Multi-objective Bayesian Optimisation

Tinkle Chugh

Department of Computer Science
University of Exeter, UK

In this lecture

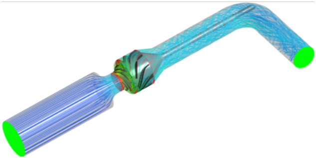
- Exploitation vs Exploration in Multi-objective BO
- Acquisition functions in multi-objective BO
- Mono-surrogate approach and Multi-surrogate approach

Expensive Multi-objective Optimization

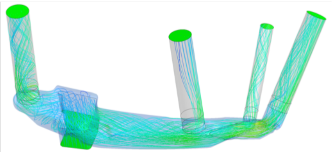
A Multi-objective optimization problem:

$$\text{minimize } (f_1(\mathbf{x}), \dots f_M(\mathbf{x}))$$
$$\text{subject to } \mathbf{x} \in \mathcal{X}$$

- Time consuming simulations
- One evaluation - hours to days
- Need to find solutions in few evaluations



Centrifugal pump



Air intake ventilation system

Multi-objective Bayesian Optimisation

Problem

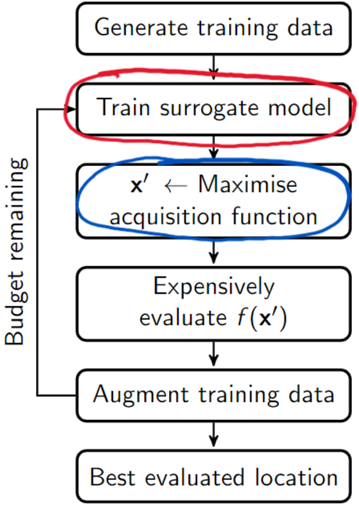
$$\min_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x}) \qquad \mathbf{x}\in\mathcal{X}\subset\mathbb{R}^d$$

Bayesian optimisation

- Build a probabilistic surrogate model of $f(\mathbf{x})$
- Maximise an *acquisition* function $\alpha(\mathbf{x})$

$$\mathbf{x}' \leftarrow \underset{\mathbf{x}\in\mathcal{X}}{\operatorname{argmax}}\ \alpha(\mathbf{x})$$

  Balances exploitation with exploration
- Expensively evaluate $f(\mathbf{x}')$



Mono-surrogate approach

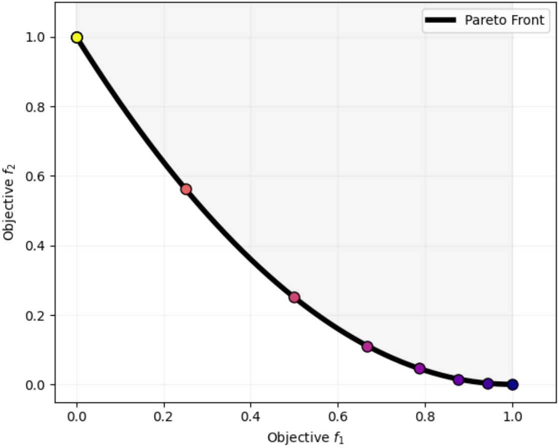
Model building in multi-objective BO

One model after aggregating objective functions
Weighted sum:

$$g(\mathbf{x}) = w_1 \cdot f_1(\mathbf{x}) + w_2 \cdot f_2(\mathbf{x}) + \cdots$$

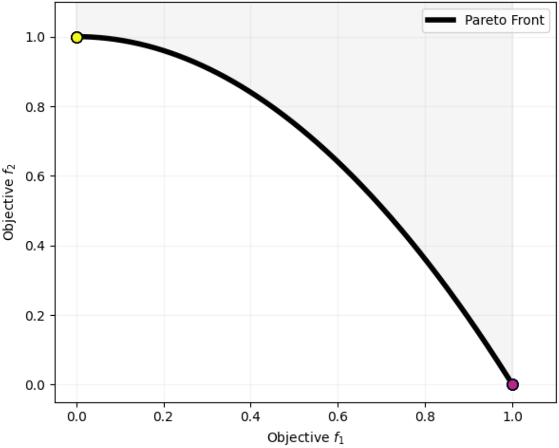
- Easy to implement and understand
- Can not handle non-convex multi-objective optimisation problems
- Small change in weights can change solution dramatically

A Toy Example



Performance of weighted sum on convex Pareto front

A Toy Example



Performance of weighted sum on concave Pareto front

Mono-surrogate approach

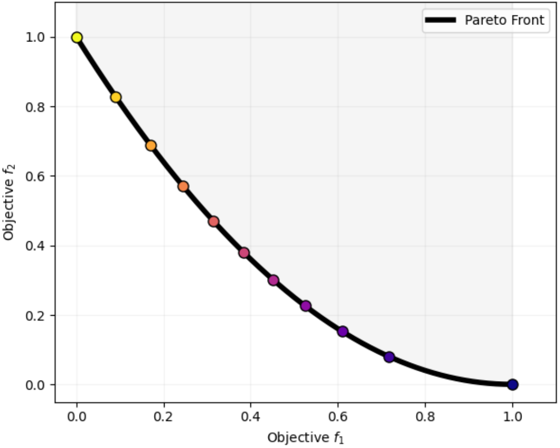
Model building in multi-objective BO

One model after aggregating objective functions
Weighted Tchebycheff:

$$g(\mathbf{x}) = \max(w_1(f_1(\mathbf{x}) - z_1), w_2(f_2(\mathbf{x}) - z_2), \cdots)$$

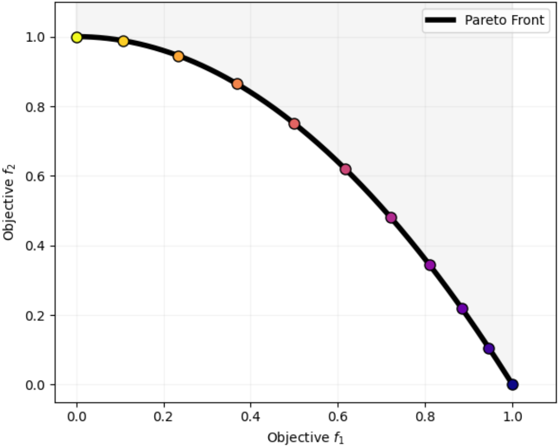
- Easy to implement and understand
- Can handle non-convex multi-objective optimisation problems

A Toy Example



Performance of weighted Tchebycheff on convex Pareto front

A Toy Example



Performance of weighted Tchebycheff on concave Pareto front

Mono-surrogate approach

ParEGO [1] uses Weighted Tchebycheff and is a mono-surrogate approach
For a weight vector $\mathbf{w}$

- Build a $\mathcal{GP}$ model on weighted Tchebycheff
- Maximise expected improvement

1 J. Knowles. ParEGO: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. IEEE Transactions on Evolutionary Computation 10, 50-66, 2006

Mono-surrogate approach

ParEGO [1] uses Weighted Tchebycheff and is a mono-surrogate approach
For a weight vector $\mathbf{w}$

- Build a $\mathcal{GP}$ model on weighted Tchebycheff
- Maximise expected improvement

1 J. Knowles. ParEGO: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. IEEE Transactions on Evolutionary Computation 10, 50-66, 2006

Advantages

- A single model instead of multiple models
- A single objective infill criterion

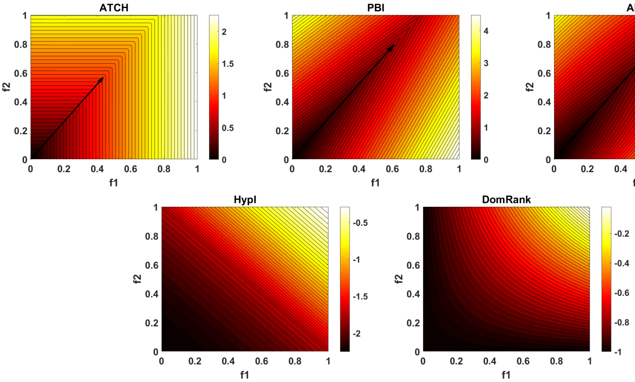
Mono-surrogate approach

Scalarising functions in Mono-surrogate approach [TC3]:

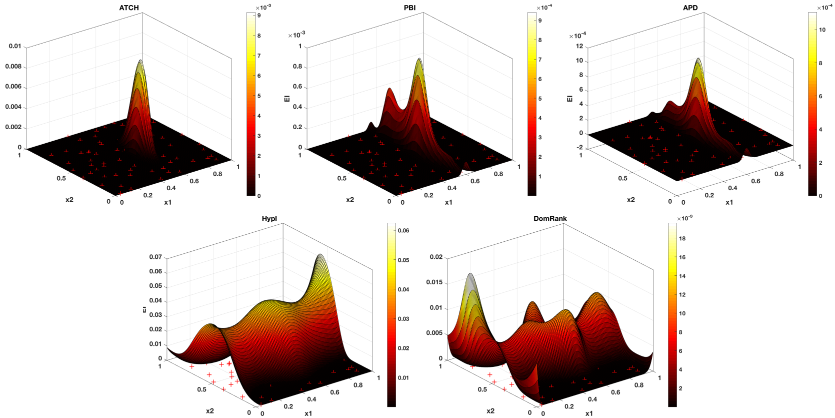
1 Weighted sum
2 Augmented Tchebycheff
3 Penalty boundary intersection
4 Angle penalised distance
5 Hypervolume improvement
6 Tchebycehff
7 Modified Tchebycheff
8 Inverted penalty boundary intersection
9 Quadratic penalty boundary intersection
10 Exponential weighted criterion
11 Weighted power
12 Weighted norm
13 Weighted product
14 Dominance ranking
15 Minimum signed distance

TC3 **T. Chugh**. Scalarizing functions in Bayesian multiobjective optimization. In Proceedings of the IEEE Congress on Evolutionary Computation (CEC), pages 1–8. IEEE, 2020

Mono-surrogate approach

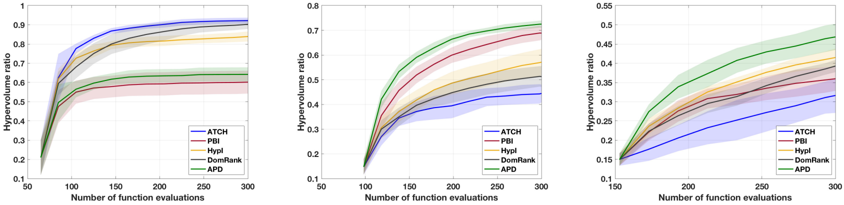


Insights into the results



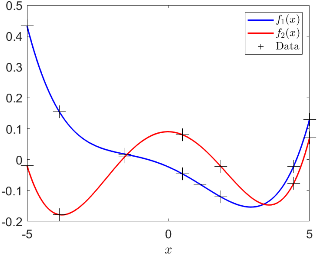
Contour plots on Expected improvement two objectives DTLZ2 using five scalarising functions

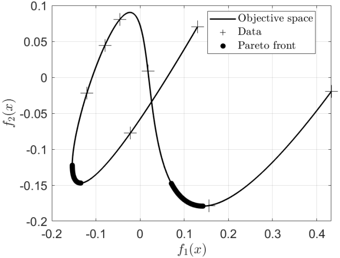
Mono-surrogate approach

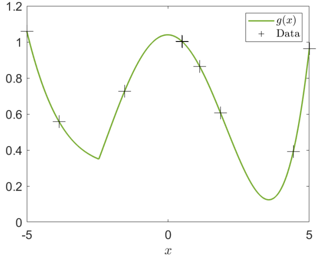


Hypervolume ratio on DTLZ2 - 2, 5 and 10 objectives

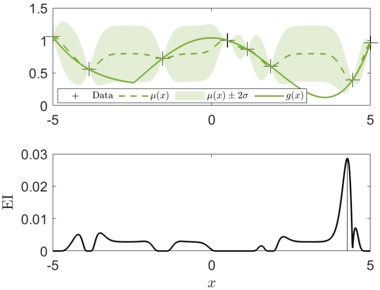
Mono-surrogate approach

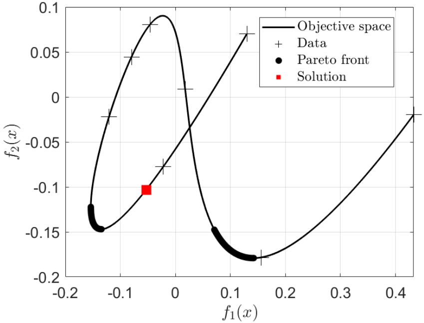






Mono-surrogate approach





Multi-surrogate approach

Mono or Multi-surrogate?

Weighted Tchebycheff

$$g = \max_i \left(w_i(f_i - z_i) \right),$$

Multi-surrogate approach

Mono or Multi-surrogate?

Weighted Tchebycheff

$$g = \max_i \left(w_i(f_i - z_i)\right),$$

In multi-surrogate

$$f_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$$

Multi-surrogate approach

Mono or Multi-surrogate?

Weighted Tchebycheff

$$g = \max_i \left(w_i(f_i - z_i) \right),$$

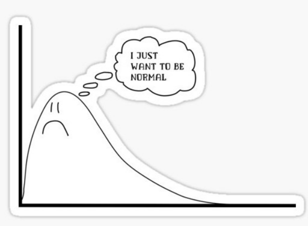
In multi-surrogate

$$f_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$$

$$g \sim \max_i \mathcal{N}\left(w_i(\mu_i - z_i), w_i^2 \sigma_i^2 \right) \tag{1}$$

Multi-surrogate approach

$$p(g) = \sum_{i=1}^{m} \frac{1}{w_i \sigma_i} \times \frac{\phi\left(\frac{g - w_i(\mu_i - z_i)}{w_i \sigma_i}\right)}{\Phi\left(\frac{g - w_i(\mu_i - z_i)}{w_i \sigma_i}\right)} \times \prod_{i=1}^{m} \Phi\Big(\frac{g - w_i(\mu_i - z_i)}{w_i \sigma_i}\Big),$$



- The distribution is NOT Gaussian
- A closed-form expression of the acquisition function cannot be used

Multi-surrogate approach

$$g \sim \text{Gumbel}\ (\alpha, \beta)$$

Gumbel distribution:

- The Generalised extreme value distribution Type 1
- Model for maximum (or the minimum) of a number of samples of distributions

Multi-surrogate approach

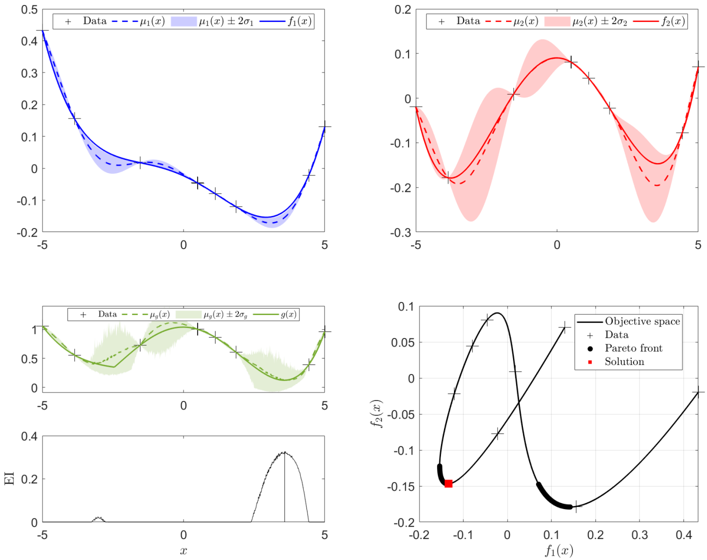
$$\alpha, \beta = \underset{\alpha,\beta}{\operatorname{argmax}} \prod_{i=1}^{N} p(g_i|\alpha, \beta) \text{ Or}$$

$$\alpha, \beta = \underset{\alpha,\beta}{\operatorname{argmax}} \left(N \log \frac{1}{\beta} - \sum_{i=1}^{N} t_i - \sum_{i=1}^{N} e^{-t_i} \right)$$

where $t_i = \frac{g_i - \alpha}{\beta}$

- No closed form expression for EI
- Laplace approximation for $p(g)$
- Monte Carlo approximation for EI

Multi-surrogate approach



Multi-Surrogate with Hypervolume

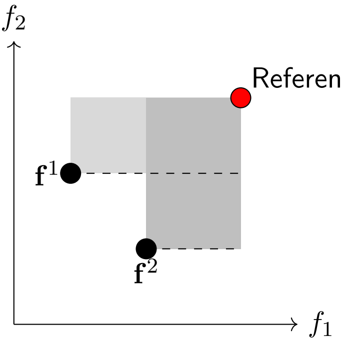
To evaluate a non-dominated Pareto front approximation without simplifying parameters (like weight vectors), we use the *Hypervolume Indicator* ($\mathcal{HV}$).

Definition

Given a non-dominated set $\mathbf{P}$ and a reference point $\mathbf{r} \in \mathbb{R}^M$ that is worse than all points in $\mathbf{P}$:

$$\mathcal{HV}(\mathbf{P}, \mathbf{r}) = \Lambda \left(\bigcup_{\mathbf{x} \in \mathbf{P}} [\mathbf{f}(\mathbf{x}), \mathbf{r}] \right)$$

where $\Lambda(\cdot)$ denotes the standard Lebesgue measure (volume) in $\mathbb{R}^M$.



Shaded multi-objective region bounded by Pareto front $\mathbf{P}$ and reference point $\mathbf{r}$.

Multi-surrogate with Hypervolume

- Measures both **convergence** (closeness to ideal front) and **diversity** (spread of solutions).
- Strictly Pareto-compliant: $\mathbf{P}_A \succ \mathbf{P}_B \implies \mathcal{HV}(\mathbf{P}_A) > \mathcal{HV}(\mathbf{P}_B)$.

Transition to Uncertainty: Defining EHVI

When individual objectives are modeled independently by Gaussian Processes ($f_m \sim \mathcal{GP}$), a candidate point $\mathbf{x}'$ yields a predictive joint distribution $P(\mathbf{f}(\mathbf{x}')) = \mathcal{N}(\mathbf{m}(\mathbf{x}'), \mathbf{\Sigma}(\mathbf{x}'))$.

Hypervolume Improvement (HI)

The marginal improvement added by a new candidate vector $\mathbf{y} = \mathbf{f}(\mathbf{x}')$ to the current non-dominated archive $\mathbf{P}$ is:

$$\mathsf{HI}(\mathbf{y}, \mathbf{P}) = \mathcal{HV}(\mathbf{P} \cup \{\mathbf{y}\}, \mathbf{r}) - \mathcal{HV}(\mathbf{P}, \mathbf{r})$$

Transition to Uncertainty: Defining EHVI

When individual objectives are modeled independently by Gaussian Processes ($f_m \sim \mathcal{GP}$), a candidate point $\mathbf{x}'$ yields a predictive joint distribution $P(\mathbf{f}(\mathbf{x}')) = \mathcal{N}(\mathbf{m}(\mathbf{x}'), \mathbf{\Sigma}(\mathbf{x}'))$.

Hypervolume Improvement (HI)

The marginal improvement added by a new candidate vector $\mathbf{y} = \mathbf{f}(\mathbf{x}')$ to the current non-dominated archive $\mathbf{P}$ is:

$$\mathsf{HI}(\mathbf{y}, \mathbf{P}) = \mathcal{HV}(\mathbf{P} \cup \{\mathbf{y}\}, \mathbf{r}) - \mathcal{HV}(\mathbf{P}, \mathbf{r})$$

Expected Hypervolume Improvement (EHVI)

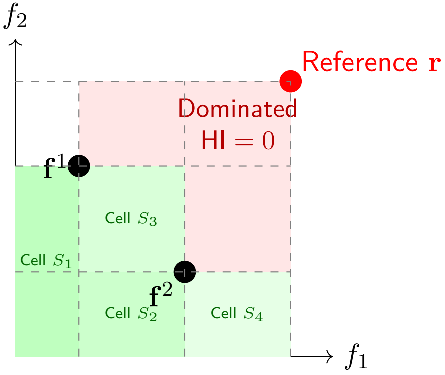
We integrate this deterministic metric across our entire multi-variate GP predictive density profile:

$$\mathsf{EHVI}(\mathbf{x}') = \int_{\mathbf{y} \in \mathbb{R}^M} \mathsf{HI}(\mathbf{y}, \mathbf{P}) \cdot \prod_{m=1}^{M} \frac{1}{\sigma_m(\mathbf{x}')} \phi\left(\frac{y_m - \mu_m(\mathbf{x}')}{\sigma_m(\mathbf{x}')}\right) d\mathbf{y}$$

Visualizing EHVI: Cell Partitioning Geometry

To evaluate the integral, the objective space is split into rectangular non-overlapping grid cells bounded by the current non-dominated points and the reference point $\mathbf{r}$.

- **Dominated Zone**: If $\mathbf{y}$ lands here (above/right of $\mathbf{f}^1$ or $\mathbf{f}^2$), it is worse than the current data, so HI$(\mathbf{y}, \mathbf{P}) = 0$.
- **Outside Zone**: If $\mathbf{y}$ lands beyond $\mathbf{r}$, it offers no valid improvement (HI $= 0$).
- **Improvement Cells (S_k)**: If $\mathbf{y}$ lands in any cell closer to the origin than $\mathbf{r}$ (and not already dominated), it expands the hypervolume.



Calculating Exact Single-Cell Integration

For a single partitioned cell bounded by lower bounds $\mathbf{l}$ and upper bounds $\mathbf{u}$, the independent Gaussian objective assumptions allow us to analytically split the integration:

Analytical Solution for Cell $S_k = [\mathbf{l}, \mathbf{u}]$

$$\mathbb{E}[\mathsf{HI}_{S_k}(\mathbf{y})] = \prod_{m=1}^{M} \Psi_m(l_m, u_m, \mu_m, \sigma_m)$$

where the linear scaling component function Ψ_m is defined as:

$$\Psi_m(l, u, \mu, \sigma) = \sigma \left[\phi\left(\frac{l-\mu}{\sigma}\right) - \phi\left(\frac{u-\mu}{\sigma}\right)\right] + (\mu - l)\left[\Phi\left(\frac{u-\mu}{\sigma}\right) - \Phi\left(\frac{l-\mu}{\sigma}\right)\right]$$

- $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard Normal PDF and CDF.
- No stochastic sampling or weight parameters are required for 2D space.

The Practical Bottleneck: Dimensional Scaling

While analytically clean, calculating exact EHVI faces computational limitations as the number of objectives increases.

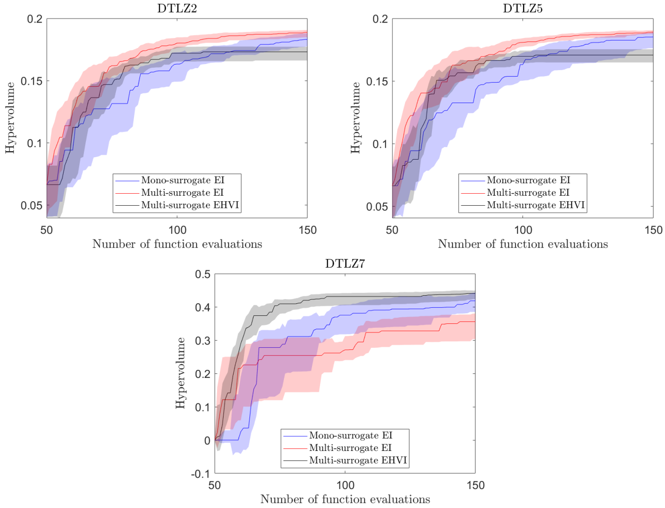
Objectives (M)	Cell Partitioning Complexity	Primary Solution Method
$M = 2$	$O(N \log N)$ (Extremely Fast)	Exact Analytical Integration
$M = 3$	$O(N^2)$	Exact Analytical Integration
$M \geq 4$	NP-hard ($O(N^M)$ growth profile)	Approximations / Monte Carlo

Computational complexity scale relative to the number of archive points N.

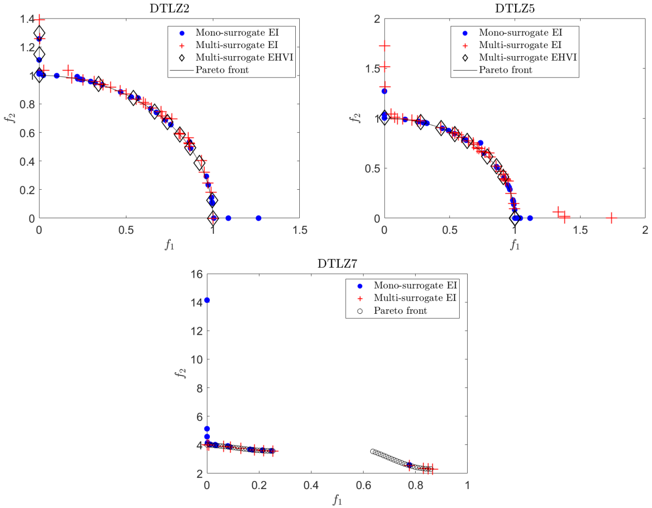
Modern Alternatives for Many-Objective Problems

- **q-EHVI (Daulton et al., 2020)**: Uses auto-differentiable Monte Carlo sampling via Box-Decomposition, enabling parallel batch evaluation ($q > 1$).

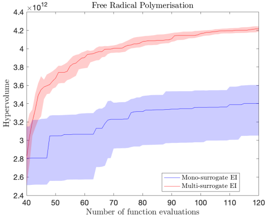
Numerical Experiments

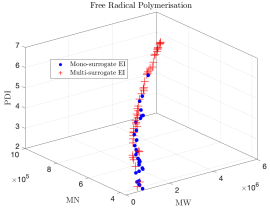


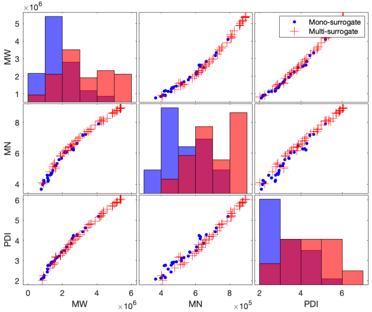
Numerical Experiments



Numerical Experiments

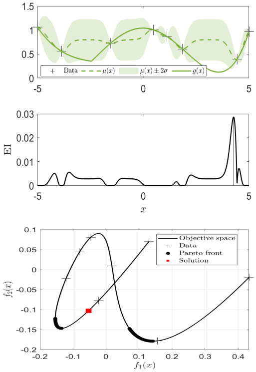




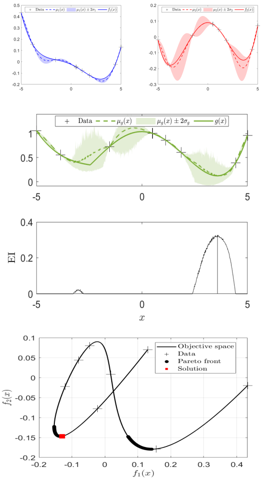


Mono- vs Multi-surrogte in BO

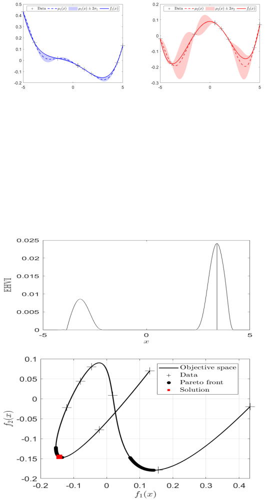
Mono-surrogate - EI



Multi-surrogate - EI



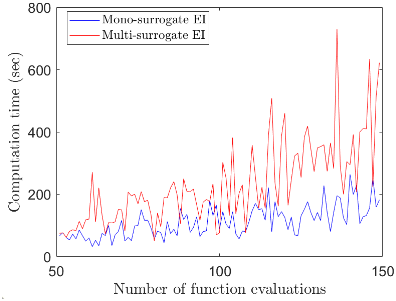
Multi-surrogate - EHVI



Summary of Mono vs Multi-Surrogate

- Mono-surrogate approach rely on the model of the scalarising function
  - May not be an accurate model
  - A Gaussian assumption may not be a viable choice
- Multi-surrogate approach rely on the model of the objective functions
  - Accurate as it doesn't depend on the model of the latent function
  - Does not assume Gaussian assumption on the latent function

- Multi-surrogate approach is time-consuming



- Laplace approximation

Summary of the course

- (Multi-objective) Optimisation is everywhere
- Most real-world problems have multiple objectives to be achieved
- Machine learning can help in alleviating the computation cost

What We Did Cover

- Multi-objective Optimisation - Pareto optimality
- Gaussian Processes
- Single-objective Bayesian Optimisation
- Multi-objective Bayesian Optimisation

What We Did NOT Cover

- Multi-task Machine Learning
    - Correlation among objectives
- Use of other Bayesian models
- Multi-objective Evolutionary Algorithms
- Performance metrics
- Constraint handling
- Robust Bayesian optimisation
- Parallel Bayesian optimisation