sequence data analysis carried out by biologists in general and also will cover few important computational algorithms in that context. Emphasis will be given to computational methods employed to identify/charecterize/analyse various biological features.

Pre-requisite : Nil

Pre-requisite (Desirable): Idea of very basic concepts in molecular biology especially with gene and genome sequence organisation and its features.

Course outcomes (CO):

1. Students will understand and grasp the rationale behind the computational mthods employed for various biological problems
2. Students will be motivated to undertake small reserach projects dealing with sequnce data
3. This course may expose them to the latest and challenging computational methods/algorithms currently undertaken by researchers in the domain of sequence data analysis

Tentative Course Plan:

Introduction and Overview of Bioinformatics , Survey of both primary and secondary Databases relatng to sequences, structures, gene expression, and regulatory and signal pathways (3 hours)

Overview of Sequence data analysis in general – Software tools and web servers   (3 hours)

Sequence alignments algorithms : Scoring matrices and scoring functions – pairwise alignment local and Global – Multiple sequence   alignment – Database search – BLAST versions (6 hours)

Statistics and serach method for Sequence patterns, profiles, motifs etc – MEME, Weight matrix, Profile, sequence Logo etc   (3 hours)

Gene Identification problem : Codon usage, fourier and markov models   (3 hours)

Overview of advanced concepts : Information theory, machine learning and probabilistic modelling (3 hours)

Phylogenetic analysis: Concept of distance – Distance, parsimony and likhlihood methos (4 hours)

RNA secondary structure – RNA folding algorithms and RNA databases   (2 hours)

Compuattional Genomics – current topics   (3 hours)

Total Theory	30 hours
Tutorials	10 hours

References:

Text Book:
1. Durbin et al (2003). Biological sequence analysis : probablistic models of Proteins and Nucleic acids. Cambridge University Press (Indian Edition)

Reference Book:
1. Jonathan Pevesner . Bioinformatics and Functional Genomics (3rd Edition) Wiley Blackwell
2. Handbook of Computational Molecular Biology Ed by Srinivas Aluru. (2005). Indian Edition Chapman & Hall/CRC

IT766: Genomics : Concepts, Methods and Applications (GCMA)
3 Credits

Examination: Midsem (10 assignments + 40) & Final (10 assignments+ 40)

Pre-requisites: Must complete the 1st semester and at least credited the course on "Computational Biology & Bioinformatics". Must have basic knowledge of Molecular Biology and Recombinant DNA Technology.

Objective: This course is an integrated presentation of genome organization, genome sequencing and characterization, comparative genomics, transcriptomics and introductory genomic data analysis. The course objective is to instill sufficient knowledge to the students to be conversant in all of the areas of genomics and provide a knowledge base that enables the student to successfully move on and master advanced topics in genomics. On the whole, the course will enhance the students' overall comprehension of the subject, improve their computational skills and eventually assist in proper planning, execution and analysis of their research work.

Course content:

Next Generation Sequencing Technologies: Methods and applications (4)

Whole Genome Sequencing and Analysis: Concept, methods, assembly (de novo and reference-based), genome annotation (structural and functional), comparative genomics (10)

High-throughput Transcriptome Profiling: Concept, methods and applications; transcriptome construction (de novo and reference-based), differential gene expression (10)

Non-coding RNAs: Small RNAs, miRNAs, long non-coding RNAs; sequencing and prediction methods; biological relevance (10)

Single nucleotide polymorphisms: Genome resequencing; data processing and SNP prediction; applications in agriculture/human health (6)

Suggested readings:
1. Review and research articles in national and international journals
2. Onlines webpages/videos of various databases and sequencing platforms.
3. Principles of genome analysis and genomics - Sandy B. Primrose, Richard M. Twyman
4. Introduction to Bioinformatics: A theoretical and practical approach - Stephen A. Krawetz, David D. Wonmble
5. Fundamentals of data mining in genomics and proteomics - Werner Dubitzky, Martin Granzow, Daniel Berrar
6. A primer of genome science - Greg Gibson, Spencer V. Muse
7. Metzker ML (2010) Sequencing Technologies—The Next Generation. Nature Reviews Genetics 11: 31 PMID: 19997069
8. Mardis ER. (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet. 2008;9:387-402.

Course Name: **Advanced Computational Methods for Optimizations (ACMO)**
Course Code: **IT775**
Credit: 3
Course offered to: M.Sc./Ph.D. (Common for CS-Track & CB-Track)
PG diploma for Data Science

Course description: The development of fast, efficient and inexpensive computers has significantly increased the range of complex problems specially biological/complex networks that can be solved reliably. Computational techniques use computers to solve problems by step-wise, repeated and iterative solution methods, which would otherwise be tedious or unsolvable by hand calculations. This course is designed to give an overview of computational methods of interest to model biological network for optimal solution(s). A part will be class room teaching and other part will be working exercise with project assigned to the students along with the demonstrations of some mathematical software(s) for problem solving (if the software is available).

Pre-requisites: Must complete the semester one and at least credited the course on Computational Biology & Bioinformatics. A part will be classroom teaching and other part will be working exercise (assignments/practicals) with project assigned to the students.

Course Outcome(s) (CO): Upon finishing the course, the student is expected to be able to:
1. Knowledge of errors analysis in computation problems.
2. Knowledge of solving system of linear equations by various mathematical methods.
3. To find the optimal solutions of complex/wide networks problems.
4. This paper helps to understand various decision making optimization techniques such as game theory, AHP, etc.

Tentative plan:

Week number(s)	Lecture topic	CO met
1-2	Introduction : Motivation and applications. Computation and Error Analysis: Accuracy and precision; Truncation and round-off errors; Binary Number System; Error propagation.	CO 1
3-5	Linear Systems and Equations: Matrix representation; Cramer's rule; Gauss Elimination; Matrix Inversion; LU Decomposition; Iterative Methods; Relaxation Methods; Eigen Values.	CO 2
6-8	Optimization – Introduction, Formulation of LPP , Geometry of LPP and Graphical Solution of LPP, Solution of LPP : Simplex Method, Big-M Method, Two-Phase Method, Special Cases in Simple Applications	CO 3
9-10	Introduction to Duality Theory, Dual Simplex Method, Post Optimaility Analysis	CO 4
11-12	Introduction to Transportation Problems & Assignment Problems Solving Various types of Transportation Problems and Assignment Problems	CO 5
12-14	Analytic Hierarchy Process, Multi Objective Decision Making,Game Theory (Introduction)	CO 6

Note: To encourage and motivate the students, topic related to further studies can be included/changed in the syllabus to maintain the quality of research.

Some Text Books and References:
1. Gupta S.K. (1995) Numerical Methods for Engineers, New Age International.
2. Chapra S.C. and Canale R.P. (2006) Numerical Methods for Engineers, 5th Ed; McGraw Hill.
3. H.A. Taha, Operations Research: An Introduction, 8th Edition, Prentice Hall, 2006.
4. G. Hadley, Linear Programming, Narosa. 1987 (2002 reprint available).
5. F.S. Hillier &G.J. Lieberm, An: Introduction to Operations Research-Concepts and Cases, 9th Edition, Tata M cG raw Hill. 2010.
6. R.E. Steuer, Multiple Criteria Optimization: Theory, Computation and Application, John Wiley, New York, 1986.
7. Ravindran A., Phillips, D.T., Solberg J.J.,. - Operations Research: Principles and Practice, 2nd ed.,2001, John Wiley & Sons.
8. Related articles from journals
9. Hyperlinks: en.wikipedia.org/wiki/Operations Research