Enhancing DGEMO with Bayesian Optimization Properties: Towards DGEBO

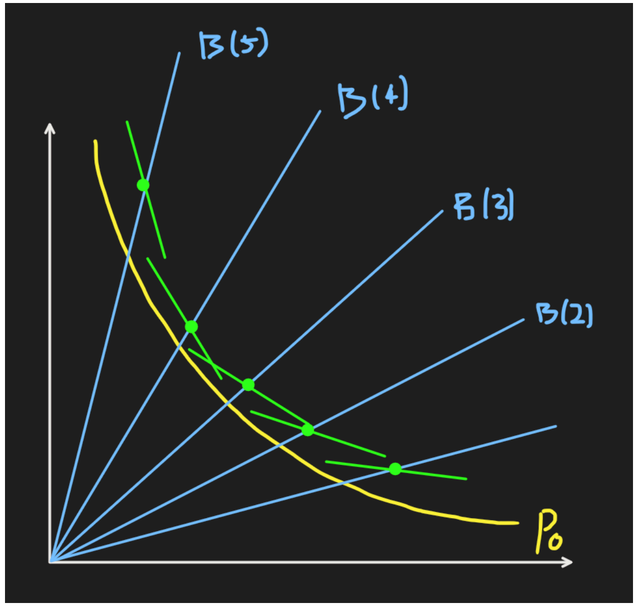
Summary

- Adding more Bayesian Optimization properties to DGEMO may improve its performance.

How?

1. Stochastic multivariate objective in First order approximation

- As is)
  - $\tilde{F} = (\mu_1, \cdots, \mu_d)$ where $\mu_j = k_j K_j^{-1} Y \forall j$
- To be)
  - $\tilde{F} \sim N(\mu, \Sigma)$



2. Modify Stochastic Sampling to use a BO approach

- As is)

  i. $\mathbf{x}_s = \mathbf{x}^j + \frac{1}{2^{\delta_p}}\mathbf{d}_p$: random sampling

- To be)
  - Exploration method from BO.
    - Acquisition functions like EI.

DGEMO Review

- MOO problem with $F = (f_1, \cdots, f_d)$
- Use GP as a surrogate model of F as
  - $\tilde{F} = (\tilde{f}_1, \cdots, \tilde{f}_d) \quad \text{where } \tilde{f}_j \sim N(m_j, k_j), \quad \begin{cases} m_j = 0 \\ k_j \text{ is a Matern Kernel,} \end{cases} \forall j$
- Use the mean function as the acquisition function.
  - $\tilde{f}_j = \mu_j = k_j K_j^{-1} Y \forall j$
- Use affine subspaces $\mathcal{A}_i$ near the samples derived with the First-Order Approximation.
  - Jacobian and Hessian of μ_j
- Use batch selection X_B to run parallel when deriving the final Pareto front.

DGEMO's Limit

1. Not fully utilizes the GP.

- Simply using the posterior μ_j for the first order approximation.
- Not fully utilizing the posterior variance Σ_j^2 might be wasting the valuable info.

2. Arbitrary Sampling procedure in the First-Order Approximation.

- From the previous candidate $\mathbf{x}_j$ in the performance buffer $B(j)$, it generates the new sample $\mathbf{x}_s$ as
  - $\mathbf{x}_s = \mathbf{x}^j + \dfrac{1}{2^{\delta_p}}\mathbf{d}_p$ where $\mathbf{d}_p$ is a uniform random unit vector that defines the stochastic direction

3. Treats $\tilde{F}$ as definitive but in reality it is stochastic.

- When optimizing the newly generated sample is uses the single objective of
  - $\mathbf{x}_o = \arg\min_{\mathbf{x}\in\mathcal{X}} \|\tilde{F}(\mathbf{x}) - \mathbf{z}(\mathbf{x_s})\|^2$

Suggestion : DGEBO

1. **What if we treat $\tilde{F} \sim N(\mu, \Sigma)$ as we did in BO.**
   - According to the assumption of the model, each f_j was independent of each other.

1-1. Since we want to define $\tilde{F}$ to be stochastic, the following optimization problem should be modified as well.

- $\mathbf{x}_o = \arg\min_{\mathbf{x}\in\mathcal{X}} \|F(\mathbf{x}) - \mathbf{z}(\mathbf{x_s})\|^2$
- Why doing this?)
    - The reason that we are optimizing this is to make our sample closer to the Pareto Front.
    - Zeleny's Compromise Programming says using various weightings and distance functions L_p norms may obtain efficient solutions close to the ideal point.
    - Schulz et al. used the L_2 Norm.
- Problem)
    - F is not deterministic anymore.

- Sol?)
  - Use Distance Metrics for Probability Distributions
    - KL Divergence : $KL(\tilde{F}, \delta(\mathbf{z(x_s)}))$
    - Mutual Information
    - Wasserstein Distance?
      - $W_2(P, Q) = \left(\inf_{\gamma \in \Pi(P,Q)} \int_{\mathcal{X} \times \mathcal{X}} \|x - y\|^2 \, d\gamma(x, y) \right)^{\frac{1}{2}}$
  - Making $\mathbf{z(x_s)}$ a probability distribution?
    - Dirac Delta : $\mathbf{z(x_s)} \sim \delta(\mathbf{z(x_s)})$
    - Gaussian : $\mathbf{z(x_s)} \sim N(\mathbf{z(x_s)}, \sigma^2)$

1-2. First order approximation should be changed as well.

- Deterministic F
  - Calculate the Jacobian and Hessian of μ
- Stochastic F
  - We should get the Jacobian and Hessian of $F \sim N(\mu, \Sigma)$
    - Is this possible? Gaussian, so yeah?

2. When sampling a new point in the performance buffer, what if we use BO acquisition function such as EI?

As is)

- $\mathbf{x}_s = \mathbf{x}^j + \frac{1}{2^{\delta_p}}\mathbf{d}_p$

To be)

- Expected Improvement with Information Gain

Possible Costs and Improvements?

1. **Treating F to be stochastic may be more expensive than treating it to be deterministic.**
    - Check if the simple kernels like low dimensional polynomials work.
2. **Treating the multivariate stochastic function $F \sim N(\mu, \Sigma)$**
    - Is this set up compatible with the performance buffer set up in DGEMO?
    - Is the new sampling scheme compatible with this?

3. Will this approach have advantage?

- More accurate approximation on the Pareto front may be available.
- More efficient sampling using the BO approach.
- DGEMO's batch selection strategy is **NOT** deteriorated by this approach.
  - Stochastic modification is applied only to the First-order approximation.
  - We do not change any of these key factors.
    - Initial LHS sampling
    - Local optimization on $\mathbf{z}(\mathbf{x_s}) = \mathbf{x}_s + \mathbf{s}(\mathbf{x_s})C(\mathbf{x}_s)$
    - First order approximation using the affine subspace
    - Use Graph-cut algorithm to achieve continuity
  - Thus, we can still take advantage of the DGEMO's efficiency.