Modeling Workgroup Meeting Quarterly Review

Optimization Update: Development of A Memetic Algorithm for Large-Scale Watershed Optimization

Kalyanmoy Deb, Pouyan Nejadhashemi, Gregorio Toscano, and Sebastian Hernandez-Suarez

6 JANUARY 2021

MICHIGAN STATE UNIVERSITY

Agenda

- Optimization Problems and Methods
  - Single-objective, multi-objective, large-scale, robust, surrogate-assisted
- Project Plan in brief
- Objective 1: Understanding the CAST system and Development of an Efficient Single-objective Hybrid Optimization Procedure
  - April 1, 2020 to September 30, 2021 (18 months)
- Current Accomplishments

Optimization Problems

- Single objective function, usually has a single optimum
- Multiple objectives, usually lead to a Pareto-optimal set
  - Choose a single preferred solution through a multi-criterion decision-making
- Large dimensions (variables and constraints)
- Uncertainties in variables and parameters
- Evaluation is computationally expensive, requiring surrogates

Optimization Methods

- Optimization methods:
  - Point-based: Fast but local approach, sensitive to initial point
  - Population-based: Global approach to near-optimality
  - No one method is provably best for all problems (NFL theorem)
- Hybrid and Customized optimization principle
  - Hybrid: Population-based methods for global perspective aided with point-based method for faster search
  - Customization: Problem-specific algorithm design
- Evolutionary optimization is flexible to be aided for handling practicalities

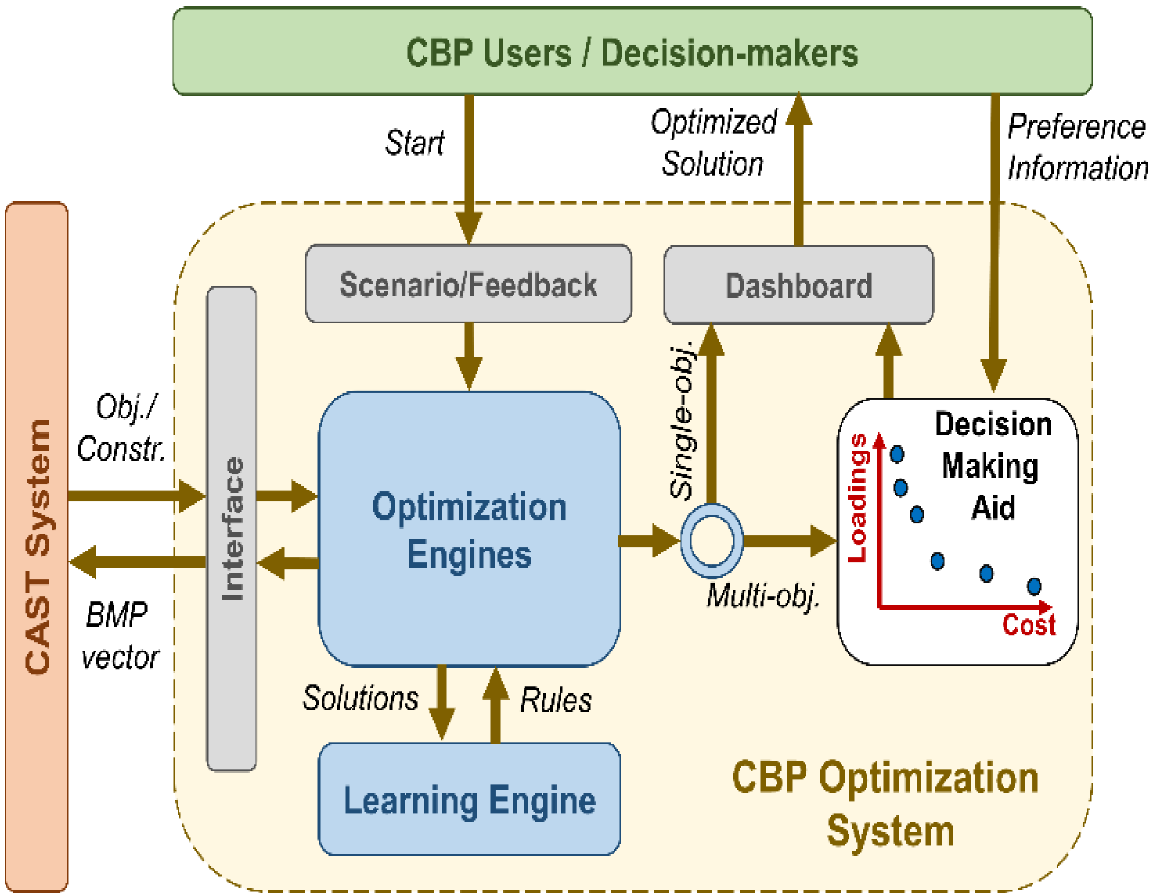
Some Past Applied Optimization Studies by the PIs

- A polynomial-time algorithm for a Billion-variable integer linear program
- Customized algorithms for Rolling mill sequencing, Manufacturing process, Airline crew scheduling, etc.
- Popular multi-objective algorithms (NSGA-II, NSGA-III)
- An astronomically large-sized Land Use Management problem with 14 objectives and involving human decision-makers
- A Gasoline Engine design for 6 objectives with 145 variables and 146 constraints
- A Water Jacket design for better heat transfer requiring 2 days of evaluation per solution involving 500+ processors

EPA Optimization Project Plan (6 years)

- Develop efficient optimization algorithms
  - Integration with CAST system
  - County, multi-county, state, multi-state, and watershed level
  - Hybrid and customized approach
- Min. Cost subject to Loading constraints
- Multi-objective: Min. (Cost, Loadings)
  - MCDM with stake-holders
  - Robust solutions
  - Knowledge-based optimization
- Implementation through validation and discussion with CBP users

Overall Plan



Objective 1: Understanding the CAST system and Development of an Efficient Single-objective Hybrid Optimization Procedure (April 1, 2020 to September 30, 2021)

- Understanding CAST modules and effect of BMPs on objectives and constraints (Achieved)
- Development of a simplified point-based structured single-objective optimization procedure
  - Gradient-based IPOPT (ongoing)
- Development of a hybrid customized single-objective optimization procedure
  - Customized Genetic Algorithms (ongoing)
- Verification and validation with CBP users and decision-makers and update of optimization procedure (planned)

Optimizing Efficiency BMPs

Single-objective Formulation:

$$\text{Minimize} \quad f(\mathbf{x}) = \sum_{s\in S}\sum_{h\in H_s}\sum_{u\in U}\sum_{b\in B_u} \tau_b x_{s,h,u,b}, \qquad \text{(Cost of BMP Implementation)}$$

$$\text{Subject to} \quad \sum_{s\in S}\sum_{h\in H_s}\sum_{u\in U}\left[\alpha_{s,h,u}\phi_{s,h,u}\Pi_{G^B\in\mathcal{G}^B}\left(1-\sum_{b\in G^B}\eta_{s,h,b}\frac{x_{s,h,u,b}}{\alpha_{s,h,u}}\right)\right] \le \Theta, \qquad \text{(Restricted Loadings)}$$

$$\sum_{b\in G^B} x_{s,h,u,b} \le \alpha_{s,h,u}, \quad \forall s\in S,\ h\in H_s,\ u\in U_s,\ G^B\in\mathcal{G}^B,$$

$$x_{s,h,u,b} \ge 0, \quad \forall s\in S,\ h\in H_s,\ u\in U_s,\ b\in B_u. \tag{1}$$

The variable $x_{s,h,u,b}$ indicates the acres used for implementing a BMP b to reduce a load resource u.

A snap picture of the whole problem

- High-dimensional: more than 2 million real variables at the watershed level
- Overlapping regions: 300 000 groups of variables
- Highly constrained: more than 100 000 constraints
- Multi-objective by nature: constraints can be naturally viewed as objectives
- Preference Incorporation and decision making

Previous work

- A local search approach based on an interior point method
- A Genetic Algorithm that finds very good solutions. However require a large number of function evaluations

Performed Improvements

- Improve the quality of solutions in both cost function and the load constraints
- Improve the execution efficiency in both the number of evaluations required and the execution time
- Solve county to multi-county problems
- Three improvements were implemented on Ipopt
  - Create an initial feasible solution
  - Use gradients
  - C++ implementation

Feasible solution to feed our local search approach

- The performance of interior-point-based approaches depends on the starting point
- Start with an initial feasible solution

Jacobian and Hessian matrix improves the convergence

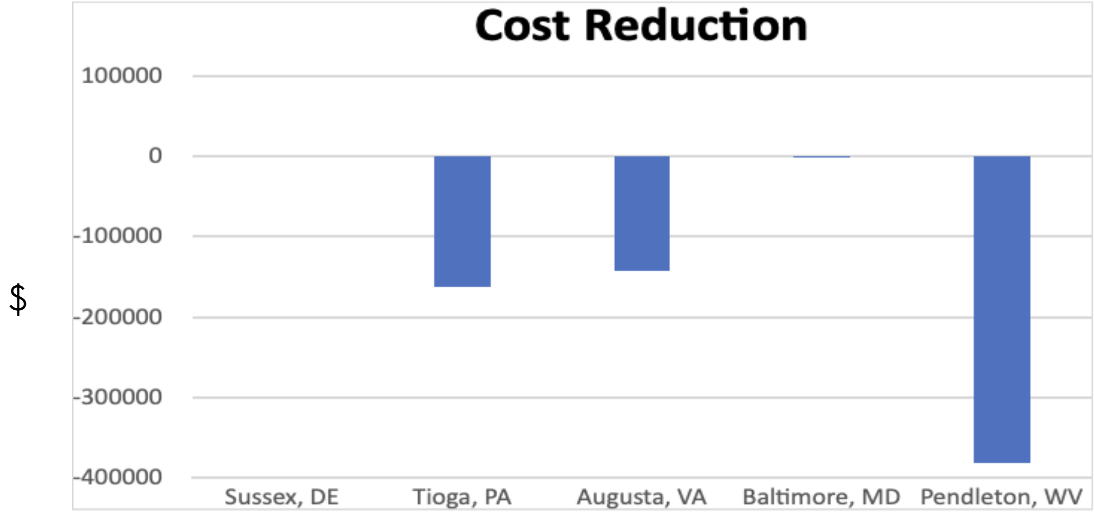
- The Jacobian and Hessian are powerful tools to improve the performance of optimization approaches
- We provide the analytical **Jacobian** and **Hessian** to our approach
- The combination of feasible start solution + analytical Jacobian and Hessian matrices reduce the fitness function evaluations to 30% from original optimization algorithm

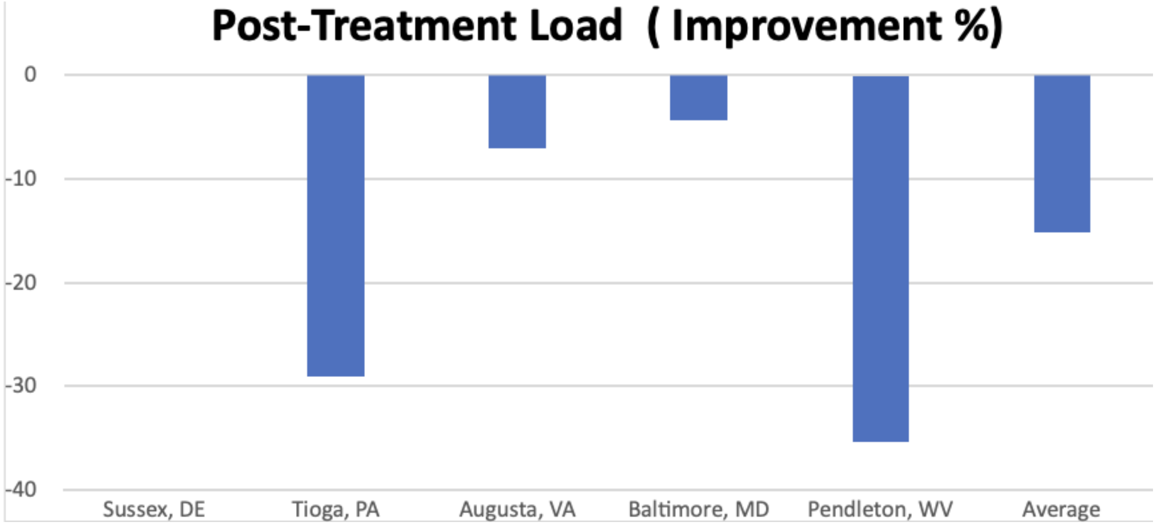
C++ programming language for fast execution

- We rewrote our interior-point-based method in C++, so it could execute faster (originally developed in Python)
- The approach computes the function up to 80 times faster (with the compiler optimization code included)

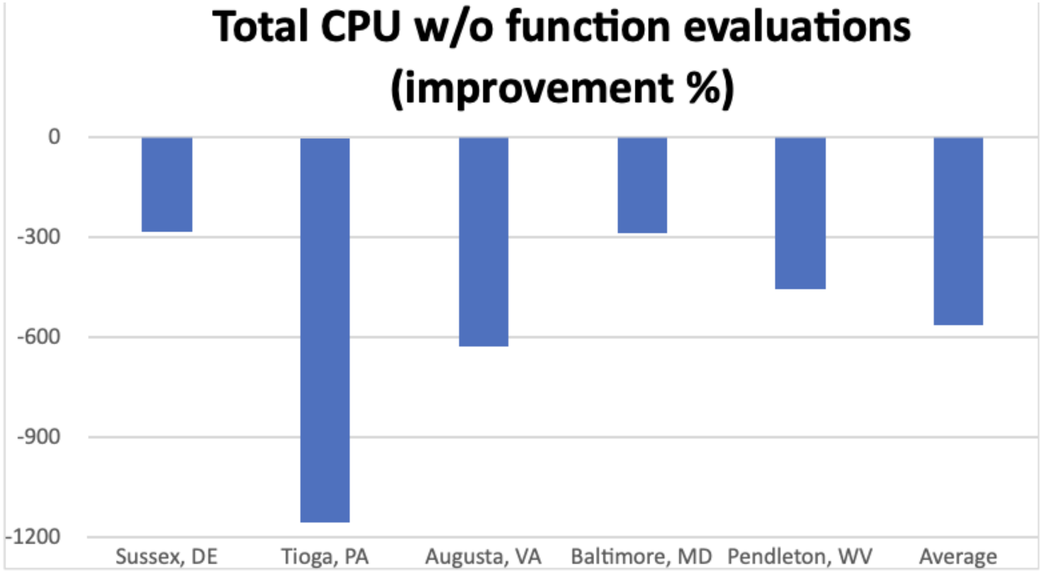
Result from several counties compared to original Ipopt

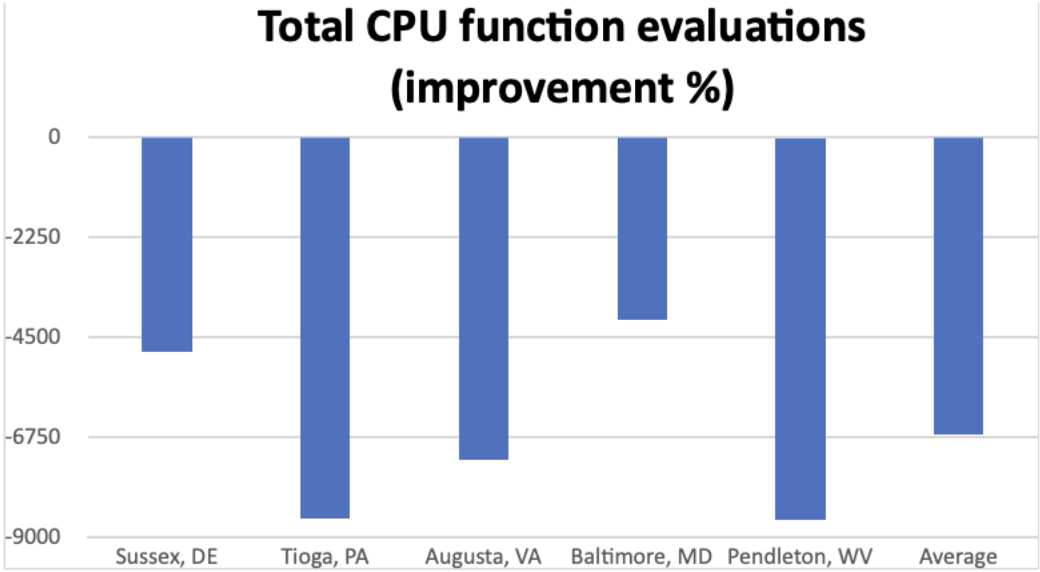
- Our new method reduced $137,377 on an average
- Our new method reduced 12% of the post-treatment load in average





Speed Improvement compared to original Ipopt



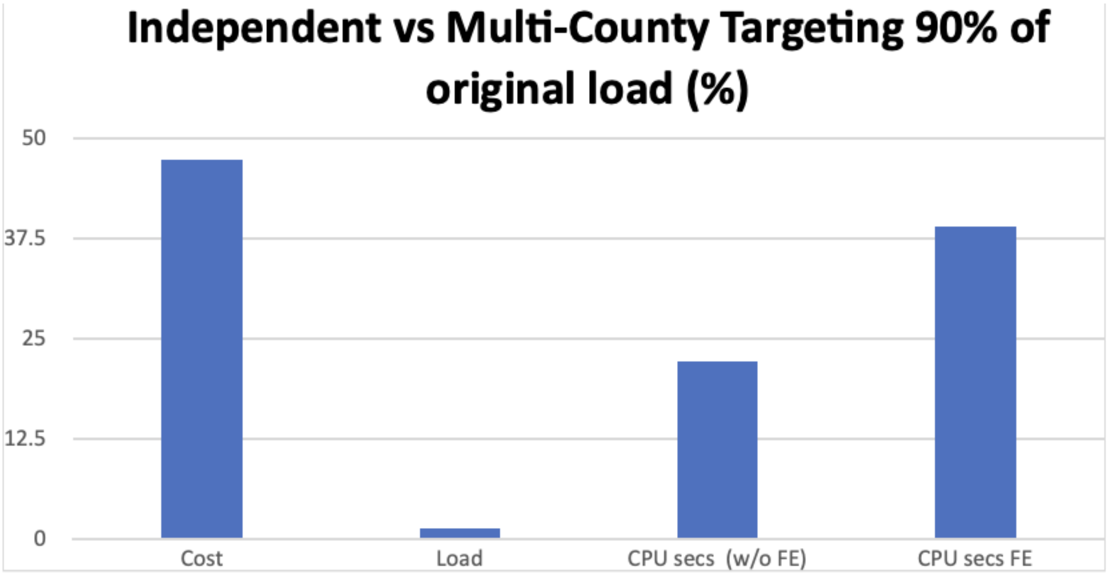


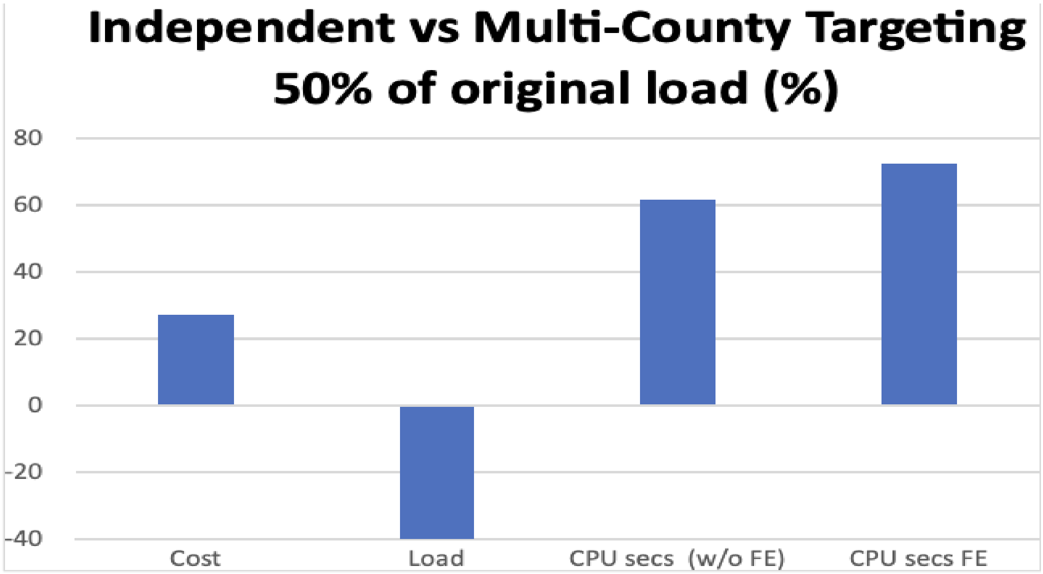
Test different Ipopt Solvers in order to Identify weaknesses and strengths

- Ipopt is able to use different solvers
- We tested 6 different solvers
- The fastest solver MA27 generally has memory problems crashing runs above 10,000 constraints
- We have not had any problem with MA47 solver and it is the second fastest algorithm for the problems we tried

Multi-County capabilities

- We compare Single County Solution with respect to a Multi-county Solution
  - Multi-county approach is promising





Parallel execution for an even faster response

- To execute even faster, we use OpenMP. Our computer system has 32 threads. In the initial experiments, we had speed up of 30X of function, constraint, Jacobian and Hessian evaluation
- Improving the execution may allow us to execute the approach with several values of the post-treatment load
- Having several executions with different post-treatment loads can help decision makers

CAST System and Optimization: Summary

- The preliminary study suggested that both **BGA** and **Ipopt** can produce good quality results and are potentially good candidates

- **Ipopt** approach can be used in a hybrid manner (for local search) within a GA
  - Taking best aspects of both optimization algorithms

CAST System and Optimization: Next Steps

- Develop a Customized GA for a single-objective opt.:
  - Developing customized genetic operators and customized initialization should improve our BGA results
  - Hybridizing GA with Ipopt
  - A scale-up study with increase in variables
  - Computational power of GPUs can help us develop a faster algorithm
- Linking optimization method with CAST