A tour d'horizon in surrogate-assisted multiobjective optimization

Michael Emmerich| Leiden University and Visiting Professor at University of Jyväskylä University, Finland, and Lviv PTU Ukraine

July 8, 2020

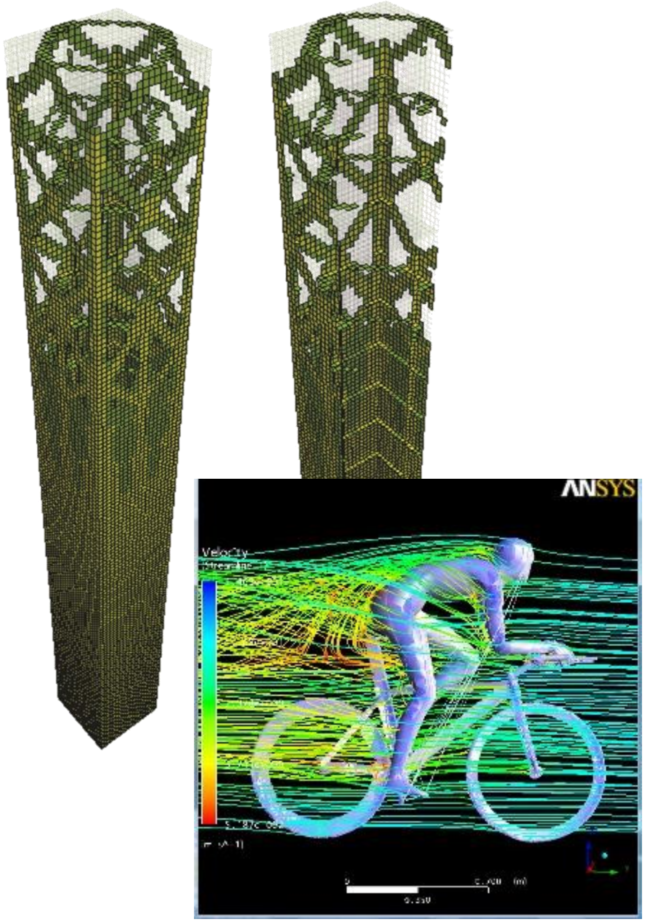
Tour d'horizon - roadmap

- Surrogate-models?!
- Infill Criteria in Single Objective Optimization
- Infill criteria in Multi Objective Optimization
- New Horizons

"Winter" by Maria Emmerich

Surrogate Models

Typical Scenario:
FEM & CFD Simulations in construction & design optimization



1. Problem:
   - Expensive Evaluations (Time, Money)
   - Optimization with Black Box Functions

     " Computer Experiments"

2. Solution:
   - Use all available information from past evaluations!

3. Method:
   - Fit **Surrogate Models** to evaluation data

… other terms, same idea …

"Metamodeling: Models of (Simulation) Models"

"Sequential Parameter Optimization"

"Statistical Model Based Optimization"

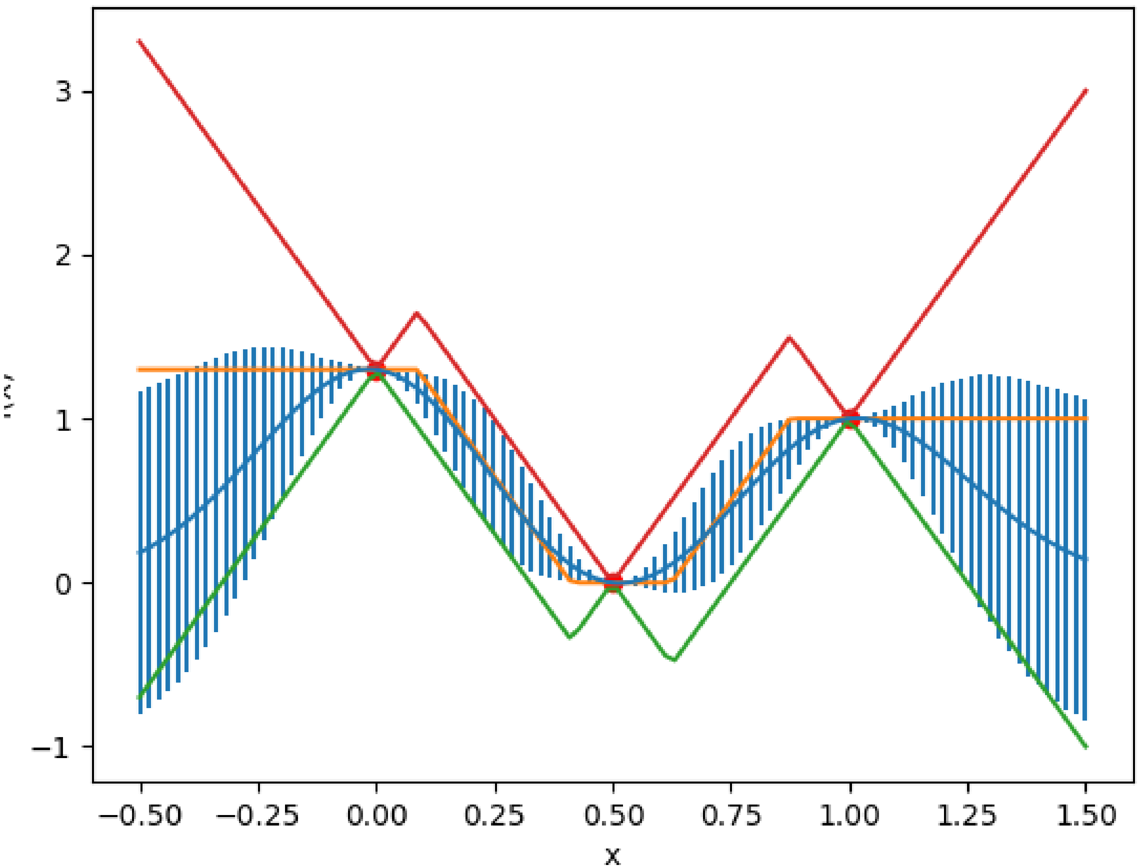
"Bayesian (Global) Optimization"

Chugh, T., Rahat, A., Volz, V., Zaefferer, M. (n.d.) Towards Better Integration of Surrogate Models and Optimizers (Ed.), *High-Performance Simulation-Based Optimization* 137 163

Function Approximation & Uncertainty quantification

- Continuity is central concept here
  - By distance dependent correlation function (Kriging , Gaussian Processes)
  - By Lipschitz Constant (bound)
  - In regression and random forest methods: cross validation error or mean squared error (global not local)
- Local Error small, if ...
  - Higher (local) density of points
  - Lower distance to points
  - Lower general amplitude and frequency of fluctuations
- Local Error estimate ~ degree of exploration

Online Python Trinket: Kriging/Lipschitz



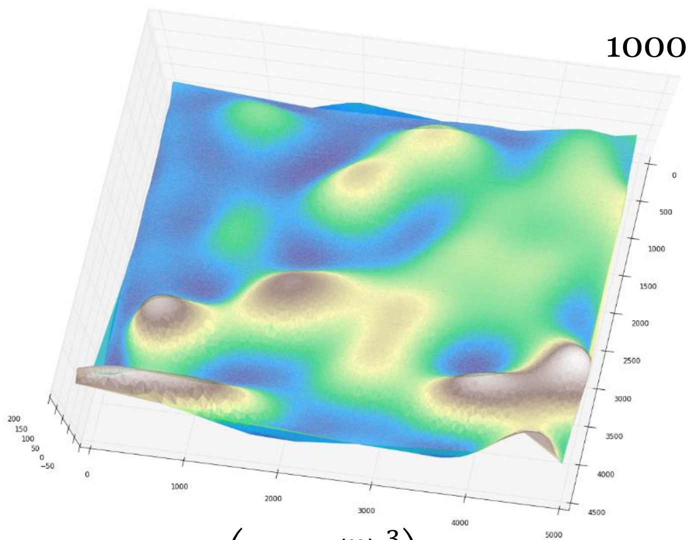
https://trinket.io/library/trinkets/c38e5ebdbc

How to approximate functions?

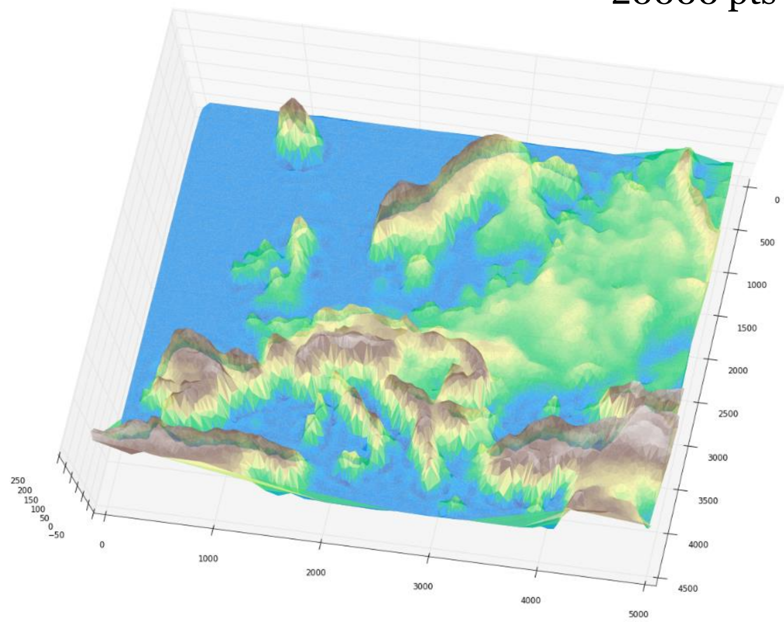
	Model	Fitting	Remarks
Kriging	Random Field combined with trend function, correlation ~distance input vectors	Best linear unbiased predictor, Mean Squared Error	Moderate dimensional, **uncertainty quantification**
Gaussian Processes	Gaussian random field	Conditional Mean, Variance	“
Lipschitz Models	Bounded change rate in distance, continuity	Distance to neighbors, min-max, diagonal	Produces exact error bounds, piecewise linear approximation
Artificial Neural Networks	Perceptron, Sigmoid or Radial Basis functios Activation functions,	Training error minimization	Many versions, topology choices and hyperparameters, Regularization theory
Splines	Piecewise defined function	Smoothness maximization	Typically only 2-D or 3-D
Regression, High Dimensional Model Representation(HDMR)	Linear, quadratic, low-degree multinomials, exponential ..	Least squares fitting, Newton/Levenbergh	Knowledge of regression function (family) crucial
Random forests	Decision trees	Training Error minimization	Typically used in discrete or mixed-integer case

Gaussian Processes for Big Data Regression Optimally Weighted Cluster Kriging

1000 pts



20000 pts



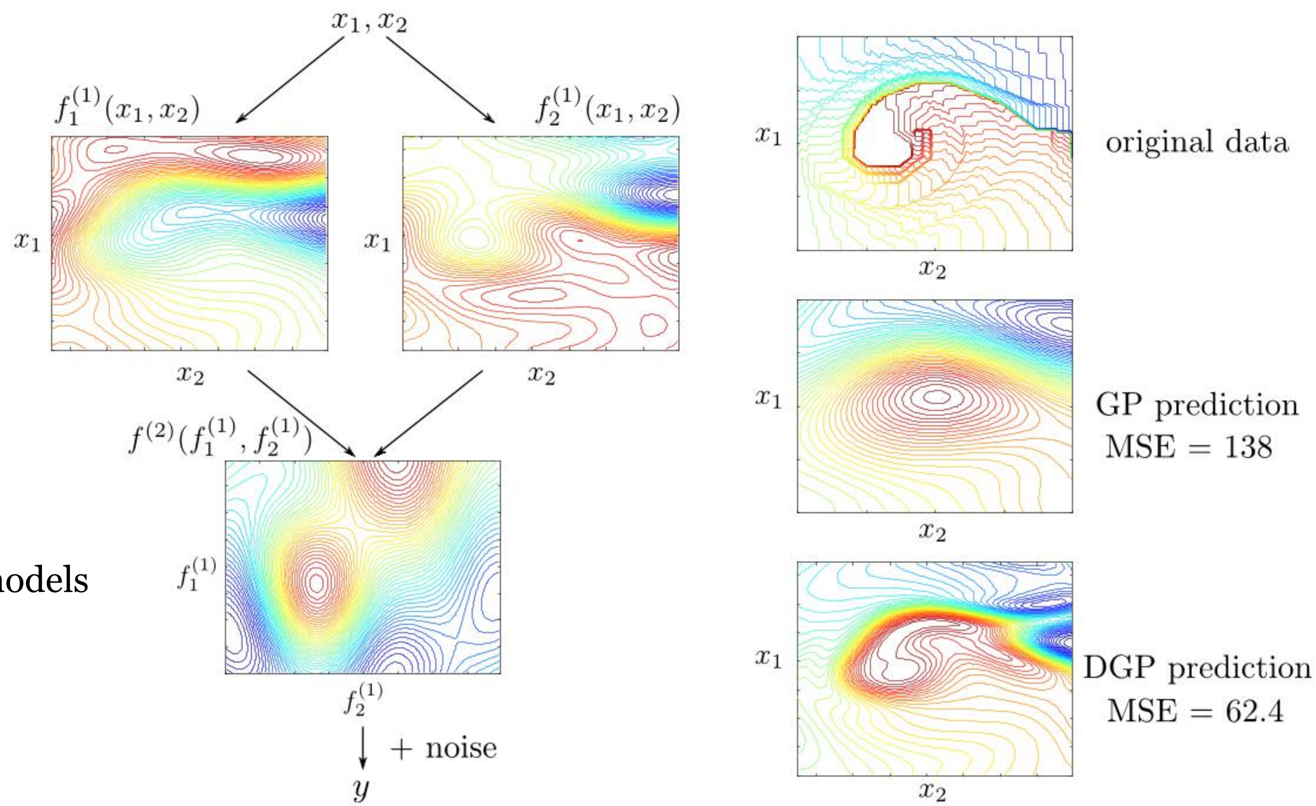
$$T(n,k) \in O\left(nk + \left(\frac{n}{k}\right)^3\right)$$

van Stein, Bas, Hao Wang, Wojtek Kowalczyk, Thomas Bäck, and Michael Emmerich. "Optimally Weighted Cluster Kriging for Big Data Regression." In International Symposium on Intelligent Data Analysis, pp. 310-321. Springer International Publishing, 2015.

Deep Gaussian Processes and Model Mixtures

Train k local models

Combine models

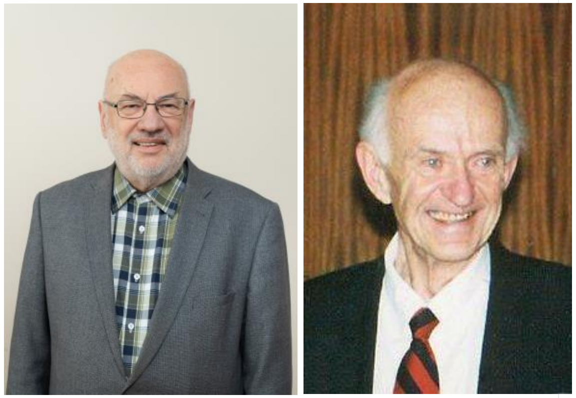


Damianou, A., & Lawrence, N. (2013, April). Deep gaussian processes. In Artificial Intelligence and Statistics (pp. 207-215).
Friese, M., Bartz-Beielstein, T., Bäck, T., Naujoks, B., & Emmerich, M. (2019, February). Weighted ensembles in model-based global optimization. In *AIP Conference Proceedings* (Vol. 2070, No. 1, p. 020003). AIP Publishing LLC.

Statistical Global Optimization*

Algorithm 1 Statistical global optimization

$D_0 \leftarrow \mathbf{evaluate}_f(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m_0)})$ {Initialize database}
$t \leftarrow m_0$ {Initialize evaluation counter}
while $t < t_{eval,max}$ **do**
    Search for $\mathbf{x}_t^* = \text{argmin}_{\mathbf{x} \in \mathbb{S}} \hat{f}_{sc}(D_t, \mathbf{x})$
    $y_t = f(\mathbf{x}_t^*)$
    if $y_t < y_{\min}^t$ **then**
        $\mathbf{x}_{\min}^t = \mathbf{x}_t^*$
        $y_{\min}^t = y_t$
    end if
    $D_{t+1} = D_t \cup \{(\mathbf{x}_t^*, y_t)\}$
end while
return $y_{\min}^t, \mathbf{x}_{\min}^t$



Antanas Žilinskas, Jonas Mockus
(Univ. Vilnius, Lithuania)

D. D. Cox and S. John. SDO: a statistical method for global optimization. In V. Hampton, editor, Multidisciplinary design optimization, volume 2, pages 315–329. SIAM, Philadelphia, PA, 1997.
H. J. Kushner. A versatile stochastic model of a function of unknown and ime varying form. Journal of Math. Anal. Appl., pages 150–167, 5 1962.
Žilinskas, A. and Mockus, J., 1972. On one Bayesian method of search of the minimum. Avtomatica i Vychislitel'naya Teknika, 4, pp.42-44.

*other names: Bayesian (Global) Optimization, Expected Improvement Algorithm, Sequential Global Optimization. Efficient Global Optimization.

Metamodel-Assisted Evolution Algorithms & Pattern Search

Functions = Fitness Landscapes

- Individual Control: Pre-select
- Generational Control: Alternate
- Similar ideas in direct deterministic methods: Model-Assisted Pattern Search
- Balancing exploration and exploitation, uncertainty quantification
- Artificial Neural Networks vs. Kriging

Yaochu Jin

Kyriakos Giannakoglou

Virginia Torczon

Andy Keane

Jin, Yaochu. "A comprehensive survey of fitness approximation in evolutionary computation." *Soft computing* 9, no. 1 (2005): 3-12.
El-Beltagy, M. A., & Keane, A. J. (2001). Evolutionary optimization for computationally expensive problems using gaussian processes. In Proc. Int. Conf. on Artificial Intelligence (Vol. 1, pp. 708-714).
Siefert, Christopher, Virginia Torczon, and Michael W. Trosset. "MAPS: Model-assisted pattern search, 1997–2000."
Giannakoglou, K. C., Giotis, A. P., & Karakasis, M. K. (2001). Low-cost genetic optimization based on inexact pre-evaluations and the sensitivity analysis of design parameters. Inverse Problems in Engineering, 9(4), 389-412.

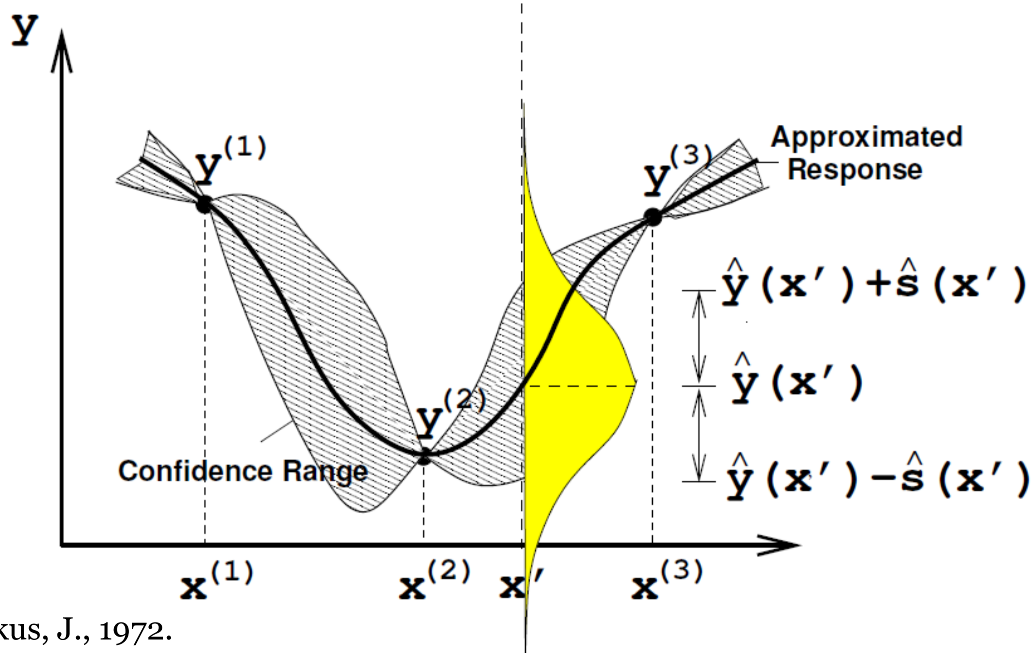
What is a good infill criterion?

- Naïve: Best predicted are best choices: $f_{sc}(x) = \hat{y}(x)$
- Better: Use local error estimate $\hat{s}(x)$ to **reward infill at unexplored regions**
  - Lower confidence bound (minimization)

$$f_{sc}(x) = \hat{y}(x) - \omega\hat{s}(x)$$

  - Expected improvement (maximization) using **improvement** $I(y(x)) = \max\left(0, \mathrm{f}_{\mathrm{min,t}} - \mathrm{y(x)}\right)$

$$f_{sc}(x) = E\big(I(x)\big) = \int I\big(y(x)\big) PDF_{\hat{s}(x),\hat{y}\{x)}(y)dy$$



Žilinskas, A. and Mockus, J., 1972.
On one Bayesian method of search of the minimum. Avtomatica i Vychislitel’naya Teknika, 4, pp.42-44

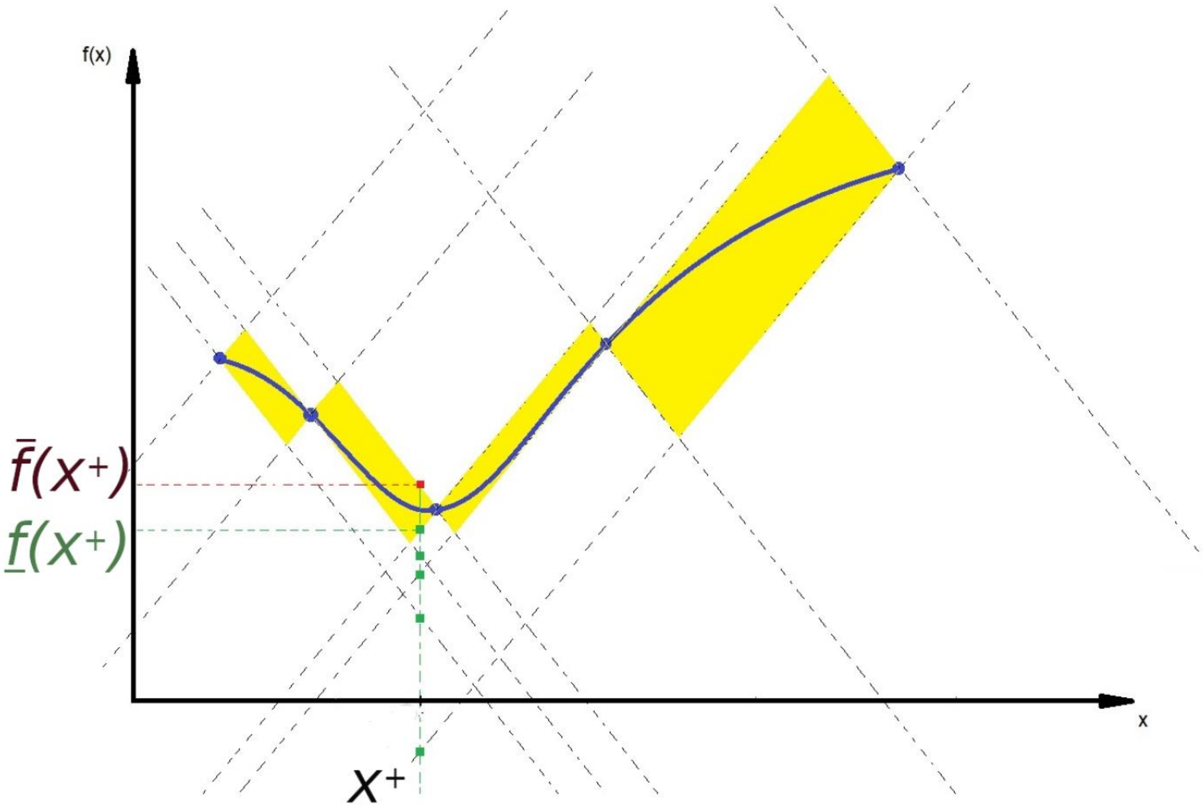
Lipschitzian optimization and Shubert's algorithm

- Classical Idea by Shubert: Use lowest Lipschitz Lower Bound
  as infill criterion

- B. O. Shubert, SIAM Journal on Numerical Analysis **9**, 379–388 (1972). https://doi.org/10.1137/0709036

- Otten, Heleen J., and Sander C. Hille. "A novel expected hypervolume improvement algorithm for Lipschitz multi-objective optimisation: Almost Shubert's algorithm in a special case." In *AIP Conference Proceedings*, vol. 2070, no. 1, p. 020031. AIP Publishing LLC, 2019.

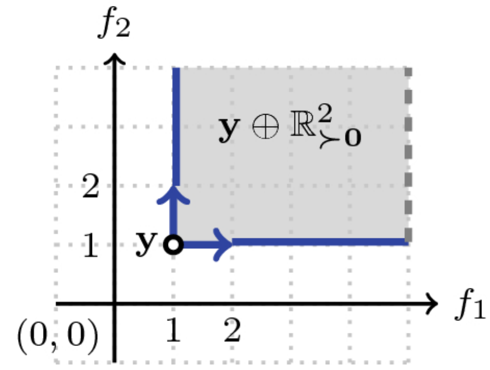
- Otten, Heleen, Hille, Sander .C. and Emmerich, M.T.M., Global optimization for Lipschitz continuous expensive black box functions. Leiden University, Technical Report, 2018



Surrogate Assisted multiobjective optimization …

What means improvement in MOO?

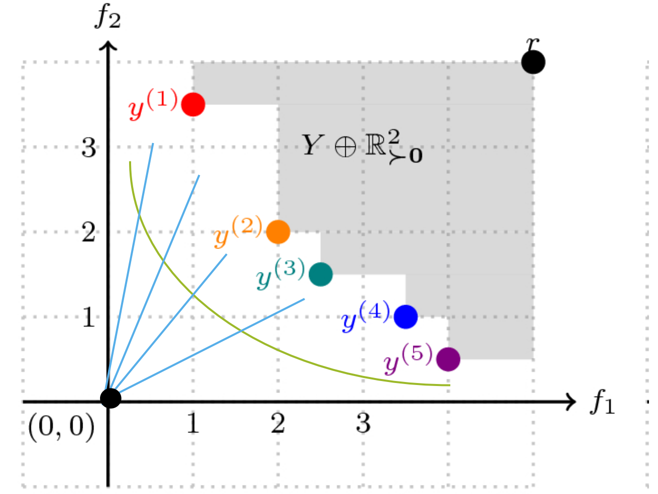
- Adding interesting non-dominated points
- Uncertainty about the preference/weights: Offer solutions for multiple weight combinations (ParEGO, S-RVEA)



Improving approximation sets to the Pareto front:

- Reducing gaps in Pareto front approximation by inverse modeling (predicting 'gap' fillers).
- Dominating more (hyper)volume or improve performance indicators (e.g. ,R2)

(R2 and Hypervolume Expected Improvement >>)



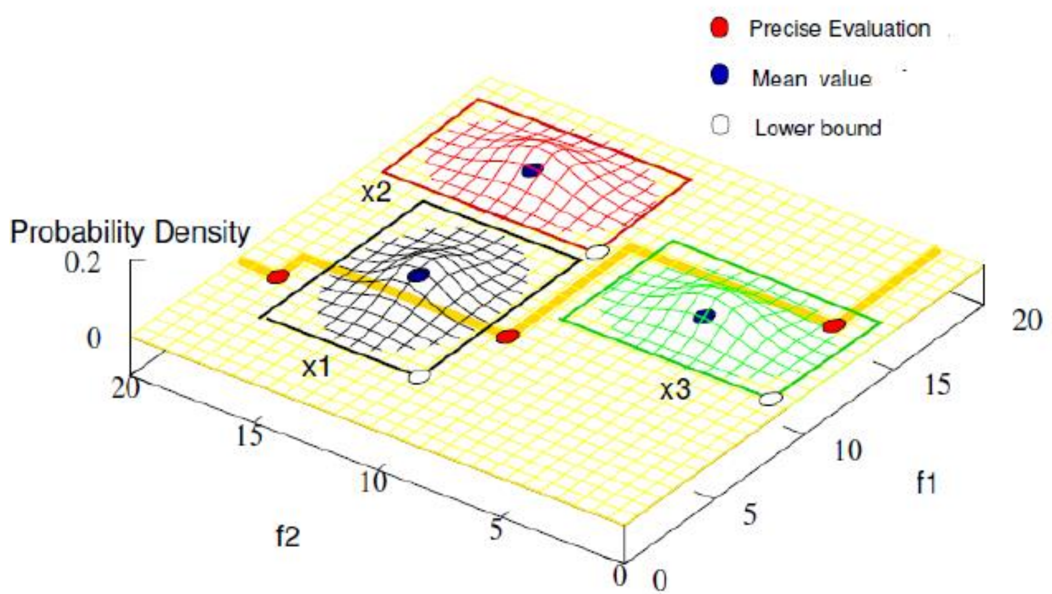
Knowles, Joshua. "ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems." *IEEE Transactions on Evolutionary Computation* 10.1 (2006): 50-66.

Emmerich, M. T., Giannakoglou, K. C., & Naujoks, B. (2006). Single-and multiobjective evolutionary optimization assisted by Gaussian random field metamodels. *IEEE Transactions on Evolutionary Computation*, *10*(4), 421-439.

Chugh, T., Jin, Y., Miettinen, K., Hakanen, J., & Sindhya, K. (2016). A surrogate-assisted reference vector guided evolutionary algorithm for computationally expensive many-objective optimization. *IEEE Transactions on Evolutionary Computation*, *22*(1), 129-142.

Cheng, R., Jin, Y., Narukawa, K. and Sendhoff, B., 2015. A multiobjective evolutionary algorithm using Gaussian process-based inverse modeling. *IEEE Transactions on Evolutionary Computation*, *19*(6), pp.838-856.

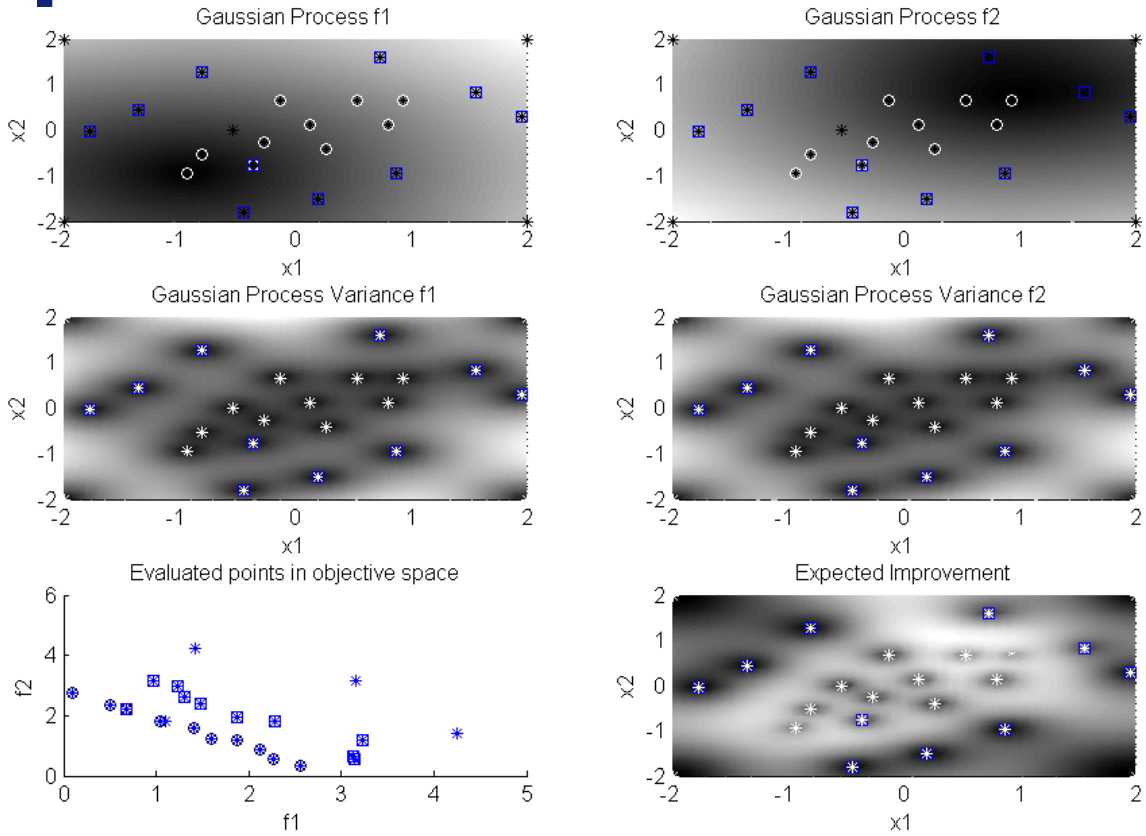
Multi-objective Surrogate-Assisted Optimization



Emmerich, Michael, and Boris Naujoks. "Metamodel assisted multiobjective optimisation strategies and their application in airfoil design." In *Adaptive computing in design and manufacture VI*, pp. 249-260. Springer, London, 2004.

Bayesian Multiobjective Global Optimization

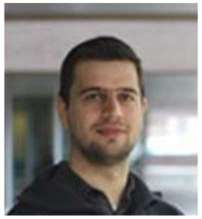
- Initial design 10 points
- 2-sphere problem
- 15 updates of archive based on maximal EHVI infill
- Infill happens in underexplored but promising regions
- Variance monotonicity (2-D), see Emmerich, Deutz, Klinkenberg (2011)

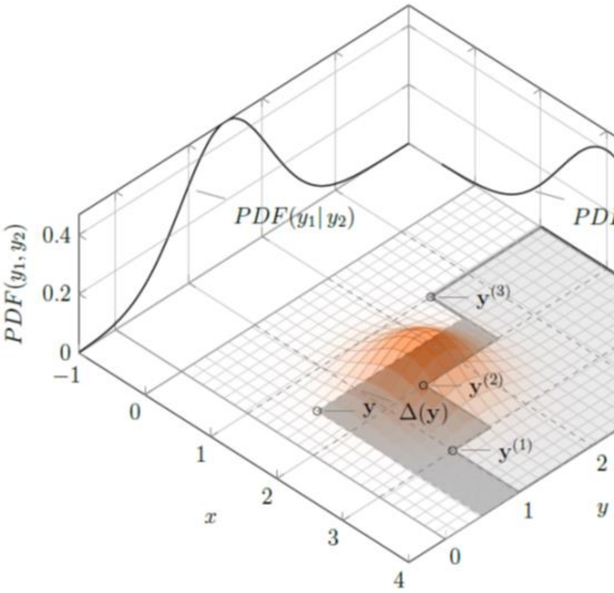


Emmerich, M., Yang, K., Deutz, A., Wang, H., & Fonseca, C. (2017). A Multicriteria Generalization of Bayesian Global Optimization In: Pardalos P and Zilinskas J. Advances in Stochastic and Deterministic Global Optimization (Springer Optimization and Its Applications)(English)". Springer, Berlin.
Emmerich, M. T., Deutz, A. H., & Klinkenberg, J. W. (2011, June). Hypervolume-based expected improvement: Monotonicity properties and exact computation. In Evolutionary Computation (CEC), 2011 IEEE Congress on (pp. 2147-2154). IEEE.

Fast Computation of Expected Hypervolume Improvement

- Integral was suggested in 2005 (my PhD thesis)
- First exact formulas in 2011 (law of Fubini, very expensive)
- Cheaper Alternatives:
  - SMS EGO (Wagner, Ponweiser)
  - S-EI (Shimoyama et al.)
- Faster Computation (Courckuyt & Deschrijver & Dhaene, Hupkens & Yang & Deutz & Emmerich)
- Fastest Computation , linear lime reduction to HV indicator
  - Yang, Fonseca, Emmerich , Deutz
  - Box decompositions Daechert, Lacour Klamroth , Fonseca et al.





$\Theta(nlogn)$ in 2D and 3D

(Emmerich, Yang, Fonseca Deutz, EMO, Munster, 2017)

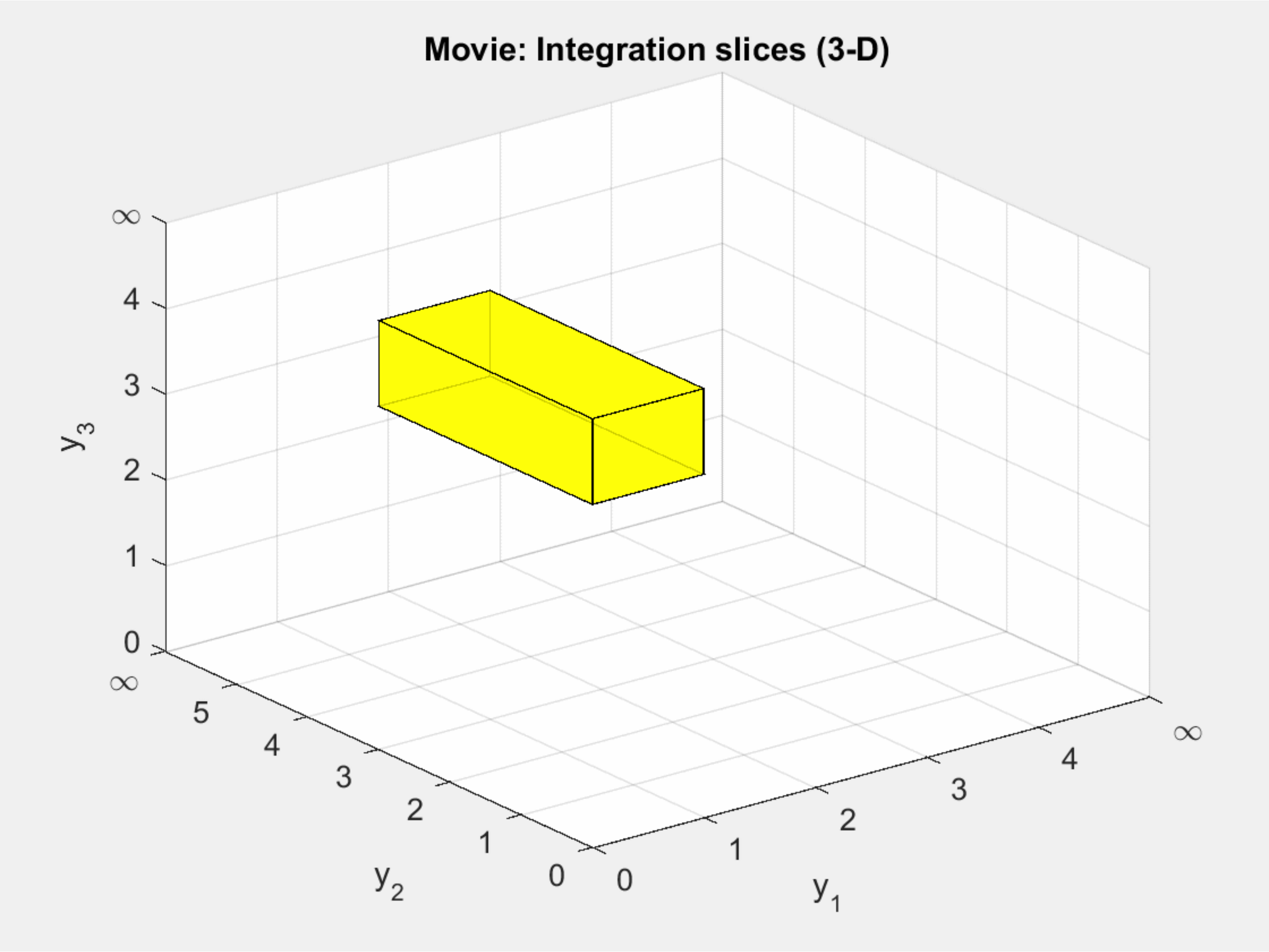
$O(n^{\lfloor \frac{d}{2} \rfloor})$ in 3D = number of partion boxes

2^d Integrals per box (not obvious at first)

Yang, Emmerich, Back, Deutz JOGO 2019

Emmerich M,. Single and Multiobjective Design Optimization, Assisted By Gaussian Random Field Metamodels, TU Dortmuntd 2005,
Couckuyt I, Deschrijver D, Dhaene T. Journal of Global Optimization. 2014 Nov 1;60(3):575-94.
Yang, K., Emmerich, M., Deutz, A., & Fonseca, C. M. (2017, March). Computing 3-D expected hypervolume improvement and related integrals in asymptotically optimal time. In *International Conference on Evolutionary Multi-Criterion Optimization* (pp. 685-700). Springer,
Yang, K., Emmerich, M., Deutz, A., & Bäck, T. (2019). Efficient computation of expected hypervolume improvement using box decomposition algorithms. *Journal of Global Optimization*, *75*(1), 3-34.

Efficient Box Decompositions of the 3-D dominated (hyper)volume

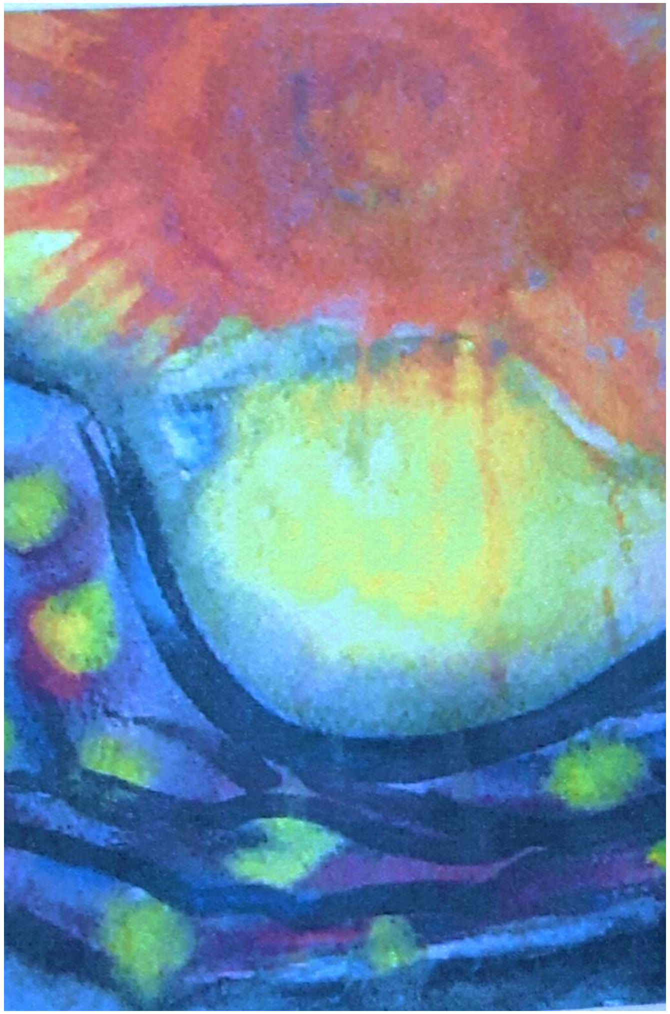


Future Horizons …

Top 10 Open Future Topics (… subjective)

1. Correlated Objectives – how to build metamodels for them
2. Constraint Modelling; NLP … beyond SQP
3. Navigation Methods; ≫3 Objectives
4. Heterogeneous Objectives
5. Batch Infill Criteria
6. Models for Big Data/Deep Learning
7. Mixed Integer Surrogate Models
8. Optimal Budget Allocation; Exact Shells
9. Explainable AI; Innovization
10. High input dimensions, discontinuity

“Sommer” Maria Emmerich



First work in these topics is already available, but there are many challenges still

Literature

Future Challenges:
Allmendinger, Richard, et al. "Surrogate-assisted multicriteria optimization: Complexities, prospective solutions, and business case." *Journal of Multi-Criteria Decision Analysis* 24.1-2 (2017): 5-24.

Correlated Objectives – how to build metamodels for them
Wang, Z., Hutter, F., Zoghi, M., Matheson, D., & de Feitas, N. (2016). Bayesian optimization in a billion dimensions via random embeddings. *Journal of Artificial Intelligence Research*, *55*, 361-387.

Constraint Modelling; NLP ... beyond SQP
Bagheri, S., Konen, W., Allmendinger, R., Branke, J., Deb, K., Fieldsend, J., ... & Sindhya, K. (2017, July). Constraint handling in efficient global optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference* (pp. 673-680).

Heterogeneous Objectives
Allmendinger, Richard, and Joshua Knowles. "On handling ephemeral resource constraints in evolutionary search." *Evolutionary computation* 21, no. 3 (2013): 497-531.

Batch Infill Criteria
Janusevskis, J., Le Riche, R., Ginsbourger, D., & Girdziusas, R. (2012, January). Expected improvements for the asynchronous parallel global optimization of expensive functions: Potentials and challenges. In *International Conference on Learning and Intelligent Optimization* (pp. 413-418). Springer, Berlin, Heidelberg.

Models for Big Data/Deep Learning
Wang, Hao, Bas van Stein, Michael Emmerich, and Thomas Bäck. "Time complexity reduction in efficient global optimization using cluster kriging." In *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 889-896. 2017.
Dutordoir, V., Knudde, N., van der Herten, J., Couckuyt, I., & Dhaene, T. (2017, December). Deep gaussian process metamodeling of sequentially sampled non-stationary response surfaces. In 2017 Winter Simulation Conference (WSC) (pp. 1728-1739). IEEE.

Mixed Integer Surrogate Models
Bartz-Beielstein, T., & Zaefferer, M. (2017). Model-based methods for continuous and discrete global optimization. *Applied Soft Computing*, *55*, 154-167.

Optimal Budget Allocation; Exact Shells
Hutter, Frank, et al. "Time-bounded sequential parameter optimization." *International Conference on Learning and Intelligent Optimization*. Springer, Berlin, Heidelberg, 2010.
Žilinskas, A., & Calvin, J. (2019). Bi-objective decision making in global optimization based on statistical models. *Journal of Global Optimization*, *74*(4), 599-609.

Explainable AI; Innovization
Gaier, Adam, Alexander Asteroth, and Jean-Baptiste Mouret. "Data-efficient design exploration through surrogate-assisted illumination." *Evolutionary computation* 26, no. 3 (2018): 381-410.
Bandaru, Sunith, Amos HC Ng, and Kalyanmoy Deb. "Data mining methods for knowledge discovery in multi-objective optimization: Part A-Survey." *Expert Systems with Applications* 70 (2017): 139-159.

High input dimensions, discontinuity
Gaudrie, David. "High-Dimensional Bayesian Multi-Objective Optimization." PhD diss., 2019.

Moo-chas gracias

UNIVERSITY OF JYVÄSKYLÄ