Predictive Entropy Search for Multi-objective Bayesian Optimization

UNIVERSIDAD AUTÓNOMA DE MADRID

Daniel Hernández-Lobato[1], José Miguel Hernández-Lobato[2], Amar Shah[3] and Ryan P. Adams [2]

[1]Universidad Autónoma de Madrid. [2]Harvard University. [3]University of Cambridge.

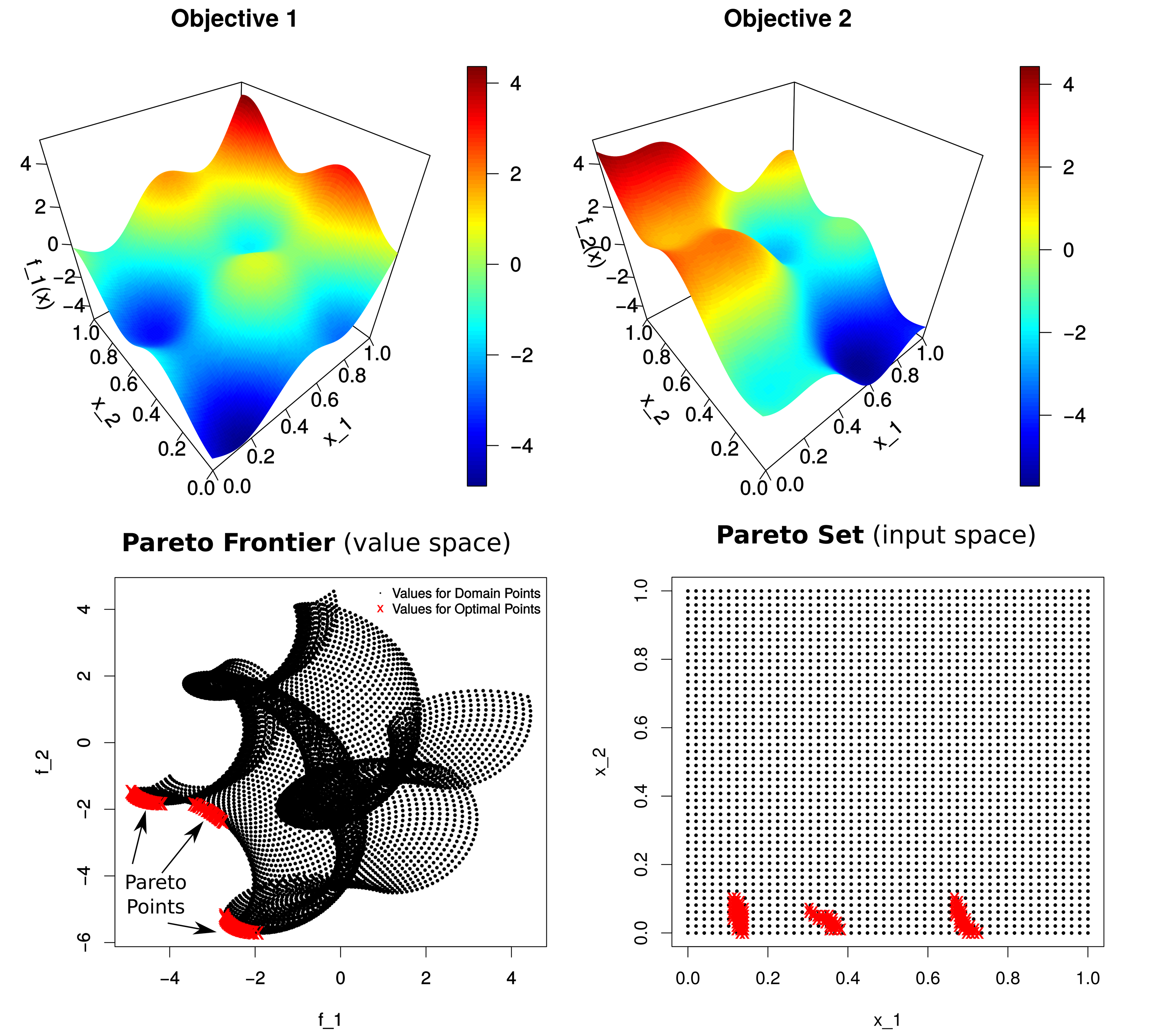
1. Introduction to Multi-objective Bayesian Optimization

We are interested in solving the **problem**:

$$\min_{\mathbf{x}\in\mathcal{X}} \quad f_1(\mathbf{x}),\ldots,f_K(\mathbf{x})\,.$$

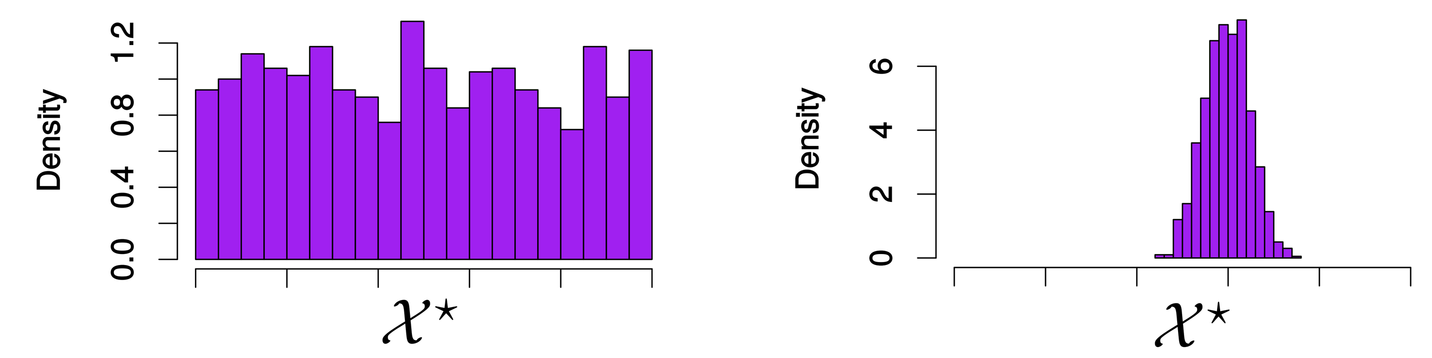
- Each $f_k(\cdot)$ is evaluated via expensive black-box queries.
- We select $\mathbf{x}$ and we observe output $\mathbf{y} = (f_1(\mathbf{x}),\ldots,f_K(\mathbf{x}))^\text{T}$
- The evaluations may be **contaminated** with Gaussian noise ϵ.

Example multi-objective optimization problem:



2. Predictive Entropy Search for Multi-objective Optimization (PESMO)

Given $\mathcal{D}_t = \{(\mathbf{x}_i,\mathbf{y}_i)\}_{i=1}^t$, we model each $f_k(\cdot)$ with a **Gaussian process**, and the **acquisition function** maximizes the reduction on the entropy of $\mathcal{X}^\star$.



The acquisition function is

$$\alpha(\mathbf{x}) = \text{H}[\mathcal{X}^\star|\mathcal{D}_t] - \mathbb{E}_\mathbf{y}\Big[\text{H}[\mathcal{X}^\star|\mathcal{D}_t\cup\{\mathbf{x},\mathbf{y}\}]\Big|\mathcal{D}_t,\mathbf{x}\Big] \quad (1)$$

How much we know about $\mathcal{X}^\star$ now.

How much we will know about $\mathcal{X}^\star$ after collecting y at $\mathbf{x}$.

Computing (1) is **very difficult in practice!**

We **swap y and $\mathcal{X}^\star$** to obtain a reformulation of the acquisition function.

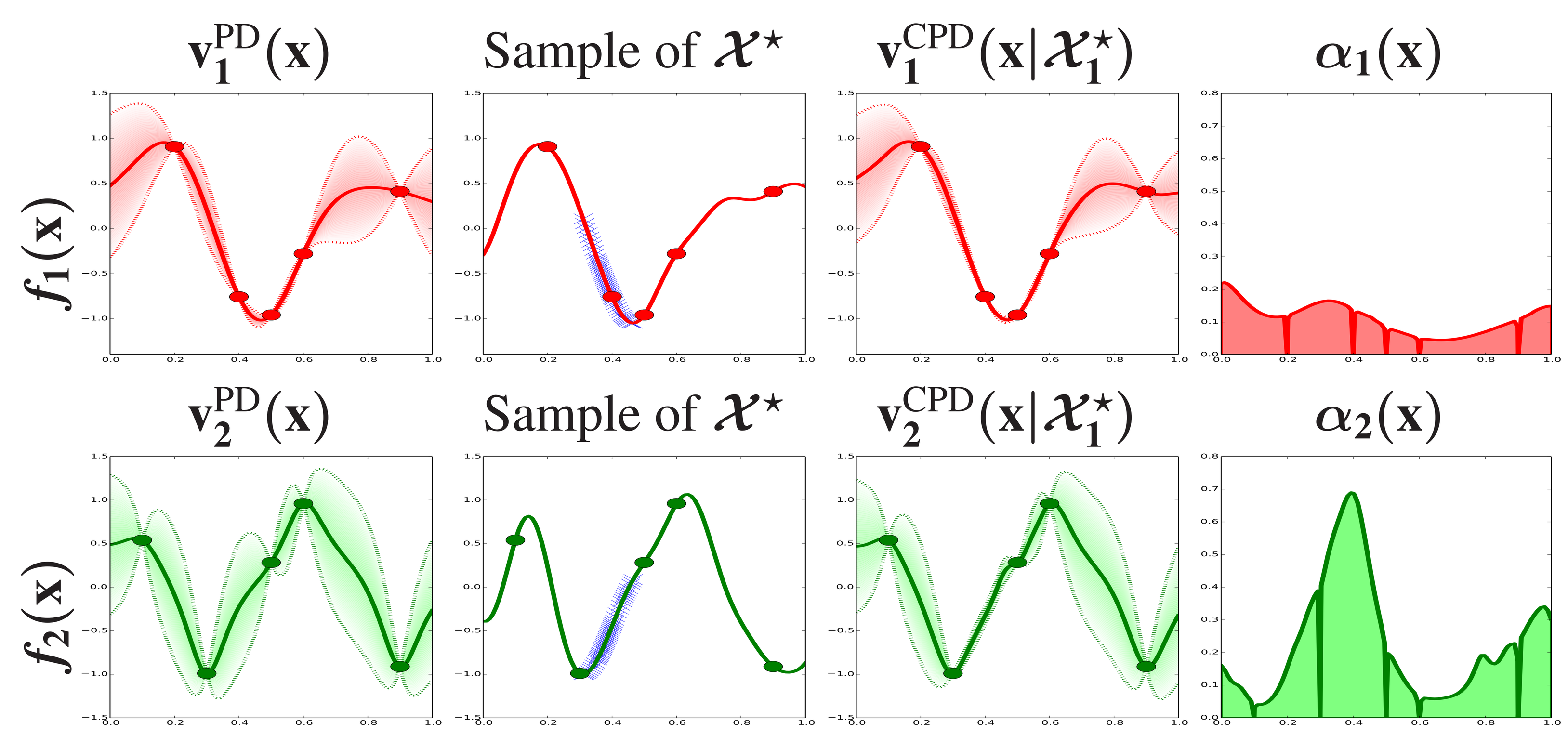
$$\text{H}[\mathcal{X}^\star|\mathcal{D}_t] - \mathbb{E}_\mathbf{y}\Big[\text{H}[\mathcal{X}^\star|\mathcal{D}_t\cup\{\mathbf{x},\mathbf{y}\}]\Big|\mathcal{D}_t,\mathbf{x}\Big] \equiv \text{MI}(\mathbf{y},\mathcal{X}^\star) \quad \text{(ESMO)}$$

$$\text{H}[\mathbf{y}|\mathcal{D}_t,\mathbf{x}] - \mathbb{E}_{\mathcal{X}^\star}\Big[\text{H}[\mathbf{y}|\mathcal{D}_t,\mathbf{x},\mathcal{X}^\star]\Big|\mathcal{D}_t,\mathbf{x}\Big] \equiv \text{MI}(\mathcal{X}^\star,\mathbf{y}) \quad \text{(PESMO)}$$

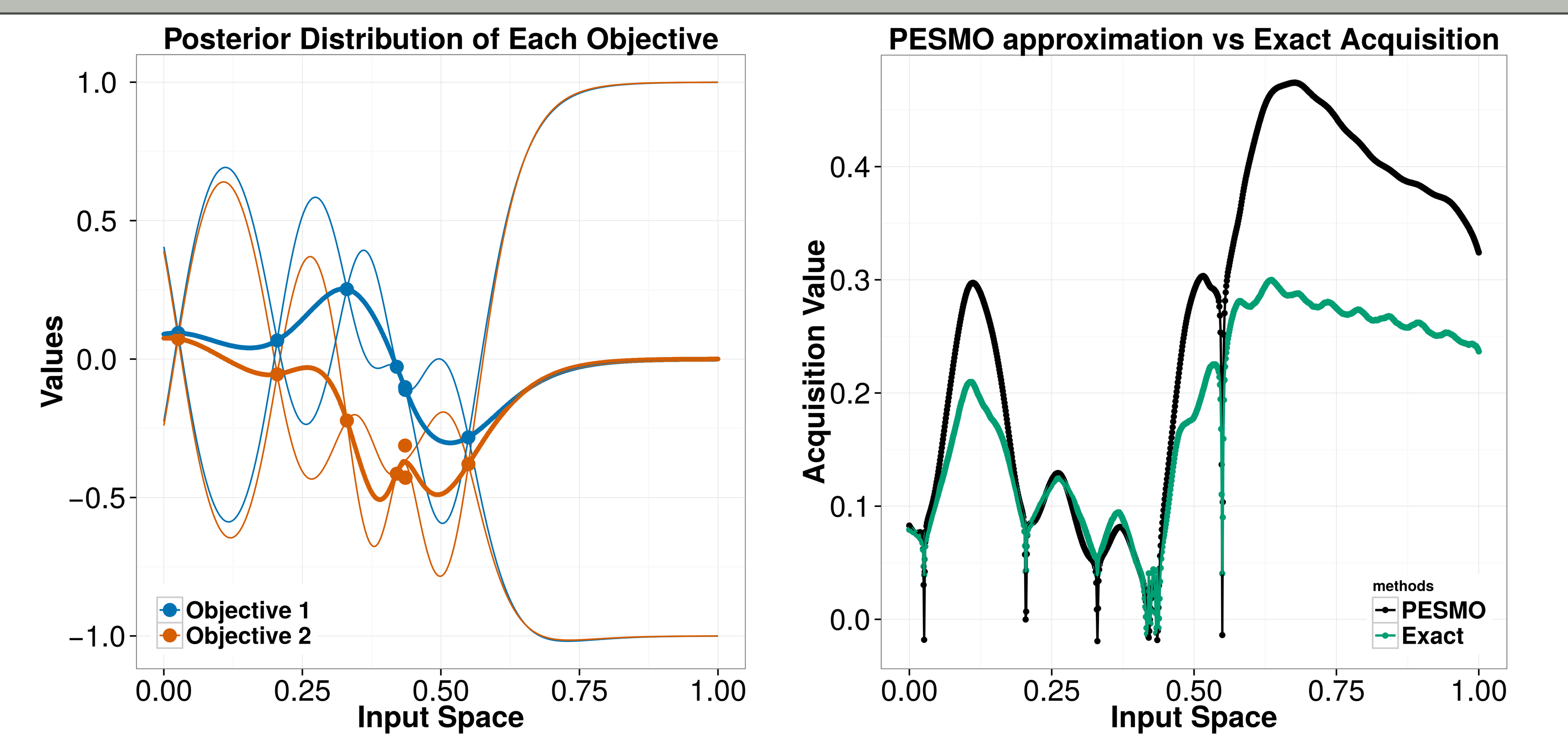
Approximated by sampling from $p(\mathcal{X}^\star|\mathcal{D}_t)$

Factorized Gaussian approximation with **expectation propagation**. $\mathcal{X}^\star$ dominates any other point in $\mathcal{X}$.

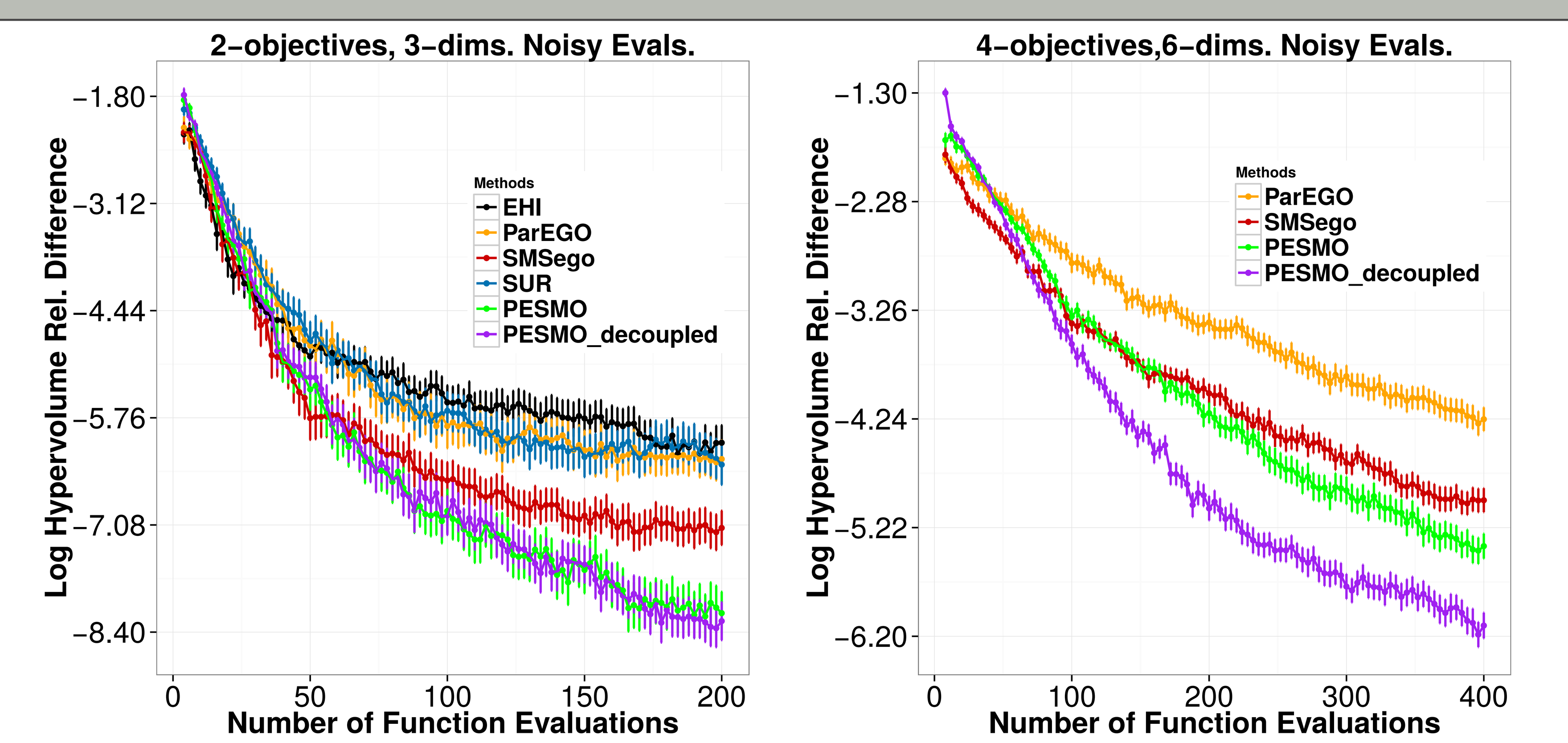
$$\alpha(\mathbf{x}) = \sum_{k=1}^K \left\{\log v_k^\text{PD}(\mathbf{x}) - \frac{1}{S}\sum_{s=1}^S \log v_k^\text{CPD}(\mathbf{x}|\mathcal{X}^\star_{(s)})\right\} = \sum_{k=1}^K \alpha_k(\mathbf{x})$$



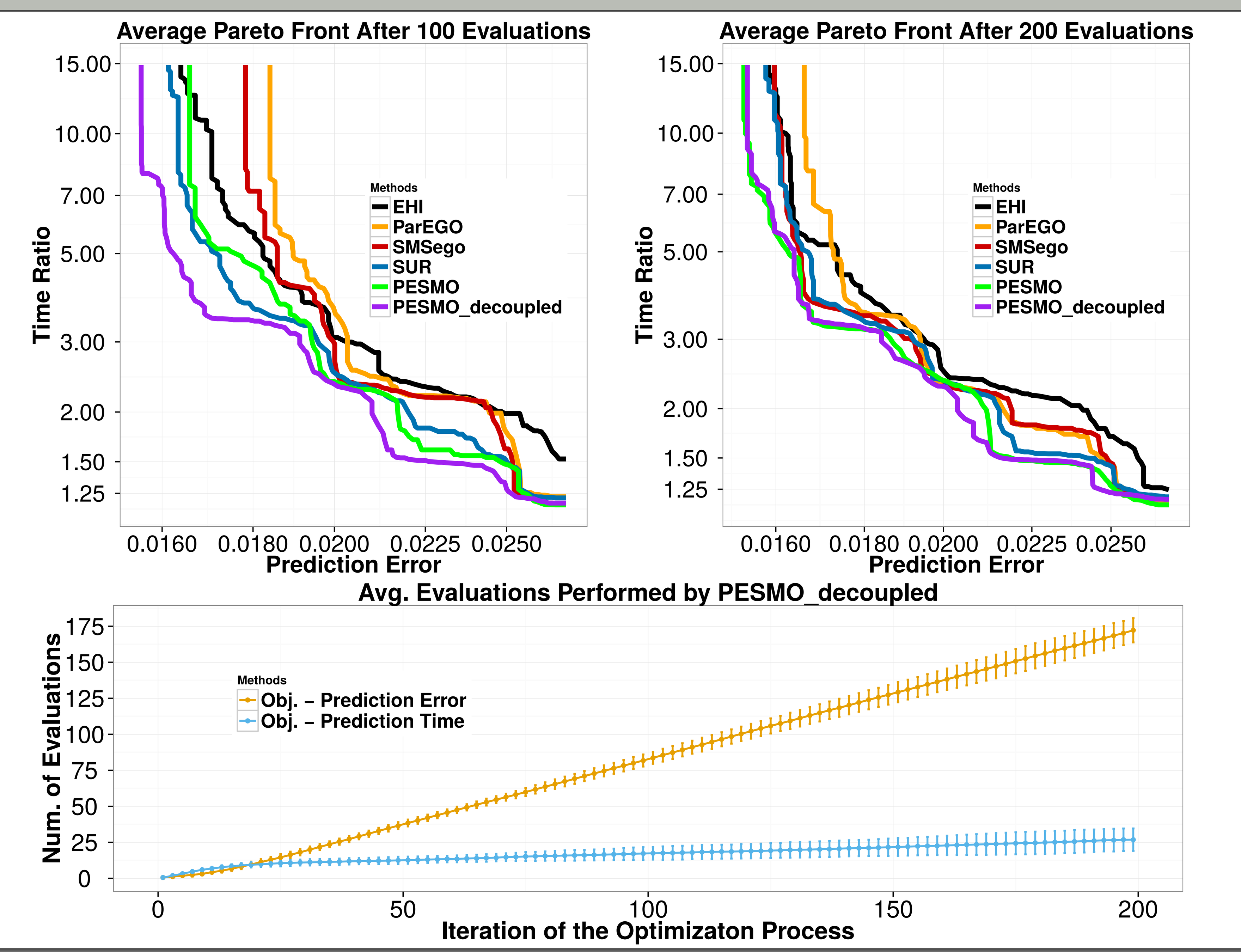
3. Accuracy of the EP Approximation



4. Experiments with Synthetic Data (objectives sampled from the GPs)



5. Finding Fast and Accurate Deep Neural Networks on MNIST



6. Conclusions

- We have described PESMO, a successful strategy to carry out Bayesian Optimization of very expensive-to-evaluate black-box functions.
- PESMO chooses the next location on which to evaluate the objectives as the one that is expected to minimize the most the entropy of the Pareto set.
- Our experiments show that PESMO has better performance than other already known strategies for multi-objective Bayesian optimization.
- PESMO allows for a decoupled evaluation scenario in which different objectives are evaluated at different input locations at each iteration.

http://dhnzl.org/ daniel.hernandez@uam.es