MA 576 Optimization for Data Science

Course Description	The objective of this course is to introduce the students to the theory and methods of optimization used in data science. The first portion of the class focuses on elements of convex analysis and subgradient calculus for non-smooth functions, optimality conditions for differentiable and for non-smooth optimization problems, and Lagrangian duality. The main part of the class discusses numerical methods for optimization with a focus and application to problems arising in data science. Approaches to large-scale/big-data optimization include decomposition methods, design of distributed and parallel methods of optimization, as well as stochastic approximation methods. Examples of optimization models in classification, clustering, statistical learning, and compressed sensing will be discussed in order to illustrate the theoretical and numerical challenges and to demonstrate the scope of applications.
Prerequisites	Undergraduate knowledge of multivariate calculus, linear algebra, and probability.

By the end of the course, you will be able to:

1. **Apply** a variety of optimization methods and techniques for the solution of a wide range of optimization problems including differentiable and non-differentiable functions, together with the verification of constraint qualification conditions.
2. **Determine** the pros and cons of different formulations and solution methods and the interaction between models and methods.
3. **Create** and Implement appropriate numerical methods for solving a particular problem formulation while being aware of their convergence properties.
4. **Utilize** and modify available code and libraries to solve optimization problems.

Course Materials

1. [B] Dimitry Bertsekas. Nonlinear programming. Athena Scientific. 2nd edition 1999.
2. [R] Andrzej Ruszczynski, Nonlinear Optimization, Princeton University Press, 2006.
3. [BV] S. Boyd and L. Vandenberghe. Convex Optimization. Cambridge University Press, 2004. http://stanford.edu/~boyd/cvxbook/
4. [SDR] Shapiro, Alexander, Darinka Dentcheva, and Andrzej Ruszczyński. Lectures on stochastic programming: modeling and theory. Society for Industrial and Applied Mathematics, 2014.
5. [S] James C. Spall, Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control, Wiley, Hoboken, 2003

Module #	Module Outcomes	Readings	Graded assessments
1 **Introduction and Prerequisites.**	By the end of this module, you will be able to do the following: **Demonstrate** Proficiency in multivariable calculus and other prerequisites. **Use** Mathematica to illustrate basic optimization problems. **Illustrate** Optimization problems in data science (e.g., matrix completion, SVM, risk minimization, clustering). **Formulate** Quadratic approximations of smooth functions. **Compute** Hessian matrices and determine if it is positive-definite. **Solve** Multivariable maxima and minima basic problems. **Solve** basic optimization problems **Find** local minima **Apply** necessary and sufficient conditions for local minima	Slides [B] Appendix A, Section 1.1 Python+colab tutorial [Python Numpy Tutorial (with Jupyter and Colab)]()	Begin Homework 1

Module #	Module Outcomes	Readings	Graded assessments
2 **Convex Sets: Definition and Representation.**	By the end of this module, you will be able to do the following: **Verify** the convexity of a given set. **Formulate** the convex hull of a given set and apply Caratheodory's theorem. **Formulate** the representation of a convex set by its extreme points and its recession cone. **Identify** extreme points of polyhedral sets. **Compute** projections over convex sets. **Determine** whether a set is a cone.	Slides [R] Sections 2.1.3; 2.2.2; 2.2.3 [B] Appendix B.1 [BV] Sections 2.1; 2.2; 2.3; 2.5	Submit Homework 1
3 **Convex Functions and Convex Problems.**	**Determine** the separability of convex sets. **Check** the convexity of a given function and the convexity of sets described by nonlinear functions. **Apply** convexity of level sets of convex functions and their properties. **Determine** if a function is convex. **Characterize** convex optimization problems. **Solve** convex problems using cvxpy.	Slides [B] Appendix B.1; B.2; B.3 [BV] Sections 3.1; 3.2; 4.1; 4.2; 4.3; 4.4 [R] Section 2.4	Begin Homework 2
4 **Optimality conditions; Gradient methods.**	**Derive** and apply conditions of optimality for linear and nonlinear unconstrained problems. **Identify** critical points and then determine the minima by using the second-order optimality conditions. **Apply** descent methods with suitable	Slides [B] Section 1.1; 1.2 Gradient Descent: Downhill to a Minimum	Submit Homework 2

Module #	Module Outcomes	Readings	Graded assessments
	directional minimization. **Select** the optimal step size of the method. **Determine** conditions for the convergence of the method.		
5 **Newton's method; Least squares problems.**	**Analyze** the rate of convergence and determine the optimal convergence rate. **Apply** Newton's method for global convergence. **Determine** its convergence rate. **Apply** variants of Newton's and determine suitability. **Apply** Gauss-Newton method for least squares problems.	Slides [B] Section 1.3; 1.4; 1.5 Newton's Method in Optimization Four Ways to Solve Least Squares Problems	Start Homework 3
6 **Conjugate gradient method; Quasi-Newton methods.**	**Apply** the conjugate gradient method and use preconditioners. **Apply** Quasi-Newton methods. **Apply** finite difference approximations.	Slides [B] Section 1.5; 1.6; 1.7 Conjugate gradient method	Submit Homework 3
7 **Optimization Over a Convex Set; Feasible Direction Methods.**	**Determine** optimality conditions over a convex set. **Apply** optimization subject to bounds and over a simplex. **Apply** the conditional gradient method. **Apply** the projected gradient method and alternatives.	Slides [B] Section 2.1; 2.2; 2.3; 2.4; 2.5	Start Homework 4 **Midterm**

Module #	Module Outcomes	Readings	Graded assessments
	Determine convergence rates.		
8 **Subgradient Calculus and Methods.**	**Apply** the definition and calculus of subgradients and their relation to directional derivatives. **Apply** the subgradient method for non-smooth optimization. **Apply** the projected subgradient method.	Slides [R] Section 7.1 Section 6.3	Submit Homework 4
9 **Constrained Optimization and Lagrange Multipliers.**	**Apply** Lagrange multipliers to solve constrained optimization problems. **Prove** the Lagrange Multiplier Theorem via the penalty method. **Verify** the constraint qualification conditions for nonlinear problems. **Apply** sensitivity analysis.	Slides [B] Section 3.1; 3.2	Start Homework 5
10 **Inequality Constraints; Interior Point Methods**	**Check** necessary and sufficient conditions for equality and inequality-constrained problems. **Apply** the Barrier method to solve constrained problems.	Slides [B] Section 3.3; 4.1	Submit Homework 5
11 **Penalty and Augmented Lagrangian Methods.**	**Apply** penalty methods to solve problems and recover the optimal Lagrange multipliers. **Apply** the multiplier method to solve constrained problems. **Apply** the multiplier method to solve constrained problems.	Slides [B] Section 4.2; 4.4 Newton's Method for constrained optimization problems	Start Homework 6

Module #	Module Outcomes	Readings	Graded assessments
	Apply Newton's method to solve the Lagrangian system of an equality-constrained problem.		
12 **Introduction to Duality; Duality Theory**	**Formulate** dual problems. **Check** duality relations. **Apply** the dual approach to solving an optimization problem. **Apply** weak and strong duality to solve an optimization problem.	Slides [B] Section 3.4; 5.1	Submit Homework 6
13 **Dual Computational Methods: Cutting Plane and ADMM; Stochastic Gradient Descent**	**Formulate** cutting planes. **Apply** the cutting plane method in its basic form. **Use** a decomposition approach like the alternating direction method of multipliers (ADMM). **Formulate** the basic form and the projected form of a stochastic method. **Choose** proper step sizes ensuring convergence of the method. **Use** the method with averaging	Slides [B] Section 6.3; 6.4	**Final**

Assessment Descriptions and Course Point Distribution

Assignment/Assessment	Competency & Behavior	Percentage of Final Grade	Due Date

Homework		35%	
Live Session and Discussion Participation		10%	
Mid-term		25%	
Final		30%	
	TOTAL:	100%	