

Multi-Objective Optimization Problems with Well-Conditioned Solutions

Harshal Bharat Dupare
18MA20015

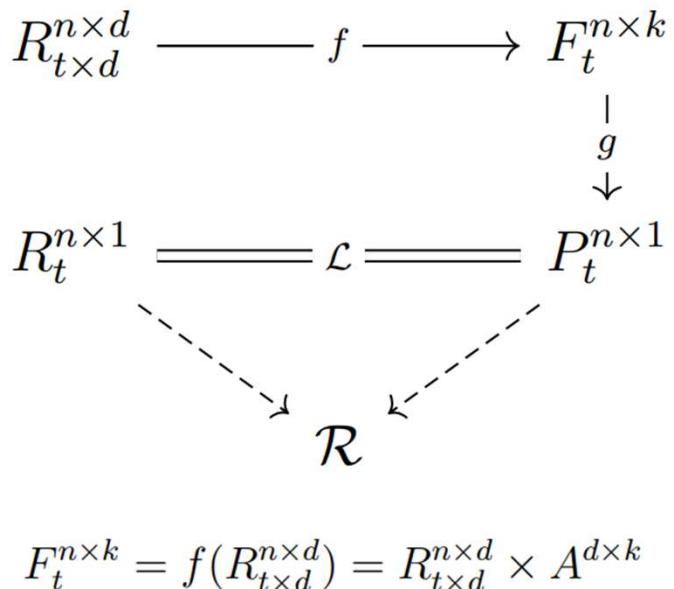
Supervised by
Dr. Geetanjali Panda
Department of Mathematics
Indian Institute of Technology, Kharagpur

Table of Contents

- Motivation and Definition
- Literature Review
- Addressing cond-fMOOP
- Observations and Methodology
- Experimental Results
- Conclusion and Future Work

Motivating Problem

Linear Factor Model



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

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

$$\mathcal{R}_{\beta^{k \times 1}, A^{d \times k}} = \frac{1}{S} \sum_{\forall t \in [T, T+S]} \frac{\langle R_t^{n \times 1}, R_{t \times d}^{n \times d} \times A^{d \times k} \times \beta^{k \times 1} \rangle}{\|R_t^{n \times 1}\|_2 \|R_{t \times d}^{n \times d} \times A^{d \times k} \times \beta^{k \times 1}\|_2}$$

$$\mathcal{L}_{\beta^{k \times 1}, A^{d \times k}} = \frac{1}{2(T-d+1)} \sum_{\forall t \in [d, T]} \|R_t^{n \times 1} - R_{t \times d}^{n \times d} \times A^{d \times k} \times \beta^{k \times 1}\|_2^2$$

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

Definitions

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

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

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

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

Induced Norm: Given a Matrix Transform $T^{m \times k} : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times k}$ 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

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

Generalization

Conditioned Functional Multi-Objective Optimization Problem:

$$\min_{f \in \mathcal{F}} \underset{\forall x \in X_1, y \in Y_1, y \equiv x}{\text{Agg}_1} \mathcal{L}(y, f(x)) \quad (1.7)$$

$$\max_{f \in \mathcal{F}} \underset{\forall x \in X_2, y \in Y_2, y \equiv x}{\text{Agg}_2} \mathcal{R}(y, f(x)) \quad (1.8)$$

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

$$\min_{f \in \mathcal{F}} \max_{x \in X_3} \text{cond}_{abs}(f, x) \quad (1.9)$$

$$\min_{f \in \mathcal{F}} \max_{x \in X_4} \text{cond}_{rel}(f, x) \quad (1.10)$$

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

$$\max_{x \in X_5} \text{cond}_{abs}(f, x) \leq \alpha \quad (1.11)$$

$$\max_{x \in X_6} \text{cond}_{rel}(f, x) \leq \alpha \quad (1.12)$$

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

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

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

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

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

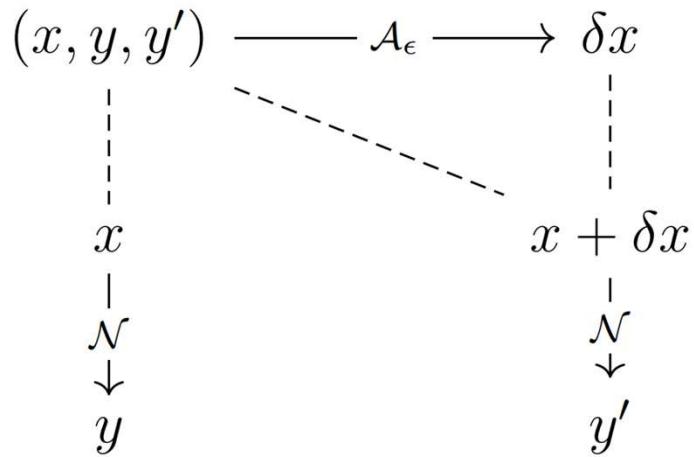
$$\mathcal{L} : Y \times Y \rightarrow \mathbb{R}, \text{ Loss function}$$

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

Other Applications

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

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



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

\mathcal{A}_ϵ , The adversary with input (x, y, y')

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

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

Multi-Objective Optimization Methods

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.

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

No Preference Methods

$$\min_{x \in X} \|\hat{F}(x) - \hat{F}^*\|$$

A Priori Methods

ϵ -Constraint Method

A Posteriori Methods

Evolutionary Algorithms (EA)

Non-dominated Sorting Genetic Algorithm-II (NSGA-II)

Particle Swarm Optimization (PSO)

Findings on Condition Number

$$\kappa_2(A) = \|A\|_2 \|A^{-1}\|_2 \quad (2.4)$$

$$\min_A \kappa_2(A) \quad (2.5)$$

2.5 is equivalent to minimizing the following objective

$$\min_A \lambda_1(A) - \kappa_2(\bar{A})\lambda_n(A) \quad (2.6)$$

$\|\cdot\|_F$ is Frobenius norm and $\sigma_i, i \in [n]$ are the singular value of A in decreasing order of their magnitude.

$$\|A\|_F = \sqrt{\sum_{i \in [n]} \sum_{j \in [m]} A_{ij}^2} \quad (2.7)$$

another interesting relation between $\kappa_2(A)$ and that of $\kappa_F(A)$ i.e. the condition number induced by Frobenius norm [8][14].

$$\frac{\kappa_F(A)}{n} \leq \kappa_2(A) \leq \kappa_F(A) \quad (2.8)$$

Pierre Maréchal and Jane J. Ye in their paper Optimizing Condition Numbers. They also proved that the problem of minimizing the condition number is non-smooth and non-convex optimization problem.

Theoretical Results

Condition Number over Composition of Function

$$cond_{abs}(f, x) \leq \lim_{\delta \rightarrow 0} \sup_{||\delta z|| \leq \delta} \frac{||\delta g(z)||}{||\delta z||} \lim_{\epsilon \rightarrow 0} \sup_{||\delta x|| \leq \epsilon} \frac{||\delta h(x)||}{||\delta x||}$$

$$cond_{abs}(f, x) \leq cond_{abs}(g, h(x)) \times cond_{abs}(h, x)$$

L_2 Regularization & Absolute Condition Number

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

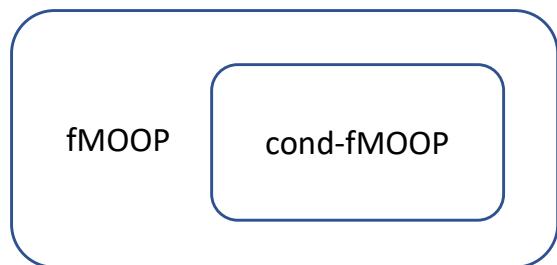
$$\min_{A \in \mathbb{R}^{1 \times d}} \frac{1}{n} \sum_{i \in [n]} ||AX_i - Y_i||_2$$

$$\min_{A \in \mathbb{R}^{1 \times d}} \max_{x \in X_3} cond_{abs}(f, x)$$

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

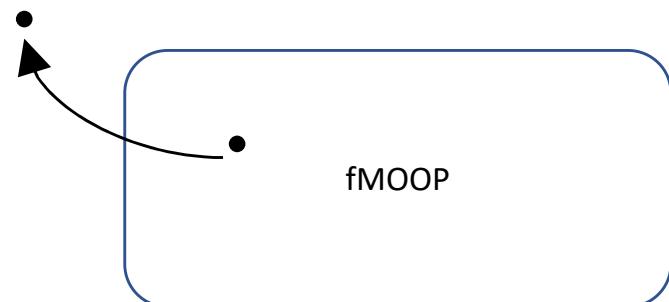
Questions on cond-fMOOP

1. For a solvable $\mathcal{P} \in$ fMOOP and $\mathcal{P}' \in$ cond-fMOOP, derived from \mathcal{P} . Is \mathcal{P}' solvable?
2. For $\mathcal{P} \in$ cond-fMOOP over X and subsets $X_5, X_6 \subseteq X$. How to determine minimum value of α for which \mathcal{P} is solvable?
3. For a $\mathcal{P} \in$ cond-fMOOP, if we transform the condition number restriction using bounds, call the new problem \mathcal{P}' . How close are the solutions of \mathcal{P}' and \mathcal{P} ?



Knowing the difference

It may not be possible to make α arbitrary small



Observations

$$\mathcal{O}_d^k = \{A | AA^\top = I_d, A \in \mathbb{R}^{d \times k}\}$$

$$\delta\mathcal{O}_d^k = \{\delta | \delta = A - B, A, B \in \mathcal{O}_d^k\}$$

Properties of \mathcal{O}_d^k

1. $A \in \mathcal{O}_d^k \implies -A \in \mathcal{O}_d^k$
2. $A \in \mathcal{O}_d^k$ and $Q \in \mathcal{O}_d^k$ we have $AQ \in \mathcal{O}_d^k$
3. $A \in \mathcal{O}_d^k$ and $Q \in \mathcal{O}_d^d$ we have $QA \in \mathcal{O}_d^k$

Properties of $\delta\mathcal{O}_d^k$

1. $\bar{\mathbf{0}} \in \delta\mathcal{O}_d^k$
2. $A, B \in \mathcal{O}_d^k$ we have $A - B, A + B \in \delta\mathcal{O}_d^k$
3. $\delta \in \delta\mathcal{O}_d^k$ and $Q \in \mathcal{O}_d^k$ we have $\delta Q \in \delta\mathcal{O}_d^k$
4. $\delta \in \delta\mathcal{O}_d^k$ and $Q \in \mathcal{O}_d^d$ we have $Q\delta \in \delta\mathcal{O}_d^k$

$$A_{i+1} = A_i + \Delta A_i$$

$$A_i Q_i \approx A_i + \Delta A_i$$

$$Q_i = (I - T_i)(I + T_i)^{-1}$$

$$2T_i = \Delta A_i^\top A_i - A_i^\top \Delta A_i \quad \beta = (A^\top \mathbf{D} A)^{-1} A^\top \mathbf{N}$$

$$\bar{A} = A Q$$

$$\bar{\beta} = (\bar{A}^\top \mathbf{D} \bar{A})^{-1} \bar{A}^\top \mathbf{N}$$

$$\bar{\beta} = (Q^\top A^\top \mathbf{D} A Q)^{-1} Q^\top A^\top \mathbf{N}$$

$$\bar{\beta} = Q^\top (A^\top \mathbf{D} A)^{-1} Q Q^\top A^\top \mathbf{N}$$

$$\bar{\beta} = Q^\top \beta$$

Methodologies

$\mathcal{A}_0[N]$: Baseline Method

```

1:  $\beta_{optimal}, A_{optimal}, \mathcal{R}_{optimal} \leftarrow \emptyset, \emptyset, -\infty$ 
2: for  $i \leftarrow 1 \dots N$  do
3:    $M_i \leftarrow random(\mathbb{R}^{d \times k})$ 
4:    $A_i \leftarrow gram\_schmidt\_algorithm(M_i)$                                  $\triangleright$  orthonormalize  $M_i$  to  $A_i$ 
5:   Compute  $\{F_{t,i}^{n \times k}\}$  for  $A_i$  over  $t \in [d, T]$ 
6:    $\beta_i \leftarrow regression(\{R_t^{n \times 1}\}, \{F_{t,i}^{n \times k}\}, t \in [d, T])$ 
7:   if  $\mathcal{R}_{\beta_i, A_i} > \mathcal{R}_{optimal}$  then           $\triangleright$  update the optimal if better solution found
8:      $\beta_{optimal}, A_{optimal}, \mathcal{R}_{optimal} \leftarrow \beta_i, A_i, \mathcal{R}_{\beta_i, A_i}$ 
9: return  $\beta_{optimal}, A_{optimal}$ 
10: complexity:  $O(N(dk + d^2k + ndS + k^3)) = O(N(k^3 + ndS))$ 
```

$\mathcal{A}_1[N, \mathcal{U}_A, \mathcal{U}_\beta]$: Orthogonal Property Iterative Scheme (OPI Scheme)

```

1:  $A_0 \leftarrow random(\mathbb{R}^{d \times k})$ 
2:  $A_0 \leftarrow gram\_schmidt\_algorithm(A_0)$                                  $\triangleright$  orthonormalize  $A_0$ , as  $A_0 \in \mathcal{O}_d^k$ 
3:  $\beta_0 \leftarrow regression(\{R_t^{n \times 1}\}, \{F_t^{n \times k}[A_0]\}, t \in [d, T])$ 
4: for  $i \leftarrow 0$  to  $N - 1$  do
5:    $\Delta A_i \leftarrow \mathcal{U}_A[\vec{\alpha}](\mathcal{R}_{\beta,A}, \mathcal{L}_{\beta,A}, \beta_i, A_i) - A_i$ 
6:    $2T_i \leftarrow \Delta A_i^\top A_i - A_i^\top \Delta A_i$ 
7:    $Q_i \leftarrow (I - T_i)(I + T_i)^{-1}$ 
8:    $A_{i+1} \leftarrow A_i Q_i$ 
9:    $\beta_{i+1} \leftarrow \mathcal{U}_\beta[\vec{\phi}](\mathcal{R}_{\beta,A}, \mathcal{L}_{\beta,A}, \beta_i, A_i, A_{i+1})$ 
10: return  $\beta_N, A_N$ 
11: complexity:  $O(N(|\mathcal{U}_A[\vec{\alpha}]| + k^2d + k^3 + k^2d + |\mathcal{U}_\beta[\vec{\phi}]|)) = O(N(|\mathcal{U}_A[\vec{\alpha}]| + k^3 + |\mathcal{U}_\beta[\vec{\phi}]|))$ 
```

$\mathcal{A}_2[N, \mathcal{U}_A, \mathcal{U}_\beta, \mathcal{U}_\mathcal{O}]$: Delayed Orthogonalization Iterative Scheme (DOI Scheme)

```

1:  $A_0 \leftarrow random(\mathbb{R}^{d \times k})$ 
2:  $\beta_0 \leftarrow regression(\{R_t^{n \times 1}\}, \{F_t^{n \times k}[A_0]\}, t \in [d, T])$ 
3: for  $i \leftarrow 0$  to  $N - 1$  do
4:    $A_{i+1} \leftarrow \mathcal{U}_A[\vec{\alpha}](\mathcal{R}_{\beta,A}, \mathcal{L}_{\beta,A}, \beta_i, A_i)$ 
5:    $\beta_{i+1} \leftarrow \mathcal{U}_\beta[\vec{\phi}](\mathcal{R}_{\beta,A}, \mathcal{L}_{\beta,A}, \beta_i, A_i, A_{i+1})$ 
6:    $\beta^*, A^* \leftarrow \mathcal{U}_\mathcal{O}[\vec{\lambda}](\mathcal{R}_{\beta,A}, \mathcal{L}_{\beta,A}, \beta_N, A_N)$ 
7: return  $\beta^*, A^*$ 
8: complexity:  $O(N(|\mathcal{U}_A[\vec{\alpha}]| + |\mathcal{U}_\beta[\vec{\phi}]|) + |\mathcal{U}_\mathcal{O}[\vec{\lambda}]|)$ 
```

$\mathcal{U}_\mathcal{O}^1[\vec{\lambda}](\mathcal{R}_{\beta,A}, \mathcal{L}_{\beta,A}, \beta^*, A^*)$: Iterative Closing Method

```

1:  $A \leftarrow random(\mathbb{R}^{d \times k})$ 
2:  $A \leftarrow gram\_schmidt\_algorithm(A)$                                  $\triangleright$  orthonormalize  $A$ 
3:  $\beta \leftarrow regression(\{R_t^{n \times 1}\}, \{F_t^{n \times k}[A]\}, t \in [d, T])$ 
4:  $\beta_{optimal}, A_{optimal}, \mathcal{R}_{optimal} \leftarrow \beta, A, \mathcal{R}_{\beta,A}$ 
5: for  $i \leftarrow 1$  to  $N$  do
6:    $\Delta A \leftarrow A^* - A$ 
7:    $Q \leftarrow ortho\_cayley\_transformation(\frac{\lambda_1}{2} \Delta A)$ 
8:    $A \leftarrow AQ$ 
9:    $\beta_{reg} \leftarrow regression(\{R_t^{n \times 1}\}, \{F_t^{n \times k}[A]\}, t \in [d, T])$ 
10:   $\Delta \beta \leftarrow (1 - \lambda_3) \frac{\partial \mathcal{R}_{\beta,A}}{\partial \beta} + \lambda_3 (\beta_{reg} - beta)$ 
11:   $\beta \leftarrow \beta + \lambda_2 \Delta \beta$ 
12:  if  $\mathcal{R}_{\beta,A} > \mathcal{R}_{optimal}$  then           $\triangleright$  update the optimal if better solution found
13:     $\beta_{optimal}, A_{optimal}, \mathcal{R}_{optimal} \leftarrow \beta, A, \mathcal{R}_{\beta,A}$ 
14: return  $\beta_{optimal}, A_{optimal}$ 
15: complexity:  $O(k^3 + N(k^3 + ndS)) = O(N(k^3 + ndS))$ 
```

Methodologies

NPM₁[$\vec{\mu}$] : No Preference Method 1

```

1:  $A \leftarrow \text{random}(\mathbb{R}^{d \times k})$ 
2:  $A \leftarrow \text{gram\_schmidt\_algorithm}(A)$ 
3:  $\beta \leftarrow \text{regression}(\{R_t^{n \times 1}\}, \{F_t^{n \times k}[A]\}, t \in [d, T])$ 
4:  $\beta_{optimal}, A_{optimal}, \mathcal{R}_{optimal} \leftarrow \beta, A, \mathcal{R}_{\beta, A}$ 
5: for  $i \leftarrow 1$  to  $N$  do
6:    $\Delta A \leftarrow (1 - \mathcal{R}_{\beta, A}) \frac{\partial \mathcal{R}_{\beta, A}}{\partial A} - \mathcal{L}_{\beta, A} \frac{\partial \mathcal{L}_{\beta, A}}{\partial A} - \frac{1}{2} \frac{\partial \|AA^\top - I\|_F^2}{\partial A}$ 
7:    $\Delta \beta \leftarrow (1 - \mathcal{R}_{\beta, A}) \frac{\partial \mathcal{R}_{\beta, A}}{\partial \beta} - \mathcal{L}_{\beta, A} \frac{\partial \mathcal{L}_{\beta, A}}{\partial \beta}$ 
8:    $Q \leftarrow \text{ortho\_cayley\_transformation}(\frac{\mu_1}{2} \Delta A)$ 
9:    $A \leftarrow AQ$ 
10:   $\beta \leftarrow \beta + \mu_2 \Delta \beta$ 
11:  if  $\mathcal{R}_{\beta, A} > \mathcal{R}_{optimal}$  then
12:     $\beta_{optimal}, A_{optimal}, \mathcal{R}_{optimal} \leftarrow \beta, A, \mathcal{R}_{\beta, A}$ 
13: return  $\beta_{optimal}, A_{optimal}$ 
14: complexity:  $O(N(ndS + k^3))$ 

```

NPM₂[$\vec{\mu}$] : No Preference Method 2

```

1:  $A \leftarrow \text{random}(\mathbb{R}^{d \times k})$ 
2:  $A \leftarrow \text{gram\_schmidt\_algorithm}(A)$ 
3:  $\beta \leftarrow \text{regression}(\{R_t^{n \times 1}\}, \{F_t^{n \times k}[A]\}, t \in [d, T])$ 
4:  $\beta_{optimal}, A_{optimal}, \mathcal{R}_{optimal} \leftarrow \beta, A, \mathcal{R}_{\beta, A}$ 
5: for  $i \leftarrow 1$  to  $N$  do
6:    $\Delta A \leftarrow (1 - \mathcal{R}_{\beta, A}) \frac{\partial \mathcal{R}_{\beta, A}}{\partial A} - \mathcal{L}_{\beta, A} \frac{\partial \mathcal{L}_{\beta, A}}{\partial A} - \frac{1}{2} \frac{\partial \|AA^\top - I\|_F^2}{\partial A}$ 
7:    $\Delta \beta \leftarrow (1 - \mathcal{R}_{\beta, A}) \frac{\partial \mathcal{R}_{\beta, A}}{\partial \beta} - \mathcal{L}_{\beta, A} \frac{\partial \mathcal{L}_{\beta, A}}{\partial \beta}$ 
8:    $A \leftarrow A + \mu_1 \Delta A$ 
9:    $\beta \leftarrow \beta + \mu_2 \Delta \beta$ 
10:  if  $\mathcal{R}_{\beta, A} > \mathcal{R}_{optimal}$  then
11:     $\beta_{optimal}, A_{optimal}, \mathcal{R}_{optimal} \leftarrow \beta, A, \mathcal{R}_{\beta, A}$ 
12: return  $\beta_{optimal}, A_{optimal}$ 
13: complexity:  $O(N(ndS + k^2))$ 

```

Methodologies

NSGA-II($N, G, p_c, p_m, \mathcal{C}, \mathcal{M}$)

```
1:  $P \leftarrow \text{random\_candidate\_solutions}(N)$ 
2: Evaluate the fitness of each candidate solution in  $P$             $\triangleright$  fitness : objective function
3: for  $t \leftarrow 1$  to  $G$  do
4:    $C \leftarrow \text{get\_offspring\_population}(P, N, \mathcal{C}, \mathcal{M})$ 
5:    $F_1, F_2, \dots, F_t \leftarrow \text{non\_dominated\_sorting}(P \cup C)$             $\triangleright F_i$  is the  $i$ 'th front
6:    $P, i \leftarrow \emptyset, 1$ 
7:   while  $|P| + |F_i| \leq N$  do
8:     Calculate crowding distance for each solution in  $F_i$ 
9:      $P \leftarrow P \cup F_i$ 
10:     $i \leftarrow i + 1$ 
11:   if  $|P| < N$  then
12:     Sort  $F_i$  in descending order of crowding distance
13:      $P \leftarrow P \cup F_i[1 : N - |P|]$ 
14: return  $P$ 
```

Methodologies

$\text{MOPSO}(N, T_{\max}, \theta, \omega, c_1, c_2, p_m, \epsilon)$ Multi-Objective Particle Swarm Optimization

```

1: Initialize the population of particles randomly:
2:  $\{\vec{x}_i\}, \{\vec{v}_i\} \leftarrow \text{initialize\_position\_and\_velocity}(N)$ 
3:  $\{\vec{f}(\vec{x}_i)\} \leftarrow \text{objective\_function}(\{\vec{x}_i\})$ 
4:  $\{\vec{p}_i\}, \{\vec{f}(\vec{p}_i)\} \leftarrow \{\vec{x}_i\}, \{\vec{f}(\vec{x}_i)\}$   $\triangleright \vec{p}_i, \vec{f}(\vec{p}_i)$ : best positions and objective for each  $\vec{x}_i$ 
5:  $A \leftarrow \text{nondominated}(\{\vec{x}_i\})$   $\triangleright A$  is the archive of non-dominated solutions
6:  $\{\vec{d}(\vec{a}_i)\} \leftarrow \text{crowding\_distance}(A)$   $\triangleright d(\vec{a}_i)$  : average distance to  $\vec{a}_i$ 's neighbors in  $A$ 
7: for  $t \leftarrow 1$  to  $T_{\max}$  do
8:   for  $i \leftarrow 1$  to  $N$  do
9:      $\vec{p}_g \leftarrow \text{get\_global\_best}[\theta](A)$   $\triangleright$  global best w.p.  $\theta$  else random point from  $A$ 
10:     $v_i \leftarrow \text{update\_velocity}[\omega, c_1, c_2](\vec{x}_i, \vec{v}_i, \vec{p}_i, \vec{p}_g)$ 
11:     $\vec{x}_i \leftarrow \text{update\_position}[\epsilon](\vec{x}_i, \vec{v}_i)$ 
12:    if True with probability  $p_m$  then
13:       $\vec{x}_i, \vec{v}_i \leftarrow \text{mutate}(\vec{x}_i, \vec{v}_i)$   $\triangleright$  Mutate the solutions particle  $\vec{x}_i, \vec{v}_i$ 
14:       $\vec{f}(\vec{x}_i) \leftarrow \text{objective\_function}(\vec{x}_i)$   $\triangleright$  Update the objective of particle  $\vec{x}_i$ 
15:       $A \leftarrow \text{nondominated\_merge}(A, \{\vec{x}_i\})$   $\triangleright$  Update the nondominated solutions
16:       $\{\vec{d}(\vec{a}_i)\} \leftarrow \text{crowding\_distance}[\text{True}](A)$ 
17:       $\{\vec{d}(\vec{x}_i)\} \leftarrow \text{crowding\_distance}[\text{False}](\{\vec{x}_i\})$ 
18:       $\{\vec{p}_i\}, \{\vec{f}(\vec{p}_i)\} \leftarrow \text{update\_personal\_best}(\{\vec{p}_i\}, \{\vec{f}(\vec{p}_i)\}, \{\vec{x}_i\}, \{\vec{f}(\vec{x}_i)\}, \{\vec{d}(\vec{x}_i)\})$ 
19: return  $A$ 
```

$\text{update_position}[\epsilon](\vec{x}_i, \vec{v}_i)$

- 1: $(\beta_i, A_i) \leftarrow x_i$
- 2: $(\Delta\beta_i, \Delta A_i) \leftarrow v_i$
- 3: $Q_i \leftarrow \text{ortho_cayley_transformation}(\frac{\epsilon}{2}\Delta A_i)$
- 4: **return** $(\beta + \epsilon\Delta\beta_i, A_i Q_i)$

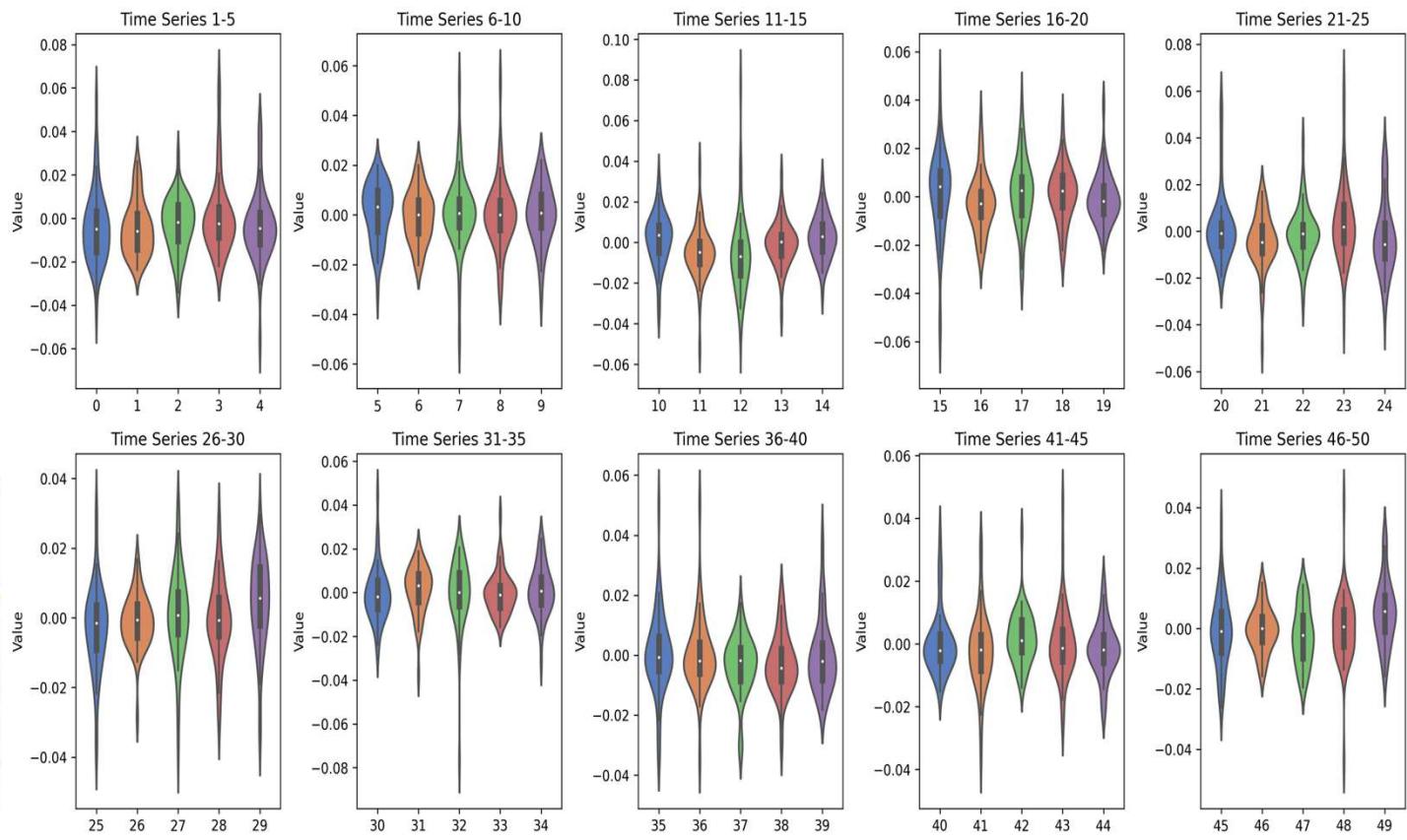
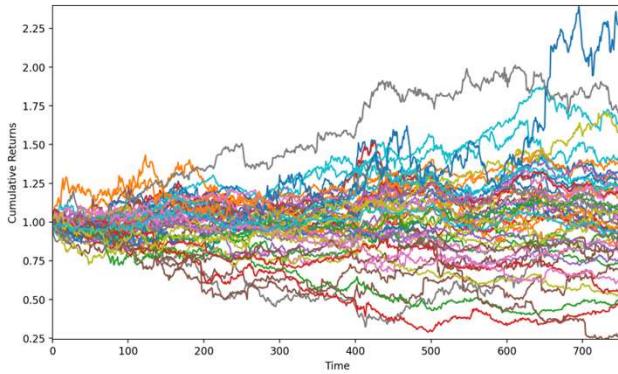
$\text{update_velocity}[\omega, c_1, c_2](\vec{x}_i, \vec{v}_i, \vec{p}_i, \vec{p}_g)$

- 1: $r_1, r_2 \sim U(0, 1)$
- 2: $v_i \leftarrow \omega v_i + c_1 r_1 (\vec{p}_i - \vec{x}_i) + c_2 r_2 (\vec{p}_g - \vec{x}_i)$
- 3: **return** v_i

Data

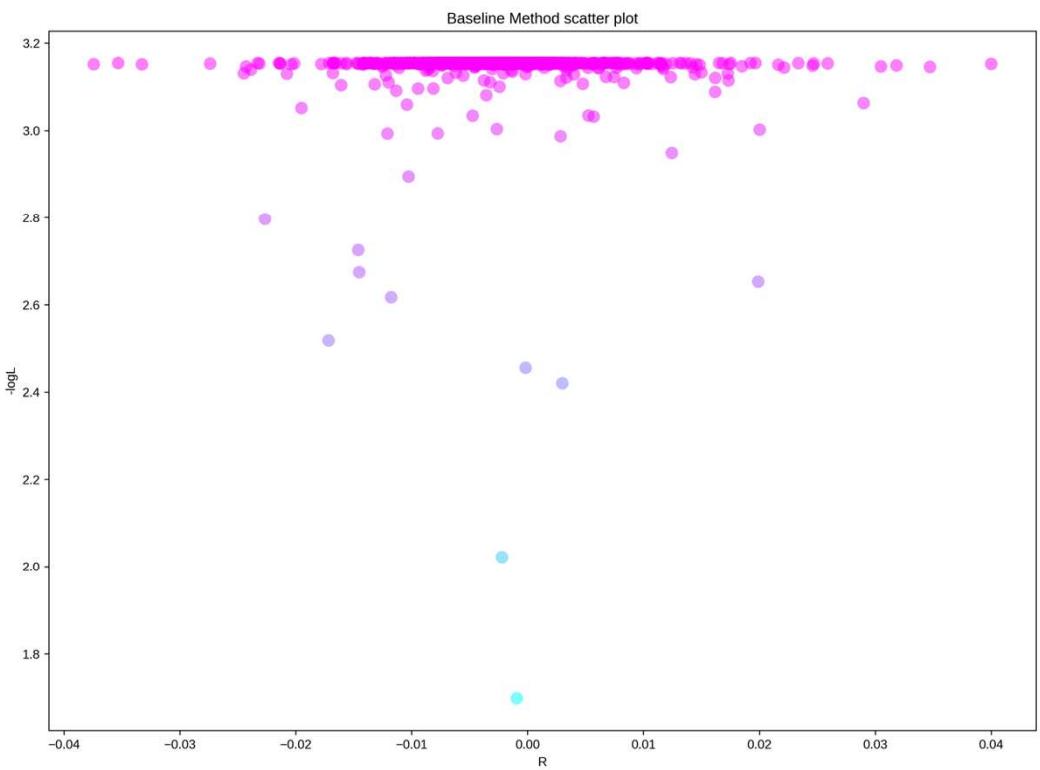
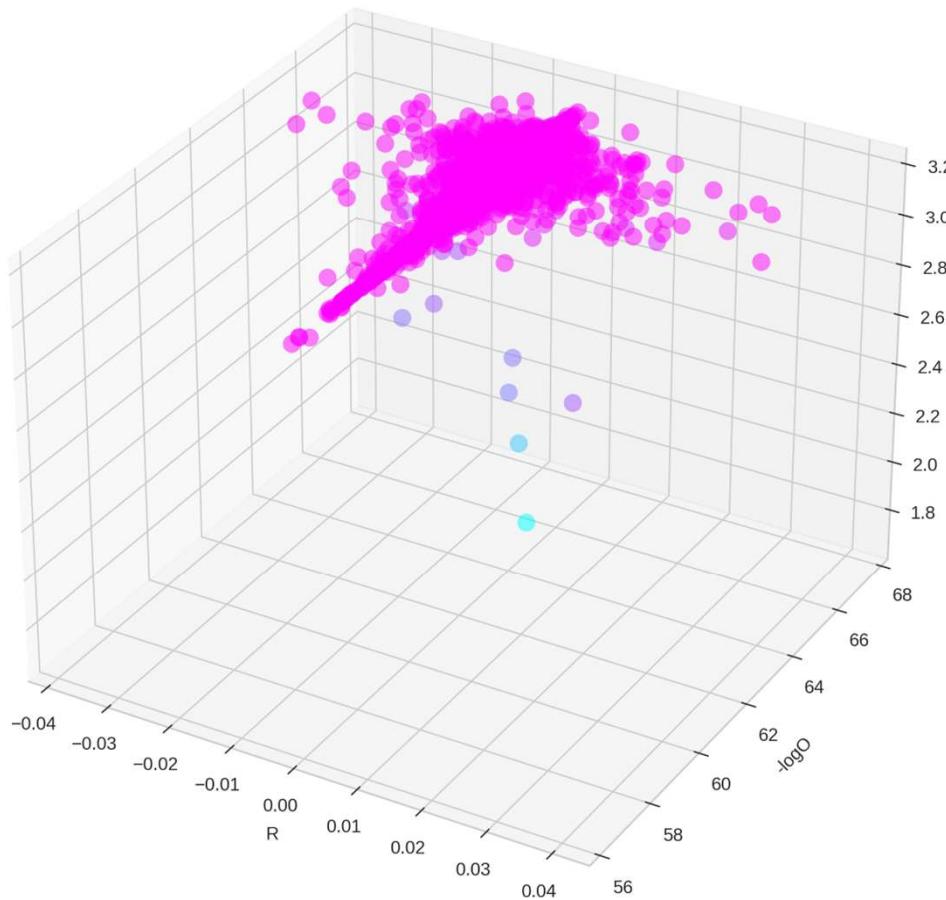
Time series returns data
For 50 time series
For 755 days

Data Source: Qube Research & Technologies (QRT) on the Challenge Data site.

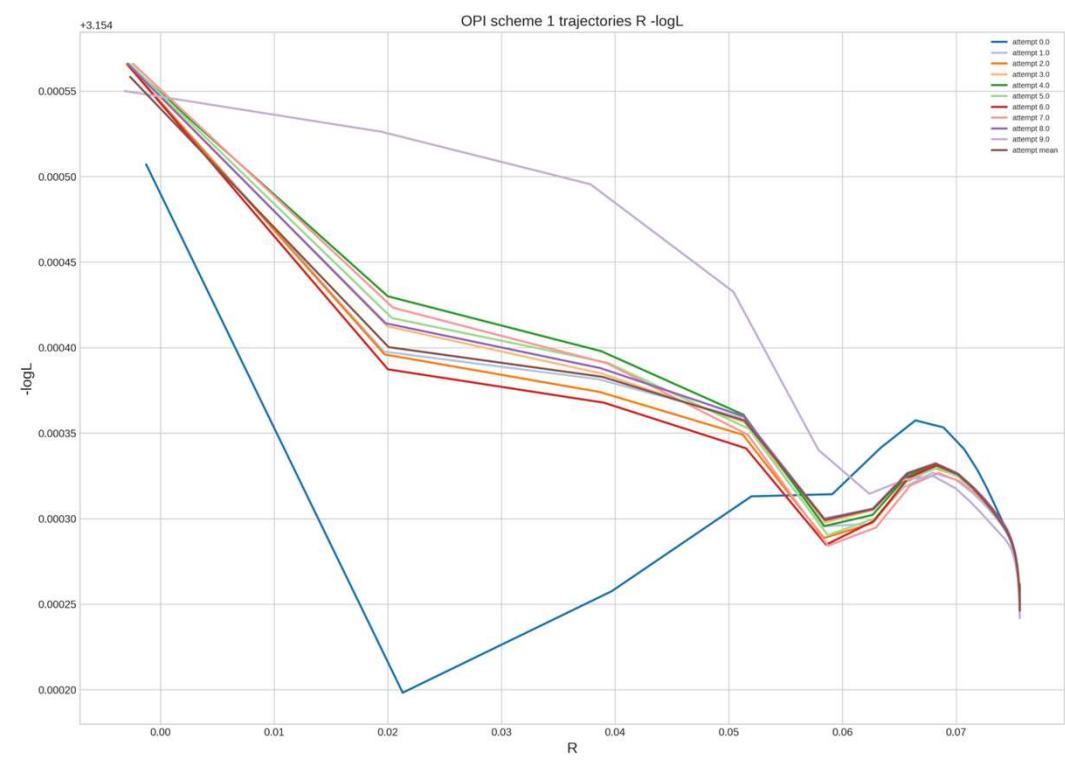
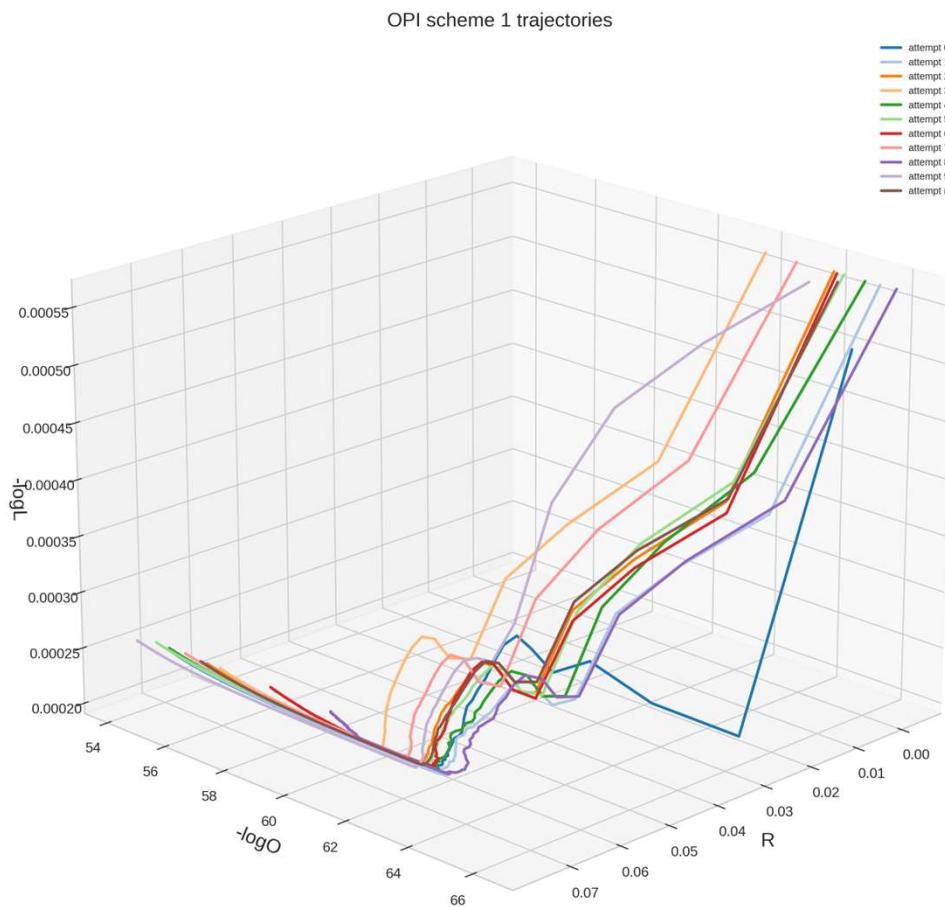


Results Visualization and Interpretation

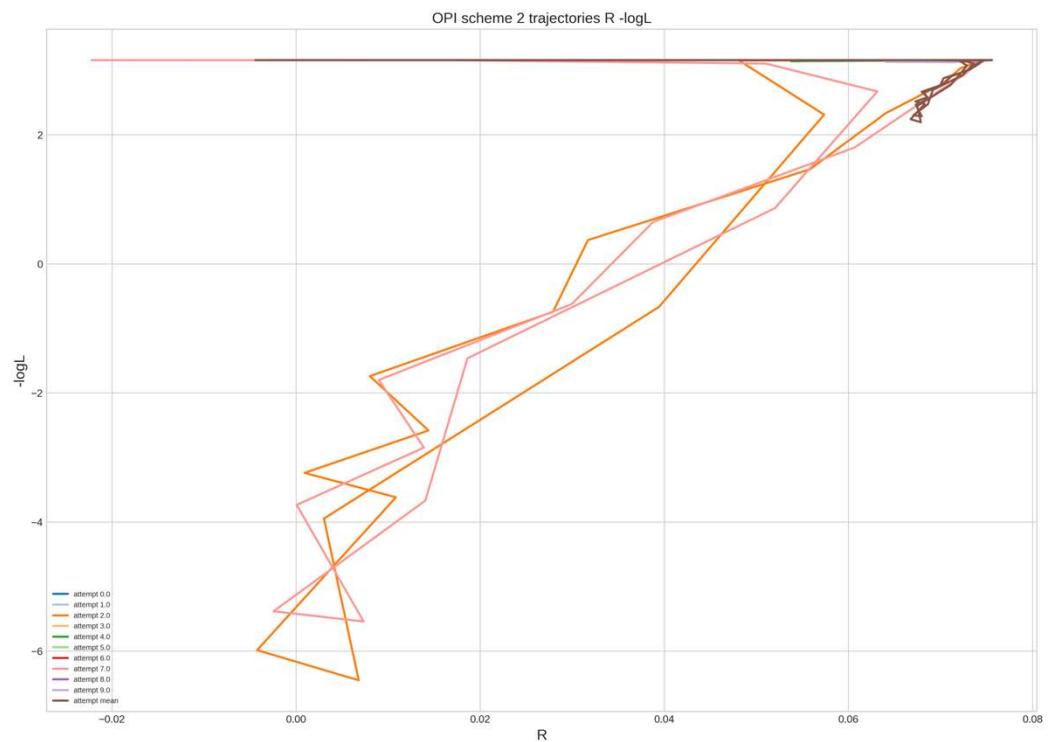
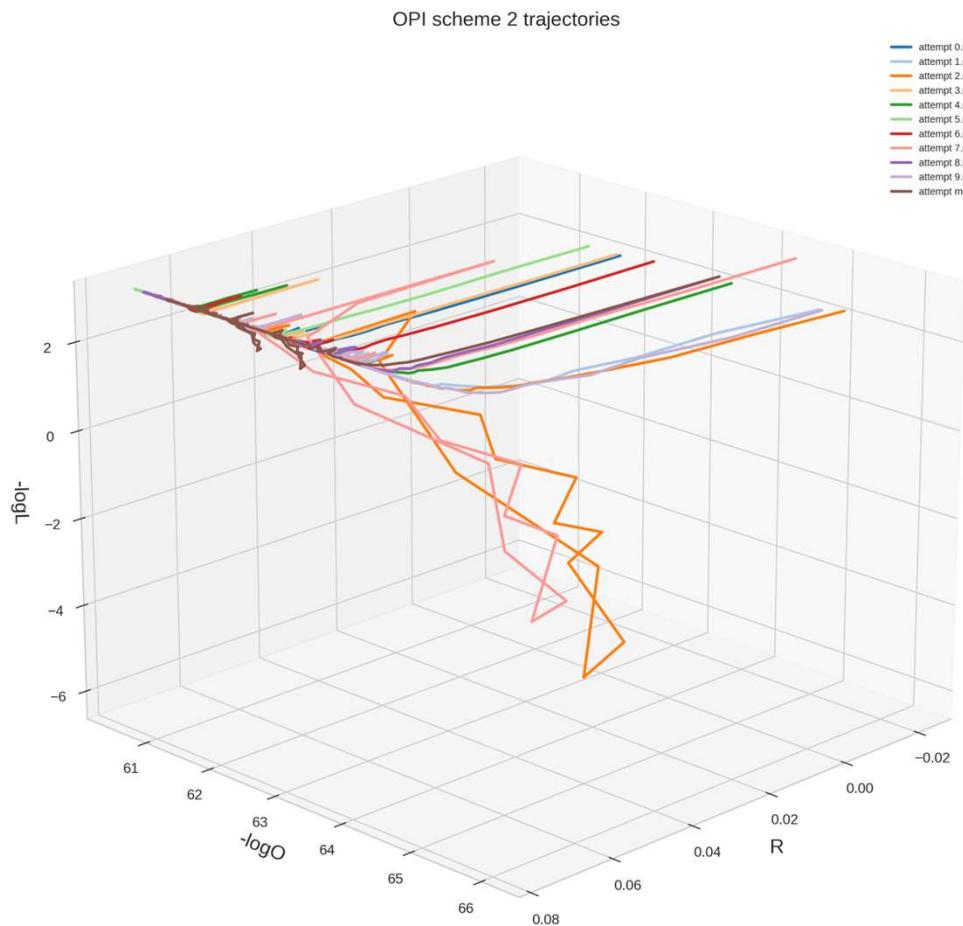
Baseline Method 3D scatter plot



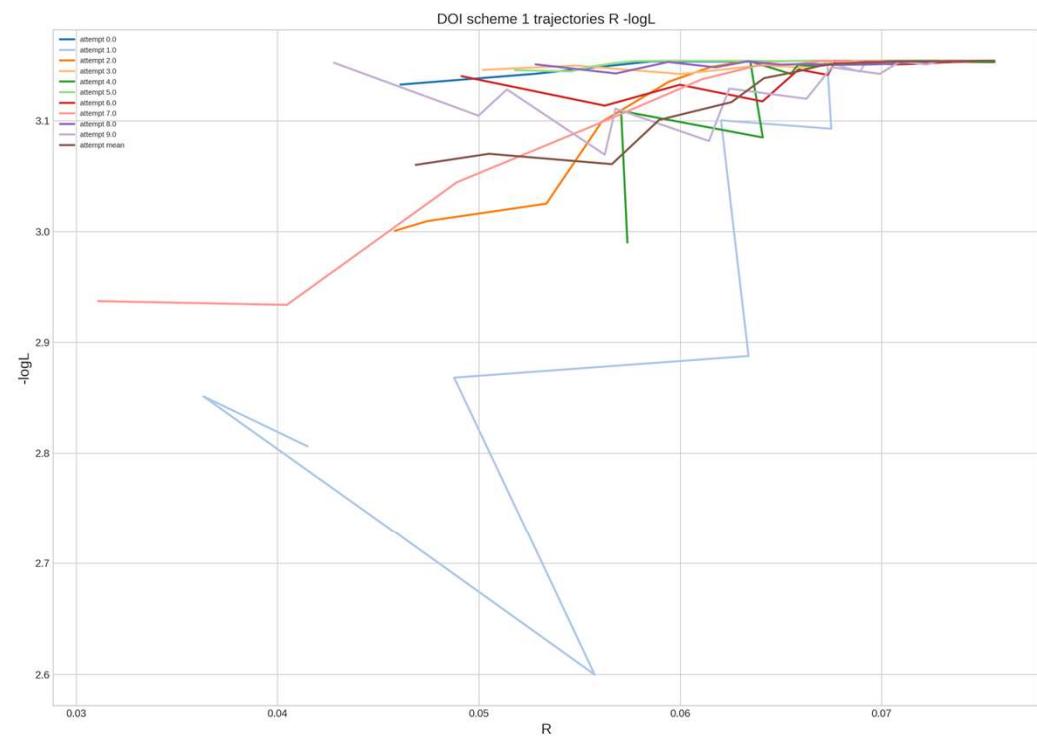
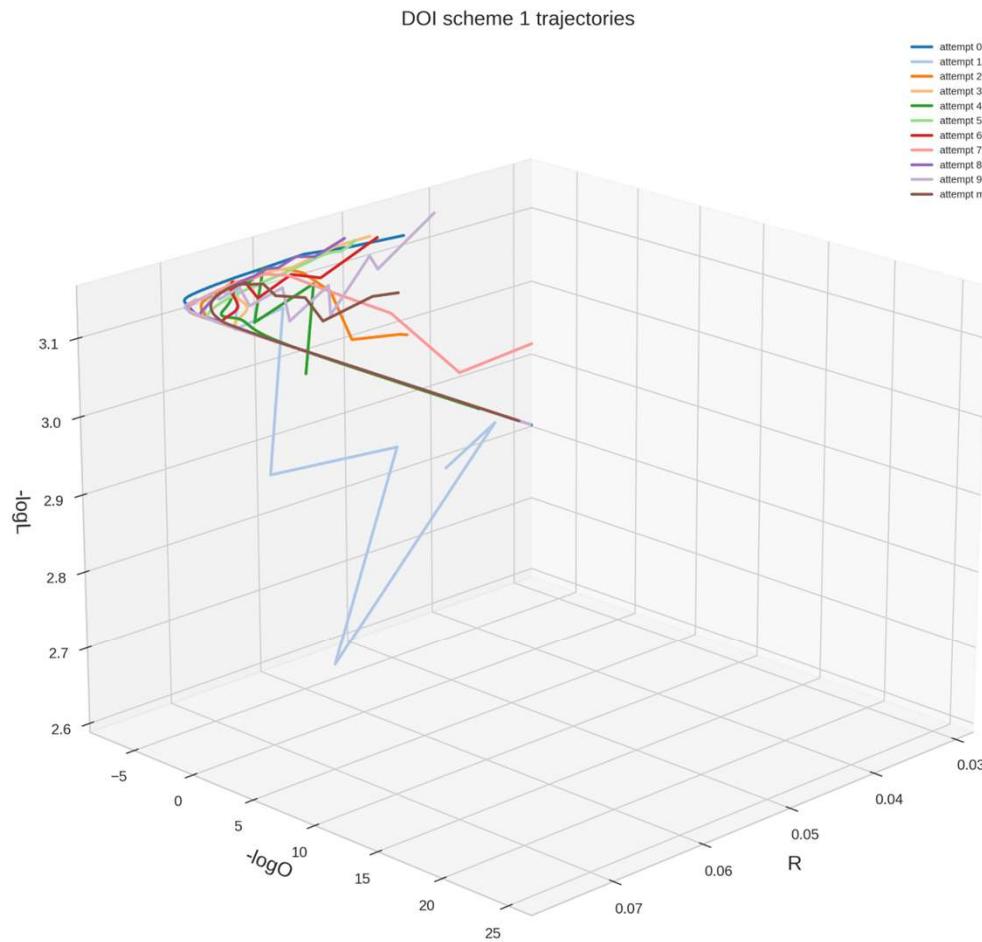
Results Visualization and Interpretation



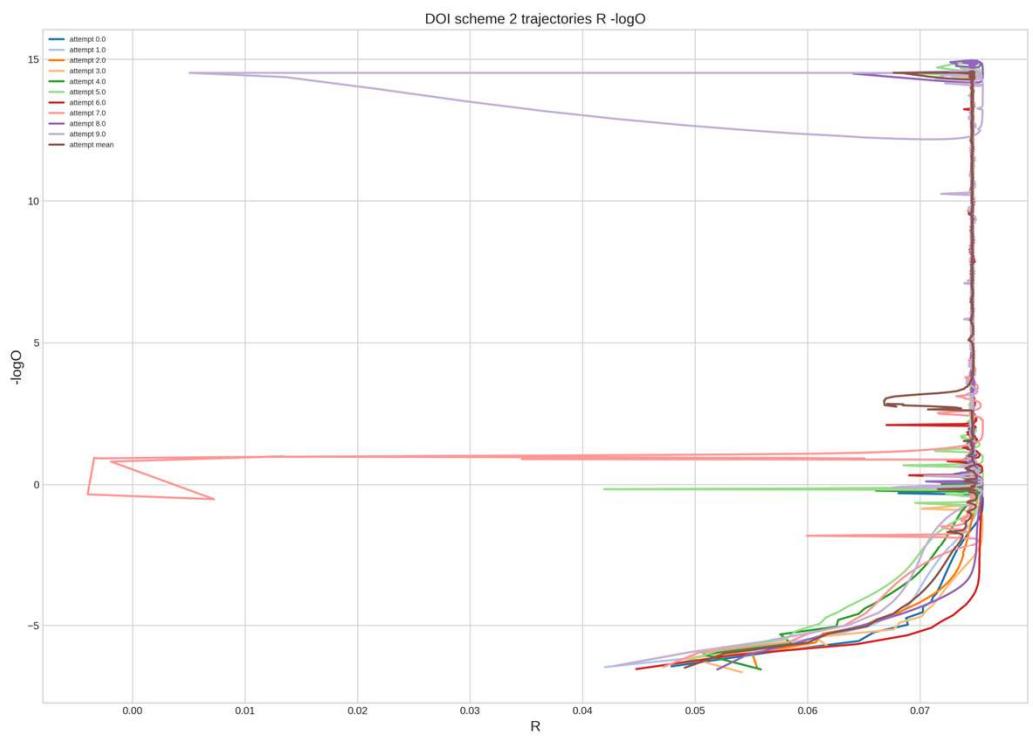
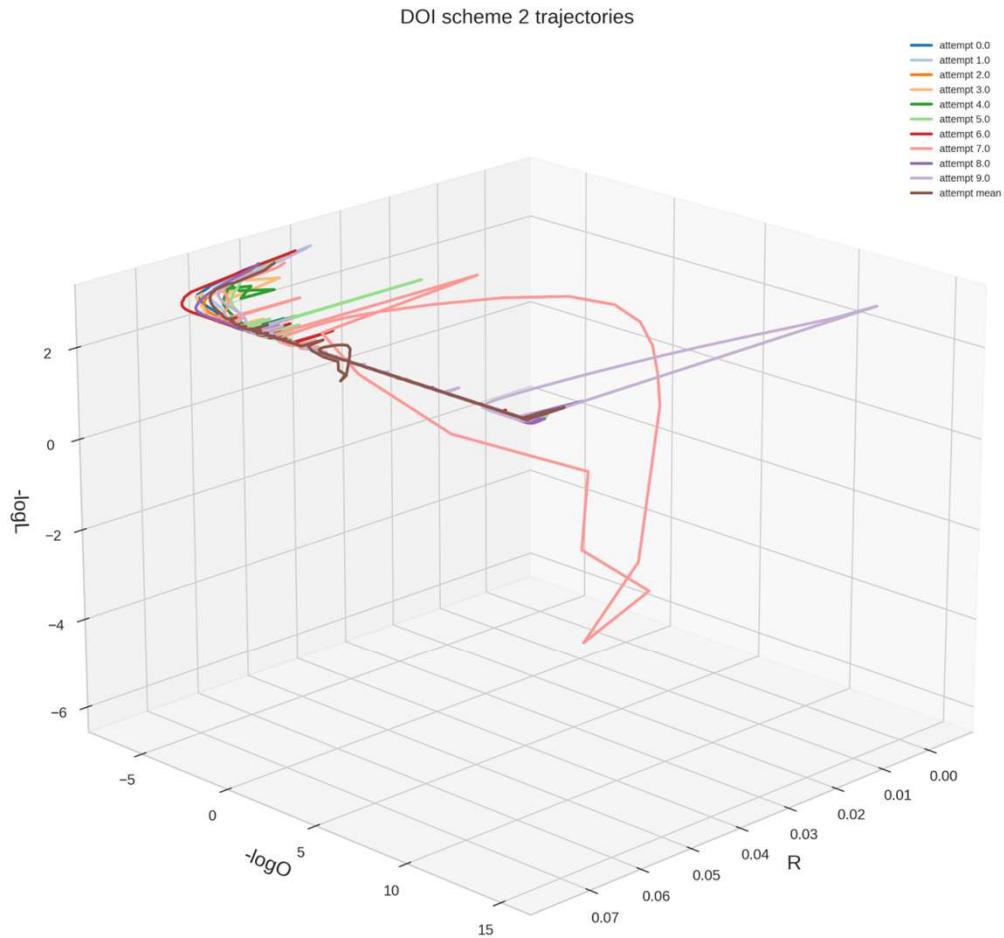
Results Visualization and Interpretation



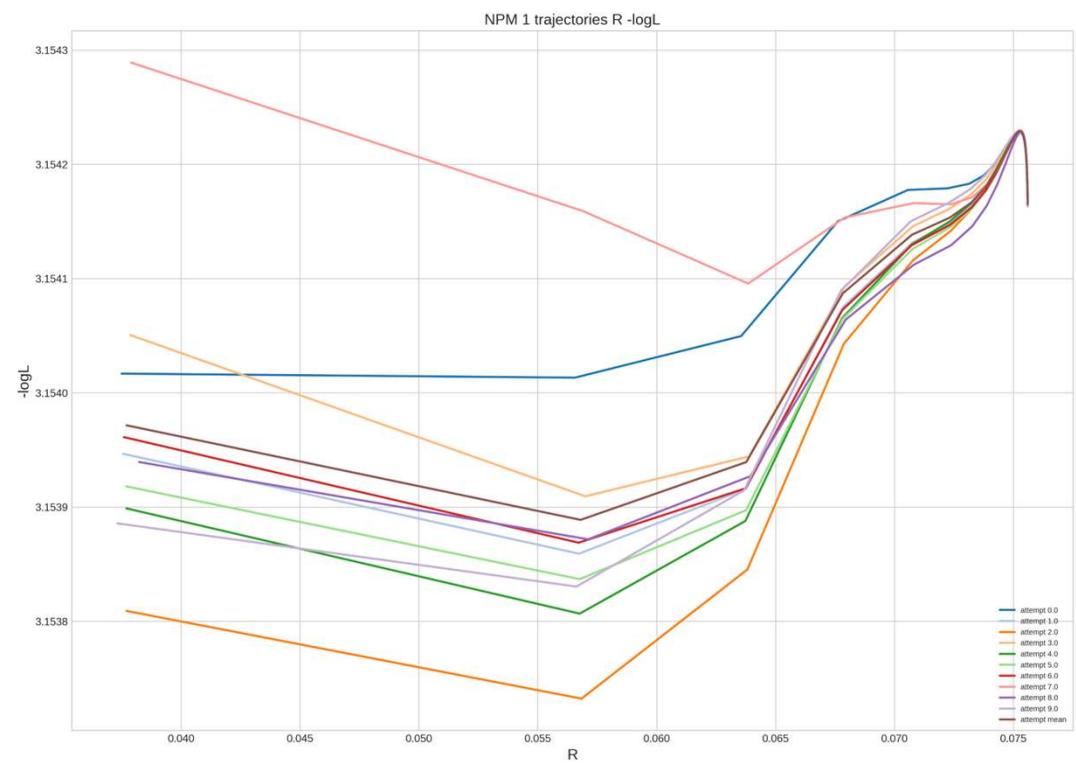
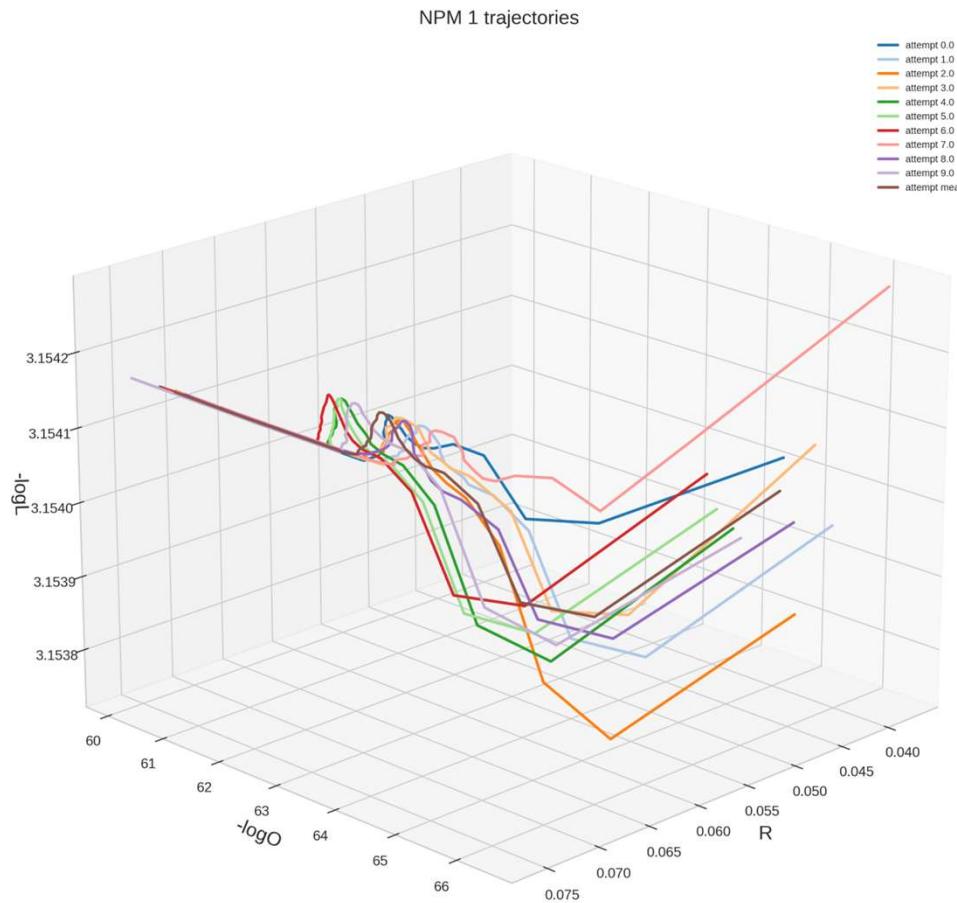
Results Visualization and Interpretation



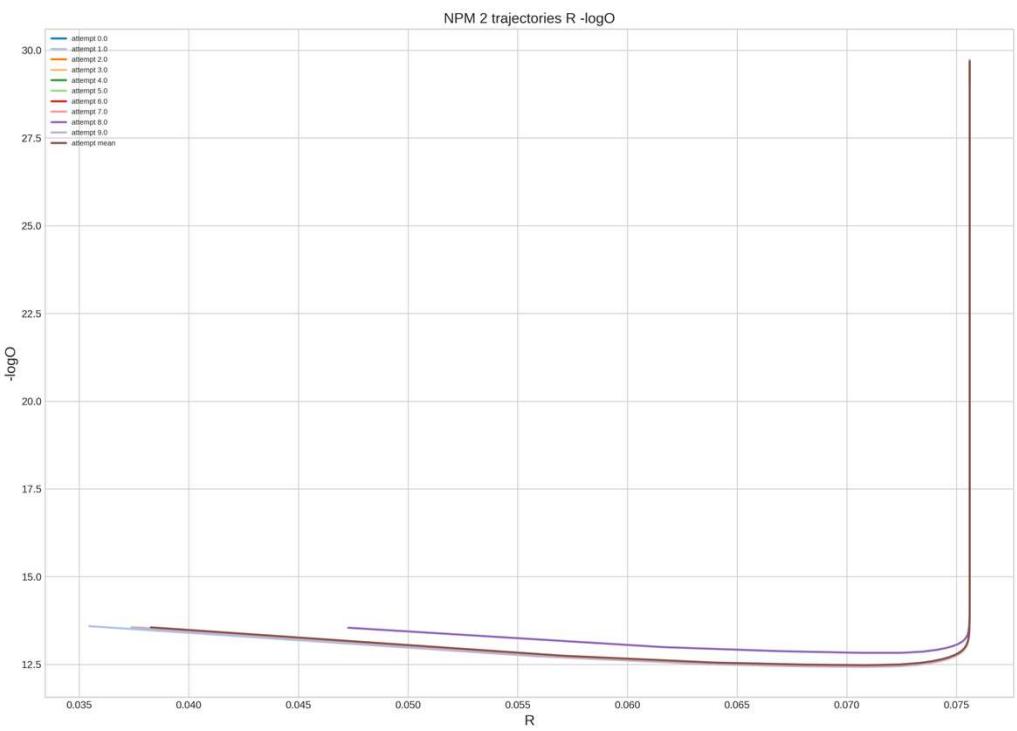
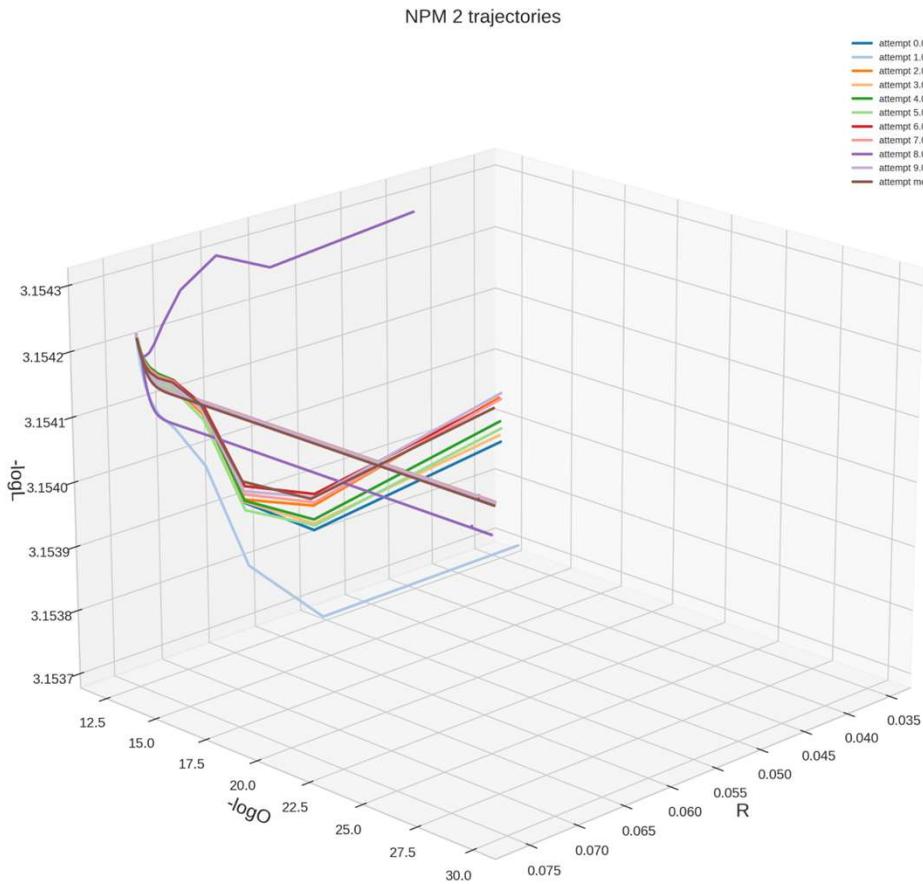
Results Visualization and Interpretation



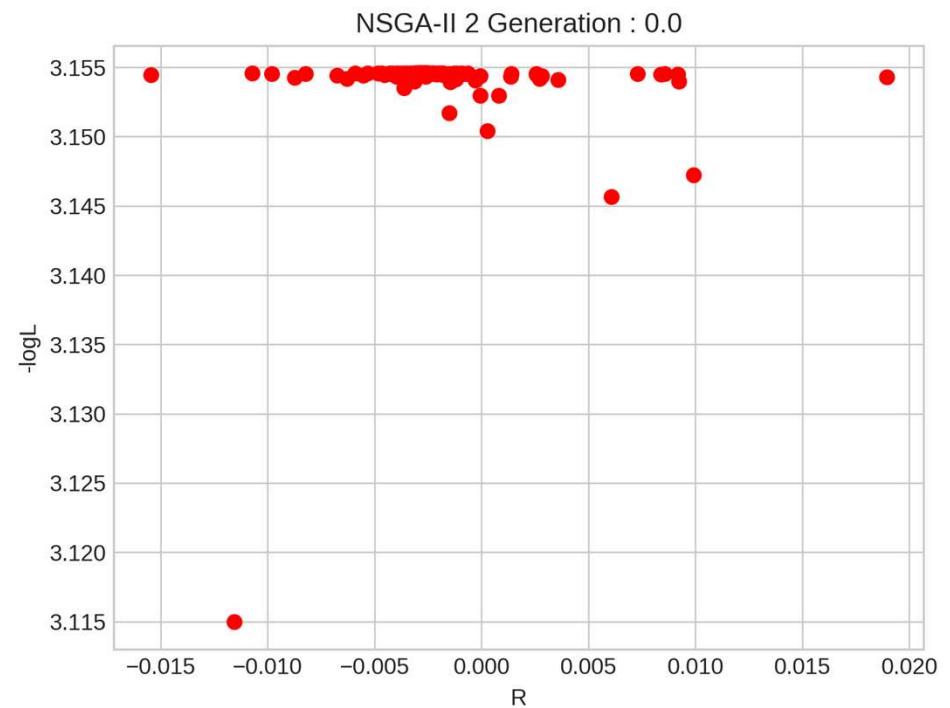
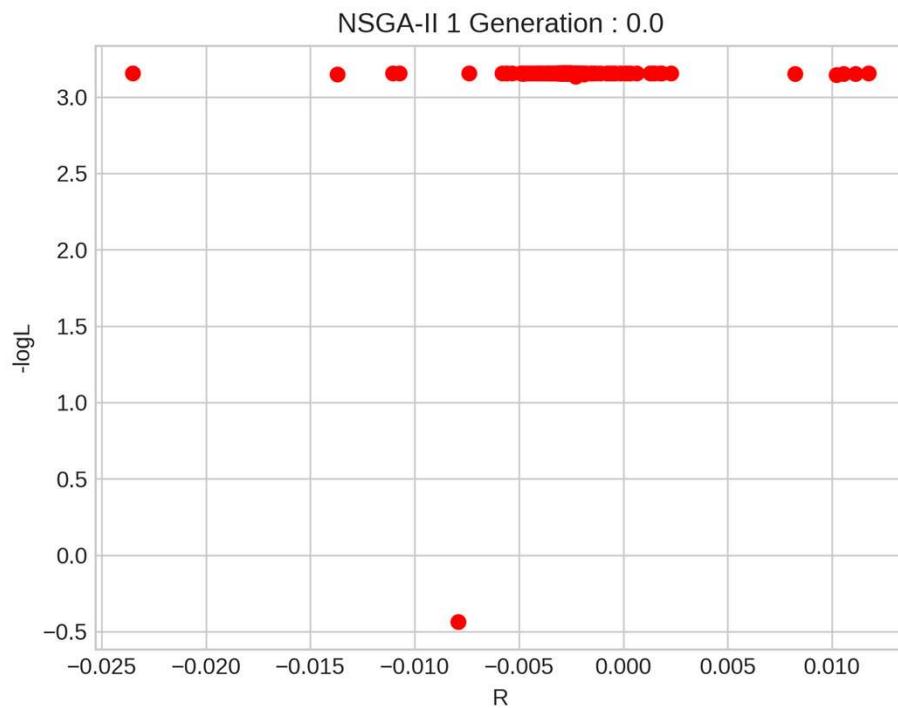
Results Visualization and Interpretation



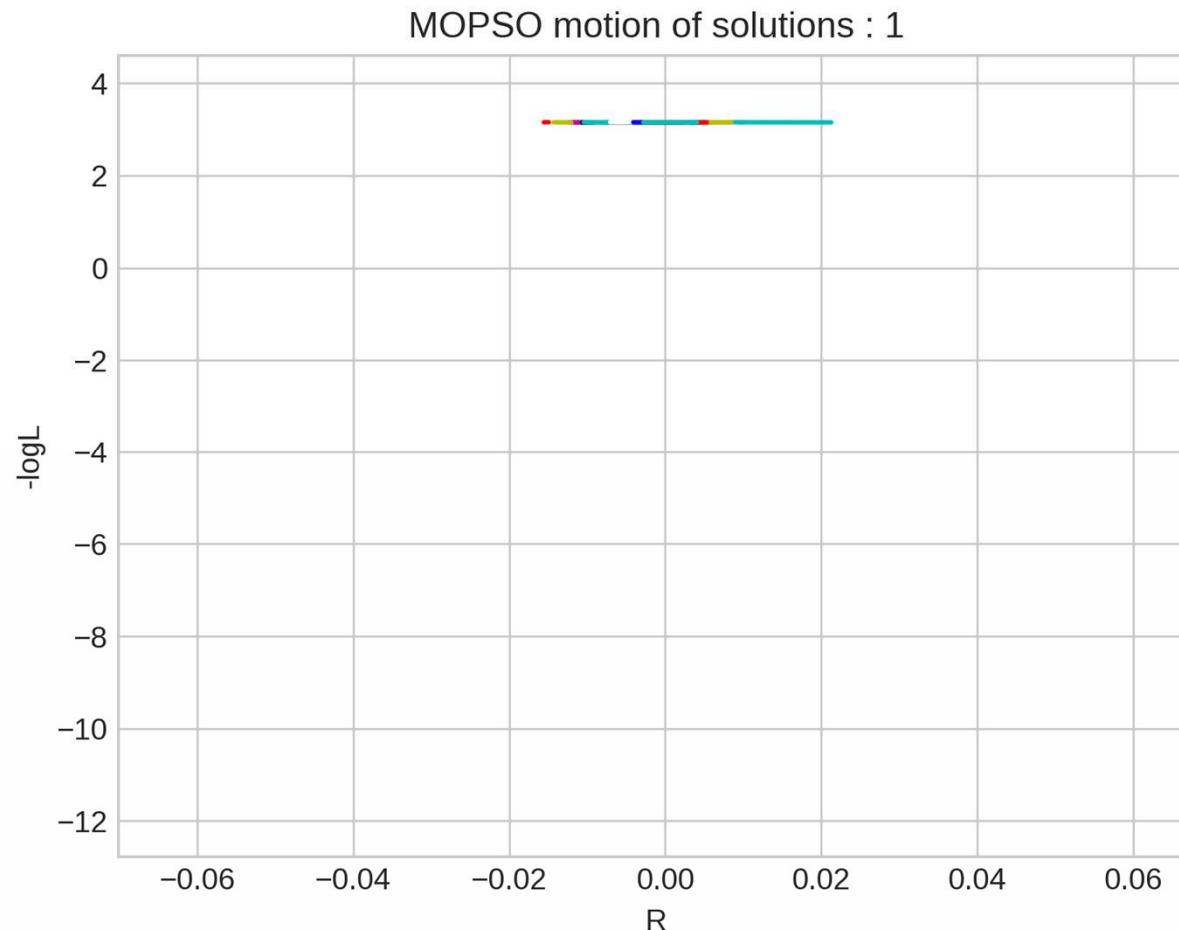
Results Visualization and Interpretation



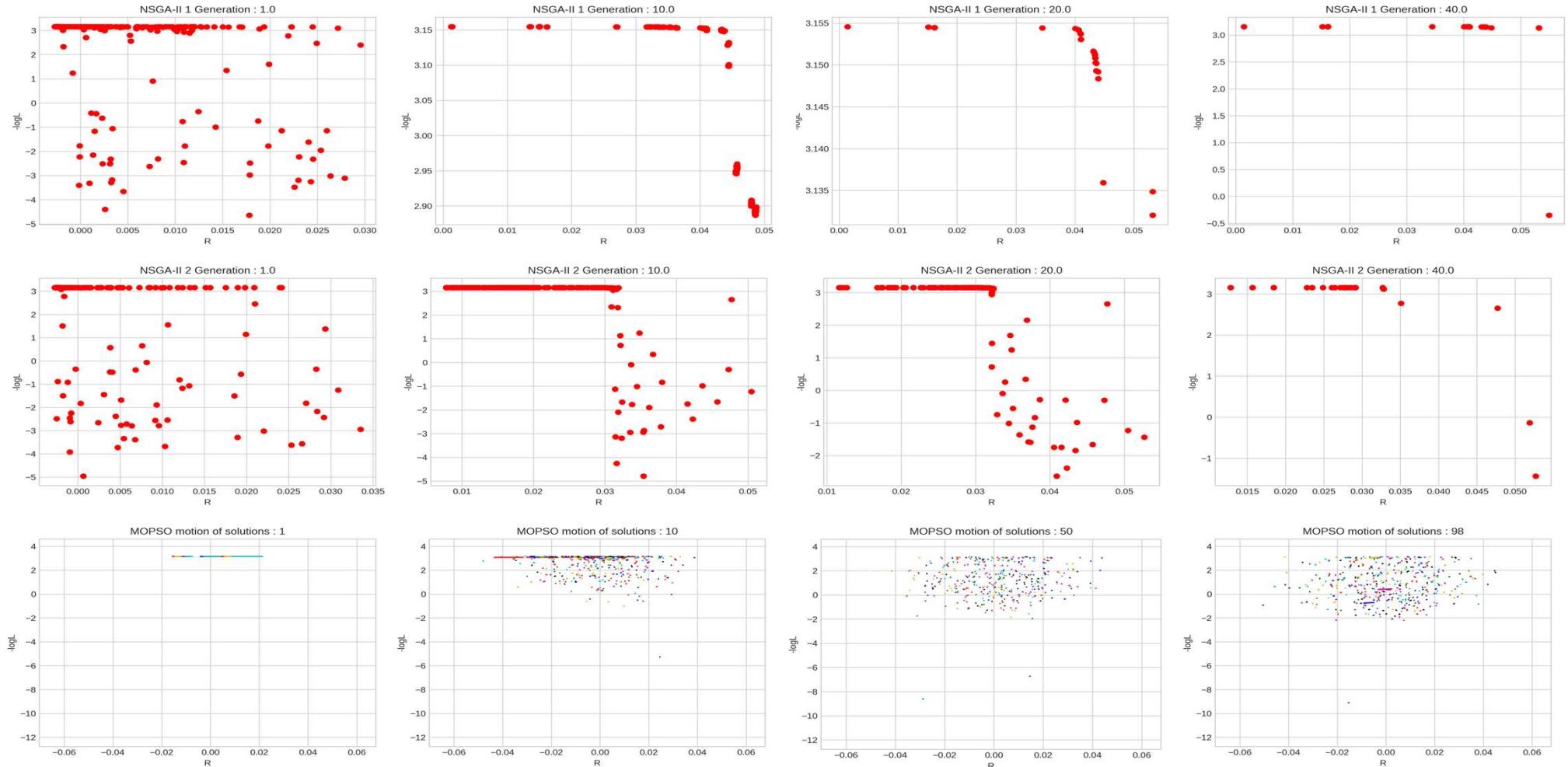
Results Visualization and Interpretation



Results Visualization and Interpretation



Results Visualization and Interpretation (Backup)



Results Table and Discussion

Objective	Statistic	Baseline	OPI		DOI		NPM		NSGA-II		MOPSO
			Type 1	Type 2	Type 1	Type 2	Type 1	Type 2	Type 1	Type 2	
R	mean (1e-2)	-0.272	7.53	7.43	7.53	7.43	7.55	7.55	1.90	2.63	0.00266
	std (1e-3)	2.36	3.43	4.48	1.85	3.23	1.42	1.41	9.38	7.60	16.2
	min (1e-2)	-3.74	-0.31	-2.22	3.11	-0.40	3.73	3.55	-2.35	-1.55	-5.86
	50% (1e-2)	-0.275	7.56	7.47	7.56	7.47	7.56	7.56	1.62	2.91	-8.96
	max (1e-2)	4.00	7.56	7.56	7.56	7.56	7.56	7.56	5.51	5.93	5.63
$-\log(L)$	mean	3.15	3.15	3.14	3.15	3.15	3.15	3.15	3.12	3.03	1.42
	std (1e-2)	2.51	0.00147	27.6	9.56	19	0.00155	0.00216	35.9	74.6	146
	min	1.70	3.15	-6.45	2.60	-6.37	3.15	3.15	-4.66	-4.96	-11.3
	50%	3.15	3.15	3.15	3.15	3.15	3.1	3.15	3.15	3.15	1.55
	max	3.15	3.15	3.15	3.15	3.15	3.15	3.15	3.15	3.15	3.15
$-\log(O)$	mean	64.46	58.34	61.80	13.82	10.06	61.58	27.22	63.60	64.15	63.38
	std	1.36	2.51	0.829	9.54	5.98	0.75	4.19	0.52	0.37	0.95
	min	56.41	54.09	60.5	-6.56	-6.65	60.06	12.43	56.32	58.47	59.65
	50%	64.68	57.84	61.6	16.21	14.4	61.46	29.01	63.54	64.09	63.42
	max	67.50	66.32	66.61	24.46	14.96	66.42	29.74	66.98	66.88	67.29

Table 3.1: Statistics on solutions generated by various method

- max achieved in R
- not 50% for some
- low std is better
- EA methods
- latter decimals
- smaller: cause R?
- DOI

Conclusion

- Defined cond-fMOOP
- Reviewed various methods to solve MOOP
- Applications of this formulation for Neural Networks
- Formulated critical question about cond-fMOOP
- Relations between cond-fMOOP and L_2 regularization
- Observed properties of original problem
- Used them to extend various methods
- Implemented those methods
- Derived Insights from those results
- Noted the advantages of using the Derived Properties.

Future Works

- Studies to determine problems in the cond-fMOOP category
- Develop a class of methods to address the cond-fMOOP problem
- Address any of the three critical questions
- Explore other properties formulations of the loss and reward functions to optimize computation of solutions
- Explore other potential applications of the presented methods beyond the Linear Factor Model problem

Thank You

You can find all the

- Report LaTeX files
- Images
- Python Code

On my github

Link: [harshal-dupare/masters_thesis_project: Masters Thesis Project, Mathematics Department, IIT Kharagpur \(github.com\)](https://github.com/harshal-dupare/masters_thesis_project)

Url: https://github.com/harshal-dupare/masters_thesis_project