Multi-Objective Optimization Problems With Well-Conditioned Solutions

Harshal Bharat Dupare
18MA20015

Under the guidance of
Dr. Swanand Ravindra Khare & Dr. Geetanjali Panda

Department of Mathematics
Indian Institute of Technology Kharagpur

November, 2022

Agenda

Preliminaries

Aggregation Scheme: Given a value function $value(\cdot)$ over x, for $x \in X$, a method to aggregate those values to one value. Denoted as follows:

$$\underset{\forall x \in X}{Agg}\ value(x) \tag{1.1}$$

e.g. Expectation over given probability distribution $p(\cdot)$ over X.

$$\underset{\forall x \in X}{Agg}\ value(x) = \int_{\forall x \in X} value(x)p(x)\, dx \tag{1.2}$$

e.g. Exponential decaying average, given a index-function $t : X \to [|X|]$, where $[n] = \{0, 1, 2, ..., n-1\}$.

$$\underset{\forall x \in X}{Agg}\ value(x) = \sum_{\forall x \in X} \lambda^{t(x)} value(x) \tag{1.3}$$

Absolute Condition Number: *Absolute Condition Number* is defined for a function $f : X \to Y$, where X and Y have a norm $||.||$ defined over them. Denoted as follows:

$$cond_{abs}(f, x) = \lim_{\epsilon \to 0} \sup_{||\delta x|| \leq \epsilon} \frac{||\delta f(x)||}{||\delta x||} \tag{1.4}$$

Relative Condition Number: *Relative Condition Number* is defined for a function $f : X \to Y$, where X and Y have a norm $||.||$ defined over them. Denoted as follows:

$$cond_{rel}(f, x) = \lim_{\epsilon \to 0} \sup_{||\delta x|| \leq \epsilon} \frac{||\delta f(x)||/||f(x)||}{||\delta x||/||x||} \tag{1.5}$$

Induced Norm: Given a Matrix Transform $T^{m \times k} : \mathbb{R}^{n \times m} \to \mathbb{R}^{n \times k}$ and a norm $|| \cdot ||$ defined over $\mathbb{R}^{n \times m}$ and $\mathbb{R}^{n \times k}$ we define norm of the matrix transform as

$$||T^{m \times k}|| = \sup_{X \in \mathbb{R}^{n \times m}/\{\bar{0}\}} \frac{||X \times T^{m \times k}||}{||X||} \tag{1.6}$$

Problem Definition

Functional Multi-Objective Optimization Problem:

$$\min_{f\in\mathcal{F}} \quad \underset{\forall x\in X_1, y\in Y_1, y\equiv x}{Agg_1} \quad \mathcal{L}(y, f(x)) \tag{1.7}$$

$$\max_{f\in\mathcal{F}} \quad \underset{\forall x\in X_2, y\in Y_2, y\equiv x}{Agg_2} \quad \mathcal{R}(y, f(x)) \tag{1.8}$$

Including other problems specific objectives, but it must have atleast one of the 1.9, 1.10 as objective

$$\min_{f\in\mathcal{F}} \max_{x\in X_3} cond_{abs}(f, x) \tag{1.9}$$

$$\min_{f\in\mathcal{F}} \max_{x\in X_4} cond_{rel}(f, x) \tag{1.10}$$

or is subject to at least one of the 1.11, 1.12 as constraints

$$\max_{x\in X_5} cond_{abs}(f, x) \leq \alpha \tag{1.11}$$

$$\max_{x\in X_6} cond_{rel}(f, x) \leq \alpha \tag{1.12}$$

$f : X \to Y \approx f^* : X \to Y$

$X_i \subseteq X, i \in [6]$

$Y_i \subseteq Y, X_i \equiv Y_i, i \in [2]$

$\mathcal{F} \equiv \mathcal{F}_\theta$

$\mathcal{R} : Y \times Y \to \mathbb{R}$, Reward function

$\mathcal{L} : Y \times Y \to \mathbb{R}$, Loss function

$\alpha \in \mathbb{R}^+$

Motivation and Applications

Linear Factor Model

$$
\begin{tikzcd}
R_{t\times d}^{n\times d} \arrow[rr, "f"] & & F_t^{n\times k} \arrow[d, dashed, "g"] \\
R_t^{n\times 1} \arrow[rr, equal, "\mathcal{L}"] \arrow[dr, dashed] & & P_t^{n\times 1} \arrow[dl, dashed] \\
& \mathcal{R} &
\end{tikzcd}
\tag{1.13}
$$

Returns data for n time series $R_t^{n\times 1} \in \mathbb{R}^{n\times 1}$, for $t \in [T]$

$R_{t\times d}^{n\times d} = [R_{t-1}^{n\times 1}, R_{t-2}^{n\times 1}, \ldots, R_{t-d}^{n\times 1}] \in \mathbb{R}^{n\times d}$

$P_t^{n\times 1} \approx R_t^{n\times 1}$

loss function $\mathcal{L}(y, \hat{y}) = ||y - \hat{y}||_2$

return function $\mathcal{R}(y, \hat{y}) = \frac{\langle y, \hat{y}\rangle}{||y||_2||\hat{y}||_2}, t \in [T, T+S]$

Lets assume that the functions $f(\cdot)$ and $g(\cdot)$ are linear w.r.t. their argument. In this case it can be written as

$$F_t^{n\times k} = f(R_{t\times d}^{n\times d}) = R_{t\times d}^{n\times d} \times A^{d\times k} \tag{1.14}$$

$$P_t^{n\times 1} = g(F_t^{n\times k}) = F_t^{n\times k} \times \beta^{k\times 1} \tag{1.15}$$

And the orthonormal condition requires

$$A^{d\times k}(A^{d\times k})^\top = I^{d\times d} \tag{1.16}$$

Further if we relax the 1.16 to bounds on condition number by $\alpha \geq 1$

$$\max_{R_{t\times d}^{n\times d}\in\mathbb{R}^{n\times d}} cond_{rel}(f, R_{t\times d}^{n\times d}) = \lim_{\epsilon\to 0} \sup_{||\delta R_{t\times d}^{n\times d}||\leq\epsilon} \frac{||\delta R_{t\times d}^{n\times d} \times A^{d\times k}||}{||\delta R_{t\times d}^{n\times d}||} \frac{||F_t^{n\times k} \times (A^{d\times k})^\top||}{||F_t^{n\times k}||} \leq \alpha \tag{1.17}$$

which implies

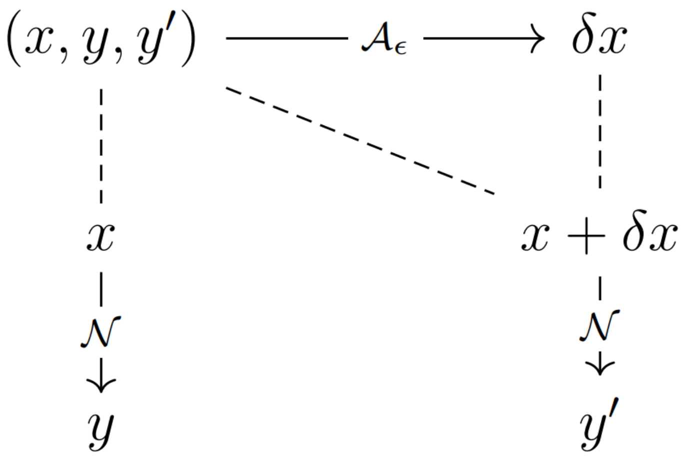
$$\max_{R_{t\times d}^{n\times d}\in\mathbb{R}^{n\times d}} cond_{rel}(f, R_{t\times d}^{n\times d}) \leq ||A^{d\times k}||||(A^{d\times k})^\top|| \leq \alpha \tag{1.18}$$

Note that $\alpha = 1$ contains the set which satisfies 1.16 equation since

$$1 = ||I^{d\times d}|| = ||A^{d\times k}(A^{d\times k})^\top|| \leq ||A^{d\times k}||||(A^{d\times k})^\top|| \leq \alpha \tag{1.19}$$

So any $\alpha \geq 1$ will contain the set specifed by the condition 1.16

Defense Against Adversarial Attacks on Neural Networks



Trained neural network, $\mathcal{N} : X \rightarrow Y$

$\alpha > 0$, $||\mathcal{N}(x + \delta x) - \mathcal{N}(x)|| = ||\delta \mathcal{N}(x)|| \leq \alpha ||\delta x||$

$$\max_{x \in X} cond_{abs}(\mathcal{N}, x) = \lim_{\epsilon \rightarrow 0} \sup_{||\delta x|| \leq \epsilon} \frac{||\delta \mathcal{N}(x)||}{||\delta x||} \leq \alpha$$

$\mathcal{A}_\epsilon$, The adversary with input (x, y, y')

outputs the perturbation δx, such that $||\delta x|| \leq \epsilon$

$$\frac{||\delta \mathcal{N}(x)||}{\alpha} \leq ||\delta x|| \longrightarrow \epsilon < \frac{||\delta \mathcal{N}(x)||}{\alpha}$$

Harshal Bharat Dupare, 18MA20015
Department of Mathematics, Indian Institute of Technology Kharagpur

Related Works

Multi-Objective Optimization

No Preference Methods

No DM is available and a neutral compromise solution is identified without any specification of the preference information.

Utopia Point : A point $F^* \in \mathbb{R}^k$ is objective space, is a utopia point iff for each $i = 1, 2..., k$, $F_i^* = \underset{x \in X}{min}\{F_i(x) | x \in X\}$.

$$\underset{x \in X}{min} \; ||\hat{F}(x) - \hat{F}^*||$$

Vincent, Thomas L / Grantham, Walter Jervis(1981): Optimality in parametric systems(Book)
Zeleny, Milan(1973): Compromise programming
Miettinen, Kaisa (1998): No-Preference Methods. Boston, MA, Springer US: 67–76

A Priori Methods

Based on the preference information given by the DM the optimal solution is found.

Utility Function Methods: Here we have a utility function $U : \mathbb{R}^k \rightarrow \mathbb{R}$, and the goal is to solve the following SOOP

$$\min_{x \in X} U(F(x)) \tag{2.2}$$

Notable methods which come under this utility model are

$$U(F(x)) = \sum_{\forall i \in [k]} w_i F_i(x) \tag{2.3}$$

Known as the Linear scalarization method, if all $\forall i \in [k], w_i > 0$ is a sufficient condition for the solution of 2.1.4 to be a pareto optima [3], but it is not a necessary condition [14].

[3] Zadeh, Lofti(1963): Optimality and non-scalar-valued performance criteria
[14] Zionts, Stanley(1989): Multiple criteria mathematical programming: an updated overview and several approaches

ϵ-Constraint Method: In this method[7] we have a single most important objective function $F_s(x)$. and the remaining objective functions are used to form additional constraints $F_i(x) \leq \epsilon_i, \forall i \in [k]/\{s\}$.

$$\min_{x \in X} F_s(x) \tag{2.6}$$

subject to the constraints

$$F_i(x) \leq \epsilon_i, \ \ \forall i \in [k]/\{s\} \tag{2.7}$$

It is proven that the by a systematic variation of ϵ_i one can generate a set of Pareto optimal solutions[37]. If the solution of 2.6,2.7 exists then it is a weakly Pareto optimal solution [41], and if the solution is unique, then it is Pareto optimal [41].

[7] Haimes, Yacov(1971): On a bicriterion formulation of the problems of integrated system identification and system optimization
[37] Hwang, Ching Lai / Masud, Abu Syed Md (1979): Methods for multiple objective decision making. Multiple objective decision making—methods and applications
[41] Miettinen, Kaisa (2012): Nonlinear multiobjective optimization. , Springer Science & Business Media.

Lexicographic Method: As the name suggests, the Objective functions are ordered as per decreasing order of importance namely $F_i(x)$ is more important than $F_j(x)$ iff $i < j$. Then, the following optimization problems are solved starting from $i = 1, 2, ..., k$.

$$\min_{x \in X} F_i(x) \tag{2.8}$$

subject to

$$F_j(x) \leq F_j(x_j^*), \forall j \in \{1, 2, ...i - 1\} \tag{2.9}$$

In 2.9 we can also have $=$ instead of $\leq$, [43]. Here x_j^* is the solution obtained at the j'th iteration, initially for $i = j = 1$ there are no constrains and $F_i(x)$ is minimized over $x \in X$.

[43] Stadler, Wolfram (1988): Multicriteria Optimization in Engineering and in the Sciences. , Springer Science & Business Media.

A Posteriori Methods

A good representative set of Pareto optimal solutions is found, and from among them the DM must choose the best solution.

Mathematical programming methods

Normal Boundary Intersection Method: Weighted sum method does provide a pareto optimal solution but its is very difficult to find evenly spread solution by varying the weights

$$\min_{x \in X, t} t$$

$$\Phi w + t\mu = F(x) - F^*$$

$$\Phi \in \mathbb{R}^{k \times k}$$

$$\Phi_{ij} = F_j(\underset{x \in X}{argmin}\, F_i(x)) - F_j^*$$

$$\mu = -\Phi e$$

$$e^\top w = \sum_{i \in [k]} w_i = 1$$

Pareto Optimal: A point, $x^* \in X$, is Pareto optimal iff there does not exist another point, $x \in X$, such that $F(x) \leq F(x)$, and $F_i(x) < F_i(x)$ for at least one function.

Weakly Pareto Optimal: A point, $x^* \in X$, is weakly Pareto optimal iff there does not exist another point, $x \in X$, such that $F(x) < F(x)$.

NBI method doesn't provide sufficient nor necessary condition for pareto optimality of the solutions

Das, Indraneel / Dennis, John E(1998): Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems
Das, Indraneel / Dennis, JE (1999): An improved technique for choosing parameters for Pareto surface generation using normal-boundary intersection.
Messac, Achille / Ismail Yahaya, Amir / Mattson, Christopher A(2003): The normalized normal constraint method for generating the Pareto frontier

Evolutionary algorithms

Evolutionary algorithms are subset of the paradigm which is inspired by nature and evolution in designing algorithms for various purposes including solving optimization problems.

Generic EA

function EA($\mathcal{I}$) ▷ gets input parameters $\mathcal{I}$

$\mathcal{P} \leftarrow initialize(\mathcal{I})$ ▷ initialize solution population $\mathcal{P}$

while $converges(\mathcal{P}) \vee terminate(\mathcal{P})$ **do** ▷ till convergence or termination

$\mathcal{P} \leftarrow evolve(\mathcal{P}, \mathcal{I})$ ▷ evolve the population to next generation

return $\mathcal{P}$

Non-dominated Sorting Genetic Algorithm-II (NSGA-II): Elitist principle based partial-order sorting of the population
Ant Colony Optimization (ACO): Ant pheromone to communicate to explore more of the promising regions
Particle Swarm Optimization (PSO): Flocking behaviour of the birds

Vikhar, Pradnya A (2016): Evolutionary algorithms: A critical review and its future prospects. 2016 International conference on global trends in signal processing
Deb, Kalyanmoy / Pratap, Amrit / Agarwal, Sameer / Meyarivan, TAMT(2002): A fast and elitist multiobjective genetic algorithm: NSGA-II
Slowik, Adam / Kwasnicka, Halina(2017): Nature inspired methods and their industry applications—Swarm intelligence algorithms
Poli, Riccardo / Kennedy, James / Blackwell, Tim(2007): Particle swarm optimization

Condition Number

Condition Number of Matrices

Pierre Maréchal and Jane J. Ye in their paper Optimizing Condition Numbers[26] showed that for a symmetric positive semi-definite n × n matrix A minimizing the condition number κ(A).

$$\kappa_2(A) = ||A||_2||A^{-1}||_2 \tag{2.13}$$

$$\min_{A} \kappa_2(A) \tag{2.14}$$

2.14 is euqivalent to minimizing the following objective

$$\min_{A} \lambda_1(A) - \kappa_2(\bar{A})\lambda_n(A) \tag{2.15}$$

[26] Maréchal, Pierre / Ye, Jane J(2009): Optimizing condition numbers, 2: 935–947.
Chen, Xiaojun / Womersley, Robert S / Ye, Jane J(2011): Minimizing the condition number of a Gram matrix, 1: 127–148.

Condition Number of Functions

Edvin Deadman and Samuel D. Relton in their paper [31] extended the taylor's theorem for a complex function of matrices

$$f : \mathbb{C} \rightarrow \mathbb{C} \longmapsto f : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}$$

$$cond_{abs}(f, A) \leq \frac{L_\epsilon}{2\pi\epsilon^2} \max_{z \in \Gamma_\epsilon} |f(z)|$$

$\epsilon > 0$

Γ_ϵ is a closed counter of length L_ϵ

A more extensive treatment and analysis of conditioning of such function is also done in the book by Nicholas J. Higham [36], based on the book a Matlab toolbox has also been made available by the name of Matrix Function Toolbox [34].

[31] Deadman, Edvin / Relton, Samuel D(2016): Taylor's theorem for matrix functions with applications to condition number estimation
[36] Higham, Nicholas J (2008): Functions of matrices: theory and computation. , SIAM.
[34] Higham, Nicholas J(2022): The Matrix Function Toolbox (https://www.mathworks.com/matlabcentral/fileexchange/20820-the-matrixfunction-toolbox).

Analysis

Condition number over composition of function: Say we have a function $f = g(h(x)) = g \circ h(x)$ where $f : X \to Y$, $h : X \to Z$, and $g : Z \to Y$ then we can write

$$cond_{abs}(f,x) = \lim_{\epsilon \to 0} \sup_{||\delta x|| \le \epsilon} \frac{||\delta f(x)||}{||\delta x||} = \lim_{\epsilon \to 0} \sup_{||\delta x|| \le \epsilon} \frac{||\delta g(h(x))||}{||\delta x||} \tag{3.1}$$

$\implies$

$$cond_{abs}(f,x) = \lim_{\epsilon \to 0} \sup_{||\delta x|| \le \epsilon} \frac{||\delta g(h(x))||}{||\delta h(x)||} \frac{||\delta h(x)||}{||\delta x||} \tag{3.2}$$

under the assumption that $\forall \epsilon \ge 0$ and $||\delta x|| \le \epsilon \; \exists \, \delta \ge 0$ such that $||\delta h(x)|| \le \delta \; \forall x$ and $\lim_{\epsilon \to 0} \delta = 0$. Then we can write

$$cond_{abs}(f,x) = \lim_{\delta \to 0} \sup_{||\delta h(x)|| \le \delta} \frac{||\delta g(h(x))||}{||\delta h(x)||} \lim_{\epsilon \to 0} \sup_{||\delta x|| \le \epsilon} \frac{||\delta h(x)||}{||\delta x||} \tag{3.3}$$

let $h(x) = z \implies \delta h(x) = \delta z$, and since z is a function of x, δz is not independent of δx hence we can write

$$cond_{abs}(f,x) \le \lim_{\delta \to 0} \sup_{||\delta z|| \le \delta} \frac{||\delta g(z)||}{||\delta z||} \lim_{\epsilon \to 0} \sup_{||\delta x|| \le \epsilon} \frac{||\delta h(x)||}{||\delta x||} \tag{3.4}$$

$\implies$

$$cond_{abs}(f,x) \le cond_{abs}(g,h(x)) \times cond_{abs}(h,x) \tag{3.5}$$

the above 3.5 bound can be used to simplify the fMOOP with constraints on condition number for complex functions which are composition of simpler function whose condition number can be bound analytically, for eg. Neural Networks come under such functions.

Relation between absolute and relative condition number: given a function $f : X \to Y$ its absolute and relative condition number are given by

$$cond_{abs}(f, x) = \lim_{\epsilon \to 0} \sup_{||\delta x|| \leq \epsilon} \frac{||\delta f(x)||}{||\delta x||} \tag{3.6}$$

$$cond_{rel}(f, x) = \lim_{\epsilon \to 0} \sup_{||\delta x|| \leq \epsilon} \frac{||\delta f(x)||/||f(x)||}{||\delta x||/||x||} \tag{3.7}$$

we can write

$$cond_{rel}(f, x) = \lim_{\epsilon \to 0} \sup_{||\delta x|| \leq \epsilon} \frac{||\delta f(x)||}{||\delta x||} \frac{||x||}{||f(x)||} \tag{3.8}$$

$\Longrightarrow$

$$cond_{rel}(f, x) = cond_{abs}(f, x) \frac{||x||}{||f(x)||} \tag{3.9}$$

L_2 **regularization and absolute condition number**: Consider the problem of fitting data with n data points $(X_i, Y_i), i \in [n]$, each column being 1 data point $(X, Y) \in (\mathbb{R}^{d\times n}, \mathbb{R}^{1\times n})$ using a function $f : \mathbb{R}^{d\times 1} \rightarrow \mathbb{R}^{1\times 1}$, for case of simplicity assume f is linear transform i.e. $f(X) = AX \approx Y$ where $A \in \mathbb{R}^{1\times d}$.
Consider the L_2 regularization formulations of this problem as follows

$$\underset{A\in\mathbb{R}^{1\times d}}{min} \frac{1}{n}\sum_{i\in[n]}||AX_i - Y_i||_2 + \lambda||A||_2^2 \tag{3.10}$$

Now, consider the same problem under fMOOP for minimizing the $||\cdot||_F$ frobenius norm of the transform A over the aggregation scheme of mean and $X_3 = \mathbb{R}^{d\times 1}$

$$\underset{A\in\mathbb{R}^{1\times d}}{min} \frac{1}{n}\sum_{i\in[n]}||AX_i - Y_i||_2 \tag{3.11}$$

$$\underset{A\in\mathbb{R}^{1\times d}}{min} \ \underset{x\in X_3}{max} \ cond_{abs}(f, x) \tag{3.12}$$

by definition

$$cond_{abs}(f, x) = \underset{\epsilon\rightarrow 0}{lim} \ \underset{||\delta x||_F\leq\epsilon}{sup} \frac{||A\delta x||_F}{||\delta x||_F} \tag{3.13}$$

note that $||A\delta x||_F = \sqrt{(\sum_{i\in[d]} A_i \delta x_i)^2}$ and by Cauchy–Schwarz inequality we can write $(\sum_{i\in[d]} A_i\delta x_i)^2 \leq (\sum_{i\in[d]} A_i^2)(\sum_{i\in[d]} x_i^2) = ||A||_F^2||\delta x||_F^2$, which implies

$$cond_{abs}(f,x) = \lim_{\epsilon\to 0} \sup_{||\delta x||_F\leq\epsilon} \frac{||A\delta x||_F}{||\delta x||_F} \leq \frac{||A||_F||\delta x||_F}{||\delta x||_F} = ||A||_F \tag{3.14}$$

note that for this case $||\cdot||_F$ is equivalent to $||\cdot||_2$, and the above upper bound implies that the problem 3.12 when minimized for the worst case using the upper bound 3.14 we get the fMOOP as follows

$$\min_{A\in\mathbb{R}^{1\times d}} (||A||_2, \frac{1}{n}\sum_{i\in[n]} ||AX_i - Y_i||_2) \tag{3.15}$$

which when scalarized with squaring the norm restriction gives us the standard L_2 regularization.

Questions

What types of fMOOPs confirm to restrictions on condition number of f?

How to determine minimum feasible value of α for a given problem and subsets $X_1, X_2, X_5, X_6 \subseteq X$?

For the fMOOPs which confirm such condition number restrictions how does the condition number behave over the domain X_5, X_6?

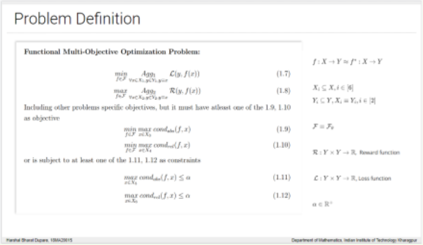
For the fMOOPs which confirm such condition number restrictions how does the condition number behave over the domain X_5, X_6?

What all computation methods/algorithms are available for condition number w.r.t. specific fMOOP?

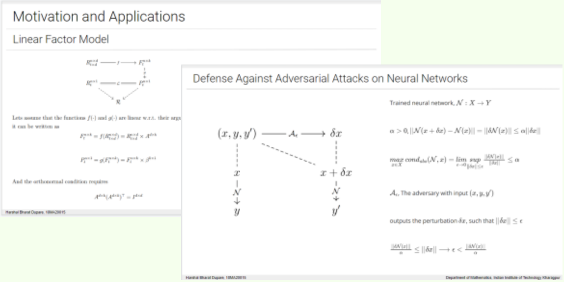
How the transformed problems with bounds on condition number behave and how close are their solutions to the original problems solutions?

Conclusion and Future Work

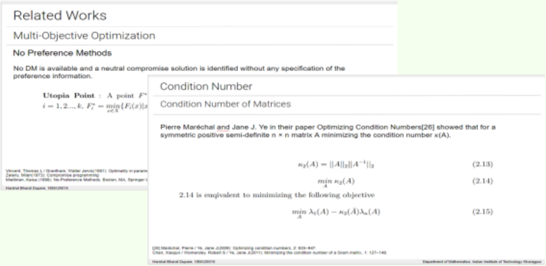
✓ Problem Definition



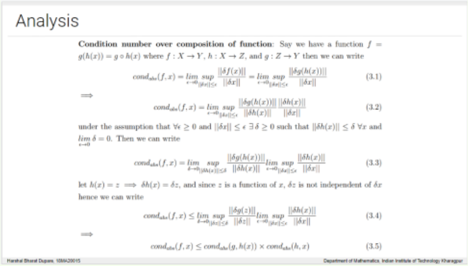
✓ Motivations & Applications



✓ Relevant Literature



✓ Preliminary Analysis



Analysis & Address Questions

Experimentations & Results

Thank You

Open For Questions