Discrete Optimization via Simulation Using COMPASS

Speaker: Weihuan Huang

School of Data Science, Fudan University

The work of L. Jeff Hong and Barry L. Nelson, *OR* (2006)

Nov. 22, 2021

Overview

Fully Constrained DOvS

- DOvS problem:

$$\min_{\mathbf{x}\in\Theta} \mathrm{E}_{\psi}[G(\mathbf{x},\psi)], \tag{1}$$

where $\Theta = \Phi \cap \mathscr{L}^d$, Φ is a closed and bounded set in $\mathbb{R}^d$, and $\mathscr{L}^d$ is the set of d-dimensional vectors with integer elements.

- Fully constrained: Φ is bounded; Discrete: integer.

- ψ: stochastic input to the simulation

- Let $g(\mathbf{x}) := \mathrm{E}_{\psi}[G(\mathbf{x},\psi)]$, $g(\mathbf{x})$ cannot be evaluated easily but the r.v. $G(\mathbf{x},\psi)$ can be observed via a simulation experiment at $\mathbf{x}$. $G_i(\mathbf{x})$ denotes the i-th observation of $G(\mathbf{x},\psi)$.

Assumption 1

Assumption 1

For every $\mathbf{x} \in \Theta$, we have

$$\Pr\left(\lim_{r\to\infty} \frac{1}{r}\sum_{i=1}^{r} G_i(\mathbf{x}) = g(\mathbf{x})\right) = 1.$$

A1: sample mean of $G(\mathbf{x}, \psi)$ is an appropriate estimator of $g(\mathbf{x})$.

- SLLN, when $G_i(\mathbf{x})$ are i.i.d.
- Ergodic thm, when $G_i(\mathbf{x})$ are ergodic

Some Notations

Neighborhood and local minimizer:

- $\mathscr{N}(\mathbf{x}) := \{\mathbf{y} \in \Theta | \|\mathbf{x} - \mathbf{y}\| = 1\}$: local neighborhood of $\mathbf{x} \in \Theta$
- $\mathbf{x}$ is a local minimizer: if $\mathbf{x} \in \Theta$ and either $\mathscr{N}(\mathbf{x}) = \emptyset$ or $g(\mathbf{x}) \leq g(\mathbf{y})$ for all $\mathbf{y} \in \mathscr{N}(\mathbf{x})$
- $\mathscr{M}$: the set of local minimizers of g in Θ

In the COMPASS algorithm,

- $\mathscr{V}_k$: the set of all solutions visited through iteration k
- $\hat{\mathbf{x}}_k^*$: the solution with the smallest aggregated sample mean among all $\mathbf{x} \in \mathscr{V}_k$ (if more than one solution having the smallest aggregated sample mean, then select $\hat{\mathbf{x}}_k^*$ randomly)
- At the end of iteration k, construct *the most promising area*:

$$\mathscr{C}_k := \{\mathbf{x} \in \Theta | \|\mathbf{x} - \hat{\mathbf{x}}_k^*\| \leq \|\mathbf{x} - \mathbf{y}\|, \forall \mathbf{y} \in \mathscr{V}_k \text{ and } \mathbf{y} \neq \hat{\mathbf{x}}_k^*\}$$

Note that $\mathscr{C}_k$ includes all feasible solutions that are at least as close to $\hat{\mathbf{x}}_k^*$ as to other solutions in $\mathscr{V}_k$. $\hat{\mathbf{x}}_k^* \in \mathscr{C}_k$, hence $\mathscr{C}_k$ is not empty.

Simulation-Allocation Rule

The COMPASS algorithm needs a simulation-allocation rule (SAR) to allocate simulation observations to solutions in $\mathscr{V}_k$.

- $a_k(\mathbf{x})$: the additional observations allocated to $\mathbf{x}$ at iteration k
- $N_k(\mathbf{x}) := \sum_{i=0}^{k} a_i(\mathbf{x})$: the total number of observations on solution $\mathbf{x}$ up to iteration k
- $\bar{G}_k(\mathbf{x})$: the sample mean of all $N_k(\mathbf{x})$ observations of $G(\mathbf{x}, \psi)$ at iteration k

Algorithm 1

COMPASS: convergent optimization via most-promising-area stochastic search

COMPASS for Fully Constrained DOvS

Step 0. Set iteration count $k = 0$. Find $\mathbf{x}_0 \in \Theta$, set $\mathscr{V}_0 = \{\mathbf{x}_0\}$ and $\hat{\mathbf{x}}_k^* = \mathbf{x}_0$. Determine $a_0(\mathbf{x}_0)$ according to the SAR. Take $a_0(\mathbf{x}_0)$ observations from $\mathbf{x}_0$, set $N_0(\mathbf{x}_0) = a_0(\mathbf{x}_0)$, and calculate $\bar{G}_0(\mathbf{x}_0)$. Let $\mathscr{C}_0 = \Theta$.

Step 1. Let $k = k + 1$. Sample $\mathbf{x}_{k1}, \mathbf{x}_{k2}, \ldots, \mathbf{x}_{km}$ uniformly and independently from $\mathscr{C}_{k-1}$. Let $\mathscr{V}_k = \mathscr{V}_{k-1} \cup \{\mathbf{x}_{k1}, \mathbf{x}_{k2}, \ldots, \mathbf{x}_{km}\}$. Determine $a_k(\mathbf{x})$ according to the SAR for every $\mathbf{x}$ in $\mathscr{V}_k$. For all $\mathbf{x} \in \mathscr{V}_k$, take $a_k(\mathbf{x})$ observations, and update $N_k(\mathbf{x})$ and $\bar{G}_k(\mathbf{x})$.

Step 2. Let $\hat{\mathbf{x}}_k^* = \arg\min_{\mathbf{x} \in \mathscr{V}_k} \bar{G}_k(\mathbf{x})$. Construct $\mathscr{C}_k$ and go to Step 1.

Remarks

- The m solutions independently from $\mathscr{C}_{k-1}$ could be repeated. The most promising areas are getting smaller in the iteration count. If $\mathscr{C}_{k-1}$ becomes a singleton, then all m solutions sampled are $\hat{\mathbf{x}}^*_{k-1}$.
- The algorithm can be stopped whenever all of the computational budget is consumed, or when $\hat{\mathbf{x}}^*_k$ does not change for many iterations and all solutions in its local neighborhood have been visited.
- Algorithm 1 may not visit every solution in Θ, even if the computational budget is infinite, which is different from globally convergent DOvS algorithms.

Assumption 2

To make Algorithm 1 converge, we make the following assumption on the SAR.

Assumption 2

The SAR guarantees that $a_k(\mathbf{x}) \geqslant 1$ if $\mathbf{x}$ is a newly visited solution at iteration k ($\mathbf{x} \in \mathscr{V}_k \backslash \mathscr{V}_{k-1}$), and $\lim_{k\to\infty} N_k(\mathbf{x}) = +\infty$ for all visited solutions ($\mathbf{x} \in \bigcup_{k=0}^{\infty} \mathscr{V}_k$).

- The simplest SAR satisfying A2 is an equal SAR that sets $N_k(\mathbf{x}) = N_k$ for all $\mathbf{x} \in \mathscr{V}_k$ and $N_k \to \infty$ as $k \to \infty$ (for instance, $N_k = k$).

Thm 1

We have the following convergence theorem.

Thm 1

If A1 and A2 are satisfied, then the infinite sequence $\{\hat{\mathbf{x}}_0^*, \hat{\mathbf{x}}_1^*, \ldots\}$ generated by Algorithm 1 converges with probability 1 to the set $\mathscr{M}$ in the sense that $\Pr\{\widehat{\mathbf{x}}_k^* \notin \mathscr{M} \text{ i.o.}\} = 0$.

- Thm 1 means, with probability 1, $\hat{\mathbf{x}}_k^*$ is not a local optimal solution only finitely many times.
- If $\mathscr{M}$ is a singleton, then the element in $\mathscr{M}$ is the global optimal solution, and Thm 1 guarantees the global convergence of Algorithm 1.

Sketch of the Proof

- Let $\mathscr{V}_\infty = \bigcup_{k=0}^{\infty} \mathscr{V}_k$,

$$\Pr\left\{\hat{\mathbf{x}}_k^* \notin \mathscr{M} \text{ i.o.}\right\} = \sum_{A \subset \Theta} \Pr\left\{\hat{\mathbf{x}}_k^* \notin \mathscr{M} \text{ i.o.} \mid \mathscr{V}_\infty = A\right\} \Pr\left\{\mathscr{V}_\infty = A\right\}.$$

It is equivalent to proving $\Pr\left\{\hat{\mathbf{x}}_k^* \notin \mathscr{M} \text{ i.o.} \mid \mathscr{V}_\infty = A\right\} = 0$ for all $A \subset \Theta$ s.t. $\Pr\left\{\mathscr{V}_\infty = A\right\} > 0$.

- If $\hat{\mathbf{x}}_k^* \notin \mathscr{M}$ i.o., then there exists an $\mathbf{x} \in A$ and $\mathbf{x} \notin \mathscr{M}$ such that $\hat{\mathbf{x}}_k^* = \mathbf{x}$ i.o. Because $\mathbf{x} \notin \mathscr{M}$, there exists a $\mathbf{y} \in \mathcal{N}(\mathbf{x})$ such that $g(\mathbf{y}) < g(\mathbf{x})$.

$$\Pr\left\{\mathbf{y} \in \mathscr{V}_{k+1} \mid \hat{\mathbf{x}}_k^* = \mathbf{x} \text{ and } \mathbf{y} \notin \mathscr{V}_k\right\} \geqslant \frac{1}{|\Theta|} > 0$$

for any iteration $k+1$. Because $\hat{\mathbf{x}}_k^* = \mathbf{x}$ i.o., $\Pr\left\{\mathbf{y} \in A \mid \mathscr{V}_\infty = A \text{ and } \hat{\mathbf{x}}_k^* = \mathbf{x} \text{ i.o.}\right\} = 1$. If $\mathbf{y} \in A$, then $g(\mathbf{x}) \neq \min_{\mathbf{z} \in A} g(\mathbf{z})$ because $g(\mathbf{y}) < g(\mathbf{x})$.

- Then, we only need to prove

$$\Pr\left\{\lim_{k \to \infty} g\left(\hat{\mathbf{x}}_k^*\right) = \min_{\mathbf{x} \in \mathscr{V}_\infty} g(\mathbf{x}) \mid \mathscr{V}_\infty = A\right\} = 1,$$

which could be proved with some inequalities.

Partially Constrained and Uncontrained DOvS

- DOvS problem:

$$\min_{\mathbf{x}\in\Theta} \mathrm{E}_{\psi}[G(\mathbf{x},\psi)], \tag{2}$$

where $\Theta = \Phi \cap \mathscr{L}^d$. If Φ is a closed but unbounded set in $\mathbb{R}^d$, then we call the problem a partially constrained or unconstrained problem.

  - If $\Phi = \mathbb{R}^d$, then the problem is unconstrained
  - Otherwise, it is partially constrained

Assumption 3

We make the following assumption on $G(\mathbf{x}, \psi)$.

Assumption 3

For any positive constant ϵ and any $\mathbf{x} \in \Theta$, there exist positive numbers r^* and ϵ^* such that for all $r \geqslant r^*$ and for all $0 < \epsilon \leqslant \epsilon^*$,

$$\mathrm{P}\left[\left|\frac{1}{r}\sum_{i=1}^{r} G_i(\mathbf{x}) - g(\mathbf{x})\right| > \epsilon\right] \leqslant \lambda(r, \epsilon)$$

where $\lambda(r, \epsilon)$ is a strictly decreasing function of r and $\lambda(r, \epsilon) \to 0$ as $r \to \infty$.

Two special cases of A3:

- $G_i(\mathbf{x})$ satisfies i.i.d. sequence with finite and uniformly bounded variance
- $G_i(\mathbf{x})$ satisfies large deviation principle

Assumption 4

To establish local convergence, we make the following assumption on the objective function

Assumption 4

For the user provided starting point $\mathbf{x}_0$, there exists a compact set Π and a positive constant δ such that $\mathbf{x}_0 \in \Pi \cap \Theta$ and $g(\mathbf{x}) \geqslant g(\mathbf{x}_0) + \delta$ for any $\mathbf{x} \in \Pi^c \cap \Theta$.

• Many DOvS problems have a *benchmark system setting*, say $\mathbf{x}_0$, which is often the current system setting. All solutions beyond some (unknown) distance from the benchmark setting will typically be inferior to the benchmark. Therefore, A4 is satisfied.

Algorithm 2

Idea: Construct bounded hyper-rectangles $\mathscr{B}_k$ and adopt Algorithm 1.

COMPASS for Partially Constrained or Unconstrained DOvS

Step 0. Set iteration counter $k = 0$. Find $\mathbf{x}_0 \in \Theta$, set $\mathscr{V}_0 = \{\mathbf{x}_0\}$ and $\hat{\mathbf{x}}_0^* = \mathbf{x}_0$. Determine $a_0(\mathbf{x}_0)$ according to the SAR. Take $a_0(\mathbf{x}_0)$ observations from $\mathbf{x}_0$, set $N_0(\mathbf{x}_0) = a_0(\mathbf{x}_0)$, and calculate $\bar{G}(\mathbf{x}_0)$. Construct $\mathscr{B}_0 = \prod_{i=1}^{d} \left[\underline{b}_0^{(i)}, \bar{b}_0^{(i)}\right]$ such that $\underline{b}_0^{(i)} < x_0^{(i)} < \bar{b}_0^{(i)}$ for $i = 1, 2, \ldots, d$. Let $\mathscr{C}_0 = \Theta \cap \mathscr{B}_0$.

Step 1. Let $k = k + 1$. Sample $\mathbf{x}_{k1}, \mathbf{x}_{k2}, \ldots, \mathbf{x}_{km}$ uniformly and independently from $\mathscr{C}_{k-1}$. Let $\mathscr{V}_k = \mathscr{V}_{k-1} \cup \{\mathbf{x}_{k1}, \mathbf{x}_{k2}, \ldots, \mathbf{x}_{km}\}$. Determine $a_k(\mathbf{x})$ according to the SAR for every $\mathbf{x}$ in $\mathscr{V}_k$. For all $\mathbf{x} \in \mathscr{V}_k$, take $a_k(\mathbf{x})$ observations, and update $N_k(\mathbf{x})$ and $\bar{G}_k(\mathbf{x})$.

Algorithm 2

COMPASS for Partially Constrained or Unconstrained DOvS

Step 2. Let $\hat{\mathbf{x}}_k^* = \arg\min_{\mathbf{x}\in\mathcal{V}_k} \bar{G}_k(\mathbf{x})$.

For each $i, i = 1, 2, \ldots, d$, let $\bar{x}_k^{(i)} = \max\left\{x_{k1}^{(i)}, \ldots, x_{km}^{(i)}\right\}$ and $\underline{x}_k^{(i)} = \min\left\{x_{k1}^{(i)}, \ldots, x_{km}^{(i)}\right\}$.

If $\bar{x}_k^{(i)} > \bar{b}_{k-1}^{(i)} - \Delta^{(i)}$, then let $\bar{b}_k^{(i)} = \bar{x}_k^{(i)} + \Delta^{(i)}$; otherwise, let $\bar{b}_k^{(i)} = \bar{b}_{k-1}^{(i)}$.

If $\underline{x}_k^{(i)} < \underline{b}_{k-1}^{(i)} + \Delta^{(i)}$, then let $\underline{b}_k^{(i)} = \underline{x}_k^{(i)} - \Delta^{(i)}$; otherwise, let $\underline{b}_k^{(i)} = \underline{b}_{k-1}^{(i)}$.

Let $\mathscr{B}_k = \prod_{i=1}^d \left[\underline{b}_k^{(i)}, \bar{b}_k^{(i)}\right]$ and $\mathscr{C}_k = \{\mathbf{x} \in \Theta \cap \mathscr{B}_k | \ \|\mathbf{x} - \hat{\mathbf{x}}_k^*\| \leqslant \|\mathbf{x} - \mathbf{y}\|, \forall \mathbf{y} \in \mathscr{V}_k \text{ and } \mathbf{y} \neq \hat{\mathbf{x}}_k^*\}$, and go to Step 1.

- $\Delta^{(i)}$ are user-provided positive numbers.

Assumption 5

To prove the convergence of Algorithm 2, we make the following assumption on the SAR.

Assumption 5

The *SAR* guarantees that there exists *a* sequence $\{r_0, r_1, \ldots\}$ such that $r_0 \geqslant 1, r_{k+1} \geqslant r_k$ for all $k \geqslant 0, r_k \to \infty$ as $k \to \infty, \min_{\mathbf{x} \in \mathscr{V}_k} N_k(\mathbf{x}) \geqslant r_k$ and $\lim_{k \to \infty} k^{d+1} \lambda(r_k, \epsilon) = 0$ for any $\epsilon < \epsilon^*$, where ϵ^* is defined in A3.

For example, $r_k = \beta k^{d+1+\gamma}$ for some positive constants β and γ.

Thm 2

Let $\mathscr{M}$ denote the set of local minimizers in Θ where $|\mathscr{M}|$ can be infinity. We have the following theorem.

Thm 2

If A3-A5 are satisfied, then the infinite sequence $\{\hat{\mathbf{x}}_0^*, \hat{\mathbf{x}}_1^*, \ldots\}$ generated by Algorithm 2 converges with probability 1 to the set $\mathscr{M}$ in the sense that $\mathrm{P}\{\hat{\mathbf{x}}_k^* \notin \mathscr{M} \text{ i.o.}\} = 0$.

Numerical Experiment

Consider the following 10-dimensional quadratic function:

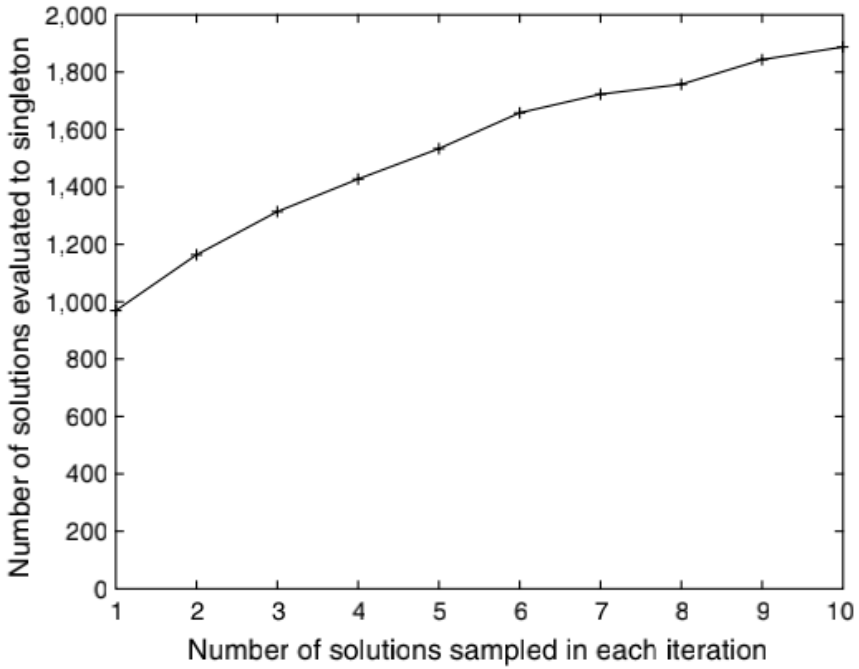
$$g(\mathbf{x}) = x_1^2 + x_2^2 + \cdots + x_{10}^2 + 1$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_{10})'$. The problem has only one local optimal solution, $\mathbf{x}^* = (0, 0, \ldots, 0)'$, with $g(\mathbf{x}^*) = 1$. We let $x_i \in [-100, 100] \cap \mathscr{Z}^{10}$ and let $\mathbf{x}_0 = (80, 80, \ldots, 80)'$. The problem has 1.08×10^{23} feasible solutions. To study local convergence of COMPASS, we let $G(\mathbf{x}, \psi) = g(\mathbf{x})$, which means the function can be evaluated without noise.

Local Convergence

From Figure 1 we see that, on average, fewer than 1,000 function evaluations are required not only to visit $\mathbf{x}^*$, but also to claim it is an optimal solution when the number of solutions sampled in each iteration is set to 1. Note that to claim that $\mathbf{x}^*$ is a global optimal solution, one has to evaluate all 1.08×10^{23} solutions. This illustrates the merit of local convergence.

Figure 1. Number of solutions evaluated to singleton.



Adaptiveness

Consider a classic (s, S) inventory problem in which the level of inventory of some discrete unit is periodically reviewed. The goal is to select s and S such that the steady-state expected inventory cost per review period is minimized. The constraints on s and S are $S - s \geqslant 10$, $20 \leqslant s \leqslant 80, 40 \leqslant S \leqslant 100$, and $s, S \in \mathscr{Z}$. The optimal inventory policy is $(20, 53)$ with expected cost/period of 111.1265.

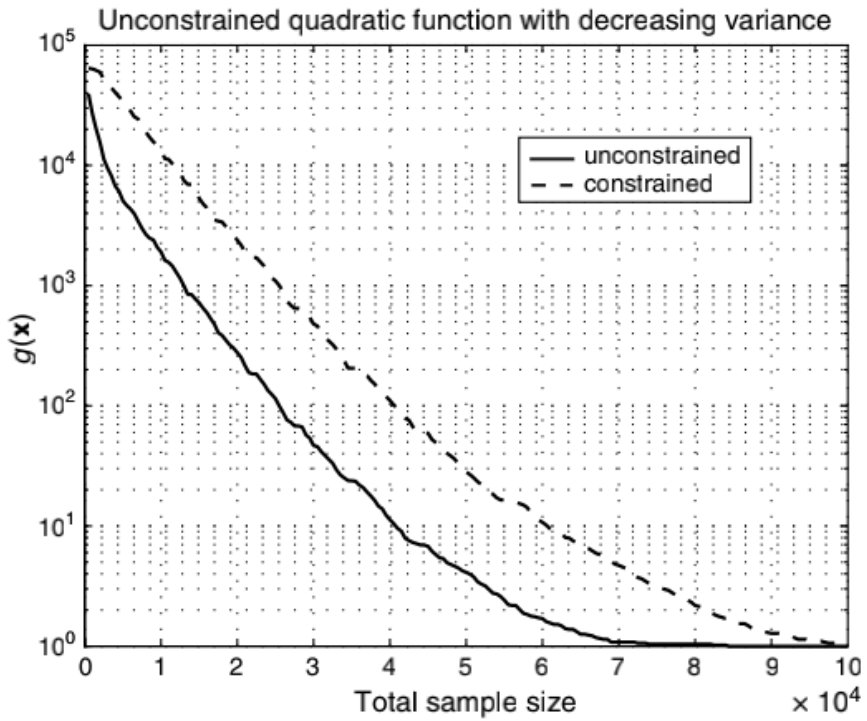
- See Figures 2 and 3 in the paper

Because COMPASS keeps allocating simulation observations to all visited solutions, it eventually escapes from the neighborhood of a nonoptimal solution and moves toward the optimal solution.

Constrained vs. Partially Constrained or Unconstrained

$G(\mathbf{x}) = g(\mathbf{x}) + \epsilon(\mathbf{x})$, letting $\epsilon(\mathbf{x})$ be normally distributed with mean 0 and standard deviation $0.1g(\mathbf{x})$. We apply Algorithm 1 to the problem by adding boundary constraints $-500 \leqslant x_i \leqslant 500, i = 1, 2, \ldots, 10$, and also apply Algorithm 2 to the unconstrained problem. Both algorithms use the same starting point $(80, 80, \ldots, 80)'$. Algorithm 2 works better than Algorithm 1.

Figure 4. Unconstrained vs. constrained COMPASS performance.



An Illustrative Example

• We study the following assemble-to-order system. The system has eight items, $I_1, I_2, \ldots, I_8$, and five types of customers, $T_1, T_2, \ldots, T_5$. Different types of customers come into the system as Poisson arrival processes with different rates, $\lambda_1, \lambda_2, \ldots, \lambda_5$, and each of them requires a set of key items and a set of nonkey items. If any of the key items are out of stock, the customer leaves. If all key items are in stock, the customer buys the product assembled from all the key items and the available nonkey items.
• Each item sold brings a profit, $p_i, i = 1, 2, \ldots, 8$, and each item in inventory has a holding cost per period, h_i, $i = 1, 2, \ldots, 8$. There are inventory capacities for each item, $C_1, C_2, \ldots, C_8$, such that $1 \leqslant x_i \leqslant C_i$, and the production time for each item is normally distributed with mean μ_i and variance $\sigma_i^2, i = 1, 2, \ldots, 8$, truncated at 0 . All parameters used are included in Tables 1 and 2, see the paper
• We are interested in finding the optimal inventory level for each item to maximize the expected total profit per period.

Results

Figure 5. Assemble-to-order system optimization.