Online Supplement to "An Adaptive Hyperbox Algorithm for High-Dimensional Discrete Optimization via Simulation Problems"

Jie Xu
Department of Systems Engineering and Operations Research, George Mason University, Fairfax, VA 22030, USA, jie.xu@u.northwestern.edu

Barry L. Nelson
Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208-3119, USA, nelsonb@northwestern.edu

L. Jeff Hong
Department of Industrial Engineering and Logistics Management, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China, hongl@ust.hk

In this online supplement, we provide the proof of Proposition 1 and color versions of the performance plots.

Proof of Proposition 1

To prove the convergence of Algorithm 1 when Conditions 1 and 2 hold, we first need to establish three lemmas. The first lemma states that if a solution $\mathbf{x}'$ is the sample best solution i.o., then all of its feasible neighbors are included in the estimation set i.o. In the following lemma, we use k_i as the index of a subsequence such that $\widehat{\mathbf{x}}^*_{k_i-1} = \mathbf{x}'$, i.e., the sample best on the iteration before iteration k_i is $\mathbf{x}'$.

Lemma 1 *When Algorithm 1 is applied to Problem (1) and Condition 1 holds, if* $\widehat{\mathbf{x}}^*_k = \mathbf{x}'$ *i.o., then for any* $\tilde{\mathbf{x}} \in \mathcal{N}(\mathbf{x}')$,

$$\Pr\{\tilde{\mathbf{x}} \in \mathcal{E}_{k_i} \text{ i.o.}\} = 1.$$

Proof: For any integer $K > 0$, let $\mathcal{R}_K = \sum_{k=K+1}^{\infty} \mathcal{I}(\mathbf{x}^*_k = \mathbf{x}')$, where $\mathcal{I}(\cdot)$ is the indicator function. Also let R be an arbitrary positive integer. We have

$$\begin{aligned}
\Pr\{\tilde{\mathbf{x}} \notin \mathcal{E}_{k_i} \forall k_i > K\} &= \sum_{r=0}^{\infty} \Pr\{\tilde{\mathbf{x}} \notin \mathcal{E}_{k_i} \forall k_i > K | \mathcal{R}_K = r\} \Pr\{\mathcal{R}_K = r\} \\
&= \sum_{r=0}^{R} \Pr\{\tilde{\mathbf{x}} \notin \mathcal{E}_{k_i} \forall k_i > K | \mathcal{R}_K = r\} \Pr\{\mathcal{R}_K = r\} + \\
&\qquad \sum_{r=R+1}^{\infty} \Pr\{\tilde{\mathbf{x}} \notin \mathcal{E}_{k_i} \forall k_i > K | \mathcal{R}_K = r\} \Pr\{\mathcal{R}_K = r\} \\
&\leq \sum_{r=0}^{R} \Pr\{\mathcal{R}_K = r\} + \sum_{r=R+1}^{\infty} (1-\epsilon)^r \Pr\{\mathcal{R}_K = r\} \\
&\leq \Pr\{\mathcal{R}_K \leq R\} + \epsilon(1-\epsilon)^{R+1}.
\end{aligned}$$

The first inequality comes from Condition 1. Since $\widehat{\mathbf{x}}^*_{k_i-1} = \mathbf{x}'$ i.o. implies that $\mathcal{R}_K = \infty$ w.p. 1, we have $\Pr\{\mathcal{R}_K \le R\} = 0$. For any $\varepsilon > 0$, we can always make R large enough such that $\epsilon(1-\epsilon)^{R+1} < \varepsilon$. Therefore, we have

$$\Pr\{\tilde{\mathbf{x}} \notin \mathcal{E}_{k_i}, \forall k_i > K\} = 0.$$

Since K is arbitrary, it means that for any given K, w.p. 1, there is an iteration $k_i > K$ on which $\tilde{\mathbf{x}}$ is included in the estimation set. Hence we conclude that

$$\Pr\left\{\tilde{\mathbf{x}} \in \mathcal{E}_{k_i} \text{ i.o.}\right\} = 1. \quad \square$$

It is not difficult to verify that Lemma 3.2 and Lemma 3.3 in Hong and Nelson (2007) still hold under Conditions 1 and 2. We present their lemmas below for reference.

Lemma 2 *Let $\widehat{\mathbf{x}}^*_k, k = 0, 1, 2, \ldots,$ be the sequence of solutions generated by the Generic Algorithm when applied to Problem (1). Suppose that Assumption 1 is satisfied. If Conditions 1 and 2 hold, then*

$$\lim_{k\to\infty}\left[g(\widehat{\mathbf{x}}^*_k) - \min_{\mathbf{y}\in\mathcal{E}_k} g(\mathbf{y})\right] = 0 \quad \text{w.p. 1.}$$

Lemma 3 *Let $\widehat{\mathbf{x}}^*_k, k = 0, 1, 2, \ldots,$ be a sequence of solutions generated by the Generic Algorithm when applied to Problem (1). Suppose that Assumption 1 is satisfied. If Conditions 1 and 2 hold, then $g(\widehat{\mathbf{x}}^*_k)$ converges w.p. 1.*

Lemma 2 states that in the limit, the algorithm is able to correctly select the best solution within the estimation set. Lemma 3 shows that the objective value of the current sample best solution converges.

Now we are ready to prove Proposition 1.

Proposition 1 *Let $\widehat{\mathbf{x}}^*_k, k = 0, 1, 2, \ldots$ be a sequence of solutions generated by Algorithm 1 when applied to Problem (1). Suppose that Assumption 1 is satisfied. If Conditions 1 and 2 hold, then $\Pr\{\widehat{\mathbf{x}}^*_k \notin \mathcal{M}$ i.o.$\} = 0$.*

Proof: Since the event $\{\widehat{\mathbf{x}}^*_k \notin \mathcal{M} \text{ i.o.}\} \subset \{\widehat{\mathbf{x}}^*_k \in \mathcal{M}^C \text{ i.o.}\}$, we have $\Pr\{\widehat{\mathbf{x}}^*_k \notin \mathcal{M} \text{ i.o.}\} \le \Pr\{\widehat{\mathbf{x}}^*_k \in \mathcal{M}^C \text{ i.o.}\}$. Suppose $\widehat{\mathbf{x}}^*_k \in \mathcal{M}^C$ i.o. Since $|\Theta|$ is finite, so is $\mathcal{M}^C$. Therefore, $\widehat{\mathbf{x}}^*_k$ has a convergent subsequence when $\widehat{\mathbf{x}}^*_k \in \mathcal{M}^C$ i.o. Notice that

$$\begin{aligned}
\Pr\{\widehat{\mathbf{x}}^*_k \in \mathcal{M}^C \text{ i.o.}\} &\le \Pr\{\widehat{\mathbf{x}}^*_k = \mathbf{x} \text{ i.o. for some } \mathbf{x} \in \mathcal{M}^C\} \\
&\le \sum_{\mathbf{x}\in\mathcal{M}^C} \Pr\{\widehat{\mathbf{x}}^*_k = \mathbf{x} \text{ i.o.}\}. \qquad (1)
\end{aligned}$$

We now consider $\Pr\{\widehat{\mathbf{x}}^*_k = \mathbf{x}' \text{ i.o.}\}$ for some $\mathbf{x}' \in \mathcal{M}^C$. Let $\tilde{\mathbf{x}}$ be a feasible neighbor of $\mathbf{x}'$ such that $g(\tilde{\mathbf{x}}) < g(\mathbf{x}')$; $\mathbf{x}'$ must have such a neighbor or it is not in $\mathcal{M}^C$. We have

$$\Pr\{\widehat{\mathbf{x}}^*_k = \mathbf{x}' \text{ i.o.}\} = \Pr\{\widehat{\mathbf{x}}^*_k = \mathbf{x}' \text{ i.o.}, g(\widehat{\mathbf{x}}^*_k) \text{ converges}\} + \Pr\{\widehat{\mathbf{x}}^*_k = \mathbf{x}' \text{ i.o.}, g(\widehat{\mathbf{x}}^*_k) \text{ diverges}\}.$$

By Lemma 3, $g(\widehat{\mathbf{x}}_k^*)$ converges w.p. 1. So $\Pr\{\widehat{\mathbf{x}}_k^* = \mathbf{x}' \text{ i.o.}, g(\widehat{\mathbf{x}}_k^*) \text{ diverges}\} = 0$. Hence we have $\Pr\{\widehat{\mathbf{x}}_k^* = \mathbf{x}' \text{ i.o.}\} = \Pr\{\widehat{\mathbf{x}}_k^* = \mathbf{x}' \text{ i.o.}, g(\widehat{\mathbf{x}}_k^*) \text{ converges}\}$. We again use k_i to denote the subsequence such that $\widehat{\mathbf{x}}_{k_i-1}^* = \mathbf{x}' \in \mathcal{M}^C$ for all $i = 1, 2, \ldots$. Consider a sample path on which $\widehat{\mathbf{x}}_k^* = \mathbf{x}'$ i.o. and $g(\widehat{\mathbf{x}}_k^*)$ converges. Since the subsequence $g(\widehat{\mathbf{x}}_{k_i-1}^*) = g(\mathbf{x}')$, and thus converges to $g(\mathbf{x}')$, we know $g(\widehat{\mathbf{x}}_k^*) \to g(\mathbf{x}')$ on that sample path. Therefore,

$$\Pr\{\widehat{\mathbf{x}}_k^* = \mathbf{x}' \text{ i.o.}\} \quad = \quad \Pr\{\widehat{\mathbf{x}}_k^* = \mathbf{x}' \text{ i.o.}, g(\widehat{\mathbf{x}}_k^*) \to g(\mathbf{x}')\}. \tag{2}$$

By Lemma 1, $\Pr\{\tilde{\mathbf{x}} \in \mathcal{E}_{k_i} \text{ i.o.}\} = 1$, so we can rewrite (2) as

$$\Pr\{\widehat{\mathbf{x}}_k^* = \mathbf{x}' \text{ i.o.}\} \quad = \quad \Pr\{\widehat{\mathbf{x}}_k^* = \mathbf{x}' \text{ i.o.}, g(\widehat{\mathbf{x}}_k^*) \to g(\mathbf{x}'), \tilde{\mathbf{x}} \in \mathcal{E}_{k_i} \text{ i.o.}\}. \tag{3}$$

Let k_{i_j} be the subsequence of the sequence k_i such that $\tilde{\mathbf{x}} \in \mathcal{E}_{k_{i_j}}$ for $j = 1, 2, \ldots$. By Condition 2, $N_{k_{i_j}}(\mathbf{x}) \to \infty$ as $k_{i_j} \to \infty$ for all $\mathbf{x} \in \mathcal{E}_{k_{i_j}}$. Since Condition 2 requires that $\mathbf{x}' = \widehat{\mathbf{x}}_{k_{i_j}-1}^* \in \mathcal{E}_{k_{i_j}}$ and by the definition of k_{i_j}, $\tilde{\mathbf{x}} \in \mathcal{E}_{k_{i_j}}$, we have $N_{k_{i_j}}(\mathbf{x}') \to \infty$ and $N_{k_{i_j}}(\tilde{\mathbf{x}}) \to \infty$ as $k_{i_j} \to \infty$.

According to Assumption 1, for all $\varepsilon > 0$, there exists a random variable K_ε such that for all $k_{i_j} > K_\varepsilon, |\bar{G}_{k_{i_j}}(\mathbf{x}') - g(\mathbf{x}')| < \varepsilon, |\bar{G}_{k_{i_j}}(\tilde{\mathbf{x}}) - g(\tilde{\mathbf{x}})| < \varepsilon$ and $K_\varepsilon < \infty$ w.p. 1. Therefore, for all $0 < \delta < 1$, there exists a constant $k_{\varepsilon,\delta}$ such that $\Pr\{K_\varepsilon < k_{\varepsilon,\delta}\} > \delta$. This means the event $\Omega = \{\omega : K_\varepsilon < k_{\varepsilon,\delta}\}$ satisfies $\Pr\{\Omega\} > \delta$. So we can rewrite (3) as

$$\begin{aligned}\Pr\{\widehat{\mathbf{x}}_k^* = \mathbf{x}' \text{ i.o.}\} \quad &= \quad \Pr\{\widehat{\mathbf{x}}_k^* = \mathbf{x}' \text{ i.o.}, g(\widehat{\mathbf{x}}_k^*) \to g(\mathbf{x}'), \tilde{\mathbf{x}} \in \mathcal{E}_{k_i} \text{ i.o.}, K_\varepsilon < k_{\varepsilon,\delta}\} + \\ &\qquad \Pr\{\widehat{\mathbf{x}}_k^* = \mathbf{x}' \text{ i.o.}, g(\widehat{\mathbf{x}}_k^*) \to g(\mathbf{x}'), \tilde{\mathbf{x}} \in \mathcal{E}_{k_i} \text{ i.o.}, K_\varepsilon \geq k_{\varepsilon,\delta}\} \\ &\leq \quad \Pr\{\widehat{\mathbf{x}}_k^* = \mathbf{x}' \text{ i.o.}, g(\widehat{\mathbf{x}}_k^*) \to g(\mathbf{x}'), \tilde{\mathbf{x}} \in \mathcal{E}_{k_i} \text{ i.o.}, K_\varepsilon < k_{\varepsilon,\delta}\} + 1 - \delta. \end{aligned} \tag{4}$$

Consider a sample path $\omega \in \Omega$ along which $\widehat{\mathbf{x}}_k^* = \mathbf{x}'$ i.o. and $\tilde{\mathbf{x}} \in \mathcal{E}_{k_i}$ i.o. Choose $\varepsilon = (g(\mathbf{x}') - g(\tilde{\mathbf{x}}))/4$. On iteration k_{i_j}, $\tilde{\mathbf{x}} \in \mathcal{E}_{k_{i_j}}$ and $\widehat{\mathbf{x}}_{k_{i_j}-1}^* = \mathbf{x}'$, we have

$$\bar{G}_{k_{i_j}}(\tilde{\mathbf{x}}) < g(\tilde{\mathbf{x}}) + \varepsilon < g(\mathbf{x}') - \varepsilon < \bar{G}_{k_{i_j}}(\mathbf{x}') < g(\mathbf{x}') + \varepsilon \tag{5}$$

for all $k_{i_j} > K_{\varepsilon,\delta}$ and $\omega \in \Omega$. So $\bar{G}(\widehat{\mathbf{x}}_{k_{i_j}}^*) \leq \bar{G}_{k_{i_j}}(\tilde{\mathbf{x}}) < \bar{G}_{k_{i_j}}(\mathbf{x}')$ for all $k_{i_j} > K_{\varepsilon,\delta}$, which means $\widehat{\mathbf{x}}_{k_{i_j}}^* \neq \mathbf{x}'$ for all $k_{i_j} > K_{\varepsilon,\delta}$. In addition, from (5) we have $|\bar{G}_{k_{i_j}}(\widehat{\mathbf{x}}_{k_{i_j}}^*) - g(\mathbf{x}')| > \varepsilon$ for all $k_{i_j} > K_{\varepsilon,\delta}$. This means that $g(\widehat{\mathbf{x}}_k^*)$ does not converge to $g(\mathbf{x}')$ along sample path ω. Therefore, there can not be any sample path ω that satisfies $\widehat{\mathbf{x}}_k^* = \mathbf{x}$ i.o., $\tilde{\mathbf{x}} \in \mathcal{E}_{k_i}$ i.o., $\omega \in \Omega$ and $g(\widehat{\mathbf{x}}_k^*) \to g(\mathbf{x}')$ simultaneously. So we conclude

$$\Pr\{\widehat{\mathbf{x}}_k^* = \mathbf{x}' \text{ i.o.}, g(\widehat{\mathbf{x}}_k^*) \to g(\mathbf{x}'), \tilde{\mathbf{x}} \in \mathcal{E}_{k_i} \text{ i.o.}, \omega \in \Omega\} = 0. \tag{6}$$

Plugging (6) into (4), we have $\Pr\{\widehat{\mathbf{x}}_k^* = \mathbf{x}' \text{ i.o.}\} \leq 1 - \delta$ for all $0 < \delta < 1$. As we drive δ towards 1, we have $1 - \delta \to 0$ and thus $\Pr\{\widehat{\mathbf{x}}_k^* = \mathbf{x}' \text{ i.o.}\} = 0$. Since $\mathbf{x}' \in \mathcal{M}^C$ is arbitrary, we have $\Pr\{\widehat{\mathbf{x}}_k^* = \mathbf{x} \text{ i.o.}\} = 0$ for all $\mathbf{x} \in \mathcal{M}^C$. Therefore, (1) gives $\Pr\{\widehat{\mathbf{x}}_k^* \notin \mathcal{M} \text{ i.o.}\} = 0$. □

Derivation of (5):

Let $u^{(d)} = \min\{\mathbf{x}_1^{(d)}, \mathbf{x}_2^{(d)}, \ldots, \mathbf{x}_m^{(d)}\}$, for $d = 1, 2, \ldots, m$. Then

$$V = \prod_{d=1}^{D} (u^{(d)} - 0).$$

Since $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ are sampled uniformly from $\Theta_1 = [0,1]^D$, we know that $\mathbf{x}_1^{(d)}, \mathbf{x}_2^{(d)}, \ldots, \mathbf{x}_m^{(d)}$ are i.i.d. $U(0,1)$ distributed. Therefore, we have $\mathrm{E}(u^{(d)}) = 1/(m+1)$ for all $d = 1, 2, \ldots, D$. Because all $u^{(d)}$'s are also independent, we have

$$\mathrm{E}(V) = \prod_{d=1}^{D} \mathrm{E}(u^{(d)}) = \left(\frac{1}{m+1}\right)^D . \square$$

Derivation of (6):

Let $l^{(d)} = \max\{-1/2, \mathbf{x}_i^{(d)}, i = 1, 2, \ldots, m : \mathbf{x}_i^{(d)} < 0\}, u^{(d)} = \min\{1/2, \mathbf{x}_i^{(d)}, i = 1, 2, \ldots, m : \mathbf{x}_i^{(d)} > 0\}$, for $d = 1, 2, \ldots, m$. Then $V = \prod_{d=1}^{D}(u^{(d)} - l^{(d)})$.

Let n_d be the number of solutions that fall within $[-1/2, 0]$ along dimension d. Clearly n_d has a binomial distribution $\mathrm{Bin}(m, 0.5)$. Conditioning on n_d, it is easy to obtain that

$$\mathrm{E}(l^{(d)}|n_d) = -\frac{1}{2(n_d+1)}, \quad \mathrm{E}(u^{(d)}|n_d) = \frac{1}{2(m-n_d+1)}.$$

Because of the independence among all directions, we have

$$\begin{aligned}
\mathrm{E}(V) &= \mathrm{E}_{n_1,n_2,\ldots,n_D}[\mathrm{E}(V|n_1, n_2, \ldots, n_D)] \\
&= \prod_{d=1}^{D} \mathrm{E}_{n_d}[\mathrm{E}(u^{(d)} - l^{(d)}|n_d)] \\
&= \prod_{d=1}^{D} \mathrm{E}_{n_d}\left[\frac{1}{2(n_d+1)} + \frac{1}{2(m-n_d+1)}\right] \\
&= \prod_{d=1}^{D} \left[\sum_{n=0}^{m} \binom{m}{n} \left(\frac{1}{2}\right)^n \left(\frac{1}{2}\right)^{m-n} \left(\frac{1}{2(n_d+1)} + \frac{1}{2(m-n_d+1)}\right)\right].
\end{aligned}$$

By symmetry of the last formula, we have

$$\mathrm{E}(V) = \left\{2\sum_{n=0}^{m} \binom{m}{n} \left(\frac{1}{2}\right)^{m+1} \frac{1}{n+1}\right\}^D . \tag{7}$$

Note that

$$\sum_{n=0}^{m} \binom{m}{n} \left(\frac{1}{2}\right)^{m+1} \frac{1}{n+1} = \frac{1}{m+1} \sum_{n=0}^{m} \frac{(m+1)!}{(n+1)!(m-n)!} \left(\frac{1}{2}\right)^{m+1} .$$

Let $k = n+1$. Then the previous equation becomes

$$\begin{aligned}
\sum_{n=0}^{m} \binom{m}{n} \left(\frac{1}{2}\right)^{m+1} \frac{1}{n+1} &= \frac{1}{m+1} \sum_{k=1}^{m} \frac{(m+1)!}{k!(m+1-k)!} \left(\frac{1}{2}\right)^{m+1} \\
&= \frac{1}{m+1} \left[\sum_{k=0}^{m} \frac{(m+1)!}{k!(m+1-k)!} \left(\frac{1}{2}\right)^{m+1} - \left(\frac{1}{2}\right)^{m+1}\right] \\
&= \frac{1}{m+1} \left[1 - \left(\frac{1}{2}\right)^{m+1}\right].
\end{aligned}$$

Therefore, by Equation (7), we have

$$\mathrm{E}(V) = \left\{ \frac{2}{m+1} \left[1 - \left(\frac{1}{2} \right)^{m+1} \right] \right\}^D . \square$$

Derivation of (8):

We start with the simplest case $m = 1$. Let $\mathbf{x}'$ be the solution sampled. Recall that the generic COMPASS constraint is $(\mathbf{x}^* - \mathbf{x}')^T(\mathbf{x} - (\mathbf{x}^* + \mathbf{x}')/2) \geq 0$. Since $\mathbf{x}^* = (0, 0, \ldots, 0)^T$, the constraint is equivalent to

$$\mathbf{x}'^T(\mathbf{x} - 1/2\mathbf{x}') \leq 0. \tag{8}$$

For Corner Case, $\mathbf{x}' = [u_1, u_2, \ldots, u_D]^T$ is sampled uniformly from $\Theta_1 = [0, 1]^D$. Since the volume of Θ_1 is 1, the expected volume of the MPA is the same as the probability that $\mathbf{x} = [w_1, w_2, \ldots, w_D]^T \sim U(0, 1)^D$ satisfies (8). That is,

$$\begin{aligned}
\mathrm{E}_{\mathbf{x}'}(V) &= \mathrm{E}_{\mathbf{x}'}\left[\mathrm{E}_{\mathbf{x}}(V|\mathbf{x}')\right] \\
&= \Pr\left\{ \mathbf{x}'^T\mathbf{x} - \frac{1}{2}\mathbf{x}'^T\mathbf{x}' \leq 0 \right\} \\
&= \Pr\left\{ \sum_{i=1}^{D} u_i w_i - \frac{1}{2}\sum_{i=1}^{D} u_i^2 \leq 0 \right\} \\
&= \Pr\left\{ \sum_{i=1}^{D} (u_i w_i - \frac{1}{2}u_i^2) \leq 0 \right\}. \qquad (9)
\end{aligned}$$

Since $\mathbf{x}$ and $\mathbf{x}'$ are i.i.d. $U(0, 1)^D$ distributed, we know that u_i, w_i are i.i.d. $U(0, 1)$ distributed, hence $w_i - 1/2u_i^2$ are i.i.d. Therefore, we can approximate (9) using Central Limit Theorem. After some calculation, it is not difficult to obtain that $\mathrm{E}(u_i w_i - 1/2u_i^2) = 1/12$ and $\mathrm{Var}(u_i w_i - 1/2u_i^2) = 7/240$. So

$$\begin{aligned}
\mathrm{E}_{\mathbf{x}'}(V) &= \Pr\left\{ \frac{D^{-1}\sum_{i=1}^{D}(u_i w_i - \frac{1}{2}u_i^2) - 1/12}{\sqrt{7/(240D)}} \leq \frac{-1/12}{\sqrt{7/(240D)}} \right\} \\
&\approx \Phi\left(\frac{-1/12}{\sqrt{7/(240D)}} \right) = \Phi(-0.49\sqrt{D}).
\end{aligned}$$

Now we extend the analysis to the general case with $m \geq 1$. Let $\mathbf{x}_i = [u_{i,1}, \ldots, u_{i,D}]$, $\mathbf{x} = [w_1, \ldots, w_D]$ and

$$Z_i = \frac{D^{-1}\sum_{d=1}^{D}(u_{i,d} w_d - \frac{1}{2}u_{i,d}^2) - 1/12}{\sqrt{7/(240D)}}.$$

The equation now is

$$\begin{aligned}
\mathrm{E}_{\mathbf{x}',\ldots,\mathbf{x}_m}(V) &= \mathrm{E}_{\mathbf{x}',\ldots,\mathbf{x}_m}\left[\mathrm{E}_{\mathbf{x}}(V|\mathbf{x}', \ldots, \mathbf{x}_m)\right] \\
&= \Pr\left\{ \mathbf{x}_i^T\mathbf{x} - \frac{1}{2} \cdot \mathbf{x}_i^T\mathbf{x}_i \leq 0, i = 1, 2, \ldots, m \right\} \\
&= \Pr\left\{ Z_i \leq -0.49\sqrt{D}, i = 1, 2, \ldots, m \right\}, \qquad (10)
\end{aligned}$$

where $\mathbf{x}, \mathbf{x}_i, i = 1, 2, \ldots, m$ are i.i.d. $U(0, 1)$ distributed and Z_i have an approximate multivariate normal distribution $MVN(\mathbf{0}, \Sigma)$. The third step follows from the preceding analysis with $m = 1$. We now claim that Z_i's are positively correlated. We first note that the sign of $\mathrm{Cov}\left(\sum_{d=1}^{D}(u_{i,d}w_d - 1/2u_{i,d}^2), \sum_{d'=1}^{D}(u_{j,d'}w_{d'} - 1/2u_{j,d'}^2)\right)$ is the same as the sign of $\mathrm{Cov}(Z_i, Z_j)$ for any $i, j \in \{1, 2, \ldots, D\}$. After some manipulations, we can write the covariance as

$$
\begin{aligned}
&\mathrm{Cov}\left(\sum_{d=1}^{D}(u_{i,d}w_d - 1/2 \cdot u_{i,d}^2), \sum_{d=1}^{D}(u_{j,d}w_{d'} - 1/2 \cdot u_{j,d'}^2)\right) = \\
&\quad \sum_{d=1}^{D}\sum_{d'=1}^{D}\mathrm{Cov}\left(u_{i,d}w_d, u_{j,d'}w_{d'}\right) - \sum_{d=1}^{D}\sum_{d'=1}^{D}\mathrm{Cov}\left(u_{i,d}^2/2, u_{j,d'}w_{d'}\right) - \\
&\quad \sum_{d=1}^{D}\sum_{d'=1}^{D}\mathrm{Cov}\left(u_{i,d}w_d, u_{j,d'}^2/2\right) + \sum_{d=1}^{D}\sum_{d'=1}^{D}\mathrm{Cov}\left(u_{i,d}^2/2, u_{j,d'}^2/2\right).
\end{aligned}
$$

It is straightforward to verify that $\mathrm{Cov}\left(u_{i,d}w_d, u_{j,d'}w_{d'}\right) = 0$ if $d \neq d'$ and $\mathrm{Cov}\left(u_{i,d}w_d, u_{j,d'}w_{d'}\right) = \mathrm{E}(w_d^2)\mathrm{E}(u_{i,d})\mathrm{E}(u_{j,d}) - \mathrm{E}(w_d)^2\mathrm{E}(u_{i,d})\mathrm{E}(u_{j,d}) = 1/48 > 0$ when $d = d'$. The last three terms are all 0 due to the independence between all random variables involved. So $\mathrm{Cov}(Z_i, Z_j) > 0$. By Slepian's inequality (Tong 1980), we have (10) $\geq \Phi(-0.49\sqrt{D})^m$ asymptotically. □

Derivation of (9): We follow the previous proof procedure for Equation (8). First, $\mathbf{x}, \mathbf{x}', \ldots, \mathbf{x}_m$ now are i.i.d. $U(-1/2, 1/2)^D$ random variables, and thus u_i, w_i are i.i.d. $U(-1/2, 1/2)$. We then have $\mathrm{E}(u_iw_i - 1/2u_i^2) = -1/24$ and $Var(u_iw_i - 1/2u_i^2) = 1/120$. So for $m = 1$,

$$
\begin{aligned}
E_{\mathbf{x}'}(V) &= P\left(\frac{D^{-1}\sum_{i=1}^{D}(u_iw_i - \frac{1}{2}u_i^2) + 1/24}{\sqrt{1/(120D)}} \leq \frac{1/24}{\sqrt{1/(120D)}}\right) \\
&\approx \Phi\left(\frac{1/24}{\sqrt{1/(120D)}}\right) = \Phi(0.46\sqrt{D}).
\end{aligned}
$$

For $m > 1$, the difference now is $\mathrm{Cov}(Z_i, Z_j) = 0$. This is because $\mathrm{Cov}\left(u_{i,d}w_d, u_{j,d'}w_{d'}\right) = \mathrm{E}(w_d^2)\mathrm{E}(u_{i,d})\mathrm{E}(u_{j,d}) - \mathrm{E}(w_d)^2\mathrm{E}(u_{i,d})\mathrm{E}(u_{j,d}) = 0$ as a result of $\mathrm{E}(u_{i,d}) = \mathrm{E}(u_{j,d'}) = 0$. Therefore, we can still apply Slepian's inequality and obtain $\mathrm{E}(V) \geq \Phi(0.46\sqrt{D})^m$. □

References

Tong, Y.L. 1980. *Probability Inequalities in Multivariate Distributions.* Academic Press, New York.

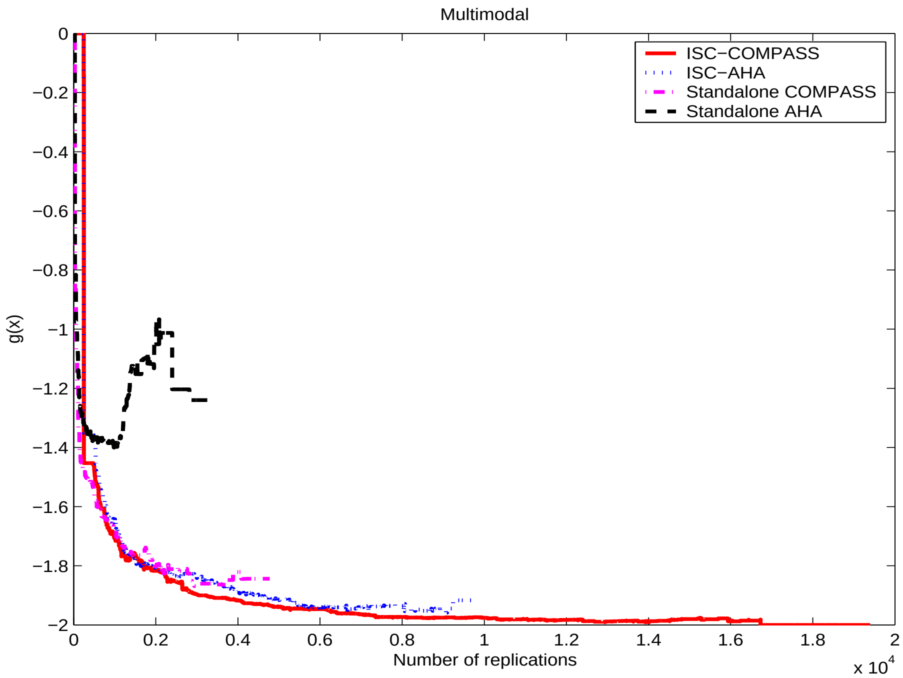

Figure 3: Performance plot for the multimodal function

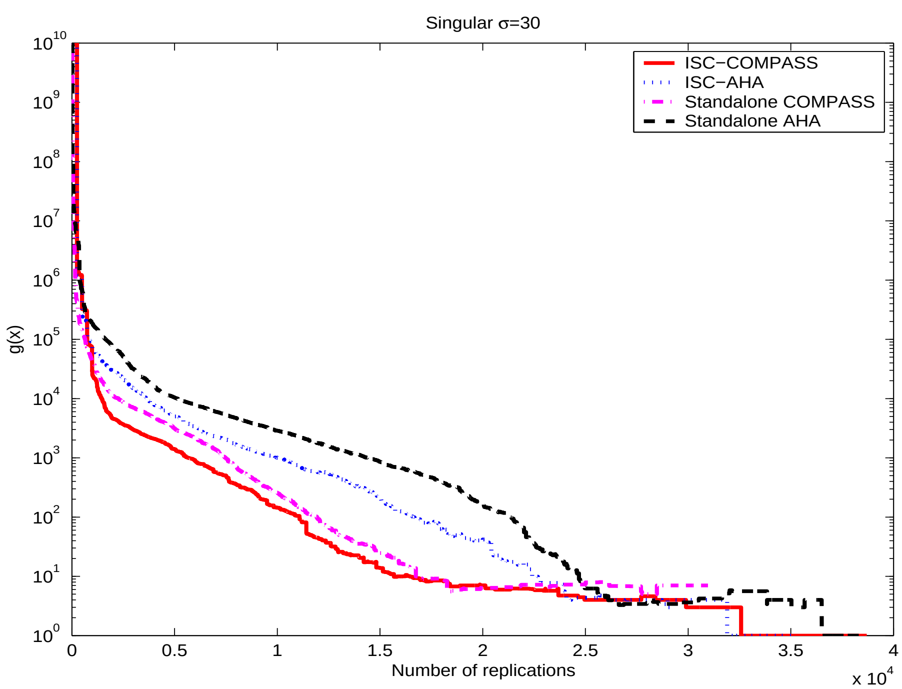

Figure 4: Performance plot for the singular function

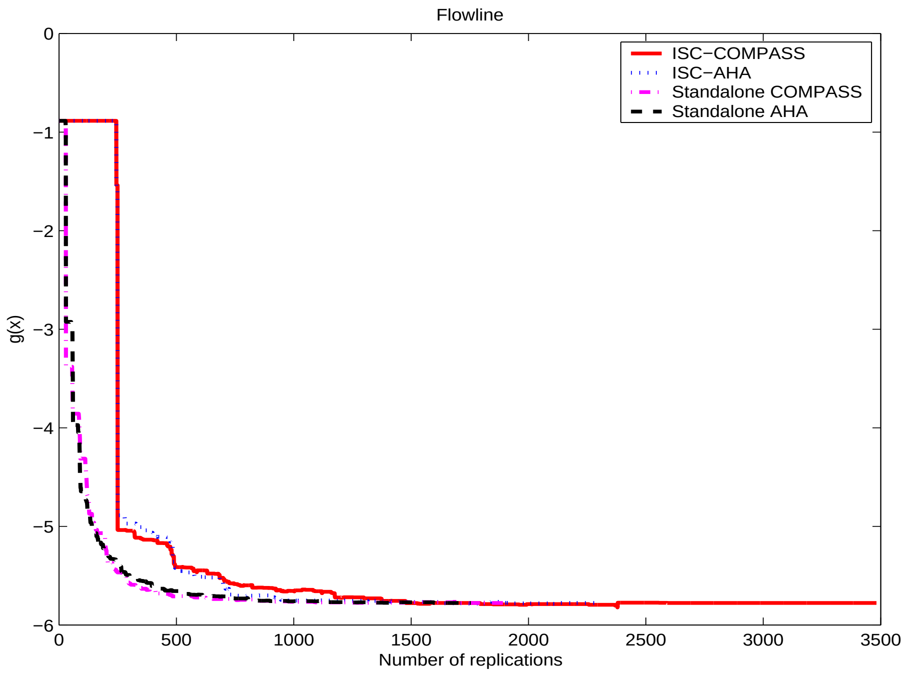

Figure 5: Performance plot for the flowline problem

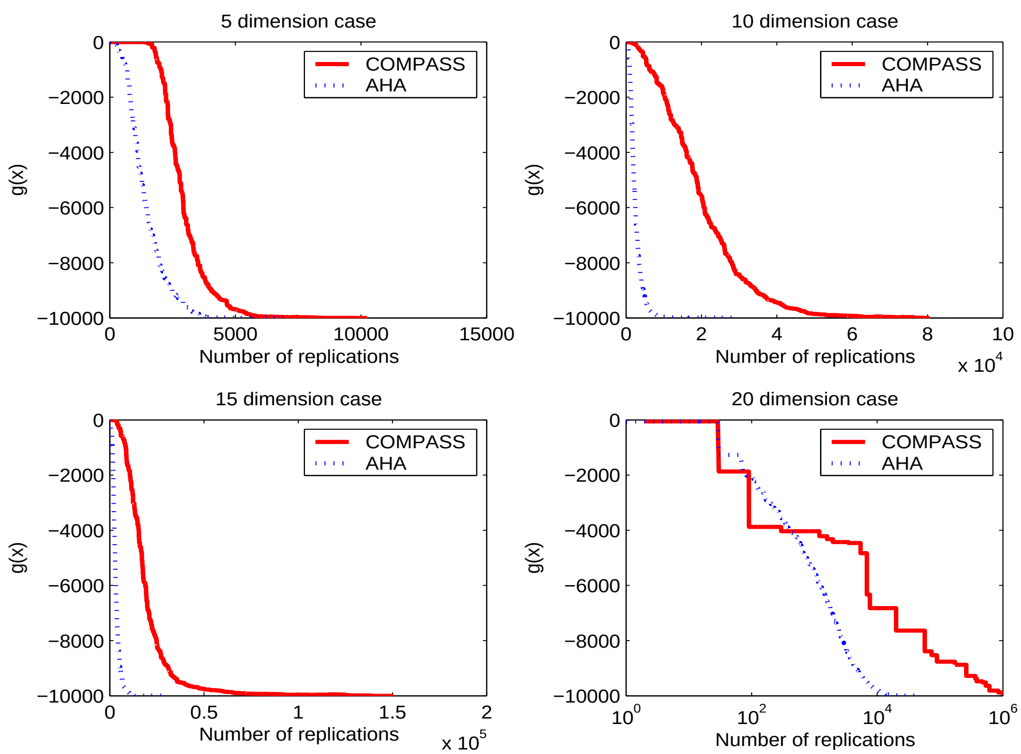

Figure 6: Performance plot for the high-dimensional test problem: D=5, 10, 15 and 20.

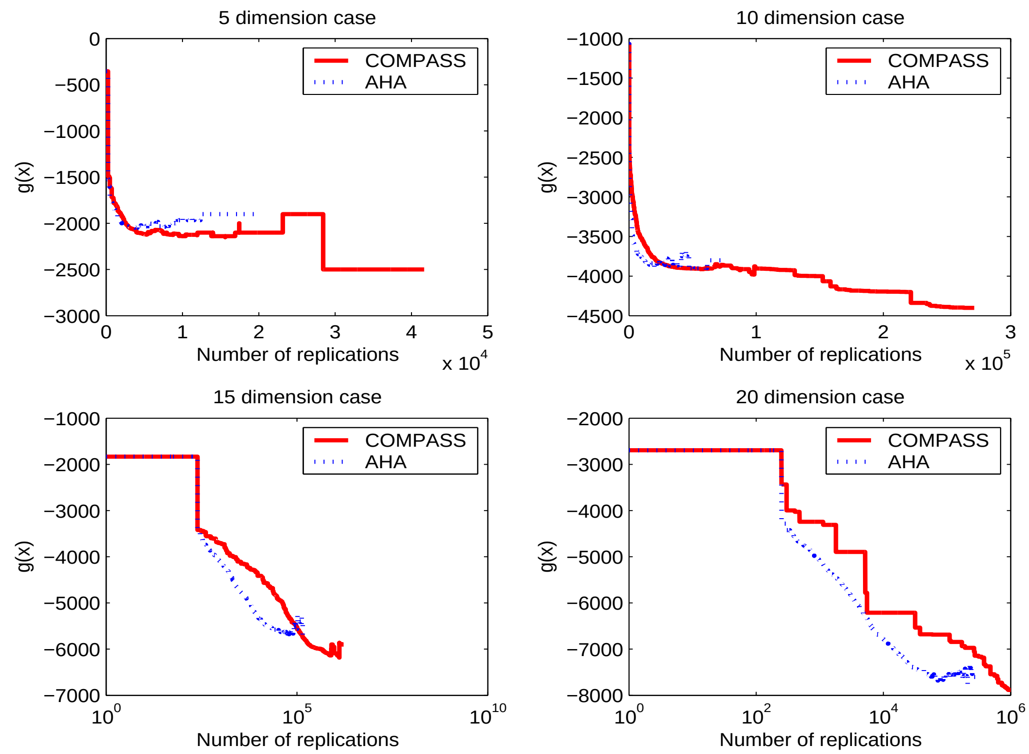

Figure 7: Performance plot for the high-dimensional multimodal test problem: D=5, 10, 15 and 20.