Bayesian Optimization : A New Sample Efficient Workflow for Reservoir Optimization under Uncertainty

Peyman Kor[*,a], Aojie Hong[a], Reidar Brumer Bratvold[a]

[a]*Energy Resources Department, University of Stavanger, Stavanger, Norway*

Abstract

Field development in E&P requires several level of decision making, where amoung the available options, the one with highest objective is decided. The decisions to make could be well configuration, well types , and well controls. The search to find th eoptimum decison, which called here optimization, a computionally demanding nonlinear problem. More challenging, in teh case of optimization under uncertainty, such problems requires large number of flow simulation given diffrent geological realizations. To alleviate this problem and to provide the fesible framework dealing with computional tcomplexity in essevor optimization, a new sworkflow is presented. The "Bayesian Optimization" in this work efficntley sample from the flow simulation based on the probalistc modeling of the objective function, asissting to perform optimization process with as few as possible flow simulation, without losing the optimum solution. The paper begins with introduction to gaussian process and building a probalistiv view of the objective function as "Surrogate" and then provides the foundation for acquistion function, which decide to the next query point from the exppensive function, here flow simulation. Further, a 1-D probelm will be optimized through the Bayesian optimization which helops to better understading of the workflow. Then, the workflow has been applied to 3D, "Egg model" rto perform optimization in the reaslistic field scenario, considering geological uncertainty. Here, the comparsion of the workflow with other two well-kow algorithm in reservoir optimization literature, namely Genetic Algorithm (GA) and Partcile Swarm Inteligence (PSO) has been performed. The results of the comparison show the the BAyesian Optimization workflow presented here could reach the same global optimum point achieved with GA and PSO, yet reduce computionaly complexity of the optimization 5X, which could b esignificant, in the case of the 3D flow simulation, potentially taking hours of running.

*Corresponding Author

**Equal contribution

Email addresses: peyman.kor@uis.no (Peyman Kor), aojie.hong@uis.no (Aojie Hong), reidar.bratvold.uis.no (Reidar Brumer Bratvold)

1. Introduction:

While developing a field, prediction of reservoir response to change in the variable is an am important factor to have an efficient reservoir management. The field of reservoir engineering is rich in the development and application of full-physics numerical simulations for forward modeling. However, the computational power needed to run such numerical simulators most of the time is huge. Especially, the framework of the Robust Optimization where uncertainty are considered through multiple of geological realization (thousand or multi-thousand), the practical applicability of such forward modeling is considerably limited. To address this challenge, the proxy-modeling for reservoir managemnet has emerged to reduce the cost of forward simulation. The root of this field goes back to the work of the (Bruce 1943) where the analogy between flow of electricty in tthorugh te electrical device and the response of the reservoir was constructed. (Albertoni and Lake 2003) Proposed the a Multivariate Linear Regression (MLR) to estimate the production from the well, where it claimed that total production at each is linear combination of injection rates with diffusitivity filter. Building on the work of (Albertoni and Lake 2003), (Yousef 2006) proposed a new procedure to quantify communication between wells, where for each injector/producer pair, two coefficients, capacitance to quantify connectivity and time constant to quantify the degree of fluid storage were considered. (Sayarpour 2008) used superposition in time to analytically solve the numerical integration in CRM. (Zhao et al. 2015) articulated that CRM models are not applicable when production rates change significantly, mainly due to those models neglect to include intearction of production-production and injector-injector pair well. Formulating the model in the framework titled INSIM (interwell numerical simulation model), it was used as efficient replacement to reservoir simulator when history-matching phase rate data or water-cut data obtained during water flooding operations. (White and Royer, n.d.)

Seperatley, in sake of utilization of recent advancemnet in the world of Information technology, couple of reaserach has been don eon the development of “Surrogate Reservoir Models” - (Mohaghegh and Guruswamy 2006) proposed the workflow for SRM model where Fuzzy Paatern Recognition (FPR) technology was dimensionality reduction, in both static and dynamic parameters of the reservir . Key Performance Indicator (KPI) was used to select the most important variable.- (Sampaio 2009) tried on use of feed-forward neural networks as nonlinear proxies of reservoir response. A few set of rock and fluid properties were considere a input (porosity, permeability in “x” direction, permeability in “y” direction, and rock compressibility) where in developing the neural network model, only one hidden layer was used. flow proxy modeling to replace full simulation in history matching, and built the proxy with a single hidden layer artificial neural network (ANN). To predict the oil production from the SAGD rocess, (Fedutenko et al. 2014) employed the RBF Neural Network to predict the cumulative oil production over the time horizon of production (10 years) .

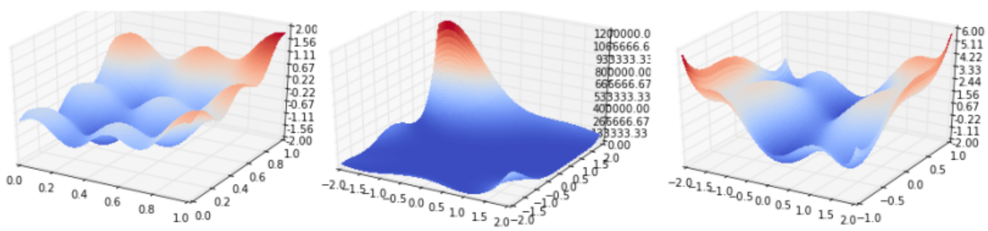
Survey of suurgate -based papers in onepetro database

2. Problem Statement

Consider a "well behaved" function $f : \chi \to \mathbb{R}$ where $\chi \subseteq \mathbb{R}^D$ is a bounded domain.

$$x_M = argmin\ f(x) \tag{1}$$

reference equation 1



f is explicitly unknown and multimodal. Evaluations of f are expensive. Expensive functions, who doesn't have one?

In reservoir optimization, we can have different objectives: Recovery Factor, Net Present Value (NPV), ...

$$u_M = \underset{u \in \text{constraints}}{\text{argmax}} \ \text{Objective Func(u)} \tag{2}$$

u is a decision variable to make: could be injection rate, well locations, order of drilling wells.

$$\text{Objective Func(u)} = J(u, G) = \sum_{k=1}^{n_T} \frac{q_o^k(u,G)P_o - q_w^k(u,G)P_{wp} - I^k(u,G)P_{wi}}{(1+b)^{t_k/D}} \tag{3}$$

$$\text{Objective Func(u)} = \overline{J}(u) = \frac{\sum_{i=1}^{n_e} J_r(u, G_i)}{n_e} \tag{4}$$

Example case:.

- Lets' say you are optimizng the well control problem with dimension, $d = 10$
- Number of 3D geological realizations is $n_e = 100$ and running each possible u^*in commercial reservoir simulator, at each realization talks ~ 1 hour. Then time it takes to calculate $\overline{J}(u^*)$ is ~ 100 hours.
- With 6 month CPU running time budget, the total available budget is to run is ~ 50.

- Availablity of only 50 $\overline{J}(u^*)$ evaluation is a small budget, considering the 10 dimentional optimization problem.

Problem $\overline{J}(u)$ is expensive function, meaning the # times we can evaluate it severely limited. Solution Bayesian Optimization (BayesOpt) propose a new workflow to conduct optimization at the small $\overline{J}(u)$ budget, without affecting the optimum solution.

3. Bayesian Optimization Workflow

Overal View

Workflow to perform global optimization of multimodal black-box functions:

- Step 1. Choose some initial design points and build a probabilistic model over the space of possible objective f, this probabilistic model serves as prior.
- Step 2. Combine prior and the likelihood to get a posterior of probabilistic model over the objective given some observations.
- Step 3. Use the posterior to decide where to take the next evaluation $\mathbf{x}^*$ according to some policy for decision making.
- Step 4. Evaluet the f at $\mathbf{x}^*$ and augment it to the initial data, in step 1.

Iterate between 2 and 4 until the evaluation budget is over. ## 3.1 Gaussian Process

Step 1. Probalistic Model as Prior

Gaussian Process (GP). reference to the book (**murphy2022?**)

Key Assumption in (GP) is that: the function values at a set of $M > 0$ inputs, $\mathbf{f} = [\mathbf{f}(\mathbf{x_1}), ..., \mathbf{f}(\mathbf{x_M})]$, is jointly Gaussian, with mean and Covariance

$$(\mu = m(x_1), ...m(x_M)) \sum_{i,j} = \kappa(x_i, x_j) \tag{5}$$

and κ is a positive definite (Mercer) kernel. Suppose we have a initial design points, set $\mathcal{D} = (x_n, y_n) : n = 1 : N$, where $y_n = f(x_n)$ is the noise-free observation of the function evaluated at x.

Now we consider the case of predicting the outputs for new inputs that may not be in $\mathcal{D}$.

$$\mathbf{f}_* = [\mathbf{f}(\mathbf{x_1}), .., \mathbf{f}(\mathbf{x_{N_*}})] \tag{6}$$

By definition of the GP, the joint distribution $p(\mathbf{f_X}, \mathbf{f}|\mathbf{X}, \mathbf{X}_*)$ has the following form:

$$\begin{bmatrix} \mathbf{f_X} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left((\begin{bmatrix} \mu_X \\ \mu_* \end{bmatrix}), \begin{bmatrix} \mathbf{K}_{X,X} & \mathbf{K}_{X,*} \\ \mathbf{K}^{\mathsf{T}}{}_{X,*} & \mathbf{K}_{*,*} \end{bmatrix} \right) \tag{7}$$

$$\begin{aligned} \mu_X &= [m(x_1), ..., m(x_N)] \\ \mu^* &= [m(x_1^*, ...m(x_N^*))] \end{aligned} \tag{8}$$

$$\begin{aligned} K_{X,X} &= \kappa(X, X; \theta), \qquad size(N \times N) \\ K_{X,*} &= \kappa(X, X_*; \theta), \qquad size(N \times N_*) \\ K_{*,*} &= \kappa(X_*, X_*; \theta), \qquad size(N_* \times N_*) \end{aligned} \tag{9}$$

Covariance Kernels	assumeing $h = \|\|x - x'\|\|$
Gaussain	$\kappa(x, x') = \sigma_f^2 exp(-\frac{h^2}{2\ell^2})$
Matern $\mu = \frac{5}{2}$	$\kappa(x, x') = \sigma_f^2(1 + \frac{\sqrt{5}\|h\|}{\ell} \frac{5h^2}{3\ell^2})exp(-\frac{-\sqrt{5}\|h\|}{\ell})$
Matern $\mu = \frac{3}{2}$	$\kappa(x, x') = \sigma_f^2(1 + \frac{\sqrt{3}\|h\|}{\ell})exp(-\frac{-\sqrt{3}\|h\|}{\ell})$
Exponetial	$\kappa(x, x') = \sigma_f^2 exp(-\frac{\|h\|}{\ell})$
Power-Exponetial	$\kappa(x, x') = \sigma_f^2 exp(-(\frac{\|h\|}{\ell})^p)$

$$\kappa(x,x';;\theta) = (1+\frac{\sqrt{5}|h|}{\theta}+\frac{5h^2}{3\theta^2})exp(-\frac{-\sqrt{5}|h|}{\theta} \tag{10}$$

Covariance Kernel, Parameter estimation.

$$p(y|\mathbf{X},\theta) = \int \mathbf{p(y|f,X)p(f|X,}\theta) \tag{11}$$

$$\log p(y|\mathbf{X},\theta) = \mathcal{L}(\zeta,\sigma_{\mathbf{f}}^{\mathbf{2}}) = -\frac{\mathbf{1}}{\mathbf{2}}(\mathbf{y}-\mu_{\mathbf{X}})^{\intercal}\mathbf{K}_{\mathbf{X},\mathbf{X}}^{-1}(\mathbf{y}-\mu_{\mathbf{X}}) - \frac{\mathbf{1}}{\mathbf{2}}\mathbf{log}|\mathbf{K}_{\mathbf{X},\mathbf{X}}| - \frac{\mathbf{n}}{\mathbf{2}}\mathbf{log(2}\pi) \tag{12}$$

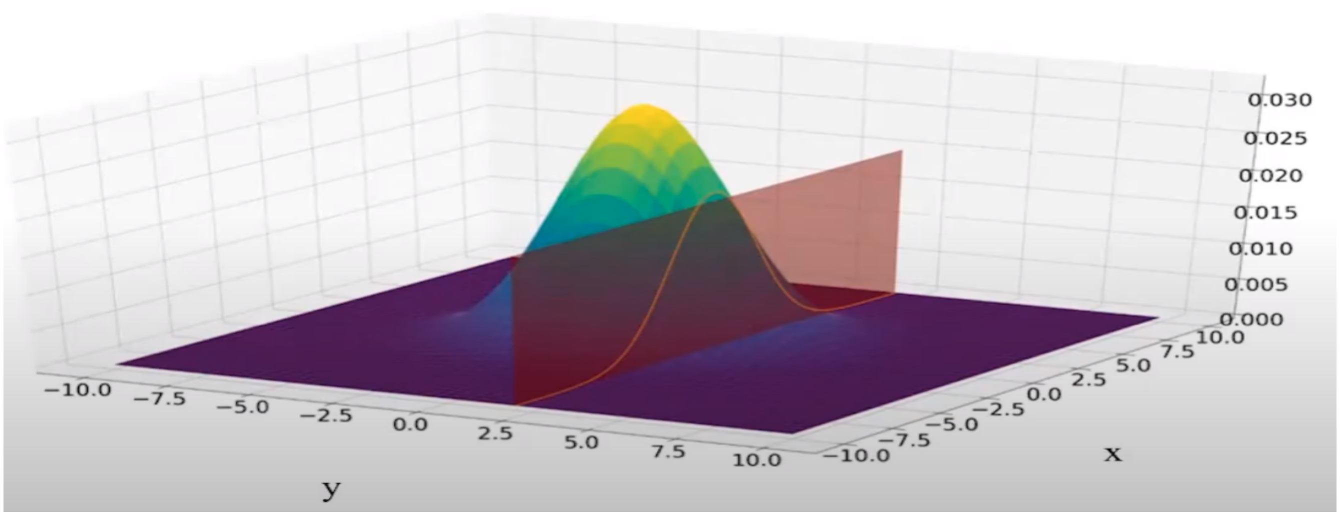
Where the dependence of the $\mathbf{K}_{\mathbf{X},\mathbf{X}}$ on θ is implicit.The gradient-based optimizer is performed in order to:

$$[\zeta^*,\sigma_f^{2*}] = argmax\mathcal{L}(\zeta,\sigma_f^2) \tag{13}$$

However, since the objective $\mathcal{L}$ is not convex, local minima can be a problem, so we may need to use multiple restarts.

Step 2. Posterior of Probabilistic Model

Posterior of Gaussain Process, (conditioning on initial data).



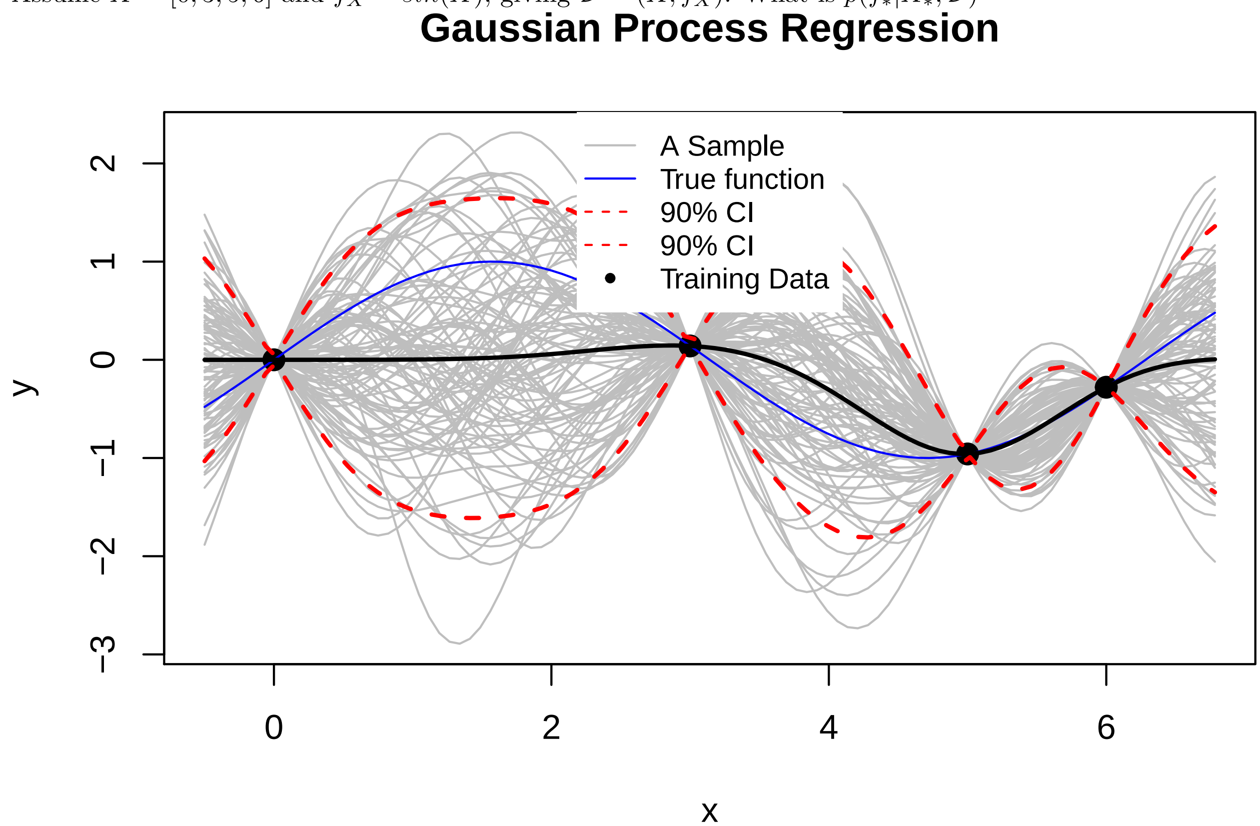
–

$$p(f_*|X_*,\mathcal{D}) = \mathcal{N}(f_*|\mu_*,\sum\nolimits_*) \tag{14}$$

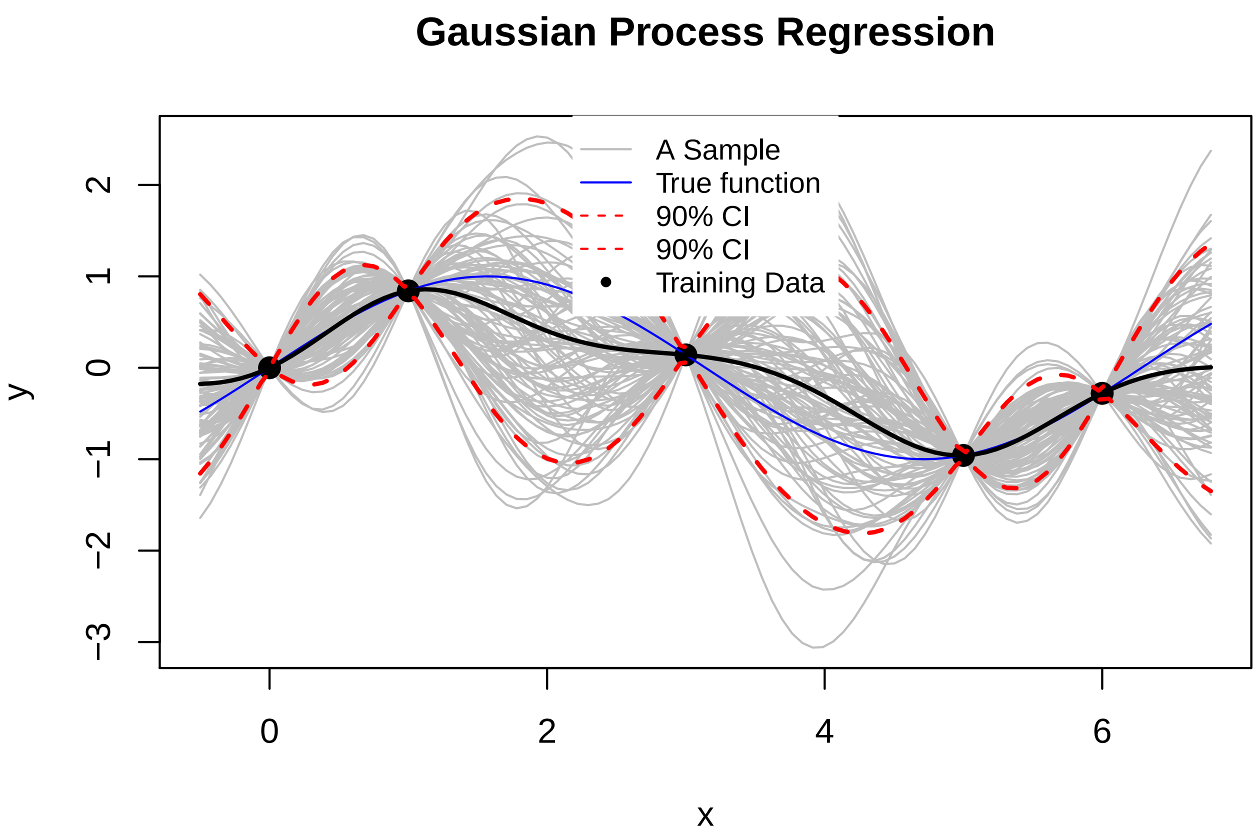
$$
\begin{aligned}
\mu_* &= m(\mathbf{X}_*) + \mathbf{K^{\intercal}}_{\mathbf{X},*}\mathbf{K^{-1}}_{\mathbf{X},\mathbf{X}}(\mathbf{f_X} - \mathbf{m}(\mathbf{X})) \\
\sum\nolimits_* &= \mathbf{K}_{*,*} - \mathbf{K}^{\intercal}_{X,*}\mathbf{K}^{-1}_{X,X}\mathbf{K}_{X,*}
\end{aligned} \tag{15}
$$

Example of Step.1 and Step.2

Assume $X = [0, 3, 5, 6]$ and $f_X = sin(X)$, giving $\mathcal{D} = (X, f_X)$. What is $p(f_*|X_*, \mathcal{D})$



Now we sample the point $X = 1$, and add to $\mathcal{D}$



Step.3 Deciding on next $\mathbf{x}^*$ *based on Posterior*

Posterior of the probalistic model quantify the uncertainty over the space of the f. The question is what is the next $\mathbf{x}^*$ to be sampled from the *expensive function*?

Define an utility function to collect new data points satisfying some optimality criterion: optimization as decision making.

There are a few of policies in the literature of Bayesopt, here the *Expected Improvement (EI)* policy will be used.

Expected Improvement as Policy for Decision Making. In Expected Improvement (EI) policy choose the next query point as the one which has the highest expected improvement over the space of the *expensive function*

$$utility(x;\theta,\mathcal{D}) = \alpha_{EI} = \int_y max(0, y - f)p(y|x;\theta,\mathcal{D}) \tag{16}$$

$$utility(x;\theta,\mathcal{D}) = \alpha_{EI} = \int_y max(0, y - f)p(y|x;\theta,\mathcal{D})\, dy$$

However, we do not have access to the *expensive function*, f, therefore we replace the f with the best available solution found so far, y^+

$$utility(x;\theta,\mathcal{D}) = \alpha_{EI} = \int_y max(0, y - y^{\dagger}) p(y|x;\theta,\mathcal{D})\, dy \tag{17}$$

y^{+} : The best solution found in the training dataset $\mathcal{D}$

The good news: The analytical form of the utility function is available for the gaussian process

$$\gamma(\mathbf{x}) = \frac{\mu(\mathbf{x};\theta,\mathcal{D}) - y^{\dagger}}{\sigma(\mathbf{x};\theta,\mathcal{D})} \tag{18}$$

$$utility(\mathbf{x};\theta,\mathcal{D}) = \alpha_{EI}(x;\theta,\mathcal{D}) = (\mu(x;\theta,\mathcal{D}) - y^{\dagger})\Phi(\gamma(x)) + \sigma(x;\theta,\mathcal{D})\phi(\gamma(x)) \tag{19}$$

Where $\Phi(.)$ and $\phi(.)$ are CDF and PDF of standard Gaussian distribution.

It is too greedy in the context of the sequential decision making. Therefore, an explorative term is added as explorative" parameter ϵ.

$$\gamma(\mathbf{x}) = \frac{\mu(\mathbf{x};\theta,\mathcal{D}) - y^{\dagger} - \epsilon}{\sigma(\mathbf{x};\theta,\mathcal{D})} \tag{20}$$

$$\alpha_{EI}(x;\theta,\mathcal{D}) = (\mu(x;\theta,\mathcal{D}) - y^{\dagger} - \epsilon)\Phi(\gamma(x)) + \sigma(x;\theta,\mathcal{D})\phi(\gamma(x)) \tag{21}$$

BO As a "mapping" between two problems. BO is an strategy to transform the problem

$$u_M = \underset{u \in \text{constraints}}{\text{argmax}} \; \overline{J}(u) \tag{22}$$

unsolvabale!

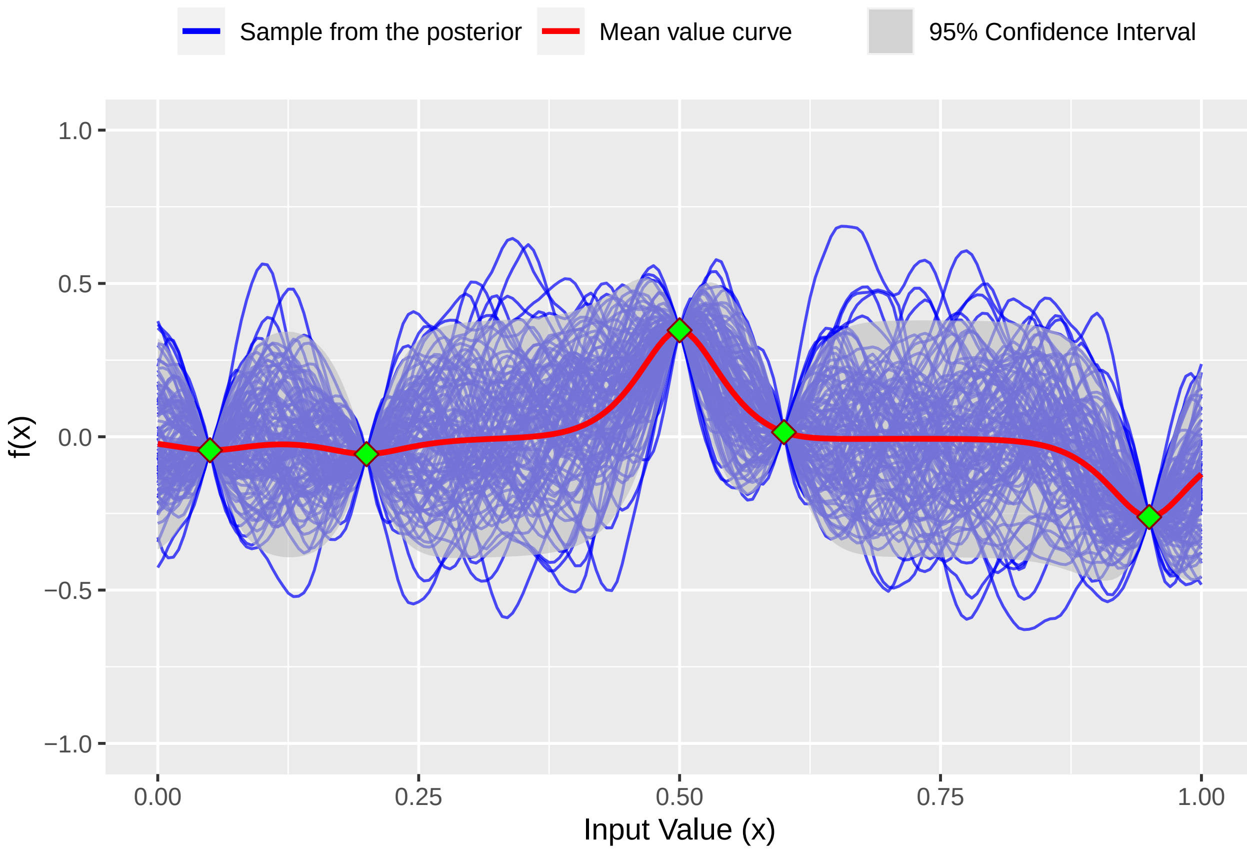
$$u^{next} = \underset{u \in \text{constraints}}{\text{argmax}} \; \alpha_{EI}(u;\mathcal{D}_n,\theta^*) \tag{23}$$

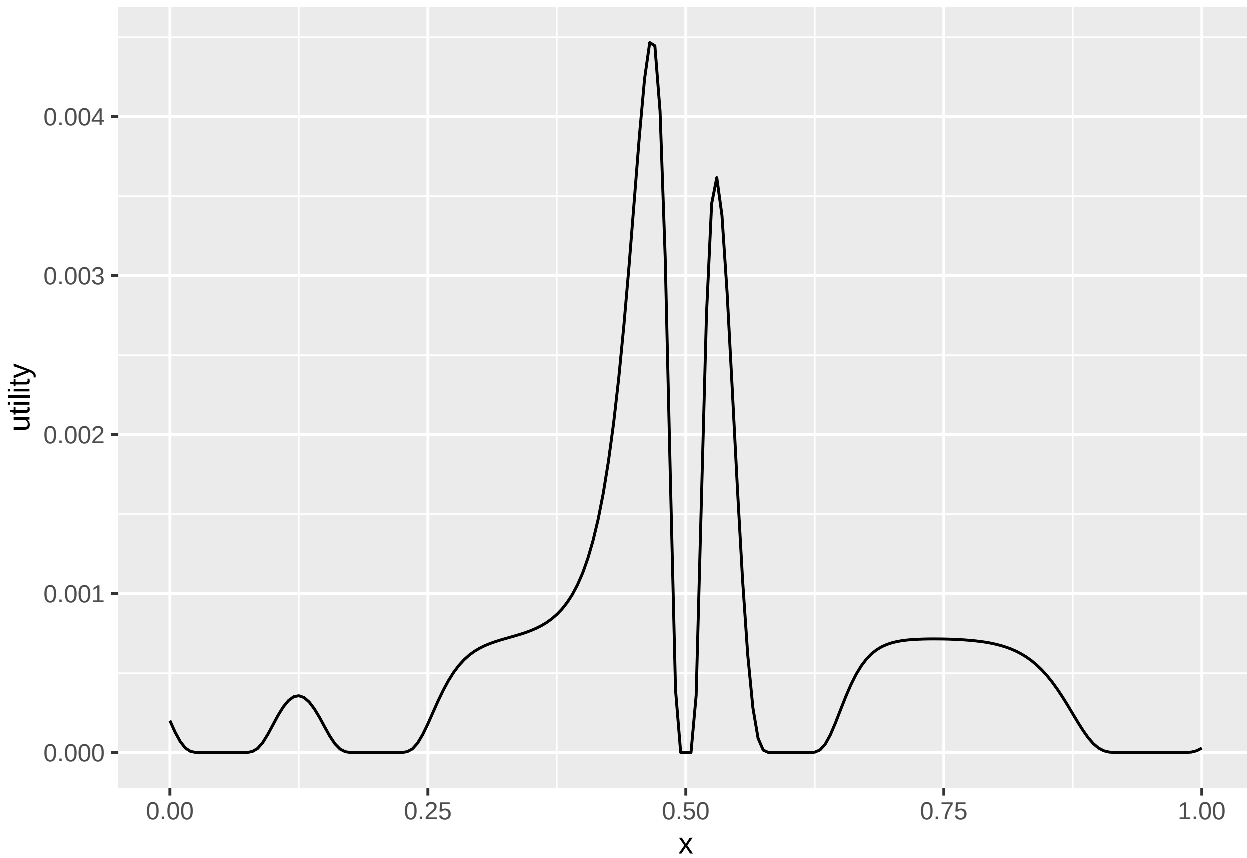
solvabale!

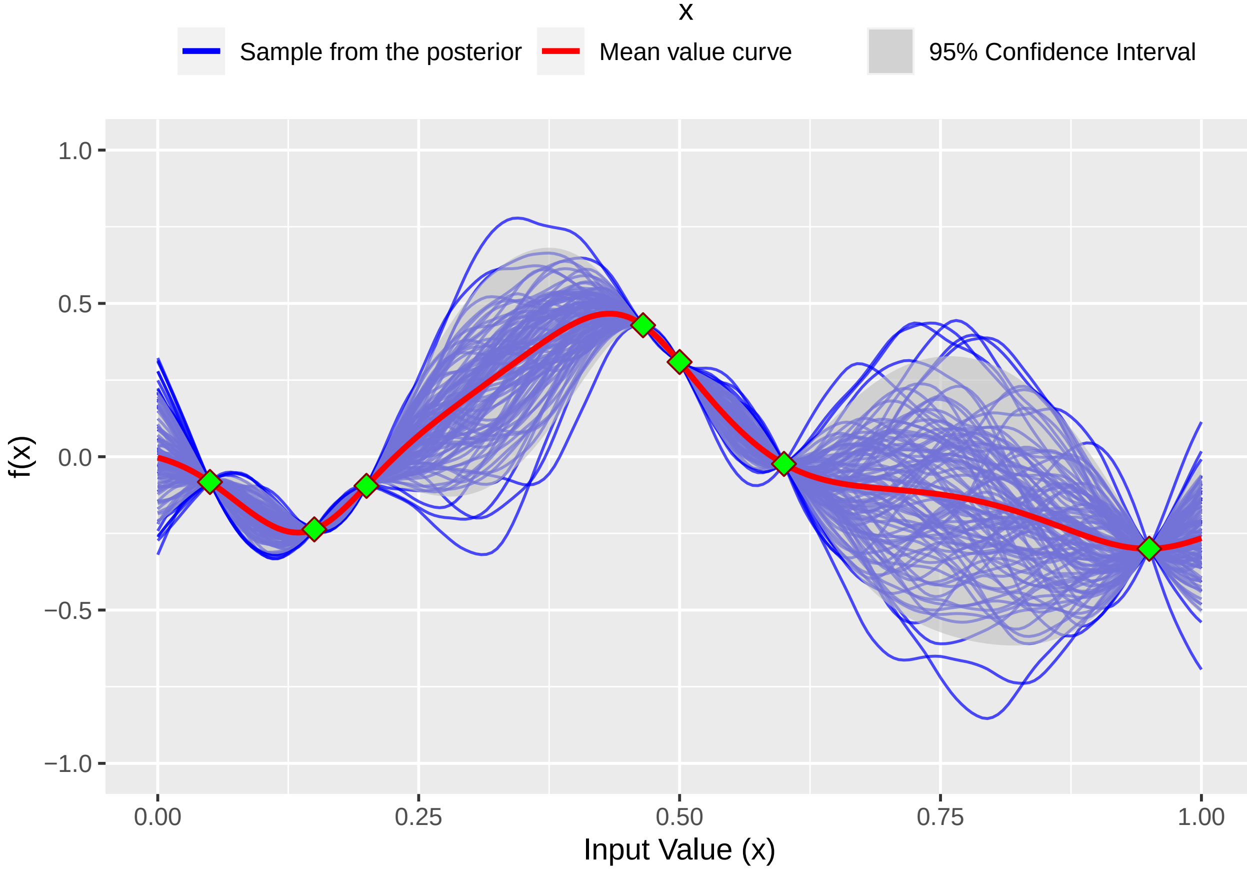
- $\alpha_{EI}(u)$ is inexpensive to evaluate.
- The analytical expression for gradient of $\alpha_{EI}(u)$ is available.
- Still need to find u^{next}, the multi-start BFGS is used for finding u^{next}.

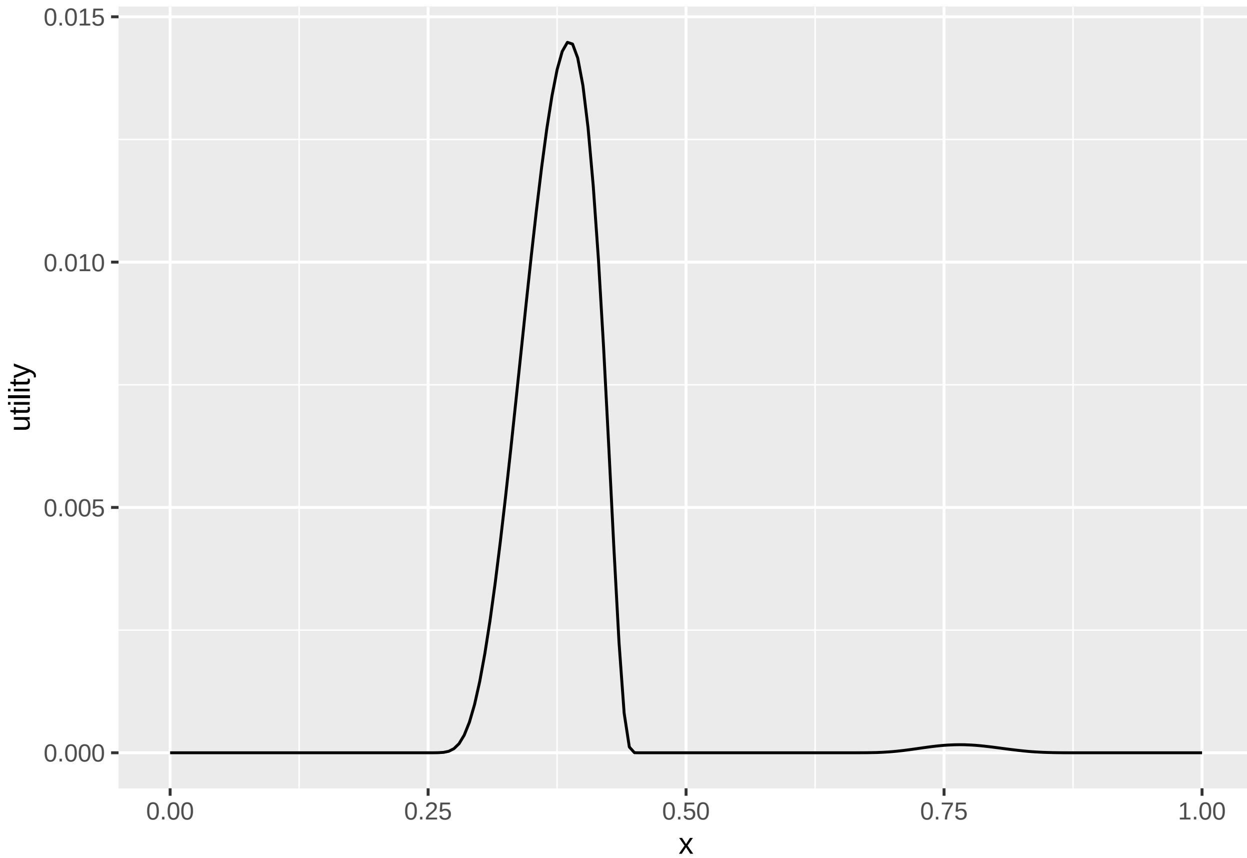
4. Example Cases

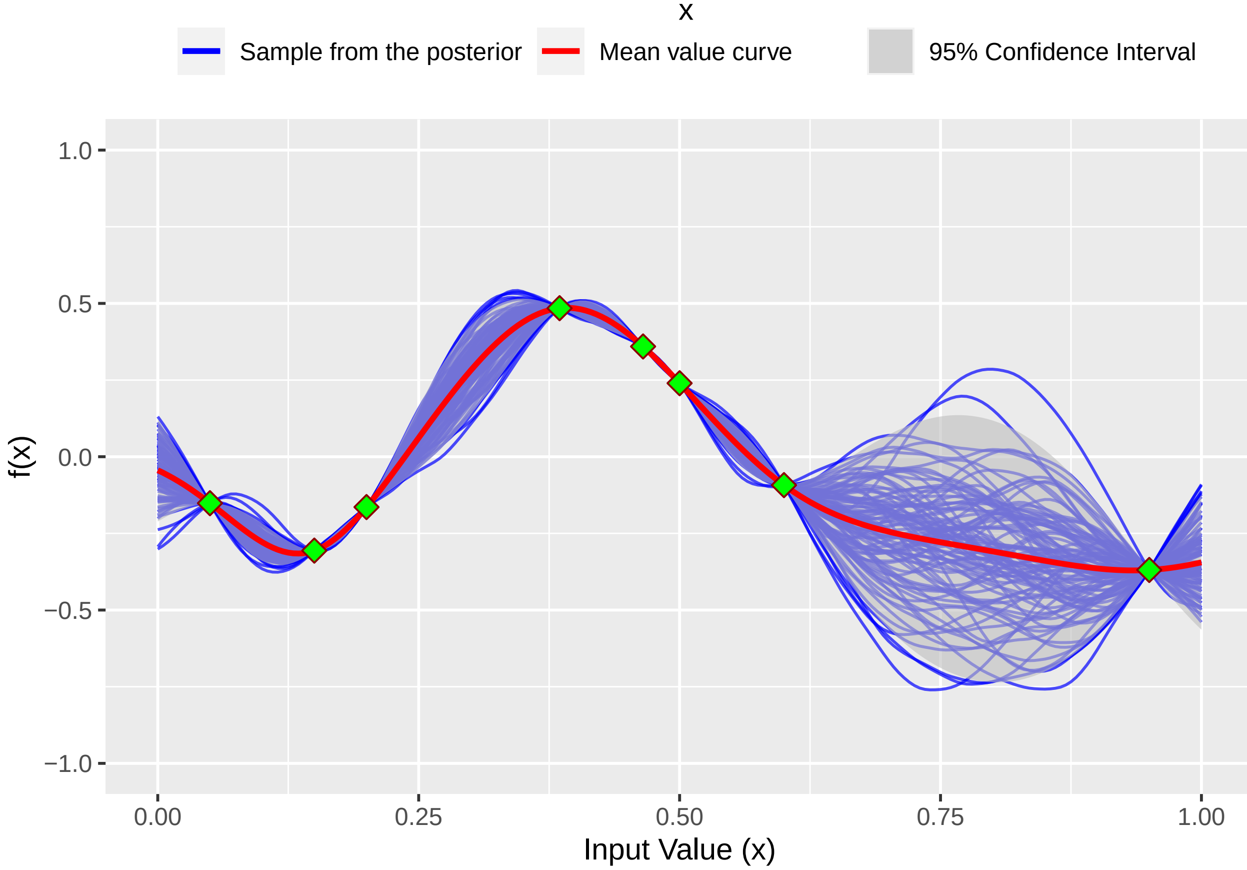
1-D toy Problem

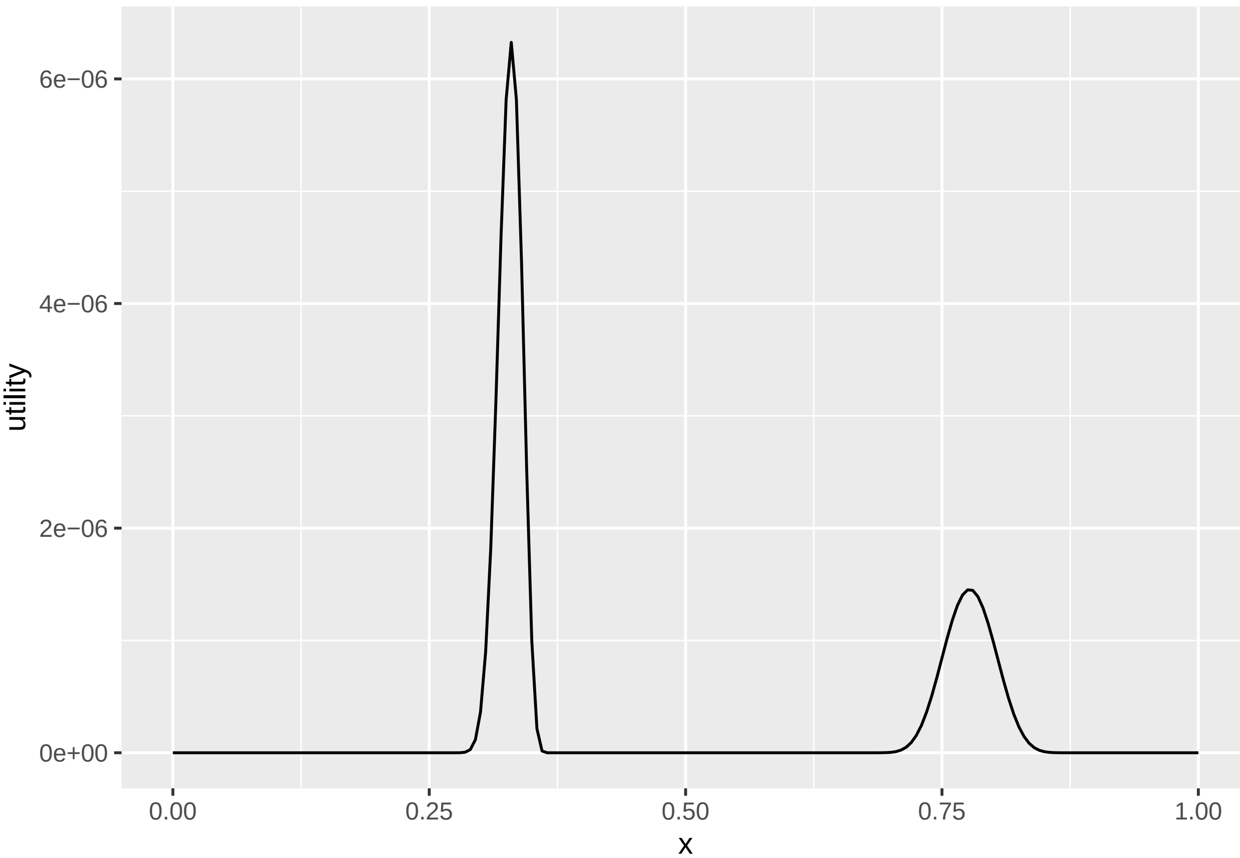



utility
0.004
0.003
0.002
0.001
0.000
0.00
0.25
0.50
0.75
1.00
x


Sample from the posterior
Mean value curve
95% Confidence Interval
f(x)
1.0
0.5
0.0
−0.5
−1.0
0.00
0.25
0.50
0.75
1.00
Input Value (x)


utility
0.015
0.010
0.005
0.000
0.00
0.25
0.50
0.75
1.00
x

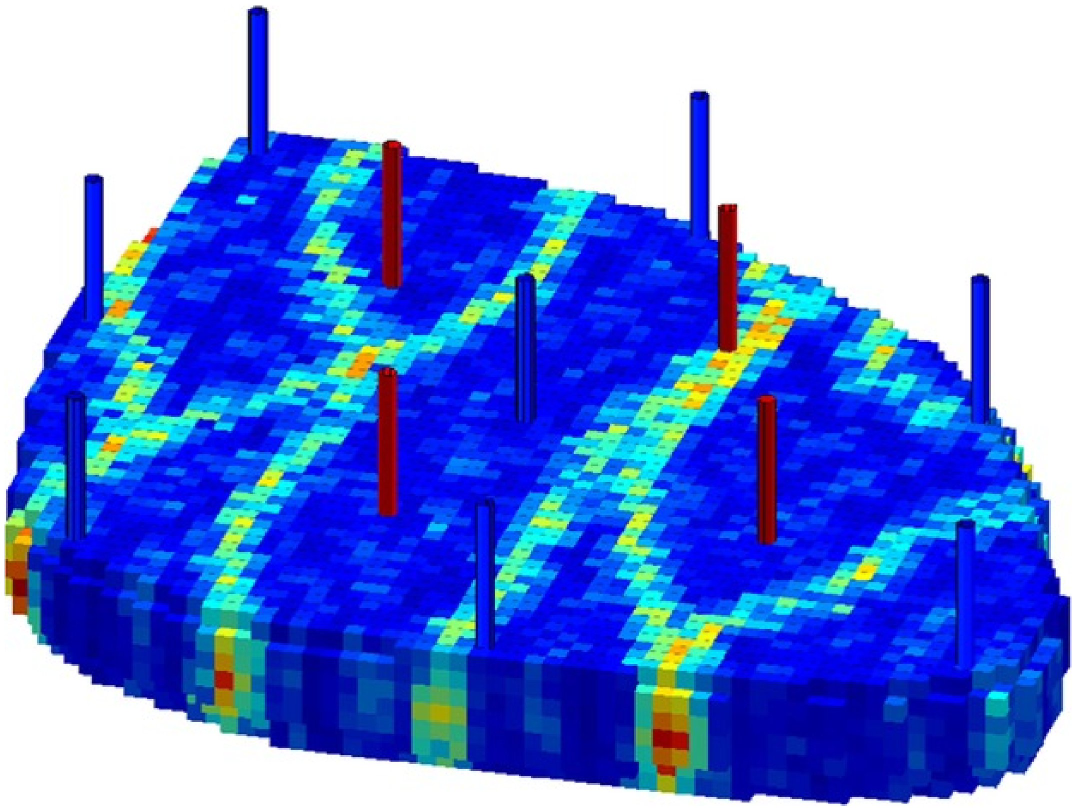

Sample from the posterior
Mean value curve
95% Confidence Interval
f(x)
1.0
0.5
0.0
−0.5
−1.0
0.00
0.25
0.50
0.75
1.00
Input Value (x)


6e−06
4e−06
2e−06
0e+00
utility
0.00
0.25
0.50
0.75
1.00
x

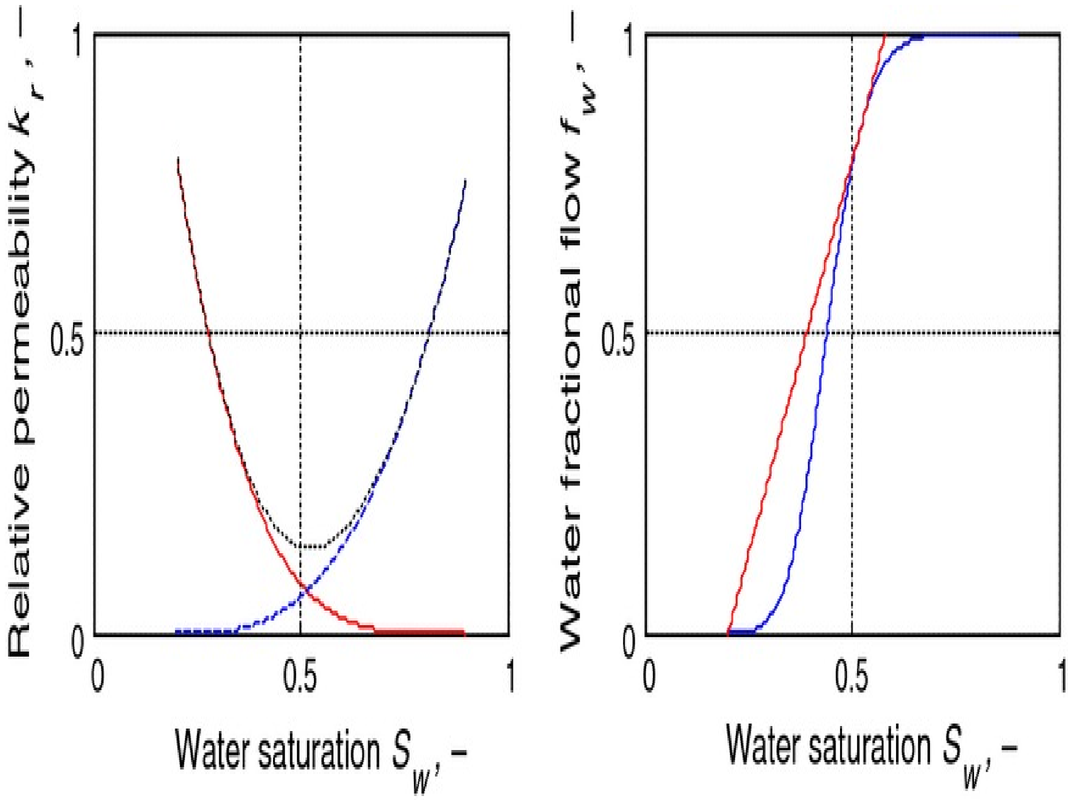
Field Scale

Egg Model (8 Injection wells, 4 Production wells)

- Egg Model



- Rel Perm Curves



BayesOpt Applied to Reservoir Case:

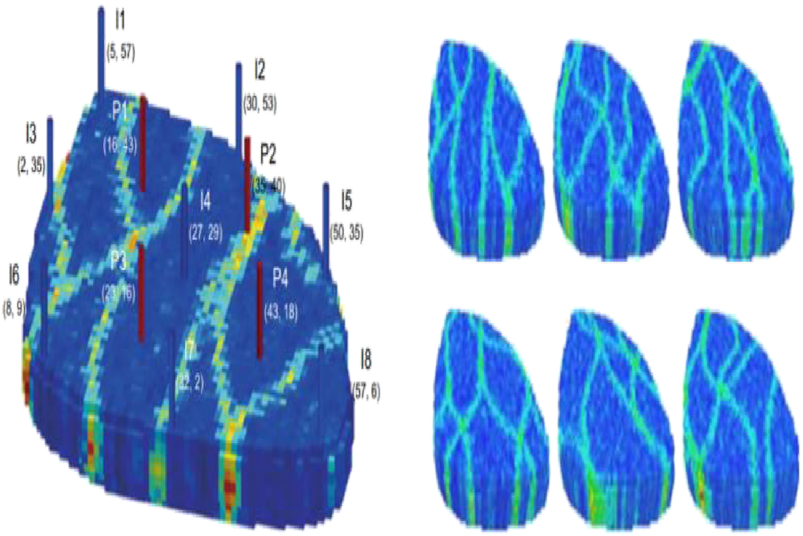
Optimization Objective

- $$J(u) = NPV(u) = \sum_{k=1}^{n_T} \frac{q_o^k(u)P_o - q_w^k(u)P_{wp} - I^k(u)P_{wi}}{(1+b)^{t_k/D}}$$

- $$\overline{J}(u) = \frac{\sum_{r=1}^{n_e} J_r(u)}{n_e}$$

- u is Injection rate for the each injection well,
- $u = [u_{inj1}, u_{inj2}, u_{inj3}, u_{inj4}, u_{inj5}, u_{inj6}, u_{inj7}, u_{inj8}]^\intercal$

Item	Pric	Items	Value
P_o	315	b	8%
P_wp	47.5	D	365
P_wi	12.5	n_e	10



I1
(5, 57)
I2
(30, 53)
I3
(2, 35)
P1
P2
I4
(27, 29)
I5
(50, 35)
I6
(8, 9)
P3
P4
(43, 18)
I7
I8
(57, 6)

5. Coparsion with other optimization algorithm

6. Conclusing Remarks

... ...

Acknowledgements

This work received support from the Research Council of Norway and the companies AkerBP, Wintershall–DEA, Vår Energy, Petrobras, Equinor, Lundin, and Neptune Energy, through the Petromaks–2 DIGIRES project (280473) (http://digires.no). We acknowledge the access to Eclipse licenses granted by Schlumberger.

Here are two sample references: (Dirac, 1953; Feynman and Vernon Jr., 1963).

References

Dirac, P.A.M., 1953. The lorentz transformation and absolute time. Physica 19, 888–896. doi:10.1016/S0031-8914(53)80099-6

Feynman, R.P., Vernon Jr., F.L., 1963. The theory of a general quantum system interacting with a linear dissipative system. Annals of Physics 24, 118–173. doi:10.1016/0003-4916(63)90068-X