East Coast Optimization Meeting 2022

Theme: Nonsmooth Optimization

Dates

March 31 - April 1, 2022

Location

Virtually hosted by Center for Mathematics and Artificial Intelligence
George Mason University
Fairfax, VA

Organizing Committee

Harbir Antil (George Mason University)
Drew P. Kouri (Sandia National Laboratories)
Denis Ridzal (Sandia National Laboratories)

Sponsors

National Science Foundation
Air Force Office of Scientific Reserach (AFOSR)
Society for Industrial and Applied Mathematics (SIAM) Washington-Baltimore Section
Center for Mathematics and Artificial Intelligence, George Mason University
Department of Mathematical Sciences, George Mason University
College of Science, George Mason University
Association for Women in Mathematics, GMU Chapter
Society for Industrial and Applied Mathematics (SIAM), GMU Chapter

Keynote Speakers

1. Amir Beck (Tel-Aviv University)

TUTORIAL:

Title. Proximal-Based Methods in Convex Optimization

Abstract. Proximal-based methods are recognized as extremely effective methods for solving large-scale problems arising from different applications areas. This tutorial will survey several proximal-based algorithms, including proximal (sub)gradient, dual proximal gradient, FISTA, smoothing, block proximal as well as various splitting methods. Particular emphasis will be placed on complexity results, and various examples will be given to demonstrate the applicability of the described methods.

PUBLIC LECTURE:

Title. On the convergence to stationary points of deterministic and randomized feasible descent directions methods

Abstract. We study the class of nonsmooth nonconvex problems in which the objective is to minimize the difference between a continuously differentiable function (possibly nonconvex) and a convex (possibly nonsmooth) function over a convex polytope. This general class contains many types of problems, including difference of convex functions (DC) problems, and as such, can be used to model a variety of applications. We develop deterministic and randomized methods based on the notion of positive spanning sets for which it is proven that accumulation points are necessarily stationary points. We also study a new optimality measure for which we establish an efficiency estimate. This talk is based on two joint works with Nadav Hallak

2. Michael Ulbrich (Technical University Munich)

TUTORIAL:

Title. Semismooth Newton Methods – Theory and Applications

Abstract. Nonsmoothness is a very powerful modelling paradigm that allows, e.g., to reformulate systems of inequalities as equations or to enforce structural properties like sparsity or low rank. The possibility to rewrite specific systems of inequalities as nonsmooth equations opens the door for rephrasing optimality conditions of (smooth) inequality-constrained nonlinear optimization or optimal control problems equivalently as nonsmooth systems and building solvers upon this formulation. Even in the case of nonsmooth objective functions, where usually set-valued subdifferentials arise in the optimality conditions, the use of proximity operators often still allows to rewrite them as nonsmooth equations.

Analyzing and solving nonsmooth problems is often significantly more challenging than in the differentiable case. A systematic and productive way of tackling nonsmooth systems of equations is provided by the concepts of Newton-differentiability and semismooth Newton methods. This class of methods is at the core of many efficient solution algorithms. The goal of this tutorial is to give an introduction to the theoretical and numerical aspects of semismooth Newton methods, with a certain focus on infinite-dimensional optimization and applications in PDE-constrained optimization. In addition, we also highlight novel developments for this class of methods in machine learning and in other fields.

PUBLIC LECTURE:

Title. An Approximation Scheme for Distributionally Robust Nonlinear Programming with Applications to PDE-Constrained Optimization under Uncertainty

Abstract. We present a sampling-free approximation scheme for distributionally robust nonlinear optimization (DRO). The DRO problem can be written in a bilevel form that involves maximal (i.e., worst case) value functions of expectation of nonlinear functions that depend on the optimization variables and random parameters. The maximum values are taken over an ambiguity set of probability measures which is defined by moment constraints. To achieve a good compromise between tractability and accuracy we approximate nonlinear dependencies of the cost / constraint functions on the random parameters by quadratic Taylor expansions. This results in an approximate DRO problem which on the lower level then involves value functions of parametric trust-region problems and of parametric semidefinite programs. These value functions are in general nonsmooth with respect to the decision variables. Using trust-region duality, a barrier approach, and other techniques we construct gradient consistent smoothing functions for the value functions and show global convergence of a corresponding homotopy method. We discuss the application of our approach to PDE constrained optimization under uncertainty and present numerical results. This is joint work with Johannes Milz.

Guest Speakers

1. Aleksandr Aravkin (University of Washington)

   Title: Relax-and-split method for structured optimization

   Abstract: We present a relaxation method for structured convex and nonconvex optimization problems. The method is easy to apply to many different kinds of problems, and as a result has been successfully used in practical applications over the last few years. We summarize the main idea, present some use cases, and highlight recent work in variable selection for mixed effects models.

2. Damek Davis (Cornell University)

   Title: Avoiding saddle points in nonsmooth optimization

   Abstract: We introduce a geometrically transparent strict saddle property for nonsmooth functions. When present, this property guarantees that simple randomly initialized proximal algorithms on weakly convex problems converge only to local minimizers. We argue that the strict saddle property may be a realistic assumption in applications since it provably holds for generic semi-algebraic optimization problems. Finally, we close the talk with an extension of the result to "perturbed" subgradient methods.

3. Jelena Diakonikolas (University of Wisconsin-Madison)

   Title: Faster Nonsmooth Empirical Risk Minimization

   Abstract: Empirical Risk Minimization (ERM) problems are central to machine learning, and their efficient optimization has been studied from different perspectives, often taking advantage of the finite sum structure present in typical problem formulations. In particular, tight oracle complexity bounds have been obtained under fairly general assumptions about the loss functions. In this talk, I will present a rather surprising and general result that takes advantage of the separability of nonsmooth convex loss functions with efficiently computable proximal operators – such as, e.g., the hinge loss and the sum of absolute errors – to obtain an algorithm that exhibits significantly lower complexity than what is predicted by the lower bounds for general nonsmooth convex losses. I will then discuss how this result can be further improved for problems that can be stated in a form that is close to a linear program, and how such a result gives faster runtimes for problems arising in distributionally robust optimization. The talk is based on joint papers with Chaobing Song, Eric Lin, and Stephen Wright.

4. Courtney Paquette (McGill University)

Title: Algorithms for stochastic nonconvex and nonsmooth optimization

Abstract. Nonsmooth and nonconvex loss functions are often used to model physical phenomena, provide robustness, and improve stability. While convergence guarantees in the smooth, convex settings are well-documented, algorithms for solving large-scale nonsmooth and nonconvex problems remain in their infancy.

I will begin by isolating a class of nonsmooth and nonconvex functions that can be used to model a variety of statistical and signal processing tasks. Standard statistical assumptions on such inverse problems often endow the optimization formulation with an appealing regularity condition: the objective grows sharply away from the solution set. We show that under such regularity, a variety of simple algorithms, subgradient and Gauss Newton like methods, converge rapidly when initialized within constant relative error of the optimal solution. We illustrate the theory and algorithms on the real phase retrieval problem, and survey a number of other applications, including blind deconvolution and covariance matrix estimation.

Contributed Talks

(1) Logan Beaver (University of Delaware)
Constraint-Driven Control for Multi-Agent Systems

(2) Matthias Chung (Virginia Tech)
Learning Regularization Parameters of Inverse Problems via Deep Neural Networks

(3) Kelsey DiPietro (Sandia National Laboratories)
Optimization based solvers for the Monge-Ampére equation with applications to mesh adaptivity

(4) Jennifer Erway (Wake Forest University)
Advances in multipoint symmetric secant methods

(5) Malena Espanol (School of Mathematical and Statistical Sciences, Arizona State University)
An ℓp Variable Projection Method for Large-Scale Separable Nonlinear Inverse Problems

(6) Evelyn Herberg (George Mason University)
An Optimal Time Variable Learning Framework for DNNs

(7) Ratna Khatri (U.S. Naval Research Laboratory)
Optimal Control Framework For Deep Autoencoders

(8) Jiaming Liang (Georgia Institute of Technology)
A unified analysis of a class of proximal bundle methods for solving hybrid convex composite optimization problems

(9) Xiao Ling (Virginia Commonwealth University)
L1-norm regularized L1-norm best fit line problem

(10) Anton Lukyanenko (George Mason University)
Sampling-based path planning: a metric approach

(11) Mark McFeaters (University of Tennessee at Chattanooga)
Design of Compact and Connected Reserve Systems for Ecological Conservation

(12) Tina Mai (Duy Tan University at Da Nang (Vietnam) and Texas A&M University at College Station (USA))
Theory of functional connections applied to quadratic and nonlinear programming under equality constraints

(13) Vivak Patel (University of Wisconsin – Madison)
Global Convergence and Stability of Stochastic Gradient Descent

(14) Ashwin Renganathan (University of Utah)
Lookahead Bayesian optimization and applications

(15) Steven Rodriguez (United States Naval Research Laboratory)
Accelerating Derivative-Free Inverse Analyses in Process Parameter Identification via Model-Reduction with Advection-Segmented Local Reduced-Order Subspaces

(16) Tim Roith (Friedrich-Alexander Universität Erlangen-Nürnberg)
A Bregman Learning Framework for Sparse Neural Networks

(17) Alessandro Scagliotti (Scuola Internazionale Superiore di Studi Avanzati)
A piecewise conservative method for unconstrained convex optimization

(18) Melanie Weber (University of Oxford)
Constrained Optimization on Riemannian manifolds

(19) Manuel Weiß (Interdisciplinary Center for Scientific Computing (Heidelberg))
Geometry Segmentation with Total Variation Regularization

(20) Roozbeh Yousefzadeh (Yale University)
Decision Boundaries and Convex Hulls in the Feature Space that Deep Learning Functions Learn from Images

Abstracts

Constraint-Driven Control for Multi-Agent Systems

Logan Beaver[1]

[1]University of Delaware

Automatic control of decentralized cyber-physical systems is an emerging field with applications in transportation, logistics, and environmental monitoring. As these systems move into physical spaces, one key challenge is ensuring their continued operation at large distance and time scales. In this presentation, I outline an approach to multi-agent autonomy using constraint-driven optimal control. By embedding the agent-agent and agent-environment interactions as constraints, it is possible to make guarantees about the emergence, or suppression, of global-scale behavior. Furthermore, the behavior of individual agents can be entirely explained by identifying which constraints are active. Finally, the size of each agent's feasible space can be incorporated into an event-driven change in behavior, e.g., an autonomous vehicle with a small feasible space may switch to an overtaking or emergency braking mode.

Learning Regularization Parameters of Inverse Problems via Deep Neural Networks

Matthias Chung[1]

[1]Virginia Tech

In this work, we describe a new approach that uses deep neural networks (DNN) to obtain regularization parameters for solving inverse problems. We consider a supervised learning approach, where a network is trained to approximate the mapping from observation data to regularization parameters. Once the network is trained, regularization parameters for newly obtained data can be computed by efficient forward propagation of the DNN. We show that a wide variety of regularization functionals, forward models, and noise models may be considered. The network-obtained regularization parameters can be computed more efficiently and may even lead to more accurate solutions compared to existing regularization parameter selection methods. We emphasize that the key advantage of using DNNs for learning regularization parameters, compared to previous works on learning via optimal experimental design or empirical Bayes risk minimization, is greater generalizability. That is, rather than computing one set of parameters that is optimal with respect to one particular design

objective, DNN-computed regularization parameters are tailored to the specific features or properties of the newly observed data. Thus, our approach may better handle cases where the observation is not a close representation of the training set. Furthermore, we avoid the need for expensive and challenging bilevel optimization methods as utilized in other existing training approaches. Numerical results demonstrate the potential of using DNNs to learn regularization parameters.

Optimization based solvers for the Monge-Ampére equation with applications to mesh adaptivity

Kelsey Dipietro[1]

[1]Sandia National Laboratories

Adaptive mesh refinement is a crucial component for accurately solving partial differential equations numerically, particularly in the case of nonlinear behaviors such as shocks and singularities. One method for providing mesh adaptation is r-adaptivity, in which a fixed number of mesh points is redistributed within the domain toward high interest areas. R-adaptive meshes can be generated by solving an optimal transport problem, whose solution gives a transportation plan for distributing a probability measure from one location to another with minimal cost. The optimal transport map can be calculated by solving the Monge-Ampére equation, a fully nonlinear partial differential equation. In this presentation, we present a low-order mixed finite element approximation for solving the Monge-Ampére equation, whose fast solution hinges on using an optimized based nonlinear solver. We will show how our low order method scales up straightforwardly and outperforms previous methods. We conclude with several examples of mesh adaptivity using the Monge-Ampére equation.

Advances in multipoint symmetric secant methods

Oleg Burdakov[1], Jennifer Erway[2], and Mostafa Rezapour[2]

[1]Linköping University, [2]Wake Forest University

In this talk, we discuss a new multipoint symmetric secant (MSS) that uses a dense initial matrix rather than a multiple of the identity initial matrix. We discuss the convergence analysis of the new method and compare the numerical results of applying the new method with the standard MSS, which uses a multiple of the identity initial matrix, on several problems from the CUTEst test problem set.

An ℓp Variable Projection Method for Large-Scale Separable Nonlinear Inverse Problems

Malena Espanol[1], Mirjeta Pasha[1]

[1]Arizona State University

Variable projection methods are among the classical and efficient methods to solve separable nonlinear least squares problems such as blind deconvolution, system identification, and machine learning. In this talk, we present a modified variable projection method for large-scale separable nonlinear inverse problems, that promotes edge-preserving and sparsity properties on the desired solution, and enhances the convergence of the parameters that define the forward problem. Specifically, we adopt a majorization minimization method that relies on constructing quadratic tangent majorants to approximate an ℓp regularization term, by a sequence of $\ell 2$ problems that can be solved by the aid of generalized Krylov subspace methods at a relatively low cost compared to the original unprojected problem. In addition, more potential generalized regularizers including total variation (TV), framelet, and wavelet operators can be used, and the regularization parameter can be defined automatically at each iteration with the aid of generalized cross validation. Numerical examples on large-scale two-dimensional imaging problems arising from blind deconvolution are used to highlight the performance of the proposed method in both quality of the reconstructed image as well as the reconstructed forward operator.

An Optimal Time Variable Learning Framework for DNNs

Harbir Antil[1], **Hugo Díaz**[2], **Evelyn Herberg**[1]

[1]George Mason University, [2]University of Delaware

Deep neural networks (DNN) are considered as time-discretizations of non-linear ordinary differential equations (ODE). The novelty lies in letting the time step-length vary from layer to layer, which needs to be learned, in an optimization framework. Thus enabling us to introduce a completely new paradigm in deep learning. As examples, residual neural networks (ResNets) and DNNs with memory (Fractional-DNNs) are considered. This framework is shown to help overcome the vanishing and exploding gradient issues. The proposed approach is applied to an ill-posed parametrized 3D-Maxwell's equation and to a standard classification problem.

Optimal Control Framework For Deep Autoencoders

Ratna Khatri[1]

[1]U.S. Naval Research Laboratory

In this talk, we introduce an optimal control and low rank tensor framework for autoencoder type deep neural networks (DNNs). The learning problem is an optimization problem subject to two differential equations, representing encoder and decoder, as constraints. This approach is mathematically rigorous and offers multiple advantages, like a compressed network due to rank reduction, significant memory savings, and the ability to train on small data. We show successful application of this network in various computer vision tasks like image denoising and image deblurring.

A unified analysis of a class of proximal bundle methods for solving hybrid convex composite optimization problems

Jiaming Liang[1]

[1]Georgia Institute of Technology

This talk presents a proximal bundle (PB) framework based on a generic bundle update scheme for solving the hybrid convex composite optimization (HCCO) problem and establishes a common iteration-complexity bound for any variant belonging to it. As a consequence, iteration-complexity bounds for three PB variants based on different bundle update schemes are obtained in the HCCO context for the first time and in a unified manner. While two of the PB variants are universal (i.e., their implementations do not require parameters associated with the HCCO instance), the other newly (as far as the authors are aware of) proposed one is not but has the advantage that it generates simple, namely one-cut, bundle models. The talk also presents a universal adaptive PB variant (which is not necessarily an instance of the framework) based on one-cut models and shows that its iteration-complexity is the same as the two aforementioned universal PB variants.

L1-norm regularized L1-norm best fit line problem

Xiao Ling[1]

[1]Virginia Commonwealth University

We develop a sparse and outlier-insensitive method for one-dimensional line fitting that can be used as the basis for outlier-insensitive machine learning methods such as principal component analysis. The method is insensitive to outlier observations by formulating procedures as optimization problems seeking the L1-norm best-fit line. It is also able to produce a small number of non-zero principal components with additional penalty term to take sparseness into account. Computational results demonstrate that this method can provide outlier-insensitive and sparse solutions.

Sampling-based path planning: a metric approach

Anton Lukyanenko[1]

[1]George Mason University

The RRT* path planning algorithm is widely used in navigation problems, but its convergence guarantees have remained unclear in many common use cases. We provide a new analysis of the algorithm that strengthens the guarantees in Euclidean space and extends them to non-Euclidean spaces. The latter include state spaces for fleets of self-driving vehicles, as well as abstract spaces like the Sierpinski gasket.

Theory of functional connections applied to quadratic and nonlinear programming under equality constraints

Tina Mai[1,2], **Daniele Mortari**[2]

[1]Duy Tan University at Da Nang (Vietnam), [2]Texas A&M University at College Station (USA)

In [Tina Mai and Daniele Mortari. Theory of functional connections applied to quadratic and nonlinear programming under equality constraints. Journal of Computational and Applied Mathematics, 406:113912, 2022], we use the theory of functional connections to introduce an efficient method for solving quadratic and nonlinear programming problems with linear equality constraints. This is accomplished without the need for a traditional Lagrange multiplier. To address the constrained quadratic programming problem as an unconstrained one for closed-form solution, two distinct formulas (completely satisfying the equality constraints) are provided. These formulas are generated using an optimization variable vector, also known as the free vector g in the theory of functional connections. In the spirit of this theory, the Newton's approach combined with an elimination scheme in optimization is employed to solve the equality constrained nonlinear programming problem. A numerical example of the proposed approach is used to support the convergence analysis.

Design of Compact and Connected Reserve Systems for Ecological Conservation

Mark McFeaters[1]

[1]University of Tennessee at Chattanooga

Protected large, compact, and connected areas are imperative in biological conservation efforts. The classical set-based models select areas that maximize the protection of the most species or minimize the cost of choosing locations to provide small regions scatted for selection. Recently, more models have been proposed to create protected areas that are large, compact, and then connected with corridors through additional steps. We demonstrate a new approach that integrates these goals to create a reserve system of one or more clusters that are efficient, extensive, compact, and connected. We identify compact groups using graph density which define as the ratio of the number of graph edges to the number of nodes. Although maximizing graph density can be formulated as a fractional optimization problem, we show it can be structured and solved as a linear integer program. Further, we proposed a breadth-first search algorithm to join clusters in an efficient manner. We demonstrate the performance of our approach using real data and test problems available in the literature.

Global Convergence and Stability of Stochastic Gradient Descent

Vivak Patel[1]

[1]University of Wisconsin – Madison

Stochastic gradient descent (SGD) is widely deployed in a number of different disciplines, often on non-convex problems with complicated noise models. However, SGD's existing convergence theory often does not apply. In this talk, we will establish the need for a convergence theory under broader assumptions with some simple examples. We will then state a global convergence result for SGD under these broad assumptions. Then, we will discuss the issue of stability, which addresses what happens when SGD's iterates diverge. If time remains, we will preview some of our recent results on local asymptotic convergence rates.

Lookahead Bayesian optimization and applications

Ashwin Renganathan[1]

[1]University of Utah

We propose a novel Bayesian optimization framework that introduces a *lookahead* acquisition principle. This includes a generalized framework which can be used to construct lookahead versions of existing (myopic) acquisition principles. Furthermore, we show that our framework can be leveraged to solve problems in a multifidelity setting, i.e., when multiple models that trade accuracy for computational expense are available. We demonstrate our method on application problems in optimization, surrogate modeling, and uncertainty quantification.

Accelerating Derivative-Free Inverse Analyses in Process Parameter Identification via Model-Reduction with Advection-Segmented Local Reduced-Order Subspaces

Steven Rodriguez[1], **John C. Steuben**[1], **Athanasios P. Iliopoulos**[1], **John G. Michopoulos**[1]

[1]United States Naval Research Laboratory

Inverse problems can greatly benefit from projection-based model-order reduction (PMOR) to significantly reduce computational costs in time and resources. However, PMORs deployed on advection-dominated problems often suffer from slow-decaying Kolmogorov N-widths resulting in suboptimal subspace dimensions that can limit cost savings. The presented work will demonstrate the development of PMORs equipped with advection-segmented local reduced- order subspaces to overcome slow-decaying Kolmogorov N-widths. A demonstration of the proposed PMOR will be presented along with its application to derivative-free inverse analyses of process parameter identification for a class of heat-deposition problems akin to additive manufacturing (AM). The resulting framework aims to further explore directions in which PMORs can enable process parameter control within the AM digital twin paradigm.

A Bregman Learning Framework for Sparse Neural Networks

Tim Roith[1]

[1]Friedrich-Alexander Universität Erlangen-Nürnberg

I will present a novel learning framework based on stochastic Bregman iterations. It allows to train sparse neural networks with an inverse scale space approach, starting from a very sparse network and gradually adding significant parameters. Apart from a baseline algorithm called LinBreg, I will also speak about an accelerated version using momentum, and AdaBreg, which is a Bregmanized generalization of the Adam algorithm. I will present a statistically profound sparse parameter initialization strategy, stochastic convergence analysis of the loss decay, and additional convergence proofs in the convex regime. The Bregman learning framework can also be applied to Neural Architecture Search and can, for instance, unveil autoencoder architectures for denoising or deblurring tasks.

A piecewise conservative method for unconstrained convex optimization

Piero Colli Franzone[1], **Alessandro Scagliotti**[2]

[1]University of Pavia, [2]Scuola Internazionale Superiore di Studi Avanzati

In the last years Theory of Dynamical Systems has been fruitfully applied to study existing accelerated optimization methods and to develop new ones. Typically, continuous-time models for accelerated methods are mechanical systems with damping. In this talk we present an optimization method based on a conservative mechanical system, where the objective function plays the role of the potential energy. Due to the absence of damping, the convergence of this method completely relies on the restart strategy: *) the initial velocity is set equal to 0; *) by the conservation of the mechanical energy, part of the initial potential energy is transformed into kinetic energy; *) when a proper restart condition is met, the velocity is reset to zero and the kinetic energy at the restart time is instantly dissipated. We prove the convergence result both for the continuous-time method and for the discrete-time version. Finally, we discuss some possible extensions to the nonsmooth case, with particular focus on l1 regularization.

Constrained Optimization on Riemannian manifolds

Suvrit Sra[1], **Melanie Weber**[2]

[1]MIT, [2]University of Oxford

Many applications involve non-Euclidean data, where exploiting Riemannian geometry can deliver algorithms that are computationally superior to standard nonlinear programming approaches. This observation has resulted in an increasing interest in Riemannian methods in the optimization and machine learning community. In this talk, we consider the problem of optimizing a function on a Riemannian manifold subject to convex constraints. We will discuss different instances of this problem on matrix manifolds and present algorithms for computing its solution efficiently. Specifically, we introduce Riemannian Frank-Wolfe (RFW) methods, a class of projection-free algorithms for constrained geodesically convex optimization and give guarantees for its efficiency. We complement our theoretical analysis with an empirical comparison of RFW against state-of-the-art Riemannian optimization methods.

Geometry Segmentation with Total Variation Regularization

Manuel Weiß[1]

[1]Interdisciplinary Center for Scientific Computing (Heidelberg)

The total variation has proven as a useful regularizer for various applications in inverse imaging and shape optimization problems. For the task of shape segmentation, we propose a model that combines normal vector data of a discrete surface and a total variation penalty and present suitable solvers.

Decision Boundaries and Convex Hulls in the Feature Space that Deep Learning Functions Learn from Images

Roozbeh Yousefzadeh[1]

[1]Yale University

Success of deep neural networks in image classification and learning can partly be attributed to the features they extract from images. It is often speculated about the properties of a low-dimensional manifold that models extract and learn from images, however, there is not adequate understanding about that low-dimensional space based on theory or empirical evidence. As a result, the narratives and speculations about that feature space have largely remained unverified. In this work, we formulate and solve a set of optimization problems involving deep learning models to study properties of the feature space in the last hidden layer of trained models. First we develop methods to investigate the decision boundaries of the models in the feature space with respect to training/testing samples and with respect to the convex hull of training set. We then solve a set of inverse optimization problems relating the feature space back to the pixel space. We report that functional task of deep networks involves extrapolation even in a 64-dimensional feature space that ResNet models learn from images. We also report that geometric arrangements of decision boundaries and convex hulls in that feature space significantly differ from the pixel space, providing insights and novel methods about adversarial vulnerabilities, image morphing, extrapolation, ambiguity in classification, and the mathematical understanding of image classification models. Our optimization problems are non-convex and highly non-linear as they involve deep learning functions. Moreover, the issue of vanishing and exploding gradients arise which we overcome using a homotopy algorithm.