Trust-Region Based Multi-Objective Optimization for Low Budget Scenarios

Proteek Roy, Rayan Hussein, Julian Blank, Kalyanmoy Deb

Department of Electrical and Computer Engineering

Michigan State University

March 12, 2019

East Lansing '19

Outlines

MICHIGAN STATE UNIVERSITY

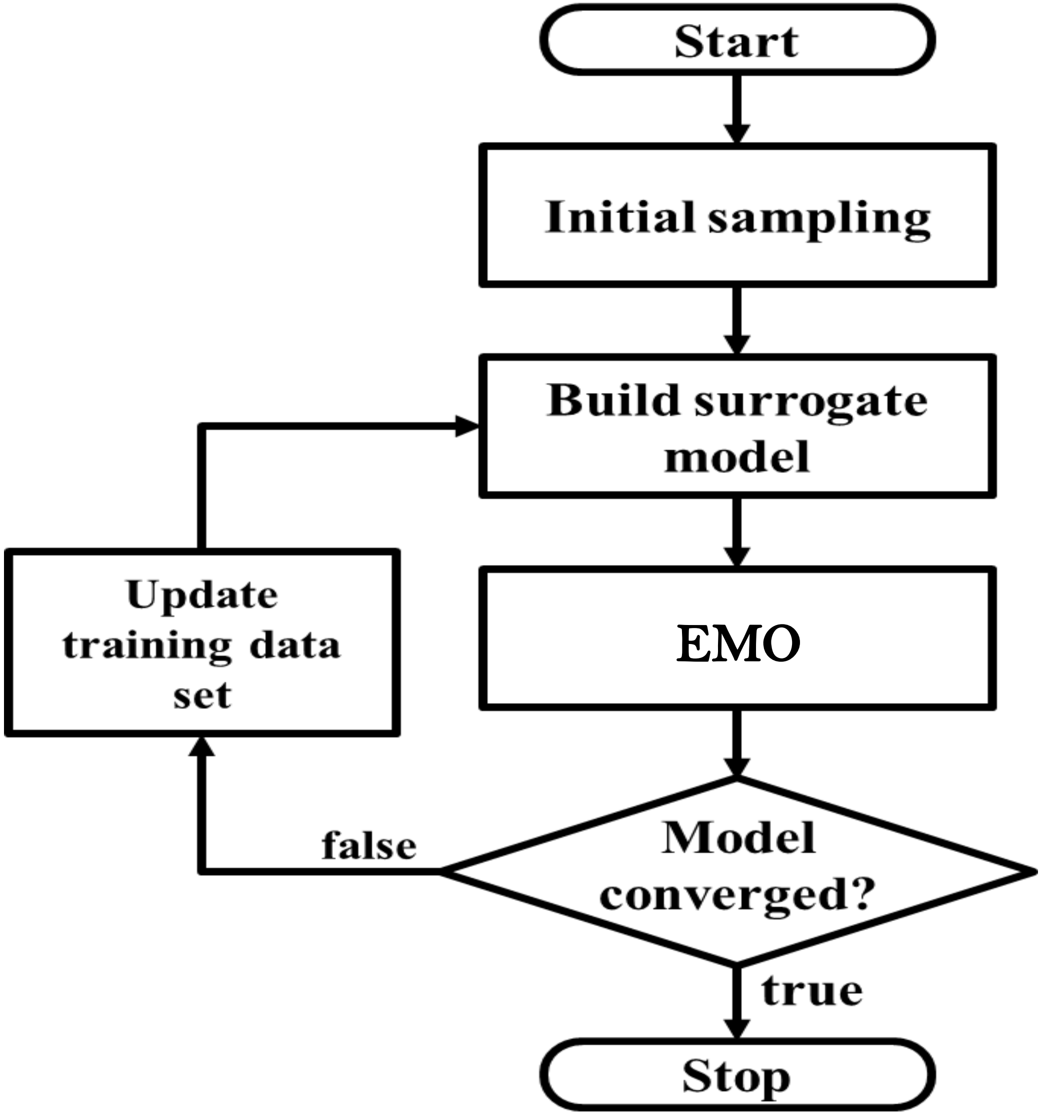
- Metamodeling for Multi-Objective Optimization
- A Taxonomy for Metamodeling Frameworks for Evolutionary Multi-Objective Optimization
- Metamodeling Framework
- Trust Region based method
- Switching Between Frameworks and Use of Trust Regions
- Results and Comparison
- Conclusions

MICHIGAN STATE UNIVERSITY

Metamodeling = Surrogate Model = Approximation Model

- Solution evaluations are computationally expensive in practice (Network flow simulation, CFD)
- Single-objective methods may not be straightforward or easy to extend to EMO
- Multiple solutions are targeted
- Metamodels are not accurate
- Multiple objectives and constraints to be meta-modeled
- Constraint handling must be integral part of metamodeling (often ignored)

1. LHS sampling & evaluation (High-Fidelity), sent to **Archive**
2. Build surrogate model(s) for objective(s) and constraint(s)
3. EMO
4. Return **one/multiple** solution(s) & evaluation (High-Fidelity), include in **Archive**
5. Go to step 2



Choices of Metamodel Based Optimization

MICHIGAN STATE UNIVERSITY

1

What functions should be metamodeled?

All Objectives?
All Constraints?
Or their aggregation?

4

Which Optimization Algorithm?

NSGA-III, MOEA/D, RVEA

2

Best Metamodel approach?

RBF, Kriging, NN?

3

How many times? Fixed or Temporal?

When to use what

Taxonomy of Metamodeling Frameworks

MICHIGAN STATE UNIVERSITY

Objectives Separately
M-Metamodels

$(f_1(x), f_2(x), \ldots, f_M(x))$

Aggregated Objective Function
1-Metamodel

ASF(x)

Combine Obj. & Constr., target 1 optimum
1-Metamodel

ASF(x)+CV(x)

Constraints Separately
J-Metamodels

$g_j(\mathbf{x}) \geq 0, \quad \forall j \in \{1, \ldots, J\}$

Constraint Violation Function
1-Metamodel

$\mathrm{CV}(\mathbf{x}) = \sum_{j=1}^{J} \langle \bar{g}_j(\mathbf{x}) \rangle$

Combine Obj. & Constr., target multiple optimum
1-Metamodel

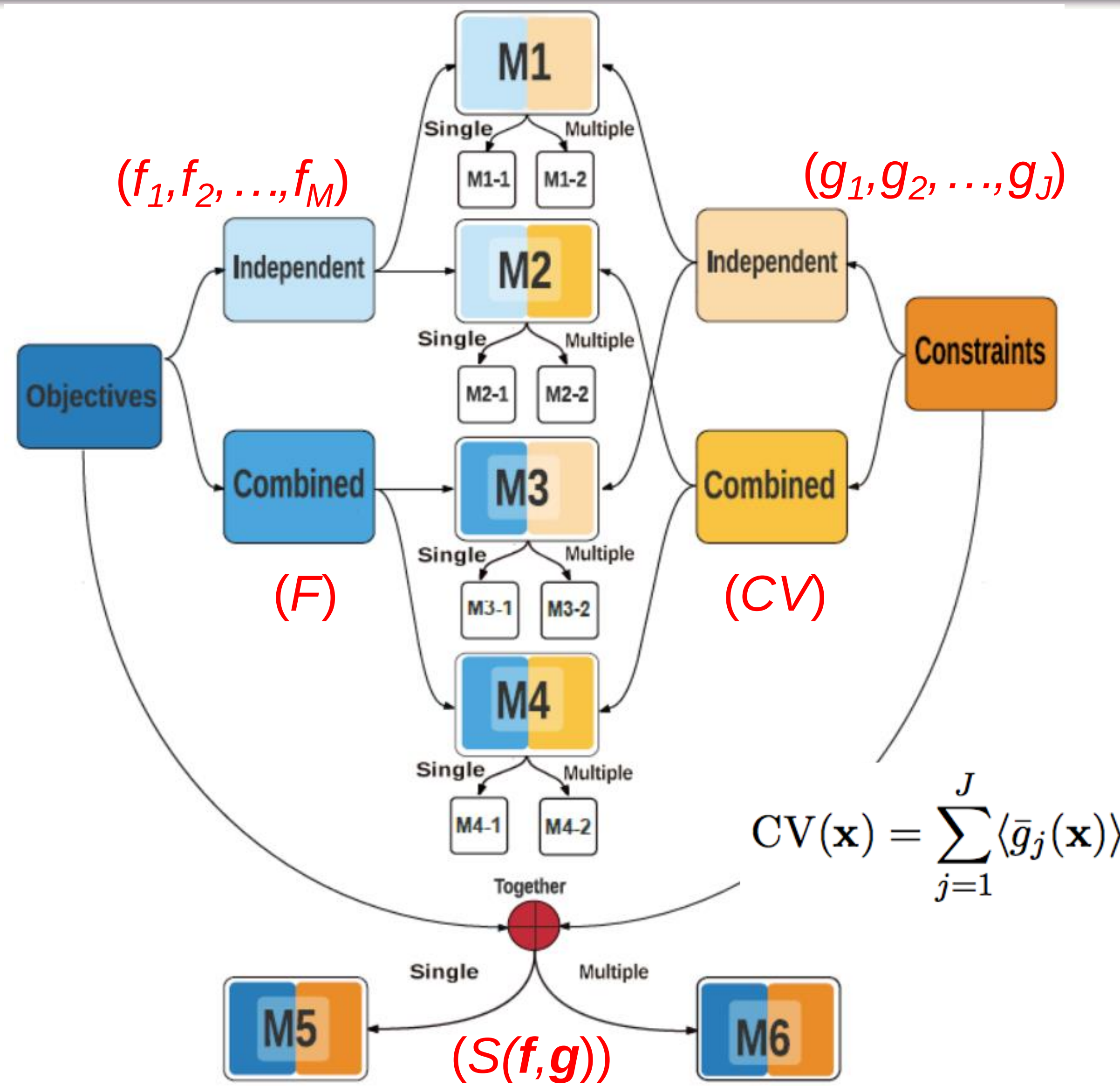
$\mathrm{Min}_h\ \mathrm{ASF}^h(x)+\mathrm{CV}(x)$

Purpose: Construct model search space with different number of metamodels

Better control over search & #metamodels, helps to improve model accuracy

A Taxonomy for Multi-Objective Constrained Problems

MICHIGAN STATE UNIVERSITY

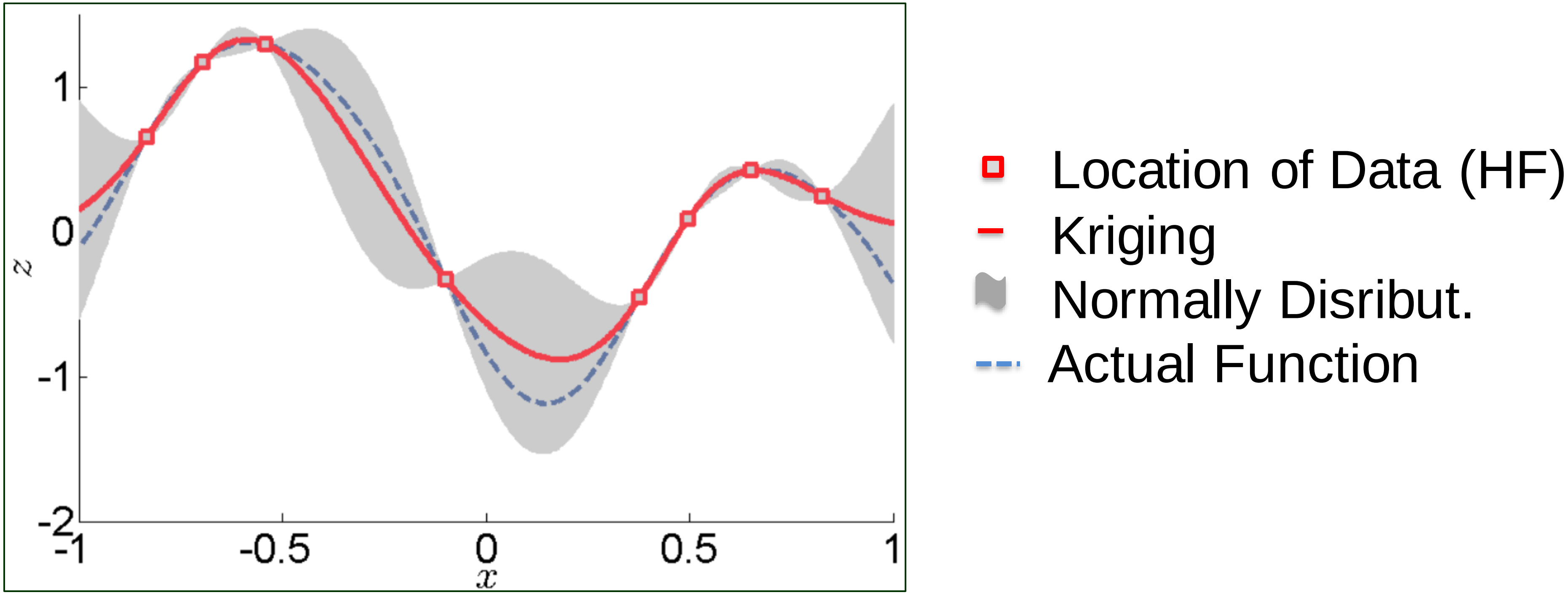


$$\mathrm{CV}(\mathbf{x}) = \sum_{j=1}^{J} \langle \bar{g}_j(\mathbf{x}) \rangle$$

[J] K. Deb, R. Hussein, P. Roy, and G. Toscano "A Taxonomy for Metamodeling Frameworks for Evolutionary Multi-Objective Optimization" , *Accepted IEEE Transactions on Evolutionary Computation,2018.*

Kriging Predictor: $\hat{y}(x) = \hat{\mu} + r(x^*, x)^T R^{-1}(y(x) - \mathbf{1}\hat{\mu})$

Error Estimate: $s^2(x) = \hat{\sigma}^2 [1 - r^T R^{-1} r + \frac{(1 - r^T R^{-1} r)^2}{\mathbf{1}^T R^{-1} 1}]$

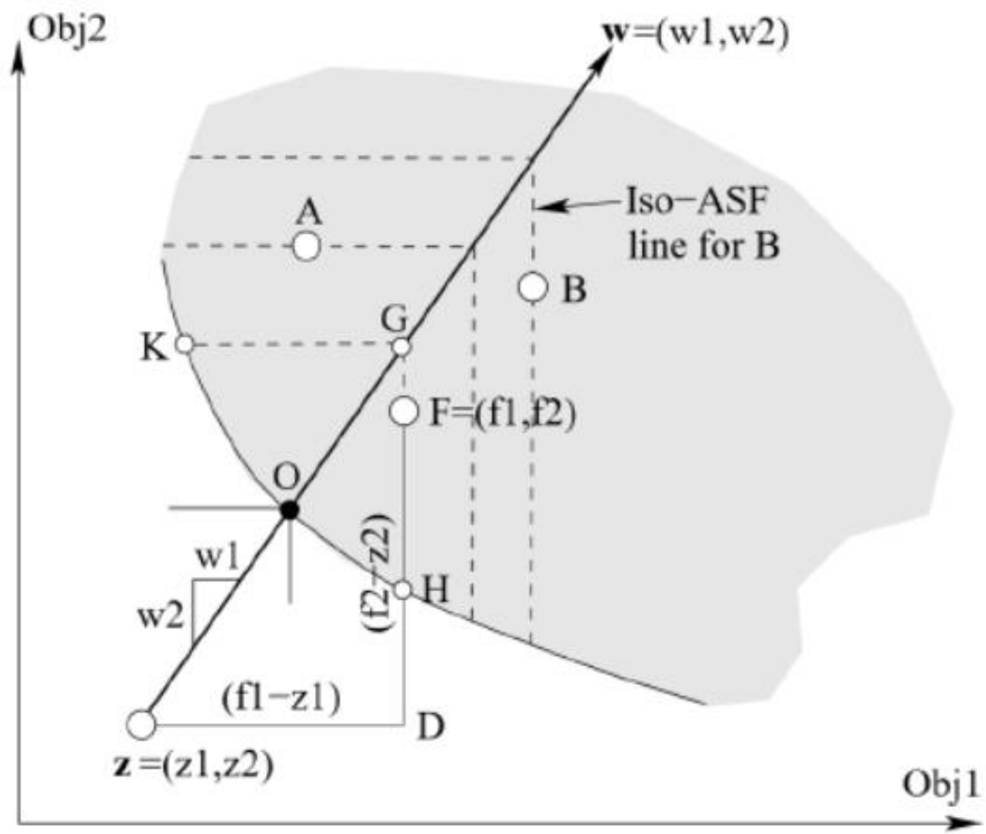


Achievement Scalarization Function (ASF)

- Achievement scalarization method (Wierzbicki, 1980)
- Reference direction ***w*** is changed, reference point ***z*** is fixed to find different PO points
- For a fixed ***z*** and changed ***w*** landscape leads to respective PO point
- It makes monotonic single objective value

$$\text{Minimize}_{(\mathbf{x})} \quad \text{ASF}(\mathbf{x}, \mathbf{z}, \mathbf{w}) = \max_{i=1}^{M} \left(\frac{f_i(\mathbf{x}) - z_i}{w_i} \right),$$

$$\text{Subject to} \quad g_j(\mathbf{x}) \leq 0, \quad j = 1, 2, \ldots, J.$$

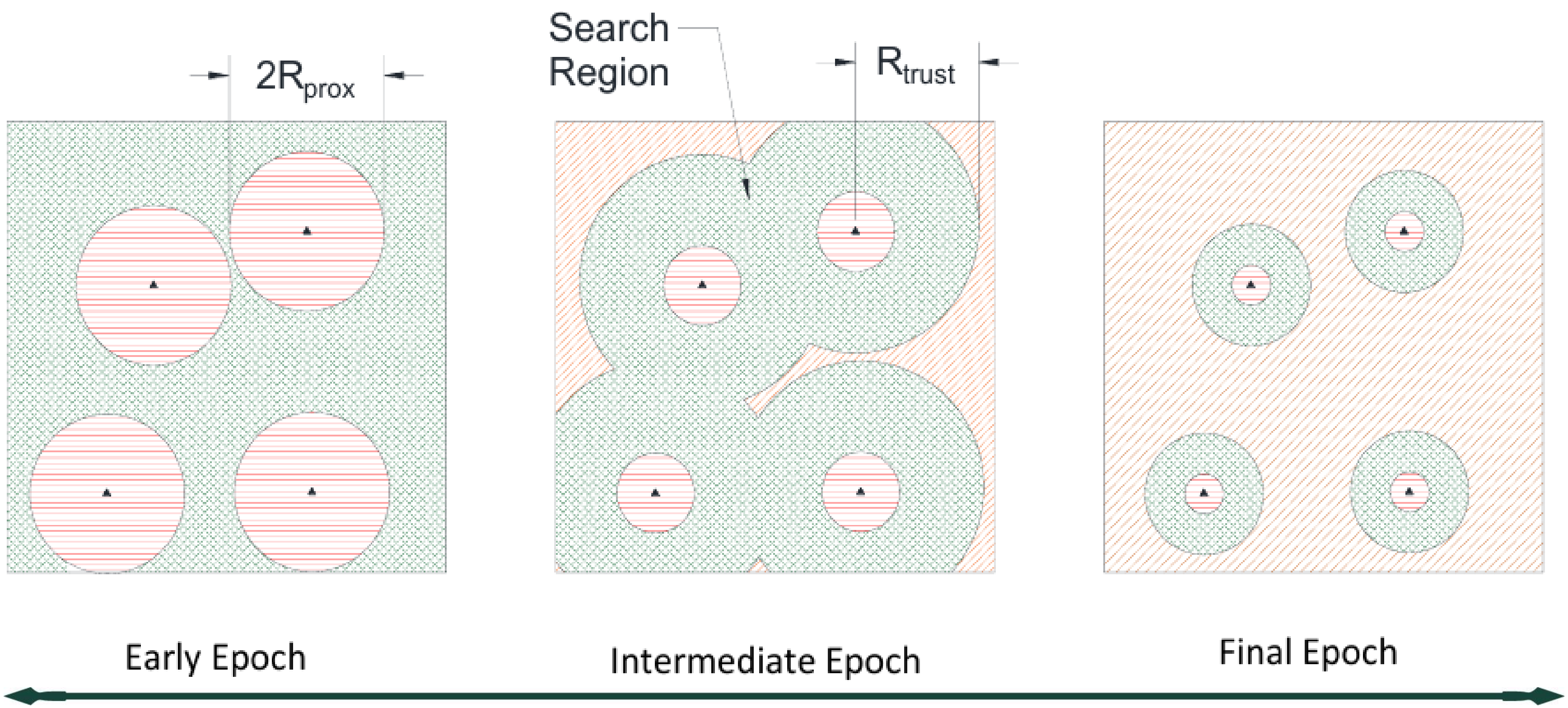


A. P. Wierzbicki, “The use of reference objectives in multiobjective optimization," in Multiple criteria decision making theory and application. Springer, 1980.

Trust Regions

- Maintain a balance between exploration versus exploitation
- Reduce the two radii (R_{trust} and R_{Prox}) after every metamodeling task by constant factors:

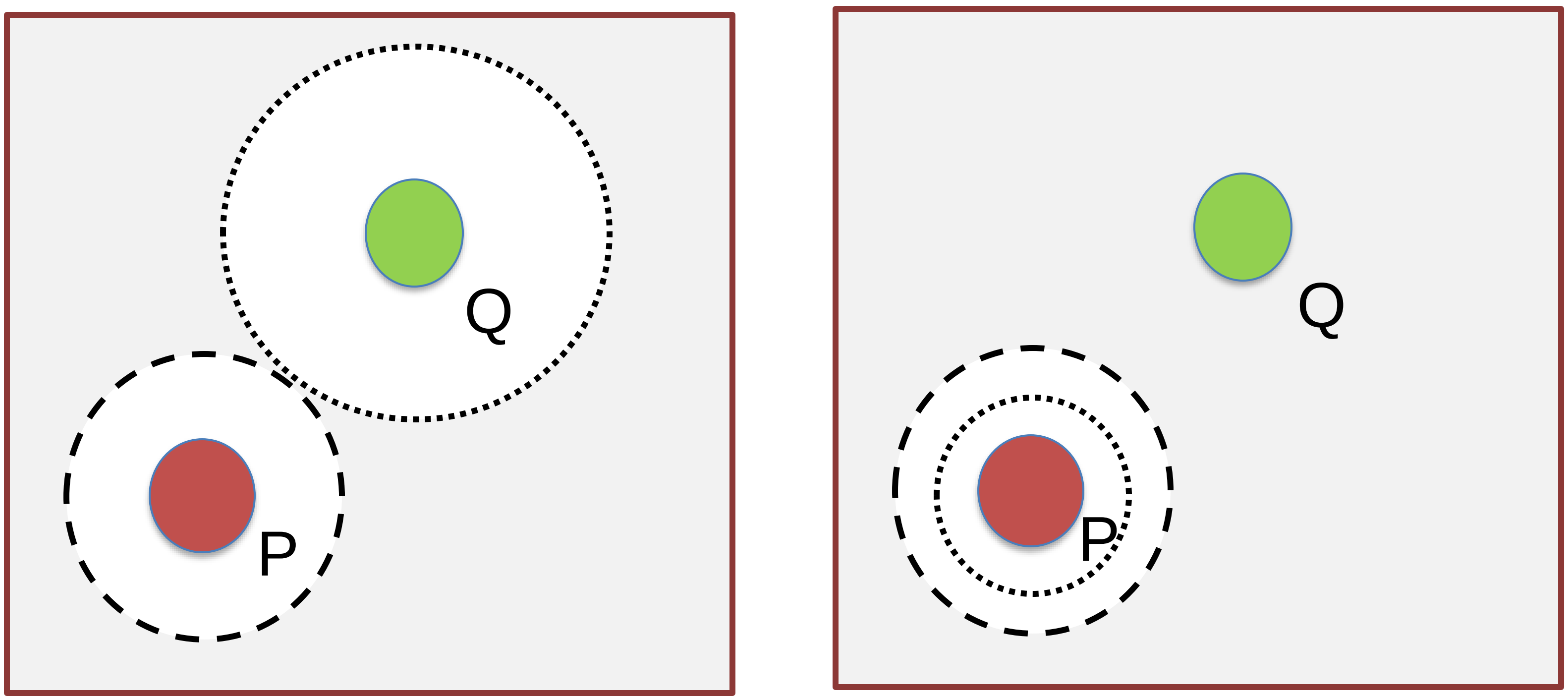
$$R_{trust}^{new} = 0.75 R_{trust}^{old} \qquad R_{prox}^{new} = 0.1 R_{trust}^{new} \qquad R_{prox} \leq R_{search} \leq R_{trust}$$



$$\text{Minimize}_q \quad \hat{f}(q), \quad \text{Subject to} \quad \|q - p\| \leq \delta_k.$$

P: The current iterate (solution).
q: The new predicted point.
δ_k: The search is restricted within a radius.

$$r = \frac{f(p) - f(q)}{f(p) - \hat{f}(q)}.$$

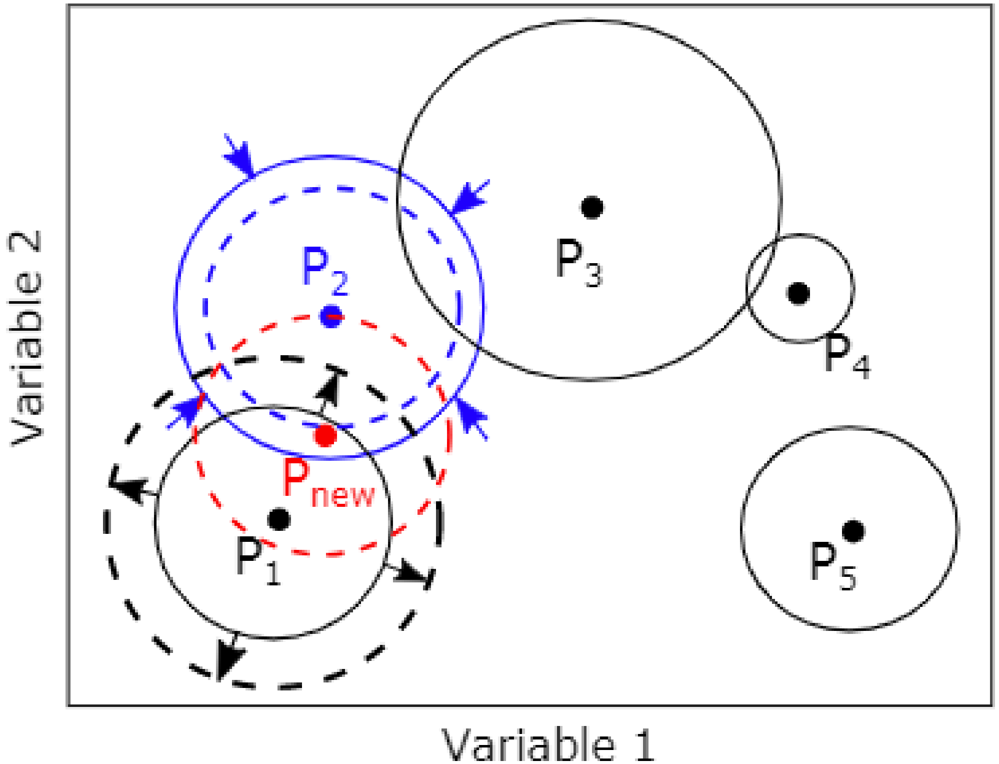


[1] Alexandrov, N.M., Dennis, J.E., Lewis, R.M., Torczon, V.: A trust-region framework for managing the use of approximation models in optimization. Structural optimization (1998)

$$\text{Minimize}_{q \in \Omega} \quad \hat{f}_1(q), \ldots, \hat{f}_M(q)$$

$$\hat{g}_j(x) \geq 0, \quad \forall j \in \{1, \ldots, J\}$$

$$\text{Subject to} \quad \|q - p\| \leq \delta_k^p, \quad \exists p \in A$$



MICHIGAN STATE UNIVERSITY

How to Apply Performance Indicator in Multi-Objective Scenario?

$$r = \frac{f(p) - f(q)}{f(p) - \hat{f}(q)}.$$

The proposed performance criteria based on ASF:

$$PI_{ASF}(q) = \frac{ASF(p) - ASF(q)}{ASF(p) - \widehat{ASF}(q)}.$$

$\widehat{ASF}$: is obtained from predicted objectives.

Performance Indicator for Constraints

A=Archive

Hypervolume based Performance Indicator (PI_{HV}):

$$PI_{HV}(q) = \frac{HV(F(A) \cup F(q)) - HV(F(A))}{HV(F(A) \cup \hat{F}(q)) - HV(F(A))}.$$

Performance Indicator for Constrained Problems:

$$PI_{CV}(q) = \frac{CV(G(p)) - CV(G(q))}{CV(G(p)) - CV(\hat{G}(q))}$$

Overall Trust Region Adaptation

$$r = \begin{cases} PI_{HV}(q) \text{ or } PI_{ASF}(q), & \text{if both } p \text{ and } q \text{ feasible}, \\ r_2 + \epsilon, & \text{if } p \text{ infeasible, } q \text{ feasible}, \\ r_1 - \epsilon, & \text{if } p \text{ feasible, } q \text{ infeasible}, \\ PI_{CV}(q), & \text{otherwise}. \end{cases}$$

$$\delta^p_{k+1} = \begin{cases} c_1 \delta^p_k & \text{if } r < r_1 \\ \min\{c_2 \delta^p_k, \Delta_{max}\} & \text{if } r > r_2 \\ \delta^p_k & \text{otherwise} \end{cases}$$

Proposed Overall Algorithm

Algorithm 1: Trust Region Based Algorithm or TR-NSGA-II

Input : Obj: $[f_1, \ldots, f_m]^T$, Constr: $[g_1, \ldots, g_J]^T$, n (vars), ρ (sample size), E (max. high-fidelity SEs), NSGA-II (multi-obj EA) with pop-size μ, number of generation for model optimization τ, other parameters of NSGA-II Γ, Constraint violation function **CV**, Trust region parameters $\delta_{init}, \Delta_{max}, c_1, c_2, r_1$ and r_2

Output: Solution set P_T

$t, e \leftarrow 0$;
$\mathrm{P}_t, \mathrm{F}_t, \mathrm{G}_t \leftarrow \varnothing$;
$\mathrm{P}_{new} \leftarrow \mathrm{LHS}(\rho, n)$`// Initial solutions`
$\delta^\ell \leftarrow \delta_{init}, \forall \ell \in \{1, \ldots, \rho\}$;
while *True* **do**
 $\mathrm{F}^i_{new} \leftarrow f_i(\mathrm{P}_{new}), \forall i \in \{1, \ldots, M\}$`// eval obj.`
 $\mathrm{G}^j_{new} \leftarrow g_j(\mathrm{P}_{new}), \forall j \in \{1, \ldots, J\}$`// eval constr.`
 if $t > 0$ **then**
  $\widehat{\mathrm{F}}^i_{new} \leftarrow \widehat{f}^i_t(\mathrm{P}_{new}), \forall i \in \{1, \ldots, M\}$`// predicted`
  $\widehat{\mathrm{G}}^j_{new} \leftarrow \widehat{g}^j_t(\mathrm{P}_{new}), \forall j \in \{1, \ldots, J\}$`// predicted`
  $\delta \leftarrow \textsc{Update\_TrustRegion}(\mathrm{F}_t, \widehat{\mathrm{F}}_{new}, \mathrm{G}_t, \widehat{\mathrm{G}}_{new}, \delta)$
 end
 $\mathrm{P}_{t+1}, \mathrm{F}_{t+1}, \mathrm{G}_{t+1} \leftarrow (\mathrm{P}_t \cup \mathrm{P}_{new}), (\mathrm{F}_t \cup \mathrm{F}_{new})$ and $(\mathrm{G}_t \cup \mathrm{G}_{new})$;
 $e \leftarrow e + |\mathrm{P}_{new}|$;
 break if $e \geq E$;
 $\widehat{f}^i_{t+1} \leftarrow \textsc{Metamodel}(\mathrm{F}^i_{t+1}), \forall i \in \{1, \ldots, M\}$`// metamodel obj.`
 $\widehat{g}^j_{t+1} \leftarrow \textsc{Metamodel}(\mathrm{G}^j_{t+1}), \forall j \in \{1, \ldots, J\}$`// metamodel constrt.`
 $\mathrm{P}_{new} \leftarrow \text{NSGA-II}(\widehat{f}_{t+1}, \widehat{g}_{t+1}, \mu, \tau, \Gamma, E - e, \mathbf{CV}, \delta)$; `// Optimize model space`
 $t \leftarrow t + 1$;
end
return $\mathrm{P}_T \leftarrow$ filter the best solutions from P_{t+1}

Parameter Settings

Used Parameters-RGA and NSGA-II

- Population size = 10n
- Crossover probability, p_c = 0.9
- Number of generations = 100
- Mutation probability, p_m = 1/n
- Distribution index for SBX, η_{c} = 2
- Distribution index for Polynomial mutation, η_{m} = 20.
- Two objectives unconstrained: ZDT1, ZDT2, ZDT3, ZDT4 and ZDT6. With 10 variables, **500 FE**, and 21 reference directions.
- Two objectives constraint: BNH, SRN, TNK, OSY, and Welded Beam. With original size variables, **500 FE**, and 21 reference directions.

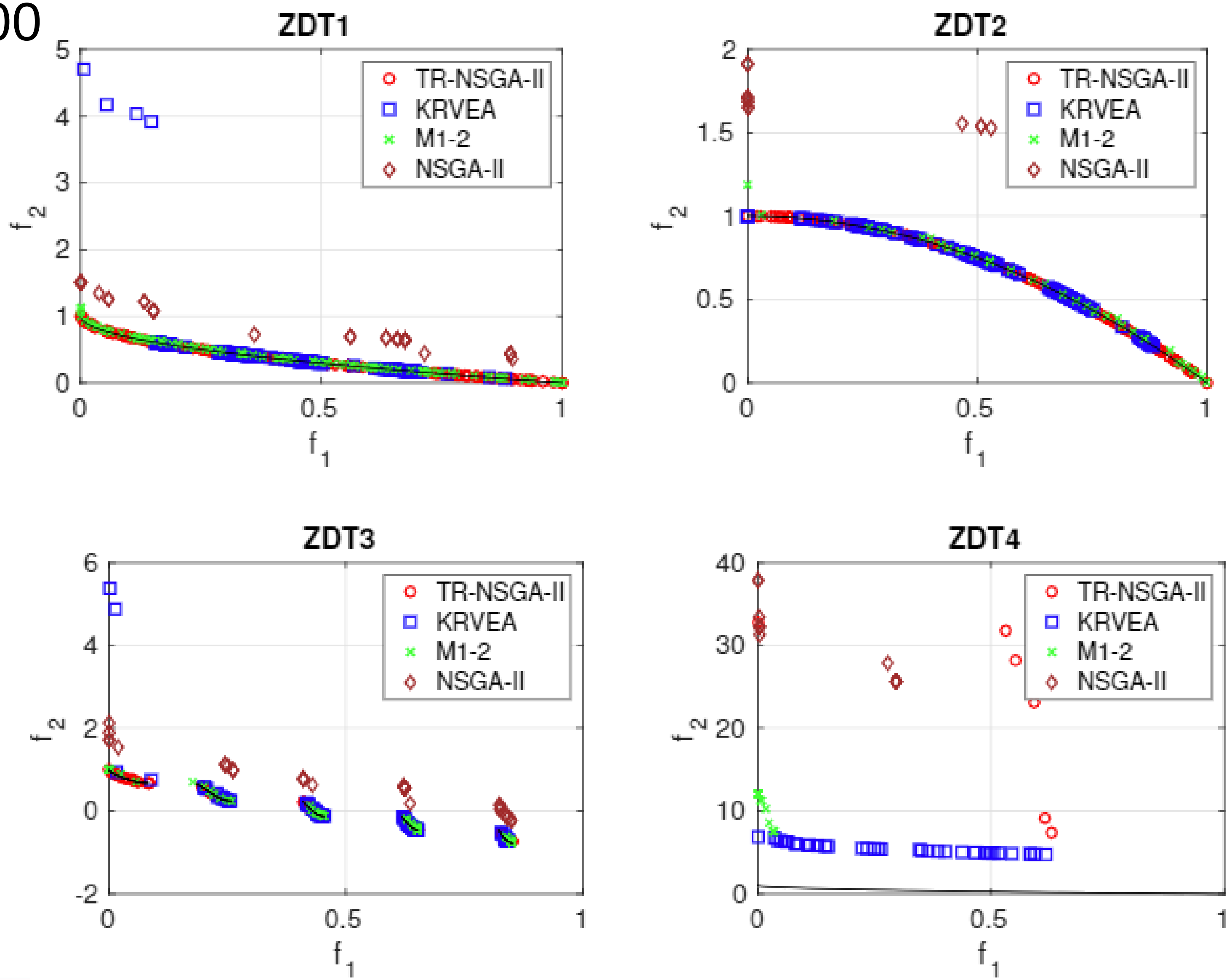
Performance Metrics

- Inverted Generational Distance (IGD)
- Wilcoxon signed-ranked (p-value)

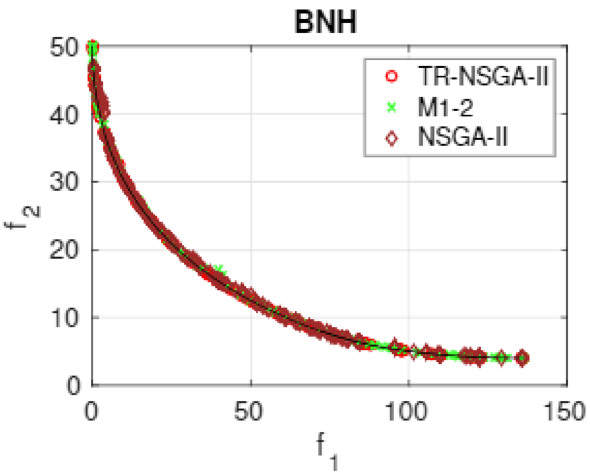
Results: Two Objective Unconstrained problems

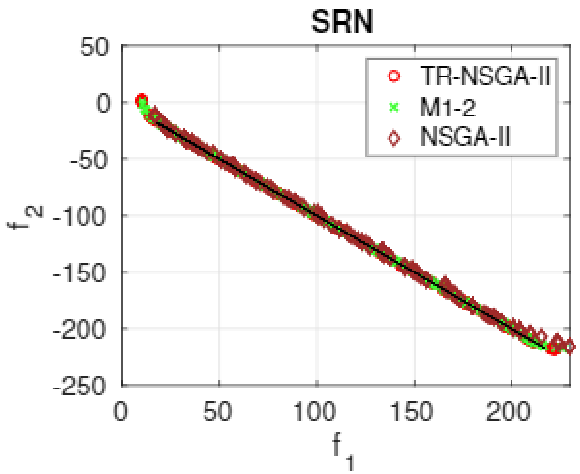
MICHIGAN STATE UNIVERSITY

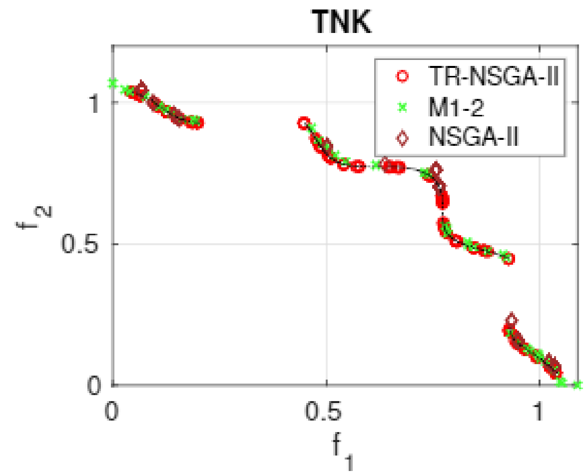
FE=500



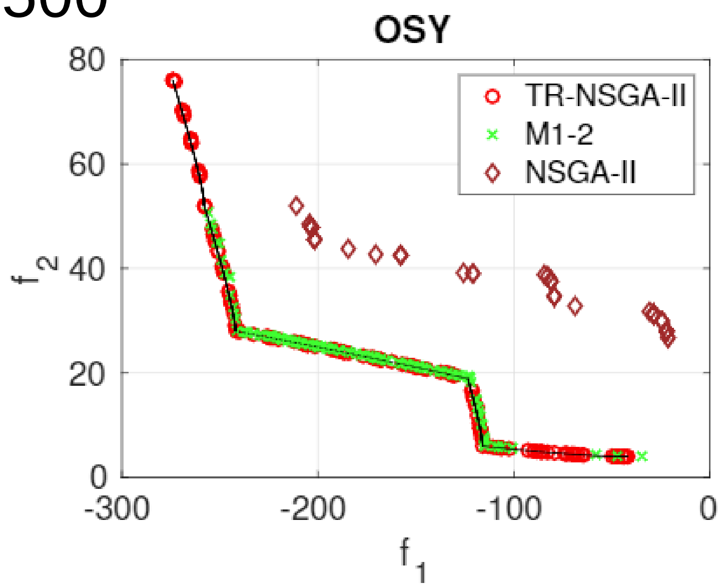
Results: Two Objective Constrained problems

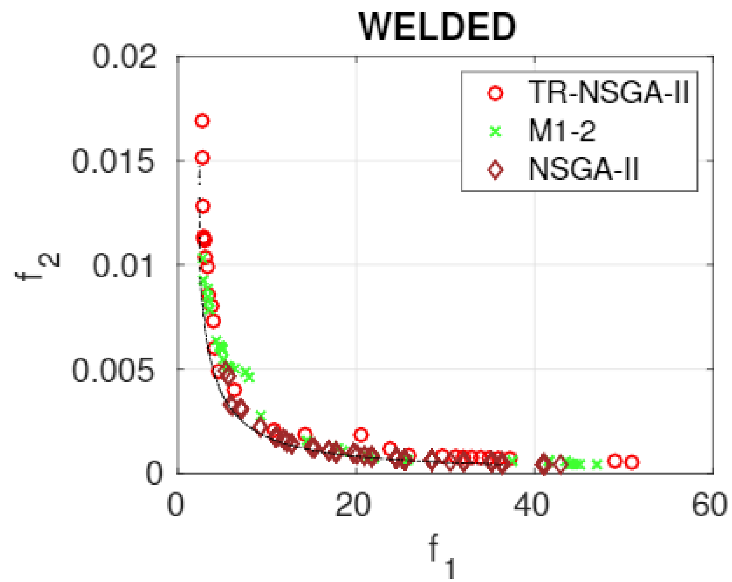






FE=500

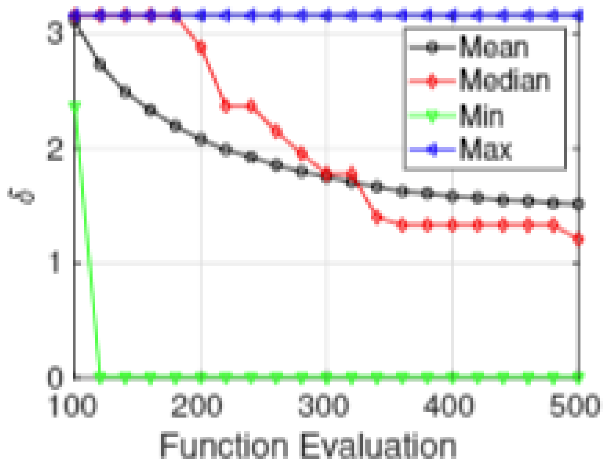




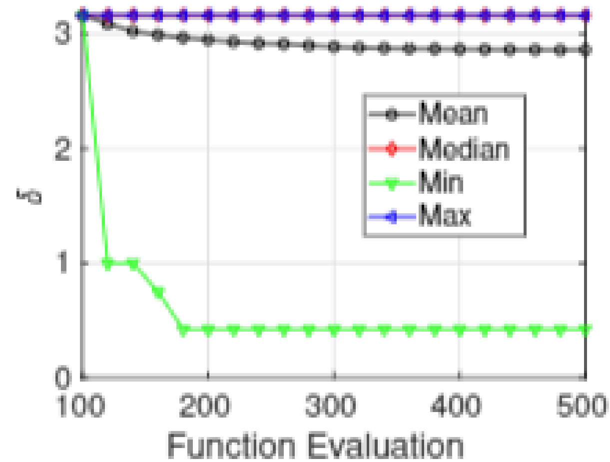
IGD and GD Comparison

Problem/Method	NSGA-II		M1-2		TR-NSGA-II		K-RVEA	
	IGD	GD	IGD	GD	IGD	GD	IGD	GD
ZDT1	0.27131	0.34582	0.01161	0.01091	**0.00121**	**0.00122**	0.07964	0.03715
	p=1.852e-05	p=1.852e-05	p=7.7801e-04	p=7.4613e-04	-	-	p=1.852e-05	p=1.852e-05
ZDT2	0.98265	0.61637	0.00975	0.00755	**0.00057**	**0.00081**	0.03395	**0.00080**
	p=1.852e-05	p=1.852e-05	p=1.852e-05	p=1.852e-05	-	p=0.2851	p=1.852e-05	-
ZDT3	0.32080	0.38940	0.01251	0.00761	**0.00870**	**0.00230**	0.02481	0.00650
	p=1.852e-05	p=1.852e-05	p=1.852e-05	p=1.852e-05	-	-	p=1.852e-05	p=1.802e-04
ZDT4	25.24040	34.43350	**7.11881**	**10.10851**	**6.97620**	**12.92170**	**4.33221**	**4.50901**
	p=1.852e-05	p=1.852e-05	p=0.7928	p=0.1007	p=0.8955	p=0.2934	-	-
ZDT6	5.00571	4.80922	1.55861	2.27535	**0.31070**	2.84941	0.65462	**1.50551**
	p=1.852e-05	p=1.852e-05	p=1.852e-05	p=0.001	-	p=0.0151	p=1.852e-05	-
BNH	0.78981	0.19842	0.45272	0.13696	**0.09651**	**0.09092**	-	-
	p=1.852e-05	p=1.852e-05	p=1.852e-05	p=1.852e-05	-	-		
SRN	1.66162	2.11235	**0.67285**	**0.75337**	1.44045	1.74951	-	-
	p=1.852e-05	p=1.852e-05	-	-	p=1.852e-05	p=1.852e-05		
TNK	0.04182	0.01341	0.01543	0.01008	**0.00141**	**0.00201**	-	-
	p=1.852e-05	p=1.852e-05	p=1.852e-05	p=1.852e-05	-	-		
OSY	35.80211	27.43991	4.78063	0.59202	**0.16731**	**0.25063**	-	-
	p=1.852e-05	p=1.852e-05	p=1.852e-05	p=1.852e-05	-	-		
Welded Beam	1.10272	**0.21092**	0.92692	1.68806	**0.07681**	1.72811	-	-
	p=1.852e-05	-	p=1.8267e-04	p= 0.0042	-	p=0.0012		
C2DTLZ2	0.13733	0.04792	**0.03355**	**0.02373**	0.06411	0.02991	-	-
	p=1.852e-05	p=1.852e-05	-	-	p=1.8267e-04	p=1.8267e-04		

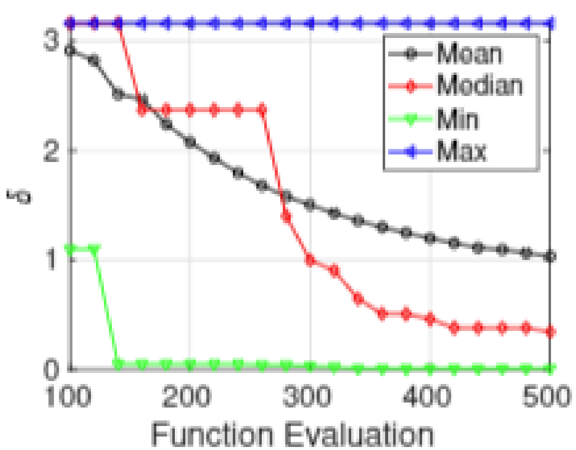
Trust Region Adaptation



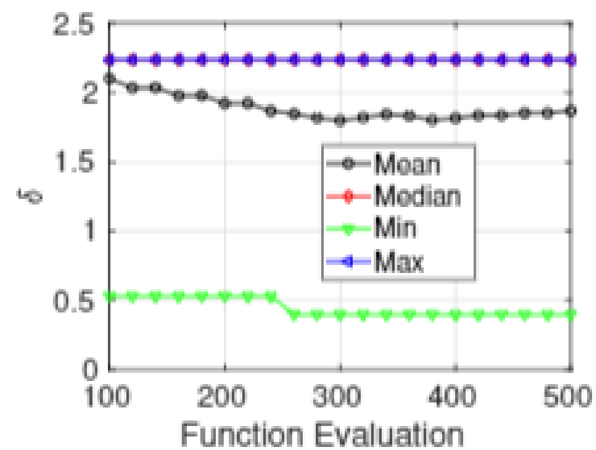
ZDT1



ZDT3



ZDT4

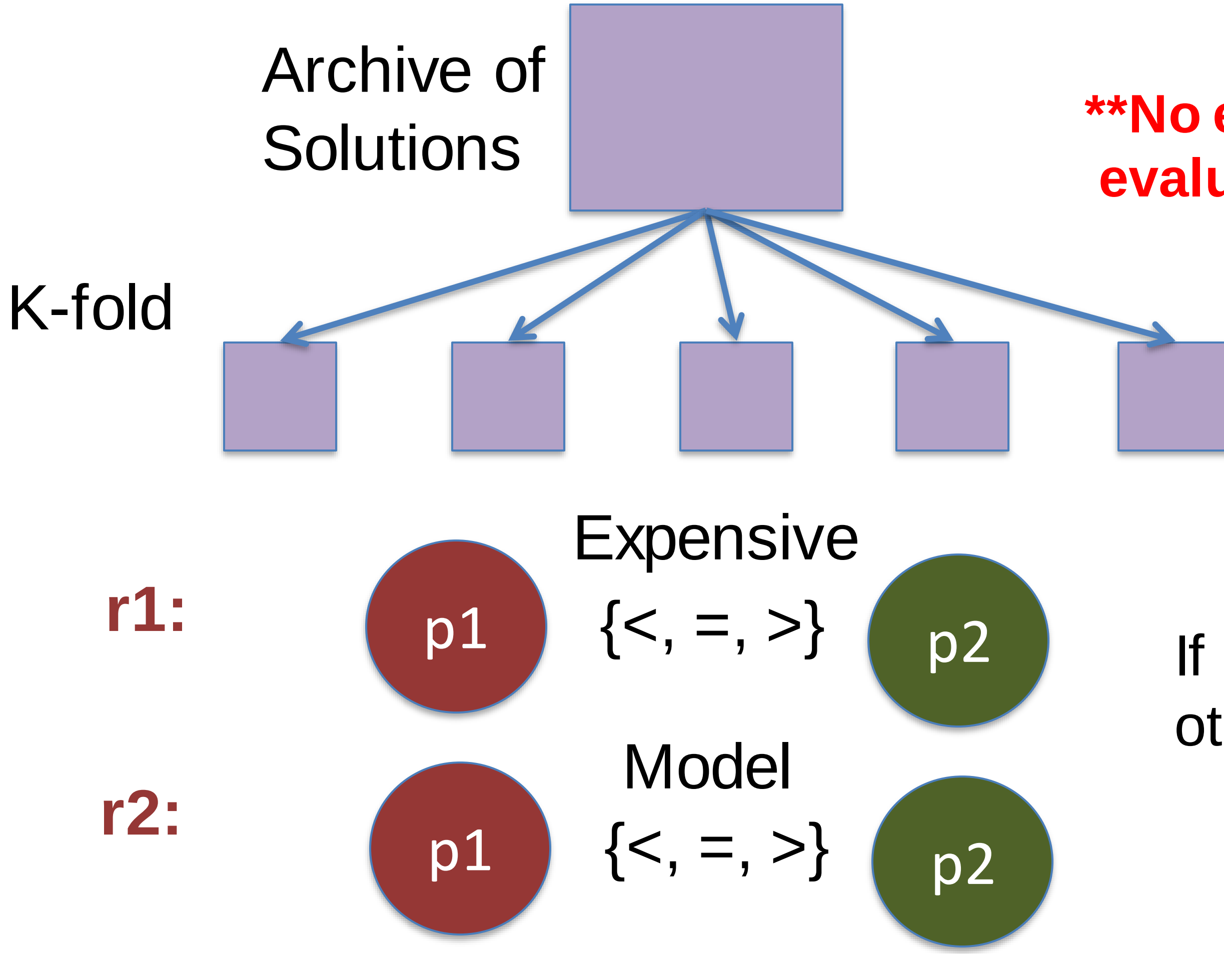


ZDT6

Switching Between Frameworks

- It is more efficient to use different metamodeling frameworks at different stages of the optimization process.
- Adaptive Switching Mechanisms: Ensemble-based method involving different metamodeling frameworks.
- Implemented the trust regions concept for getting more robust solutions and reduce the uncertainty as well.

Adaptive Switching Method

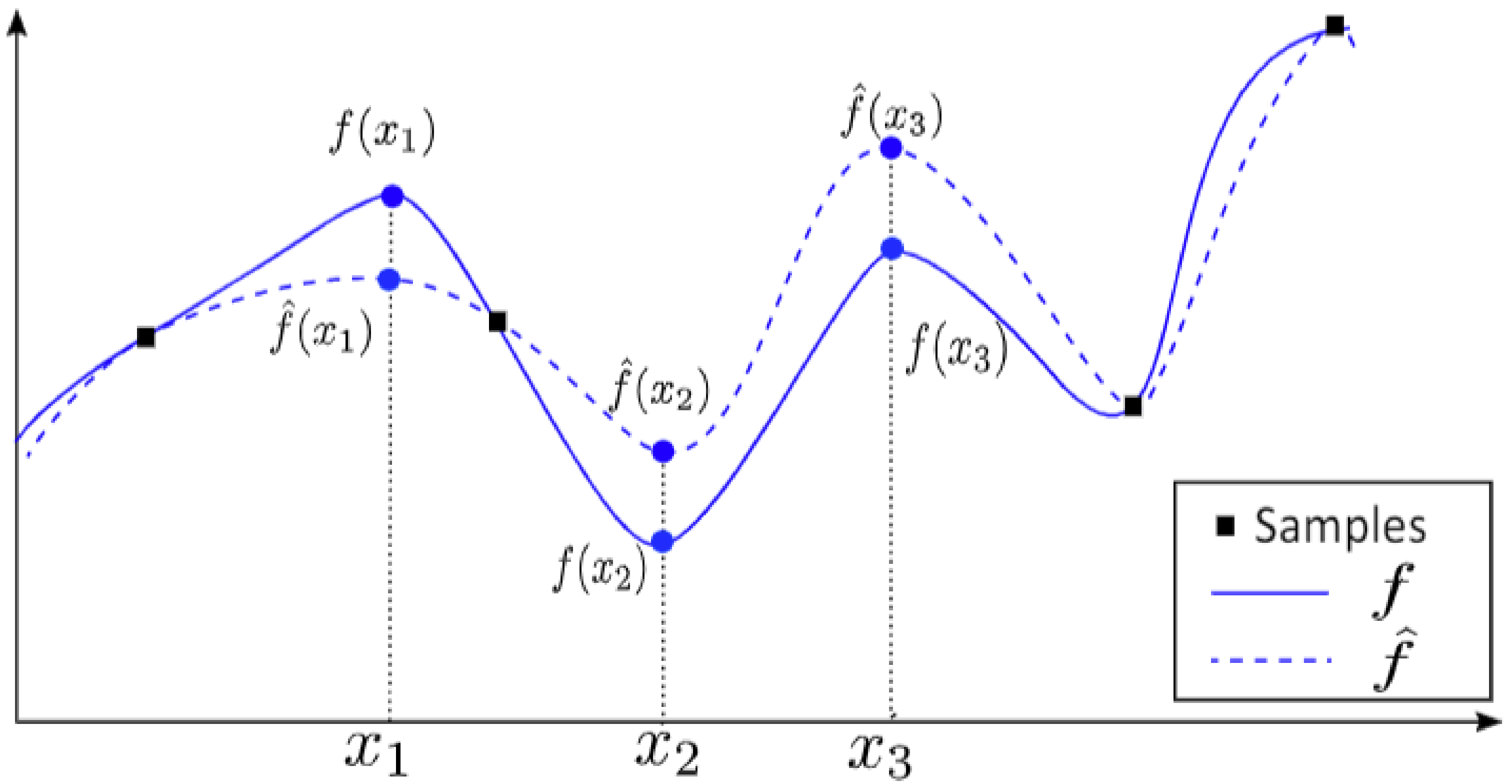


****No exact solution evaluation needed**

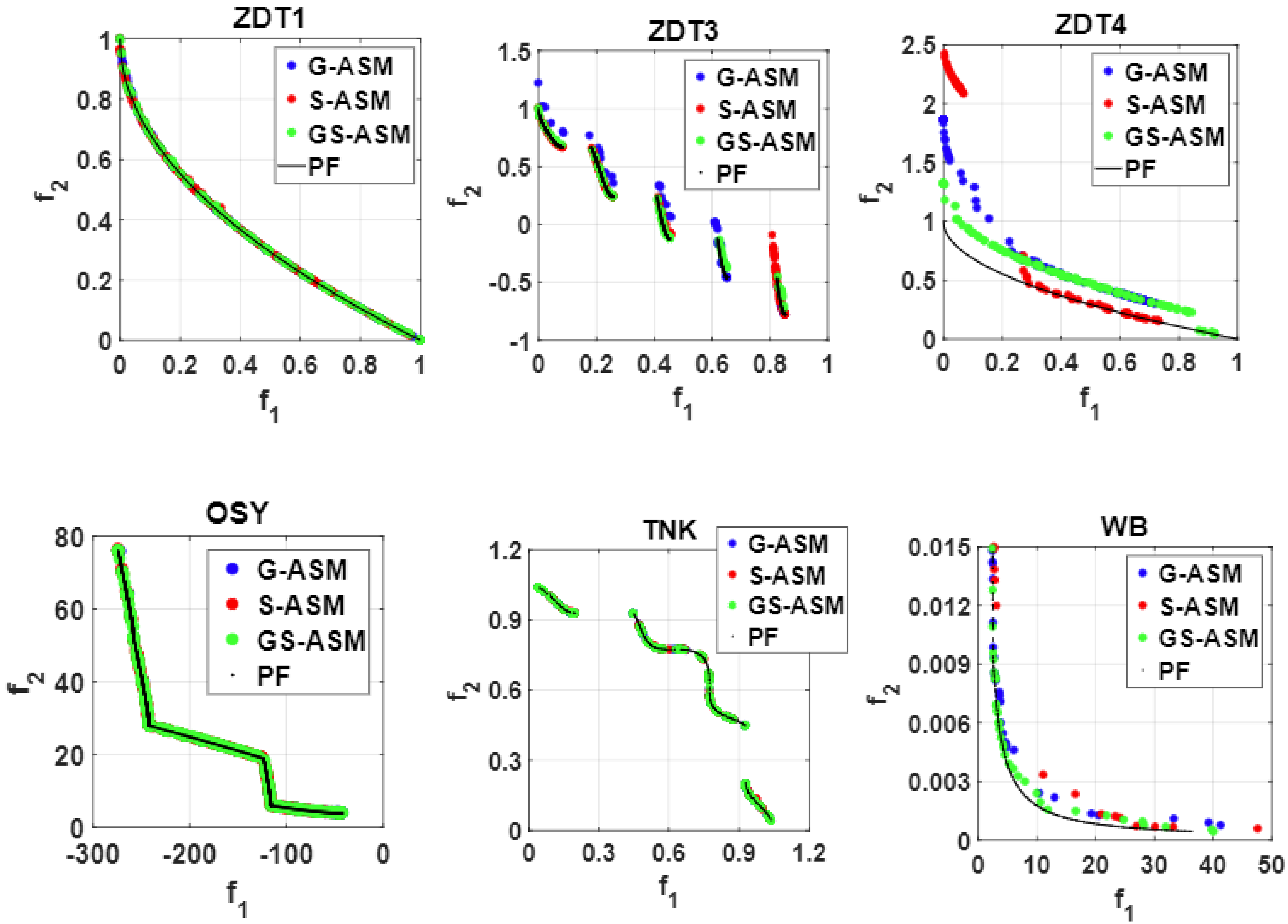
If **r1** =**r2,** no error, otherwise error

$$\text{Selection Error Probability (SEP)} = \frac{Error\ Count}{\binom{|Test\ Data|}{2}}$$

Selection Error Probability: Pairwise comparison between high-fidelity and prediction values (metamodeling)



Results of Adaptive Switching Method



Median IGD run for ZDT3 test problem

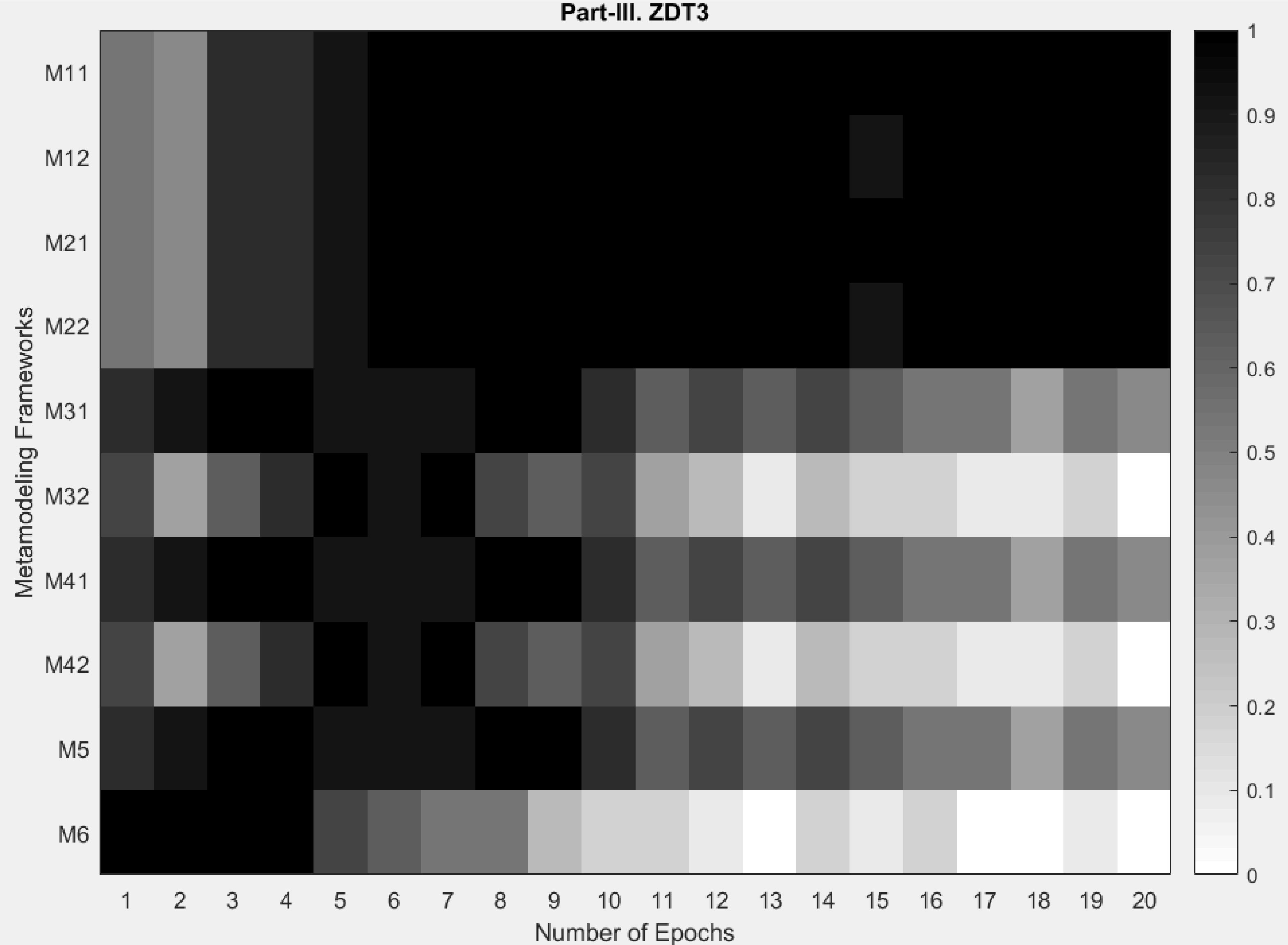
MICHIGAN STATE UNIVERSITY

	ZDT3-500 FEs ,IGD: 0.0039823, GD: 0.0015807																			
Frameworks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
M1-1																				
M1-2																				
M2-1																				
M3-1																				
M4-1																				
M5																				

	ZDT3-500 FEs ,IGD: 0.0039823, GD: 0.0015807																			
Frameworks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
M1-2																				
M2-2																				
M3-2																				
M4-2																				
M6																				

	ZDT3-500 FEs ,																			
Frameworks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
M1-1																				
M1-2																				
M2-1																				
M2-2																				
M3-1																				
M3-2																				
M4-1																				
M4-2																				
M5																				
M6																				

Mean run for ZDT3 test problems: Part-III



Results of IGD for Adaptive Switching Method

Problem		MOEA/D-EGO	K-RVEA	CSEA	GS-ASM
ZDT1		0.05611	0.07964	0.95330	**0.00130**
		p=8.1e-5	p=8.1e-5	p=8.1e-5	p=0.0910
ZDT2		0.04922	0.03395	1.01060	**0.00055**
		p=8.1e-5	p=8.1e-5	p=8.1e-5	-
ZDT3		0.30380	0.02481	0.94840	**0.00391**
		p=8.1e-5	p=8.1e-5	p=8.1e-5	-
ZDT4		73.25920	4.33221	12.71600	**0.39992**
		p=8.1e-5	p=8.1e-5	p=8.1e-5	-
ZDT6		0.51472	0.65462	5.42620	**0.24440**
		p=8.1e-5	p=8.1e-5	p=8.1e-5	p= 0.0612
DTLZ2		0.33170	0.0548	0.11420	**0.03701**
		p=8.1e-5	p=8.1e-5	p=8.1e-5	p=0.157
DTLZ4		0.64533	**0.0449**	0.08110	0.07934
		p=8.1e-5	-	p=0.0022	p=0.0380
DTLZ5		0.26203	0.0164	0.03081	**0.01252**
		p=8.1e-5	p=8.1e-5	p=8.1e-5	p=0.211
DTLZ7		5.33220	**0.0531**	0.70520	**0.06529**
		p=8.1e-5	-	p=8.1e-5	p=0.1930

Median IGD on unconstrained problems using GS-ASM and MOEA/D-EGO, K-RVEA, and CSEA algorithms.

Conclusions

- Trust regions are used as a constraint in the variable space during optimization to deal with uncertainties of metamodels.
- Proposed two performance indicators based on ASF & Hypervolume to adapt trust regions.
- A constraint handling scheme is presented to handle the trust region adaptation for constrained problems
- A multiple trust regions implemented with multiple trade-off solutions.
- Our results on several test multiobjective optimization problems have shown that we can achieve better convergence using the proposed method than that without a trust region.

Reference

- "A Taxonomy for Metamodeling Frameworks for Evolutionary Multiobjective Optimization"- K. Deb, R. Hussein, PC Roy, G. Toscano-Pulido
- "Adaptive Switching Strategy for Metamodeling Based Multi-objective Optimization: Part I, Generative Frameworks" R. Hussein, K. Deb and PC Roy
- "Adaptive Switching Strategy for Metamodeling Based Multi-objective Optimization: Part II, Simultaneous and Combined Frameworks"- PC Roy, R. Hussein, K. Deb
- Github Repo: https://github.com/proteekroy

Questions and Comments?