Laboratories

SAND2020-7665C

A sequential regularized constrained multi-objective Bayesian optimization for design applications

ASME IDETC-CIE 2020. August 17–29, 2020, Virtual Conference.

Anh Tran, Mike Eldred (SNL), Scott McCann (Xilinx), Yan Wang (Gatech)

7/27/20

ratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia,
neywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract

Acknowledgment

Joint work with Mike Eldred (SNL), Scott McCann (SNL), and Yan Wang (Gatech).

Figure: Anh Tran (SNL)

Figure: Mike Eldred (SNL)

Figure: Scott McCann (Xilinx)

Figure: Yan Wang (Georgia Tech)

Funded by DOE/NNSA/ASC.

Advantages/Disadvantages

Advantages:

- optimize with uncertainty consideration
- active machine learning (accumulate history)
- derivative free (avoid computing Jacobian)
- global optimization (convergence in probability)
- good convergence rate (provably asymptotic regret)

Disadvantages:

- high-dimensionality
- scalability: computational bottleneck $\mathcal{O}(n^3)$ when the number of observations $n \geq \mathcal{O}(10^3)$

BO features

Sandia National Laboratories

very **versatile** (adaptive for methodological extensions)

- acquisition functions: PI, EI, UCB, Thompson sampling, entropy-based, knowledge-gradient, and much more,
- constrained on objectives (known + unknown constraints) ✓
- **multi-objective** (Pareto frontier/optimal, domination) ✓
- multi-output ✗
- multi-fidelity (couple multiple low-, high-fidelity models) ✓
- batch parallelization ✓ → asynchronous parallel ✓
- stochastic ✗
- time-series (forecasting, e.g. causal kernel) ✗
- mixed-integer (discrete/categorical + continuous) ✓
- scalable ✓
- latent variable model ✗
- gradient-enhanced ✓
- physics-constrained: monotonic, discontinuous, symmetry, bound ✗
- outlier: student-t distribution ✗
- non-stationary ✗

Methodology:

- Anh Tran et al. "srMO-BO-3GP: A sequential regularized multi-objective constrained Bayesian optimization for design applications". In: *Proceedings of the ASME 2020 IDETC/CIE*. vol. Volume 1: 40th Computers and Information in Engineering Conference. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. American Society of Mechanical Engineers. Aug. 2020
- Anh Tran et al. "aphBO-2GP-3B: A budgeted asynchronously-parallel multi-acquisition for known/unknown constrained Bayesian optimization on high-performing computing architecture". In: *arXiv preprint arXiv:2003.09436* (2020)
- Anh Tran, Tim Wildey, and Scott McCann. "sMF-BO-2CoGP: A sequential multi-fidelity constrained Bayesian optimization for design applications". In: *Journal of Computing and Information Science in Engineering* 20.3 (2020), pp. 1–15
- Anh Tran, Tim Wildey, and Scott McCann. "sBF-BO-2CoGP: A sequential bi-fidelity constrained Bayesian optimization for design applications". In: *Proceedings of the ASME 2019 IDETC/CIE*. vol. Volume 1: 39th Computers and Information in Engineering Conference. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. V001T02A073. American Society of Mechanical Engineers. Aug. 2019
- Anh Tran, Minh Tran, and Yan Wang. "Constrained mixed-integer Gaussian mixture Bayesian optimization and its applications in designing fractal and auxetic metamaterials". In: *Structural and Multidisciplinary Optimization* (2019), pp. 1–24
- Anh Tran et al. "pBO-2GP-3B: A batch parallel known/unknown constrained Bayesian optimization with feasibility classification and its applications in computational fluid dynamics". In: *Computer Methods in Applied Mechanics and Engineering* 347 (2019), pp. 827–852

Applications:

- Anh Tran et al. "Multi-fidelity machine-learning with uncertainty quantification and Bayesian optimization for materials design: Application to ternary random alloys". In: *arXiv preprint arXiv:2006.00139* (2020)
- Anh Tran et al. "An active-learning high-throughput microstructure calibration framework for process-structure linkage in materials informatics". In: *Acta Materialia* 194 (2020), pp. 80–92
- Stefano Travaglino et al. "Computational optimization study of transcatheter aortic valve leaflet design using porcine and bovine leaflets". In: *Journal of Biomechanical Engineering* 142 (1 2020)
- Anh Tran et al. "WearGP: A computationally efficient machine learning framework for local erosive wear predictions via nodal Gaussian processes". In: *Wear* 422 (2019), pp. 9–26
- Anh Tran, Lijuan He, and Yan Wang. "An efficient first-principles saddle point searching method based on distributed kriging metamodels". In: *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering* 4.1 (2018), p. 011006

Classical GP

Let $\mathcal{D}_n = \{\boldsymbol{x}_i, y_i\}_{i=1}^n$ denote the set of observations and $\boldsymbol{x}$ denote an arbitrary test points

$$\mu_n(\boldsymbol{x}) = \mu_0(\boldsymbol{x}) + \boldsymbol{k}(\boldsymbol{x})^T(\boldsymbol{K} + \sigma^2\boldsymbol{I})^{-1}(\boldsymbol{y} - \boldsymbol{m}) \quad (1)$$

$$\sigma_n^2(\boldsymbol{x}) = k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}(\boldsymbol{x})^T(\boldsymbol{K} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{k}(\boldsymbol{x}) \quad (2)$$

where $\boldsymbol{k}(\boldsymbol{x})$ is a vector of covariance terms between $\boldsymbol{x}$ and $\boldsymbol{x}_{1:n}$.

Classical GP

Formulation:

- stationary: only depends on $r = ||\boldsymbol{x} - \boldsymbol{x'}||$
- the covariance matrix: symmetric positive-semidefinite matrix made up of pairwise inner products

$$\boldsymbol{K}_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{3}$$

- unknown function is *presumably* smooth
- variables are *presumably* continuous, i.e. $f : \mathbb{R}^d \to \mathbb{R}$.

Implementation:

- MLE to estimate the hyper-parameter $\theta \in \mathbb{R}^d$: compute $\boldsymbol{K}^{-1}$ at the cost of $\mathcal{O}(n^3)$
- size of $\boldsymbol{K} = n \times n$

Ingredients: some data, GP kernel, acquisition function.

Classical GP

Common kernels:

- $k_{\text{Matérn1}}(\boldsymbol{x}, \boldsymbol{x'}) = \theta_0^2 \exp(-r)$
- $k_{\text{Matérn3}}(\boldsymbol{x}, \boldsymbol{x'}) = \theta_0^2 \exp(-\sqrt{3}r)(1 + \sqrt{3}r)$
- $k_{\text{Matérn5}}(\boldsymbol{x}, \boldsymbol{x'}) = \theta_0^2 \exp(-\sqrt{5}r)\left(1 + \sqrt{5}r + \frac{5}{3}r^2\right)$
- $k_{\text{sq-exp}}(\boldsymbol{x}, \boldsymbol{x'}) = \theta_0^2 \exp\left(-\frac{1}{2}r^2\right)$

Log-likelihood function:

$$\begin{aligned} \log p(\boldsymbol{y}|\boldsymbol{x}_{1:n}, \theta) \quad = \quad & -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{K}^\theta + \sigma^2 \boldsymbol{I}| \\ & -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{m}_\theta)^T(\boldsymbol{K}^\theta + \sigma^2\boldsymbol{I})^{-1}(\boldsymbol{y} - \boldsymbol{m}_\theta) \end{aligned} \tag{4}$$

Formulation & Implementation: multi-objective GP

Let:

- $\boldsymbol{x} = \{x_i\}_{i=1}^{d} \in \mathcal{X} \subseteq \mathbb{R}^d$ be input in d-dimensional space,
- $\boldsymbol{y} = \{y_j\}_{j=1}^{s}$ as s outputs.

$$\underset{\boldsymbol{x} \in \mathcal{X}}{\operatorname{argmax}} (f_1(\boldsymbol{x}), \cdots, f_s(\boldsymbol{x})) \tag{5}$$

subjected to $\boldsymbol{c}(\boldsymbol{x}) \leq \mathbf{0}$.

Formulation & Implementation: multi-objective GP

Pareto definition:

- $\boldsymbol{x}_1$ is said to dominate $\boldsymbol{x}_2$, denoted as $\boldsymbol{x}_1 \preceq \boldsymbol{x}_2$, if and only if $\forall 1 \leq j \leq s$, such that $y_j(\boldsymbol{x}_1) \leq y_j(\boldsymbol{x}_2)$, and $\exists 1 \leq j \leq s$, such that $y_j(\boldsymbol{x}_1) < y_j(\boldsymbol{x}_2)$.
- $\boldsymbol{x}_1$ is said to strictly dominate $\boldsymbol{x}_2$, denoted as $\boldsymbol{x}_1 \prec \boldsymbol{x}_2$, if and only if $\forall 1 \leq j \leq s$, such that $y_j(\boldsymbol{x}_1) < y_j(\boldsymbol{x}_2)$.

Scalarization: multi-objective $\rightarrow$ single-objective

1. weighted Tchebycheff $y = \max_{1 \leq i \leq s} w_i(y_i(\boldsymbol{x}) - z_i^*)$,
2. weighted sum $y = \sum_{i=1}^{s} w_i y_i(\boldsymbol{x})$,
3. augmented Tchebycheff
   $y = \max_{1 \leq i \leq s} w_i(y_i(\boldsymbol{x}) - z_i^*) + \rho \sum_{i=1}^{s} w_i y_i(\boldsymbol{x})$,

where z_i^* denotes the ideal value for the i-th objective, the weights $0 \leq w_i \leq 1$, $\sum_{i=1}^{m} w_i = 1$, ρ is a small positive constant ($\rho = 0.05$).

Formulation & Implementation: multi-objective GP

What is the idea?

- learn Pareto frontier through *a uncertain binary classifier*
- GP comes out naturally as an uncertain classifier
- push the learned Pareto frontier to the true Pareto frontier
- use acquisition function in BO to promote richness (exploration) and diversity (exploitation)
- once learned Pareto frontier $\rightarrow$ true Pareto frontier: the uncertain classifier stabilizes
- some regularization to help (a) fitting objective GP and (b) reduce noise in stochastic settings

Formulation & Implementation: multi-objective GP

Acquisition function:

$$a(\boldsymbol{x}) = \underbrace{a_{\text{obj}}(\boldsymbol{x})}_{\text{objective GP}} \cdot \underbrace{a_{\text{Pareto}}(\boldsymbol{x})}_{\text{uncertain Pareto}} \cdot \underbrace{\Pr(\boldsymbol{x}|c(\boldsymbol{x}) = 1)}_{\text{unknown constraints}} \cdot \underbrace{\mathcal{I}(\boldsymbol{x})}_{\text{known constraints}} \quad (6)$$

- $a_{\text{obj}}(\boldsymbol{x})$: objective GP fitted through augmented Tchebycheff with random weights
- $a_{\text{Pareto}}(\boldsymbol{x})$: Pareto GP classifier (Pareto/non-Pareto)
- $\Pr(\boldsymbol{x}|c(\boldsymbol{x}) = 1)$: constrained classifier (feasible/infeasible)
- $\mathcal{I}(\boldsymbol{x})$: indicator function if $\boldsymbol{c}(\boldsymbol{x}) \leq \mathbf{0}$

Formulation & Implementation: multi-objective GP

Algorithm 1 srMO-BO-3GP algorithm

Input: dataset $\mathcal{D}_n$ consisting of input, observation, feasibility $(\boldsymbol{x}, \boldsymbol{y}, c)_{i=1}^n$
Input: multi-objective $(\boldsymbol{x}_i, \boldsymbol{y}_i)_{i=1}^n$, constraint GP $(\boldsymbol{x}, c_i)_{i=1}^n$,

for $n = 1, 2, \dots,$ **do**
  randomize a weight vector $\boldsymbol{w}$
  combine $\{y_j\}_{j=1}^s$ to y ▷ multi- to single-objective
  construct single-objective GP ▷ GP #1: $\mu_{\text{obj}}(\boldsymbol{x})$, $\sigma^2_{\text{obj}}(\boldsymbol{x})$
  construct Pareto front ▷ GP #2: $\mu_{\text{Pareto}}(\boldsymbol{x})$, $\sigma^2_{\text{Pareto}}(\boldsymbol{x})$
    find current Pareto front
    construct Pareto classifier GP
  construct constraints classifier GP ▷ GP #3: $\mu_{\text{feasible}}(\boldsymbol{x})$, $\sigma^2_{\text{feasible}}(\boldsymbol{x})$
  locate the next sampling point $\boldsymbol{x}_{n+1}$
   query for $\boldsymbol{y}_{n+1} = \{y_j\}_{j=1}^s$, feasibility c_{n+1}
   augment dataset $\mathcal{D}_{n+1} = \{\mathcal{D}_n, (\boldsymbol{x}_{n+1}, \boldsymbol{y}_{n+1}, c_{n+1})\}$
end for

Benchmark: 18 variants

- EI, PI, UCB for objective
- EI, PI, UCB for Pareto frontier
- regularized vs. non-regularized

Acquisition function: How to pick the next point(s)

Sandia National Laboratories

- dictates how to pick the next point: **exploitation** (focus on the promising region) or **exploration** (focus on the uncertain/unknown region)
- different flavors:
  1. **probability of improvement** (PI)

$$a_{\text{PI}}(\boldsymbol{x}; \{\boldsymbol{x}_i, y_i\}_{i=1}^n, \theta) = \Phi(\gamma(\boldsymbol{x})), \tag{7}$$

  where

$$\gamma(\boldsymbol{x}) = \frac{\mu(\boldsymbol{x}; \{\boldsymbol{x}_i, y_i\}_{i=1}^n, \theta) - f(\boldsymbol{x}_{\text{best}})}{\sigma(\boldsymbol{x}; \{\boldsymbol{x}_i, y_i\}_{i=1}^n, \theta)}, \tag{8}$$

  2. **expected improvement** (EI) scheme

$$a_{\text{EI}}(\boldsymbol{x}; \{\boldsymbol{x}_i, y_i\}_{i=1}^n, \theta) = \sigma(\boldsymbol{x}; \{\boldsymbol{x}_i, y_i\}_{i=1}^n, \theta) \cdot (\gamma(\boldsymbol{x})\Phi(\gamma(\boldsymbol{x})) + \phi(\gamma(\boldsymbol{x})) \tag{9}$$

Acquisition function: How to pick next point(s)

- dictates how to pick the next point: **exploitation** (focus on the promising region) or **exploration** (focus on the uncertain/unknown region)
- different flavors:

3. **upper confidence bound** (UCB) scheme

$$a_{\text{UCB}}(\boldsymbol{x}; \{\boldsymbol{x}_i, y_i\}_{i=1}^{n}, \theta) = \mu(\boldsymbol{x}; \{\boldsymbol{x}_i, y_i\}_{i=1}^{n}, \theta) + \kappa\sigma(\boldsymbol{x}; \{\boldsymbol{x}_i, y_i\}_{i=1}^{n}, \theta), \quad (10)$$

where κ is a hyper-parameter describing the exploitation-exploration balance.

4. **pure exploration***:
    - maximal MSE σ^2
    - maximal IMSE $\int_{\boldsymbol{x} \in \mathcal{X}} \sigma^2(\boldsymbol{x})$

ZDT1: 12d

$$f_1 = x_1, f_2 = gh, \tag{11}$$

where

$$g = 1 + \frac{9}{11}\sum_{p=2}^{12} x_p, h = 1 - \sqrt{\frac{f_1}{g}}. \tag{12}$$

$1 \leq g \leq 10 \Rightarrow$ Pareto frontier is obtained when $g = 1$. More precisely, $\{f_1, f_2\}$, where $f_2 = 1 - \sqrt{f_1}$.

ZDT1: 12-d

Figure: ZDT1 Pareto frontier comparison.

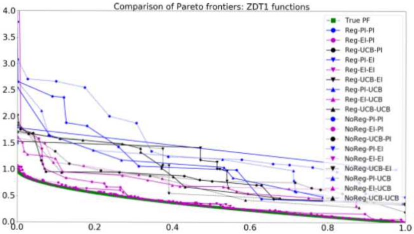
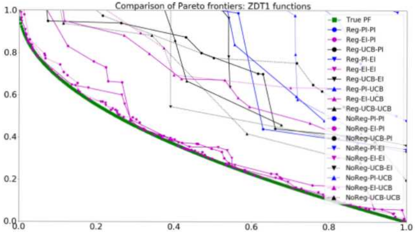


Figure: Zoom at $[0,1] \times [0,1]$.



DTLZ1: 12-d

$$f_1 = 0.5(1+g)x_1, \tag{13}$$

$$f_2 = 0.5(1+g)(1-x_1), \tag{14}$$

where $g = 10011 + \sum_{p=2}^{12} \langle (x_p - 0.5)^2 - \cos\left[20\pi(x_p - 0.5)\right] \rangle$.

DTLZ1: 12-d

Figure: DTLZ1 Pareto frontier comparison.

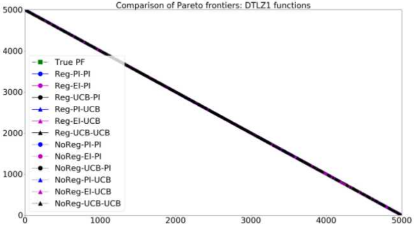
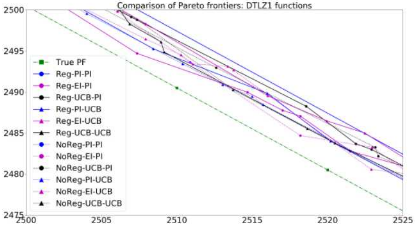


Figure: Zoom at $[2500, 2525] \times [2475, 2500]$.



Flip-chip BGA package design

Table: Design variables for the FCBGA design optimization.

Variable	Design part	Lower bound	Upper bound	Optimal value
x_1	die	20000	30000	20702
x_2	die	300	750	320
x_3	substrate	30000	40000	35539
x_4	substrate	100	1800	1614
x_5	substrate	$10 \cdot 10^{-6}$	$17 \cdot 10^{-6}$	$17 \cdot 10^{-6}$
x_6	stiffener ring	2000	6000	4126
x_7	stiffener ring	100	2500	1646
x_8	stiffener ring	$8 \cdot 10^{-6}$	$25 \cdot 10^{-6}$	$8.94 \cdot 10^{-6}$
x_9	underfill	1.0	3.0	1.52
x_{10}	underfill	0.5	1.0	0.804
x_{11}	PCB board	$12.0 \cdot 10^{-6}$	$16.7 \cdot 10^{-6}$	$16.7 \cdot 10^{-6}$

Flip-chip BGA package design

Figure: Warpage at -40°C

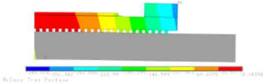


Figure: Warpage at 20°C

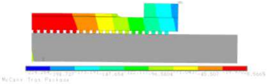
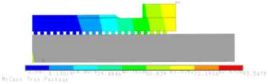


Figure: Warpage at 200°C



Flip-chip BGA package design

Figure: Correlation between objectives and joint densities.

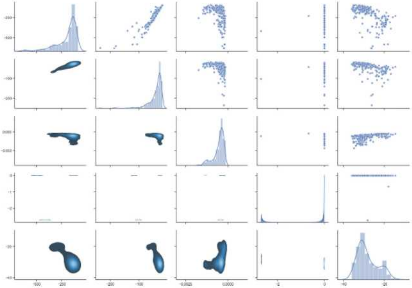
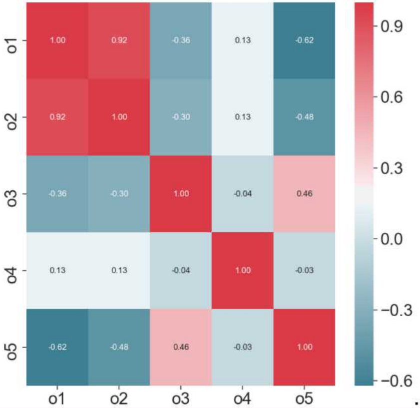


Figure: Correlation heatmap between objectives.



Additive manufacturing: inverse process-structure mapping: asynchronous parallel

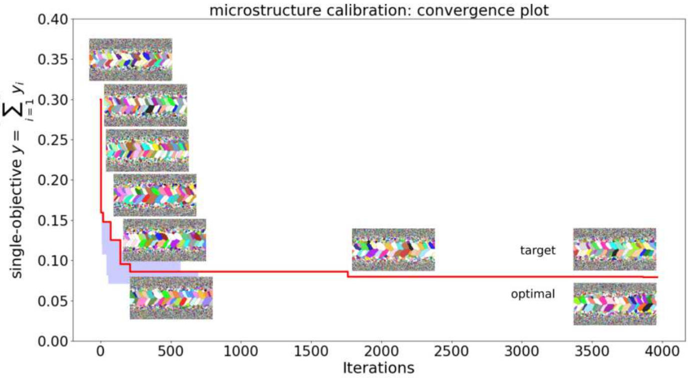


Figure: Reverse engineering an additively manufactured specimen through kinetic Monte Carlo (Sandia/SPPARKS). Tran et al., "An active-learning high-throughput microstructure calibration framework for process-structure linkage in materials informatics".

Sequential multi-fidelity: searching optimal chemical composition

Tran et al., "Multi-fidelity machine-learning with uncertainty quantification and Bayesian optimization for materials design: Application to ternary random alloys": coupling DFT and MD. Multi-fidelity for multi-scale ICME.

Figure: Iteration 4: 2 LF + 2 HF

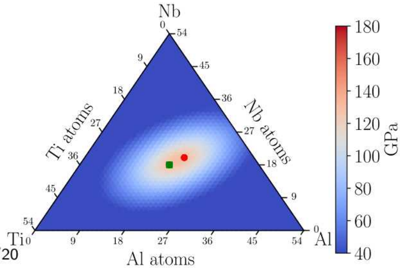
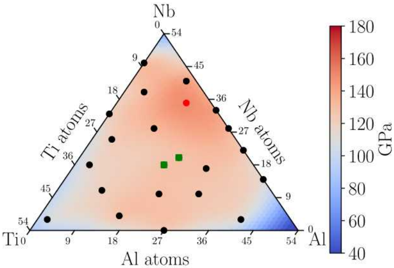


Figure: Iteration 24: 21 LF + 3 HF



Sequential multi-fidelity: searching optimal chemical composition

Tran et al., "Multi-fidelity machine-learning with uncertainty quantification and Bayesian optimization for materials design: Application to ternary random alloys": coupling DFT and MD. Multi-fidelity for multi-scale ICME.

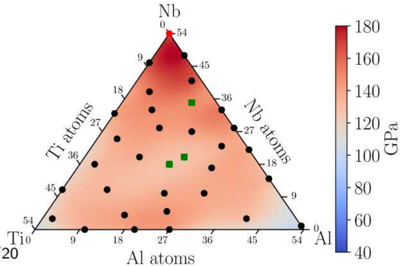


Figure: Iteration 35: 31 LF + 4 HF

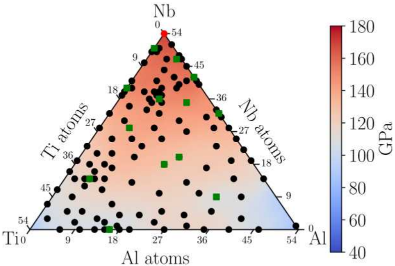


Figure: Iteration 130: 116 LF + 14 HF

Welded-beam design optimization (2d+4d) (mixed-integer)

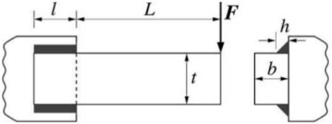


Figure: Welded-beam design

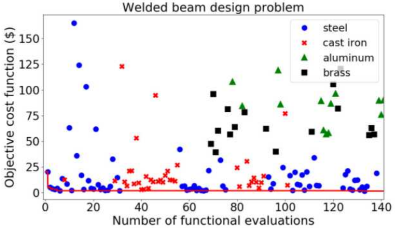


Figure: Convergence plot of welded beam design.

Speed reducer design optimization (1d+6d) (mixed-integer)

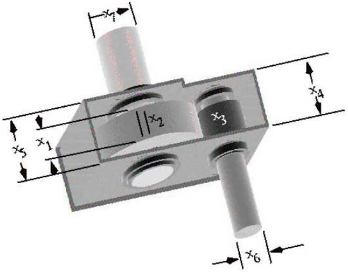


Figure: Speed reducer design

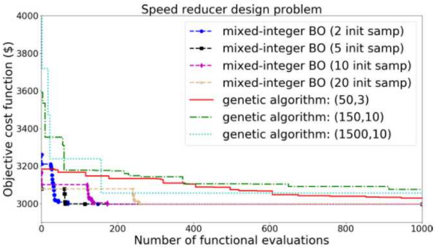


Figure: Comparison against GA.

High-dimensional discrete sphere function (5d+50d) (mixed-integer)

$f(\boldsymbol{x}^{(d)}, \boldsymbol{x}^{(c)}) =$
$f(x_1, \cdots, x_n, x_{n+1}, \cdots, x_m) =$
$\prod_{i=1}^{n} |x_i| \left(\sum_{j=n+1}^{m} x_j^2 \right)$
where
$1 \leq x_i \leq 2 (1 \leq i \leq n)$ are
n integer variables and
$-5.12 \leq x_j \leq$
$5.12 (n+1 \leq j \leq m)$ are
$m - n$ continuous
variables.

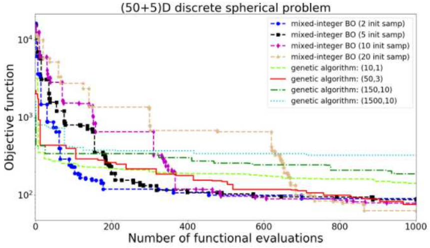


Figure: Comparison against GA.

High-dimensional discrete sphere function (5d+100d) (mixed-integer)

$f(\boldsymbol{x}^{(d)}, \boldsymbol{x}^{(c)}) =$
$f(x_1, \cdots, x_n, x_{n+1}, \cdots, x_m) =$
$\prod_{i=1}^{n} |x_i| \left(\sum_{j=n+1}^{m} x_j^2 \right)$
where
$1 \leq x_i \leq 2 (1 \leq i \leq n)$ are n integer variables and $-5.12 \leq x_j \leq 5.12 (n+1 \leq j \leq m)$ are $m - n$ continuous variables.

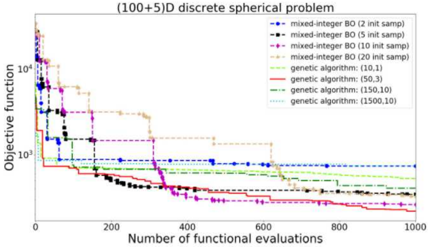


Figure: Comparison against GA.

(Fractal) mechanical metamaterials/AM (4d+5d) (mixed-integer)

Figure: Hierarchical multiscale structure of octahedral (second-order). Printed in Georgia Tech Invention Studio.

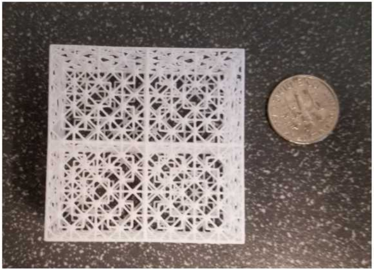


Figure: Design optimization of fractal cube. Printed in Georgia Tech Invention Studio.

(Fractal) mechanical metamaterials/AM (4d+5d) (mixed-integer)

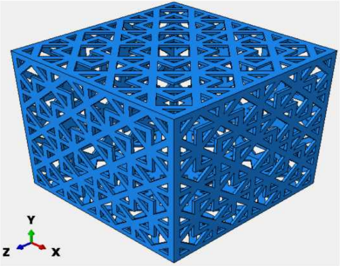


Figure: Parametric design.

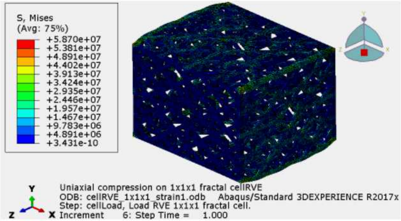


Figure: ABAQUS FEA.

Impeller design optimization using CFD (33d) (pBO-2GP-3B: parallel + blind constraints)

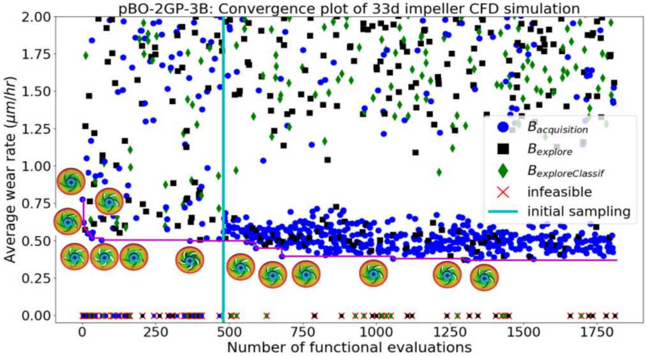
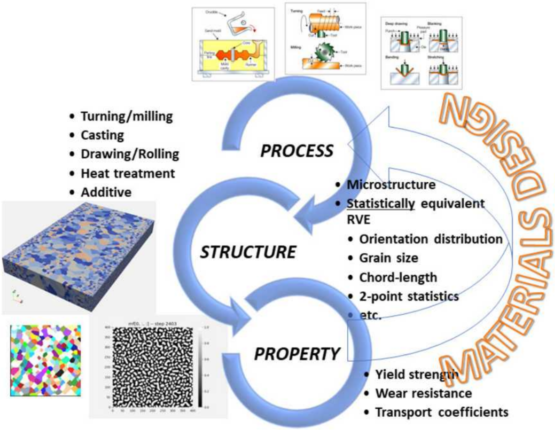


Figure: Multiphase CFD simulation for design optimization of 33d slurry pump impeller: Convergence plot .

Materials Design

process-structure-property linkage in materials



Inverse problems: find some processes such that *one* or *multiple* properties are optimized

Figure: Process-Structure-Property linkage.

Conclusion

This talk: **multi-objective** Bayesian optimization

- an uncertain Pareto classifier is proposed
- augment the acquisition function to help convergence on Pareto frontier
- design engineering and materials science applications

Thank you for listening.