Bayesian Target-Vector Optimization for Efficient Parameter Reconstruction

M. Plock[1] M. Hammerschmidt[1,2] S. Burger[1,2] P.-I. Schneider[1,2]

[1] Zuse Institute Berlin, Berlin, Germany

[2] JCMwave GmbH, Berlin, Germany

5th MathMet 2022, Paris, November 2-4, 2022

Optical metrology: the inverse problem

Problems in optical metrology often **inverse problems**

- Results of experimental measurement $\boldsymbol{y}^*$
- Parameterized model $\boldsymbol{f}(\boldsymbol{p})$ of measurement process (e.g. using FEM)
- Find model parameters $\boldsymbol{p}^*$ that explain measurement results

$$\text{Minimize} \quad \chi^2(\boldsymbol{p}) = \sum_i^N \left(\frac{y_i^* - f_i(\boldsymbol{p})}{\eta_i} \right)^2$$

$\rightarrow$ Minimization algorithm

Optical metrology: the inverse problem

Problems in optical metrology often **inverse problems**

- Results of experimental measurement $\boldsymbol{y}^*$
- Parameterized model $\boldsymbol{f}(\boldsymbol{p})$ of measurement process (e.g. using FEM)
- Find model parameters $\boldsymbol{p}^*$ that explain measurement results

$$\text{Minimize} \quad \chi^2(\boldsymbol{p}) = \sum_i^N \left(\frac{y_i^* - f_i(\boldsymbol{p})}{\eta_i} \right)^2$$

$\rightarrow$ Minimization algorithm

Optical metrology: the inverse problem

Problems in optical metrology often **inverse problems**

- Results of experimental measurement $\boldsymbol{y}^*$
- Parameterized model $\boldsymbol{f}(\boldsymbol{p})$ of measurement process (e.g. using FEM)
- Find model parameters $\boldsymbol{p}^*$ that explain measurement results

$$\text{Minimize} \quad \chi^2(\boldsymbol{p}) = \sum_{i}^{N} \left(\frac{y_i^* - f_i(\boldsymbol{p})}{\eta_i} \right)^2$$

$\rightarrow$ Minimization algorithm

Optical metrology: the inverse problem

Problems in optical metrology often **inverse problems**

- Results of experimental measurement $\boldsymbol{y}^*$
- Parameterized model $\boldsymbol{f}(\boldsymbol{p})$ of measurement process (e.g. using FEM)
- Find model parameters $\boldsymbol{p}^*$ that explain measurement results

$$\text{Minimize} \quad \chi^2(\boldsymbol{p}) = \sum_i^N \left(\frac{y_i^* - f_i(\boldsymbol{p})}{\eta_i} \right)^2$$

$\rightarrow$ Minimization algorithm

Requirements for minimization schemes

- Can handle the discrete results of measurement and model
  - E.g. Levenberg-Marquardt
- Uses available resources sparingly (forward model may be expensive)
  - E.g. methods that use surrogate models (Kriging, GPR, BO)

Requirements for minimization schemes

- Can handle the discrete results of measurement and model
  - E.g. Levenberg-Marquardt
- Uses available resources sparingly (forward model may be expensive)
  - E.g. methods that use surrogate models (Kriging, GPR, BO)

Requirements for minimization schemes

- Can handle the discrete results of measurement and model
  - E.g. Levenberg-Marquardt
- Uses available resources sparingly (forward model may be expensive)
  - E.g. methods that use surrogate models (Kriging, GPR, BO)

Combining both aspects:

Bayesian target vector optimization[1,2]

- Least-square minimization in a Bayesian optimization framework
  - ✓ Responsible use of resources
  - ✓ Direct use of each data channel

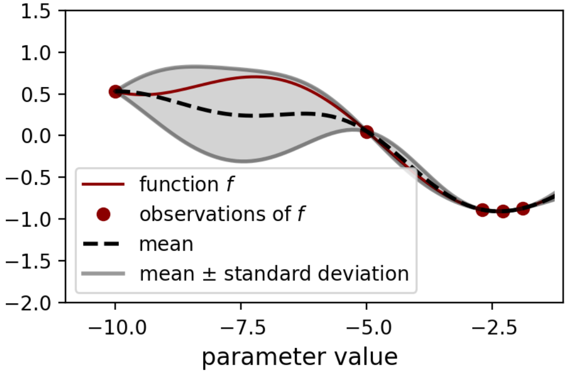
[1] Uhrenholt and Jensen, PMLR (2019)

[2] Plock, et al., Adv. Theory Simul. 5, 2200112 (2022)

Bayesian Optimization, least-square problems, and a new approach

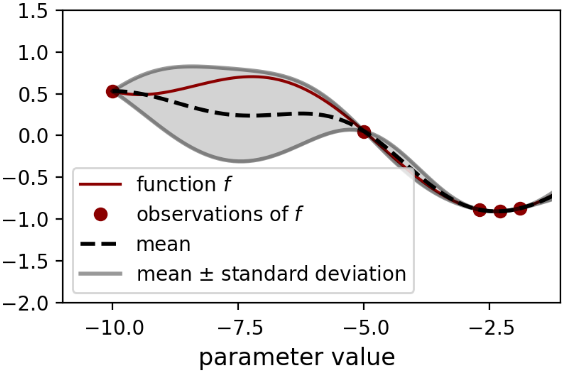
Bayesian optimization fundamentals

- Sequential optimization methods
- Use model observations to create surrogate model (Gaussian process (GP) regression)
- GP completely specified by
  - Mean function $m(\boldsymbol{p})$ (often just m_0)
  - Covariance kernel function $k(\boldsymbol{p}, \boldsymbol{p}')$
  - Model observations



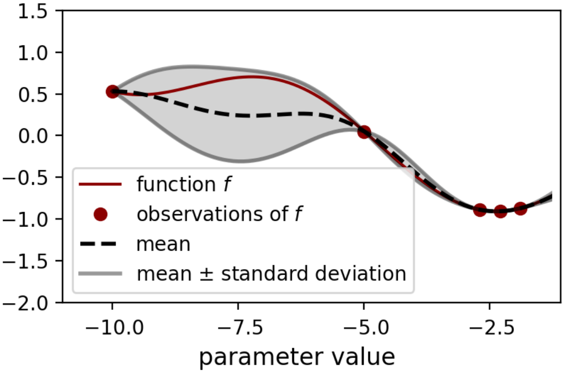
Bayesian optimization fundamentals

- Sequential optimization methods
- Use model observations to create surrogate model (Gaussian process (GP) regression)
- GP completely specified by
  - Mean function $m(\boldsymbol{p})$ (often just m_0)
  - Covariance kernel function $k(\boldsymbol{p}, \boldsymbol{p}')$
  - Model observations



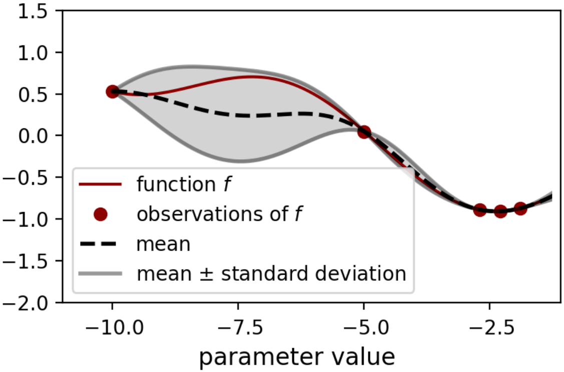
Bayesian optimization fundamentals

- Sequential optimization methods
- Use model observations to create surrogate model (Gaussian process (GP) regression)
- GP completely specified by
  - Mean function $m(\boldsymbol{p})$ (often just m_0)
  - Covariance kernel function $k(\boldsymbol{p}, \boldsymbol{p}')$
  - Model observations



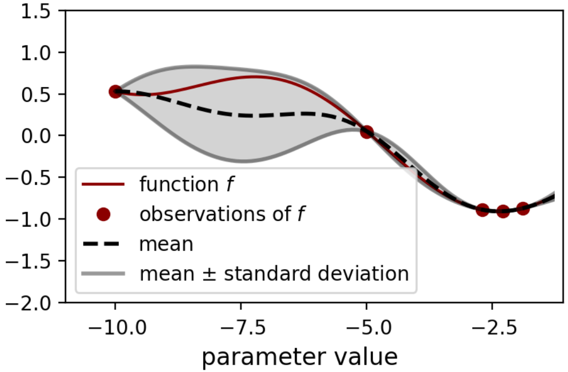
Bayesian optimization fundamentals

- GP predicts $\hat{f}(\boldsymbol{p}) \sim \mathcal{N}(\mu(\boldsymbol{p}), \sigma^2(\boldsymbol{p}))$ for all $\boldsymbol{p} \in \mathcal{X}$
- *Acquisition function* uses $\mu(\boldsymbol{p})$ and $\sigma^2(\boldsymbol{p})$
- Finds useful points in parameter space to achieve goal of optimization

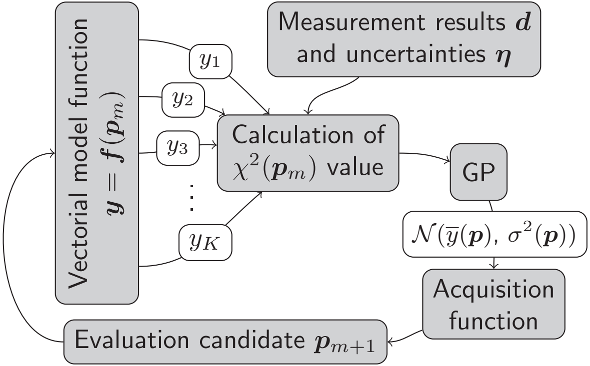


Bayesian optimization fundamentals

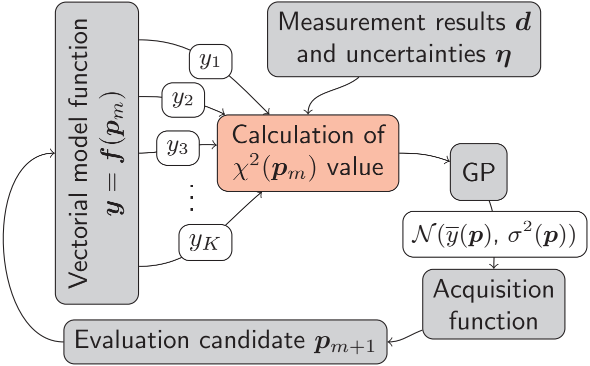
- GP predicts $\hat{f}(\boldsymbol{p}) \sim \mathcal{N}(\mu(\boldsymbol{p}), \sigma^2(\boldsymbol{p}))$ for all $\boldsymbol{p} \in \mathcal{X}$
- *Acquisition function* uses $\mu(\boldsymbol{p})$ and $\sigma^2(\boldsymbol{p})$
- Finds useful points in parameter space to achieve goal of optimization



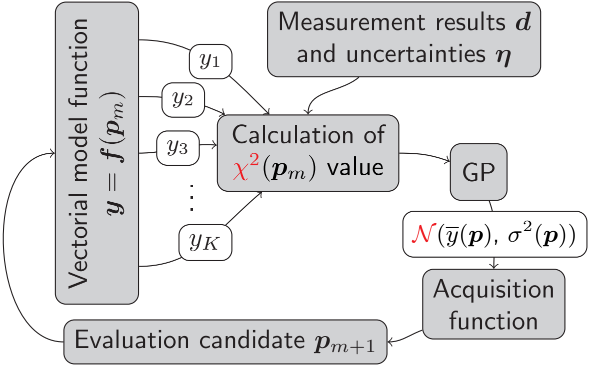
Least-squares using a conventional Bayesian optimization scheme



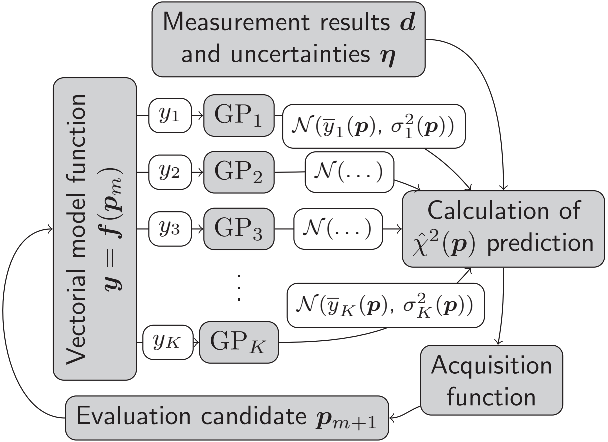
Least-squares using a conventional Bayesian optimization scheme



Least-squares using a conventional Bayesian optimization scheme



The Bayesian target-vector optimization scheme

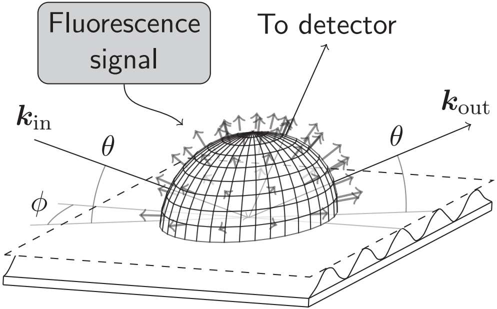


[1] Uhrenholt and Jensen, PMLR (2019)
[2] Plock, et al., Adv. Theory Simul. 5, 2200112 (2022)

Application of the method

Application of the method: the experimental dataset

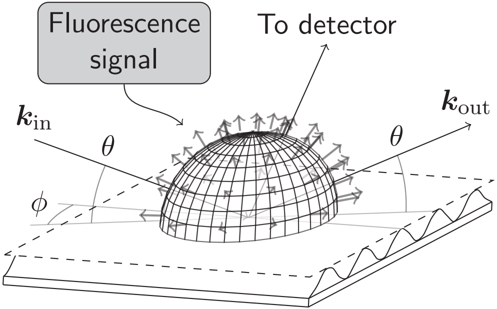
- GIXRF (Grazing Incidence XRay Fluorescence)
- Measured at PTB/BESSY

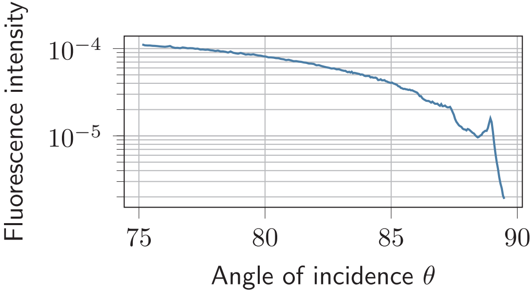


- 208 discrete measurements with uncertainties
- Measured for angles $\theta = 75.13°$ to $\theta = 89.48°$

Application of the method: the experimental dataset

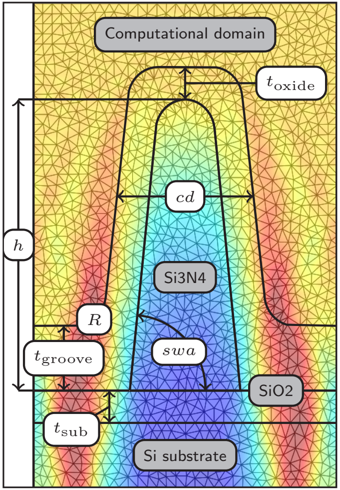
- GIXRF (Grazing Incidence XRay Fluorescence)
- Measured at PTB/BESSY





- 208 discrete measurements with uncertainties
- Measured for angles $\theta = 75.13°$ to $\theta = 89.48°$

Application of the method: the forward model



- Model created using FEM Maxwell Solver (JCMsuite)
- 10 free parameters
  - 7 geometrical
  - 3 auxiliary

Application of the method: Performance measurement

- Comparing Bayesian Target Vector optimization method (BTVO) against ...
  - Levenberg-Marquardt (LM) least-squares (`scipy` implementation)
  - Conventional Bayesian optimization (BO) method
- Determine mean and standard deviation of six optimization runs
- Metric is distance to best reconstruction result $\boldsymbol{p}_{\mathrm{LSQE}}$ in terms of Gaussian reconstruction uncertainties

$$d(\boldsymbol{p}) = \sqrt{\sum_{i}^{N} \left(\frac{p_i - p_{\mathrm{LSQE},i}}{\epsilon_{\mathrm{LSQE},i}} \right)^2}, \quad \epsilon_i \text{ from Jacobian at } \boldsymbol{p}_{\mathrm{LSQE}}$$

Application of the method: Performance measurement

- Comparing Bayesian Target Vector optimization method (BTVO) against ...
  - Levenberg-Marquardt (LM) least-squares (`scipy` implementation)
  - Conventional Bayesian optimization (BO) method
- Determine mean and standard deviation of six optimization runs
- Metric is distance to best reconstruction result $\boldsymbol{p}_{\mathrm{LSQE}}$ in terms of Gaussian reconstruction uncertainties

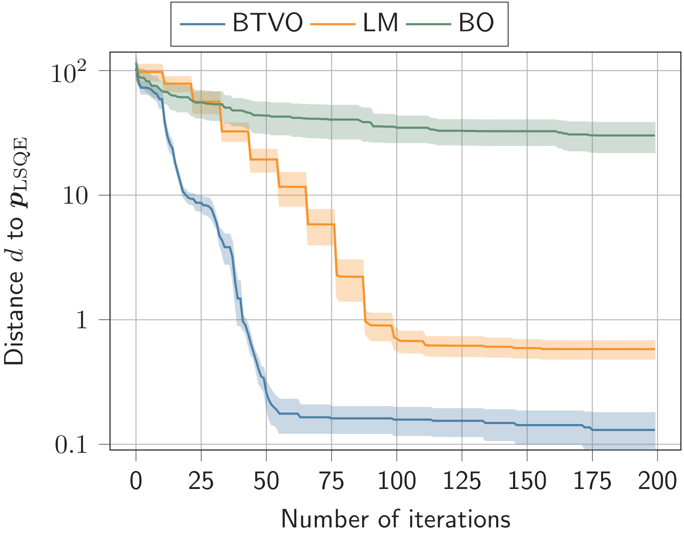
$$d(\boldsymbol{p}) = \sqrt{\sum_i^N \left(\frac{p_i - p_{\mathrm{LSQE},i}}{\epsilon_{\mathrm{LSQE},i}}\right)^2}, \quad \epsilon_i \text{ from Jacobian at } \boldsymbol{p}_{\mathrm{LSQE}}$$

Application of the method: Performance measurement

- Comparing Bayesian Target Vector optimization method (BTVO) against . . .
  - Levenberg-Marquardt (LM) least-squares (`scipy` implementation)
  - Conventional Bayesian optimization (BO) method
- Determine mean and standard deviation of six optimization runs
- Metric is distance to best reconstruction result $\boldsymbol{p}_{\mathrm{LSQE}}$ in terms of Gaussian reconstruction uncertainties

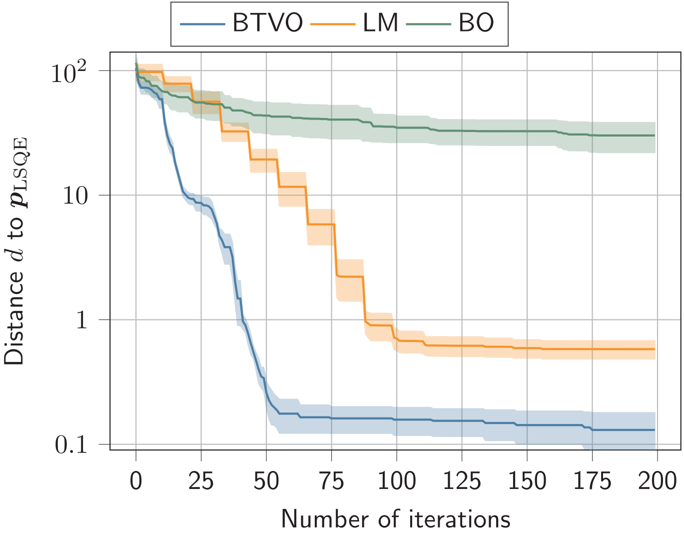
$$d(\boldsymbol{p}) = \sqrt{\sum_{i}^{N} \left(\frac{p_i - p_{\mathrm{LSQE},i}}{\epsilon_{\mathrm{LSQE},i}} \right)^2}, \quad \epsilon_i \text{ from Jacobian at } \boldsymbol{p}_{\mathrm{LSQE}}$$

Reconstruction results



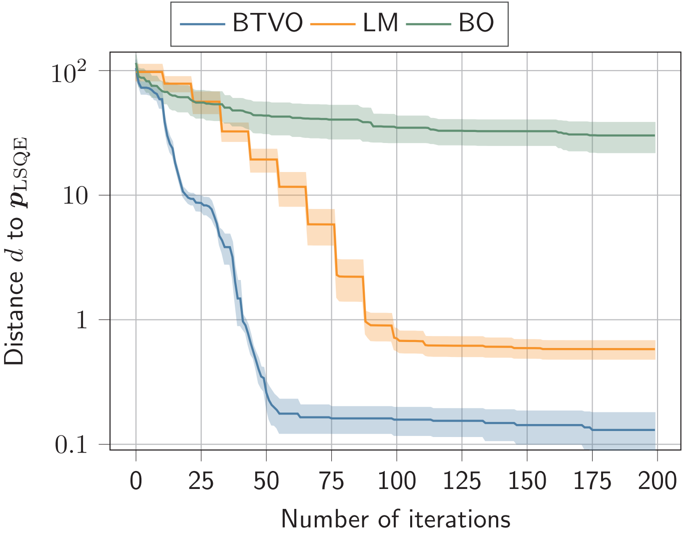
- Conventional BO fails to reconstruct parameters
- BTVO and LM reach distance of less than 1 standard deviation
- BTVO outperforms LM: faster and closer to $\boldsymbol{p}_{\mathrm{LSQE}}$

Reconstruction results



- Conventional BO fails to reconstruct parameters
- BTVO and LM reach distance of less than 1 standard deviation
- BTVO outperforms LM: faster and closer to $\boldsymbol{p}_{\mathrm{LSQE}}$

Reconstruction results



- Conventional BO fails to reconstruct parameters
- BTVO and LM reach distance of less than 1 standard deviation
- BTVO outperforms LM: faster and closer to $\boldsymbol{p}_{\mathrm{LSQE}}$

Parameter uncertainty estimation using MCMC on a trained surrogate model

MCMC with a trained multi output surrogate model

- Markov Chain Monte Carlo (MCMC) often used to determine parameter uncertainties and correlations
- Fitting an error model to the data
- Requires many samples of objective function

- BTVO and BO automatically train surrogate model
- Sampling of surrogate model cheaper than objective function
- Use trained surrogate models as stand-in for objective function

$\rightarrow$ Perform MCMC on trained surrogate model

MCMC with a trained multi output surrogate model

- Markov Chain Monte Carlo (MCMC) often used to determine parameter uncertainties and correlations
- Fitting an error model to the data
- Requires many samples of objective function

- BTVO and BO automatically train surrogate model
- Sampling of surrogate model cheaper than objective function
- Use trained surrogate models as stand-in for objective function

→ Perform MCMC on trained surrogate model

MCMC with a trained multi output surrogate model

- Markov Chain Monte Carlo (MCMC) often used to determine parameter uncertainties and correlations
- Fitting an error model to the data
- Requires many samples of objective function

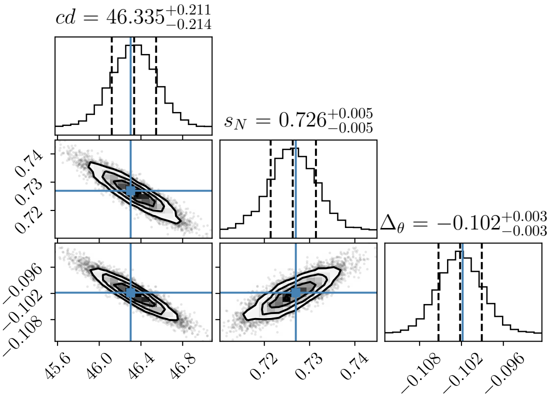
- BTVO and BO automatically train surrogate model
- Sampling of surrogate model cheaper than objective function
- Use trained surrogate models as stand-in for objective function

$\rightarrow$ Perform MCMC on trained surrogate model

GIXRF MCMC analysis

Applied to GIXRF forward problem[1]

- Reveal correlations between parameters (also discussed in [2])
- MLE and uncertainties confirm optimization reconstruction results



[1]Plock, et al., Adv. Theory Simul. 5, 2200112 (2022)
[2]Soltwisch et al., (2018) Nanoscale 10 6177

Summary

- Bayesian Target Vector optimization performs very well
- Manages to outperform a traditional least-square algorithm for the considered parameter reconstruction problem
- Multi-output GP yields fast & inexpensive MCMC option

Thank you for your interest!

Contact: plock@zib.de

Also thank you to our collaboration partners: PTB (Braunschweig, Berlin) and the siMLopt project (Martin Weiser, Phillip Semler, FHI-IISB (Andreas Erdmann, A. Rosskopf), FAU, Synopsis, JCMwave)

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

This project is funded by the German Federal Ministry of Education and Research (BMBF, project number 05M20ZAA, siMLopt; project number 01IS20080A, SiM4diM; Forschungscampus MODAL, project number 05M20ZBM), by the German Federal Ministry for Economic Affairs and Energy (BMWi, project number 50WM2067, Optimal-QT), and by the European Union's Horizon 2020 research and innovation programme (EU H2020, grant number 101007319, AI-TWILIGHT).

Gefördert durch:

Bundesministerium
für Wirtschaft
und Energie

aufgrund eines Beschlusses
des Deutschen Bundestages

Parameter uncertainty estimation, using MCMC on a trained surrogate model

MCMC with a trained multi output surrogate model

- Often used to determine parameter uncertainties and correlations
- Fitting an error model to the data
- Requires many samples of objective function

- Use trained surrogate models as stand-in for objective function
- Sampling of surrogate model cheaper than objective function

$\rightarrow$ Perform MCMC on surrogate model

MCMC with a trained multi output surrogate model

- Often used to determine parameter uncertainties and correlations
- Fitting an error model to the data
- Requires many samples of objective function

- Use trained surrogate models as stand-in for objective function
- Sampling of surrogate model cheaper than objective function

$\rightarrow$ Perform MCMC on surrogate model

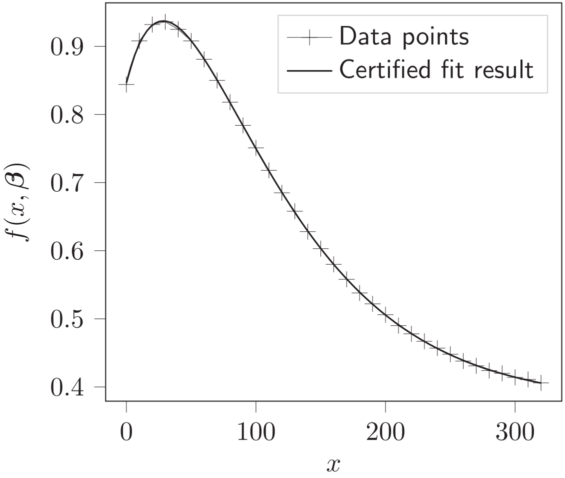
MCMC with a trained multi output surrogate model

- Often used to determine parameter uncertainties and correlations
- Fitting an error model to the data
- Requires many samples of objective function

- Use trained surrogate models as stand-in for objective function
- Sampling of surrogate model cheaper than objective function

$\rightarrow$ Perform MCMC on surrogate model

The MCMC demonstration dataset (MGH17)

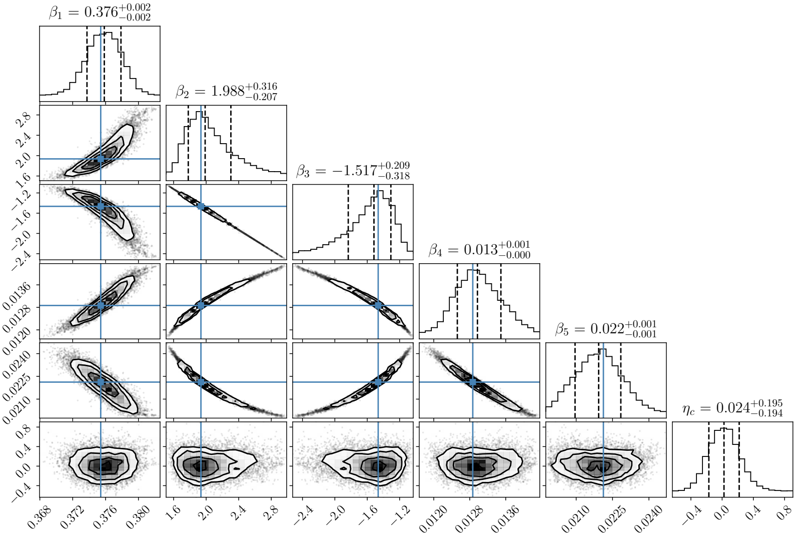


- MGH17 dataset[1] from NIST Standard Reference Database
- 33 discrete data points and 5 free parameters $(\beta_1, \ldots, \beta_5)$

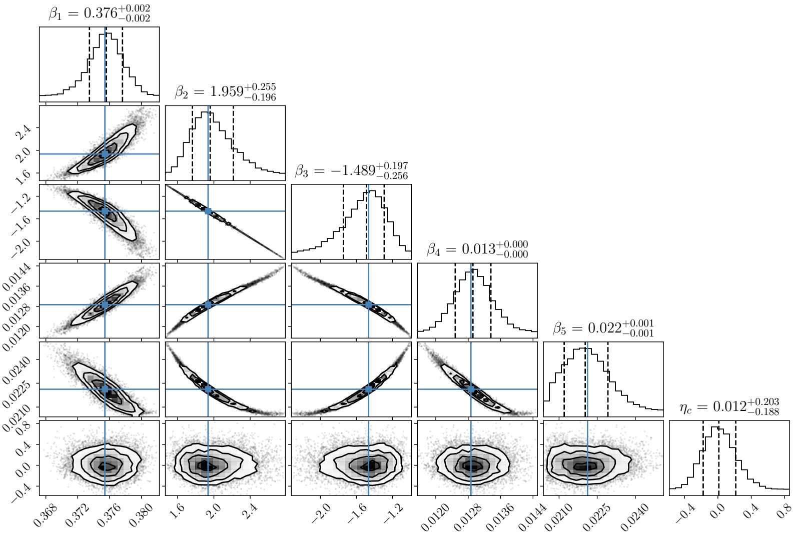
$$f(x,\boldsymbol{\beta}) = \beta_1 + \beta_2 \mathrm{e}^{-x\cdot\beta_4} + \beta_3 \mathrm{e}^{-x\cdot\beta_5}$$

[1] https://www.itl.nist.gov/div898/strd/nls/data/mgh17.shtml

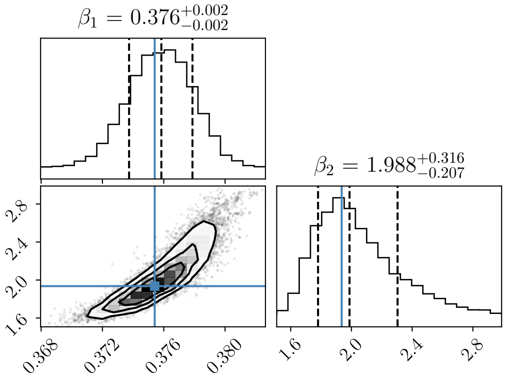
MCMC for MGH17 (50000 obs. of model function / 32 walkers)



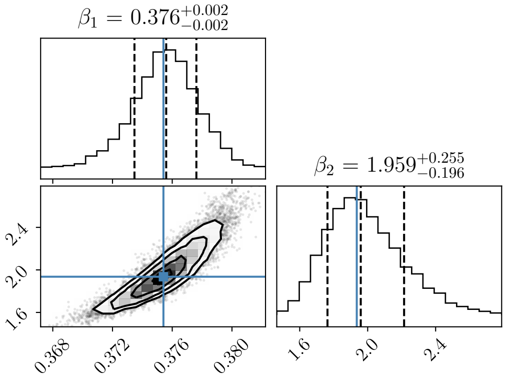
MCMC for MGH17 (50000 obs. of surrogate model / 32 walkers)



MCMC for MGH17 (side by side comparison)

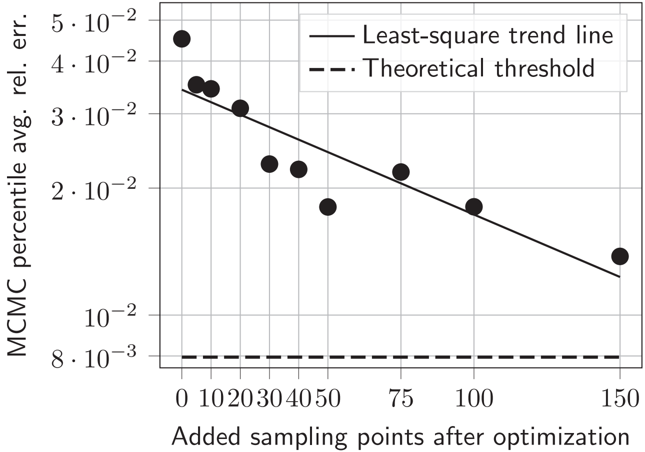


Using objective function.



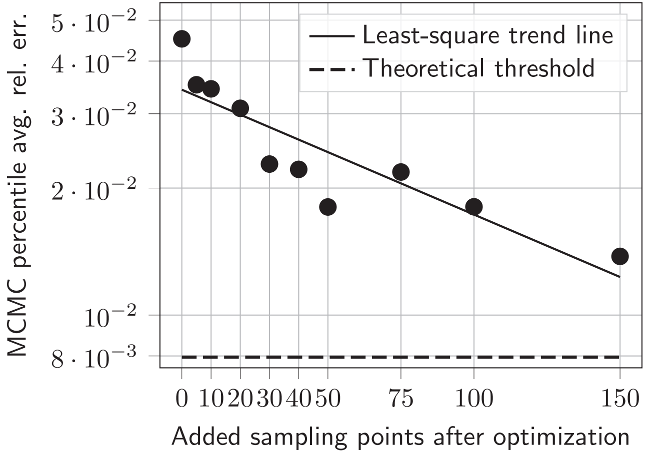
Using surrogate model.

GP MCMC convergence



- After optimization: GPs not refined enough
- Add data points around the least-square minimum
- Surrogate becomes more accurate
- Measure convergence by looking at percentiles
- Relative error of percentile positions decreases with added data points

GP MCMC convergence

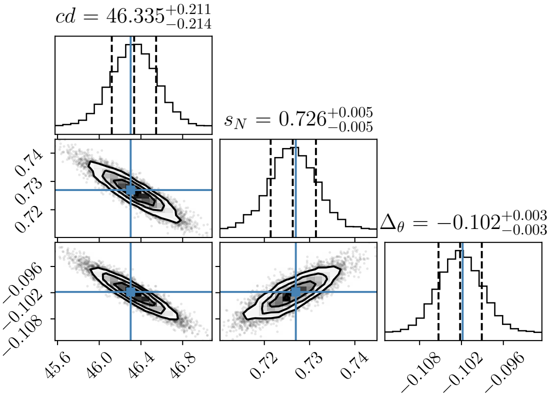


- After optimization: GPs not refined enough
- Add data points around the least-square minimum
- Surrogate becomes more accurate

- Measure convergence by looking at percentiles
- Relative error of percentile positions decreases with added data points

GIXRF MCMC analysis

- Apply to GIXRF forward problem
- Reveal correlations between parameters (also discussed in [1])
- MLE and uncertainties confirm optimization reconstruction results



[1] Soltwisch et al., (2018) Nanoscale 10 6177