HiMOSS: A Novel High-Dimensional Multi-Objective Optimization Method via Adaptive Gradient-Based Subspace Sampling for Analog Circuit Sizing

Tianchen Gu[1], Ruiyu Lyu[1], Zhaori Bi[1,*], Changhao Yan[1], Fan Yang[1], Dian Zhou[1,4,†], Tao Cui[2], Xin Liu[2], Zaikun Zhang[3], and Xuan Zeng[1,*]

[1] State Key Laboratory of Integrated Chips and Systems, School of Microelectronics, Fudan University, Shanghai, China
[2] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, and the University of Chinese Academy of Sciences, Beijing, China
[3] Department of Applied Mathematics, the Hong Kong Polytechnic University, Hong Kong, China
[4] Department of Electrical Engineering, the University of Texas at Dallas, Richardson, Texas, USA

ABSTRACT

This study presents a novel high-dimensional multi-objective optimization method via adaptive gradient-based subspace sampling for analog circuit sizing. To handle constrained multi-objective optimization, we exploit promising regions from a non-crowded Pareto front, with lightweight Bayesian optimization (BO) based on a novel approximate constrained expected hypervolume improvement. This lightweight BO is computational efficient with constant complexity concerning simulation numbers. To tackle high-dimensional challenges, we reduce the effective dimensionality around promising regions by sampling candidates in an adaptive subspace. The subspace is constructed with gradients and previous success steps with their significance decaying over iterations. The gradients are approximated by sparse regression without additional simulations. The experiments on synthetic benchmarks and analog circuits illustrate advantages of the proposed method over Bayesian and evolutionary baselines.

1 INTRODUCTION

As technology scales down, the demands for high-performance analog circuits intensifying, analog circuit automation tools encounter challenges arising from circuit scale, complex specification trade-offs, and expensive simulation cost [1]. In the recent decades, many efforts have been made to address analog circuit sizing as a single-objective optimization (SOO) problem, relying on manually defined figure of merits (FOM) [2]. For a new circuit lacking prior knowledge, automated multi-objective optimization (MOO) methods are more practical [3, 4]. However, these methods are unable to handle high-dimensional challenges from large numbers of design parameters. An insightful sample-efficient high-dimensional MOO approach is urgently needed for current analog circuit design.

Generally, current MOO methods can be classified into three categories: non-dominance sorting-based, decomposition-based, and indicator-based algorithms. The first category employs non-dominance sorting mechanism along with a diversity maintenance technique to select the elitists. For example, NSGA-II [5] uses a crowding distance mechanism to select the non-crowded pareto front points as parent points. Algorithms in the second category decompose the MOO problem into multiple SOO sub-problems, and solutions of all the sub-problems forms the final Pareto set. MOEA/D [6] solves these sub-problems with a set of predefined weights to combine the comflicting objectives into a set of FOMs. Indicator-based algorithms transform MOO problems into SOO problems using multi-objective performance indicators. Hypervolume [7] of the Pareto front is a widely used indicator, especially in real-world problems. Most of the Bayesian MOO methods fall into the this category, employing expected hypervolume improvement (EHVI) as acquisition function [8].

To address the curse of dimensionality, state-of-the-art sample-efficient high-dimensional SOO methods stand for two main approaches: region segmentation and dimension reduction. Region segmentation methods concentrate on promising local regions, avoiding inefficient exploration across the entire space. TuRBO [9, 10] utlizes trust region technique to enhance the high-dimensional capability of Bayesian optimization (BO). cVTS [11] uses Voronoi diagrams to segment the previous region into a promising part and an inferior part, and employs a decision tree to avoid exploiting inferior regions. Dimension reduction methods reduce the dimensionality based on assumptions on the intrinsic effective dimensionality. Random embedding BO [12] constructs a global low-dimensional embedding through randomly generated patterns. Additive Gaussian process BO [13] globally decouples dimensions into several groups, allowing for independent model training and optimization. Another research provides a 2-dimensional subspace based on approximate gradients for efficient high-dimensional exploitation, incorporated with globa BO exploration [14].

Concerning high-dimensional MOO problems, practical sample-efficient methods are still lacking. The state-of-the-art Bayesian method for high-dimensional MOO problems, MORBO [15, 16], integrates the trust region technique with EHVI acquisition function.

*Corresponding authors:{zhaori_bi, xzeng}@fudan.edu.cn
†Emeritus Professor, the University of Texas at Dallas.

DAC '24, June 23–27, 2024, San Francisco, CA, USA

ACM ISBN 979-8-4007-0601-1/24/06…$15.00
https://doi.org/10.1145/3649329.3657318

LoCoMOBO [17] employs Thompson sampling and NSGA-II, collarborated with the trust region technique for analog circuit sizing. However, these HV-based trust region BO methods face at least three weakness. Firstly, exploitation within trust regions remains a challenging high-dimensional MOO problem. Secondly, these methods calculate constrained EHVI with sampling estimators, unable to accurately estimate extremely small EHVI values and their differences. This limitation makes them unsuitable to handle tight constraints, especially in circuits with strict performance requirements. Moreover, EHVI computation complexity is prohibitively high, scaling as $O(m^2N_p^m)$ with the number of objectives m and the quantity of Pareto front points N_p, posing a scalability issue [7].

To address high-dimensional MOO challenges in analog circuit sizing, we propose a novel method employing adaptive gradient-based subspace sampling, named as HiMOSS. The overall algorithm comprise three key steps: promising region selection, adaptive subspace sampling, and subspace optimization steps. To maintain the Pareto front diversity, we select the promising region centered at a pruned Pareto front and allocate simulation budget with probabilities proportional to their hypervolume contributions. Subsequently, we reduce the dimensionality around promising regions using adaptive gradient-based subspace sampling with truncated Gaussian distributions. We incrementally update the covariance of distribution, incorporating new features from gradients and success steps while allowing previous features to decay. The gradient features are approximated based on directional derivatives without additional simulations. Subspace optimization involves a lightweight BO method that uses approximate EHVI based on a constant-scale Pareto front set, avoiding the increase of $O(m^2N_p^m)$ complexity. It also utilizes adaptive probability of feasibility for constraint compatibility. The overall contributions of this study can be summarized as follows.

- We propose a novel MOO framework incorporated with adaptive gradient-based subspace sampling. We employ the promising region selection and lightweight BO based on approximate EHVI to solve the MOO problem. By adaptively updating the sampling covariance matrix, we construct the effective local subspace around promising regions with gradients and success steps, combining the advantages of region segmentation and dimension reduction methods to solve the high-dimensional challenges.
- We propose a novel gradient approximation method based on directional derivatives and locally weighted sparse regression without additional simulations.
- We approximate EHVI to prevent the increase of $O(m^2N_p^m)$ complexity, incorporated with adaptive probability of feasibility to handle constraints.

The proposed HiMOSS method exhibits advantages over state-of-the-art methods in experimental comparisons. In synthetic benchmarks, HiMOSS achieves up to 20.00× evaluation speedups and over 5.26× algorithm complexity reduction compared to Bayesian baselines. In real-world analog circuits, HiMOSS achieves up to 7.72× simulation number speedups over compared algorithms.

2 BACKGROUND

2.1 Problem Formulation

Given a circuit topology, a d-dimensional normalized design parameter vector denoted as $\boldsymbol{x}$, the analog circuit sizing problem can be formulated as the constrained multi-objective optimization (MOO)

$$\begin{aligned} &\underset{\boldsymbol{x}\in[0,1]^d}{\text{minimize}} \quad \boldsymbol{F}(\boldsymbol{x}), \\ &\text{s.t.} \quad \boldsymbol{Cons}(\boldsymbol{x}) \le \boldsymbol{0}, \end{aligned} \tag{1}$$

where $\boldsymbol{F}(\cdot) = [f^{(1)}(\cdot), ..., f^{(m)}(\cdot)]^T$ denotes the objective functions and $\boldsymbol{Cons}(\cdot) = [con^{(1)}(\cdot), ..., con^{(n)}(\cdot)]^T$ denotes the constraint functions, m and n are the number of objectives and constraints respectively. In MOO problems, a point $\boldsymbol{x}_1$ is said to dominate another point $\boldsymbol{x}_2$, denoting as $\boldsymbol{x}_1 \prec \boldsymbol{x}_2$, if $f^{(i)}(\boldsymbol{x}_1) \le f^{(i)}(\boldsymbol{x}_2)$ for all $i \in \{1, ..., m\}$, and there exists $j \in \{1, ..., m\}, f^{(j)}(\boldsymbol{x}_1) < f^{(j)}(\boldsymbol{x}_2)$.

As shown in Fig.1(a), a Pareto front set $\mathcal{S}$ is the set of points which is not dominated by any element of the whole set $\boldsymbol{X}$

$$\mathcal{S}(\boldsymbol{X}) = \{\boldsymbol{x} \in \boldsymbol{X} \mid \nexists \boldsymbol{x}' \in \boldsymbol{X}, \boldsymbol{x}' \prec \boldsymbol{x}\}. \tag{2}$$

For constrained MOO problems, the Pareto front should meet the constraints $\mathcal{S}_c(\boldsymbol{X}) = \mathcal{S}(\{\boldsymbol{x} \in \boldsymbol{X} \mid \boldsymbol{Cons}(\boldsymbol{x}) \le \boldsymbol{0}\})$. Denoting the optimal Pareto front as rank 1, $\mathcal{S}_1(\cdot) := \mathcal{S}(\cdot)$, we can sort dominated Pareto front with rank $i > 1$ as $\mathcal{S}_i(\boldsymbol{X}) = \mathcal{S}(\boldsymbol{X} \backslash \bigcup_{j=1}^{i-1} \mathcal{S}_j(\boldsymbol{X}))$.

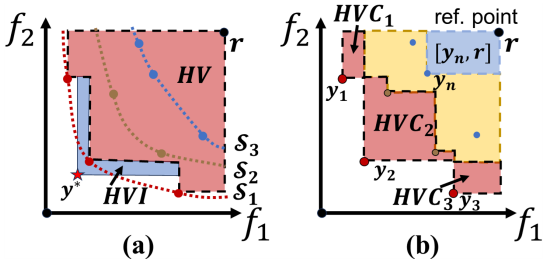

Figure 1: The illustration of (a) Pareto front with rank 1 to 3, hypervolume (HV), hypervolume improvement (HVI) and (b) hypervolume contribution (HVC).

2.2 Bayesian Optimization

Bayesian optimization (BO) is a global optimization method incorporated with uncertainty models. Given a dataset $D = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^N$, an uncertainty model gives predictive distribution of the fitness $y \sim \pi(y \mid \boldsymbol{x}, D)$. The widely used Gaussian process (GP) model gives prediction as Gaussian distributions. BO methods select the candidate points $\boldsymbol{x}^*$ for simulation by optimizing an acquisition function (AF) α as

$$\boldsymbol{x}^* := \arg\max_{\boldsymbol{x}} \ \alpha(\boldsymbol{x}). \tag{3}$$

Majority of MOO Bayesian methods focus on the hypervolume (HV) indicator-based approach

$$HV(\boldsymbol{Y}, \boldsymbol{r}) = \lambda_d(\bigcup_{i=1}^{|Y|}[\boldsymbol{y}_i, \boldsymbol{r}]), \tag{4}$$

where $\boldsymbol{Y}$ is the objective dataset, $\boldsymbol{r}$ is the reference point, λ_d is the d-dimensional Lebesgue measure, and $[\boldsymbol{y}_i, \boldsymbol{r}]$ denotes the hyperrectangle bounded by $\boldsymbol{y}_i$ and $\boldsymbol{r}$. As shown in Fig.1, the incremental hypervolume improvment (HVI) and exclusive hypervolume contribution (HVC) of a point $\boldsymbol{x}$ are given as

$$HVI/HVC(\boldsymbol{Y}, \boldsymbol{r}, \boldsymbol{x}) = HV(\boldsymbol{Y} \cup \{\boldsymbol{y}_x\}, \boldsymbol{r}) - HV(\boldsymbol{Y} \backslash \{\boldsymbol{y}_x\}, \boldsymbol{r}). \tag{5}$$

Most of HV-based Bayesian MOO method use the expected hypervolume improvement (EHVI) [8] as acquisition function,

$$\alpha_{EHVI}(\boldsymbol{x}) = \mathbb{E}[HVI(\boldsymbol{Y}, \boldsymbol{r}, \boldsymbol{x})] = \int HVI(\boldsymbol{Y}, \boldsymbol{r}, \boldsymbol{x}) \pi(\boldsymbol{y}_x) d\boldsymbol{y}_x, \tag{6}$$

where $\mathbb{E}[\cdot]$ denotes the expectation operation.

Algorithm 1: HiMOSS: high-dimensional multi-objective optimization method via adaptive gradient-based subspace sampling

Input: Objectives $\boldsymbol{F}$, constraints $\boldsymbol{Cons}$, # simulation N_{max}, batch size λ, # promising regions μ.
Output: Pareto set $\mathcal{S}$.

Initialization. Dataset $D = \{(\boldsymbol{x}_i, \boldsymbol{F}_i, \boldsymbol{Cons}_i)\}$ with random intial samples, distribution dictionary $\mathbb{C} = \{\boldsymbol{x}_i : (0.3, \boldsymbol{I})\}$;
while $|D| < N_{max}$ **do**
  Promising region selection. Select promising region centers from pruned non-crowded Pareto front by Algo.2; Allocate each simulation budget to promising region centers by a softmax probability as (9); Prepare the empty simulation queue $\boldsymbol{X} = \emptyset$;
  for $i = 1 \ldots \lambda$ **do**
    Adaptive subspace sampling. Find the center of parent subspace $\boldsymbol{x}_p$ and its distribution parameters $(\sigma_p, \boldsymbol{C}_p)$; Sample random points in truncated Gaussian distribution $\boldsymbol{X}_p^{(s)} \sim \mathcal{N}(\boldsymbol{x}_p, \sigma_p^2 \boldsymbol{C}_p)$ as (12);
    Subspace optimization. Select the candidate $\boldsymbol{x}_i$ with the best a-CEHVI value calculated by (15) ; Update the simulation queue $\boldsymbol{X} \leftarrow \boldsymbol{X} + \{\boldsymbol{x}_i\}$;
  end
  Simulation. $D \leftarrow D + \{(\boldsymbol{x}_i, \boldsymbol{F}_i, \boldsymbol{Cons}_i) \mid \boldsymbol{x}_i \in \boldsymbol{X}\}$;
  Subspace update. For $\boldsymbol{x}_i \in \boldsymbol{X}$: Calculate approximate gradients $\boldsymbol{G}_i$ as (20); Update the covariance incrementally as (11a); Expand $\sigma_i \leftarrow \eta\sigma_i$ for new region and its parent if success, otherwise $\sigma_i \leftarrow \sigma_i/\eta$;
end
return the pareto set $\mathcal{S}$ of the dataset D.

2.3 Gradient-Based Subspace Technique

The gradient-based subspace technique for the high-dimensional SOO problem, originated from Yuan [18], iteratively exploits the 2-dimensional subspace around current center $\boldsymbol{x}_i$ spanned by current gradient $\boldsymbol{g}_i$ and last descent step $\boldsymbol{d}_i$ as

$$subS_i = \text{span}\{\boldsymbol{g}_i, \boldsymbol{d}_i\}, \text{ with } \boldsymbol{d}_{i+1} := \arg\min_{\boldsymbol{d} \in subS_i} f(\boldsymbol{x}_i + \boldsymbol{d}). \quad (7)$$

This subspace is inspired from the conjugate gradient since the conjugate gradient step is contained in this subspace. It is also extended to a multi-dimensional form inspired by quasi-Newton methods [19].

$$subS_i = \text{span}\{\boldsymbol{g}_i, \boldsymbol{g}_{i-1}, ..., \boldsymbol{d}_i, \boldsymbol{d}_{i-1}, ...\}. \quad (8)$$

3 PROPOSED METHOD

In this section, we present the proposed high-dimensional multi-objective optimization method via adaptive gradient-based subspace sampling (HiMOSS).

3.1 Algorithm framework

We propose the MOO framework as Algo.1, which comprises initialization, promising region selection, adaptive subspace sampling, subspace optimization, and subspace update stages.

Initialization. We initialize the dataset D by uniformly random sampling across the entire search space. We employ truncated

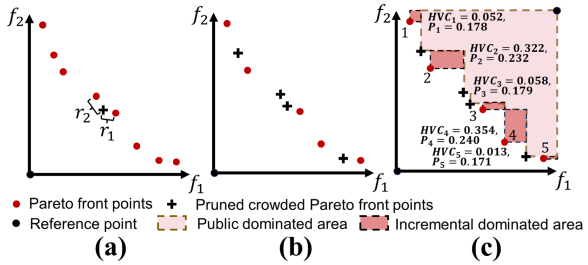

Figure 2: The illustrations of (a) pruning the most crowded Pareto front point, (b) pruned non-crowded Pareto front, and (c) stochastic budget allocation with probability w.r.t. HVCs.

Gaussian distributions to realize the adaptive gradient-based subspace sampling. A dictionary $\mathbb{C}$ stores the parameters of these Gaussian distributions. The initial deviation σ is set to 0.3 with normalized search space, and initial covariance $\boldsymbol{C}$ is set to the identity matrix. We will remove the distributions for dominated inferior regions from the dictionary to save memory.

Promising region selection. We select promising regions centered around non-crowded Pareto front points to maintain efficiency and diversity. The selection process involves two steps: non-crowded Pareto front pruning and stochastic budget allocation. We prune the Pareto front based on average distance of the nearest neighbours in objective space as in Algo.2 and Fig.2(a-b). We iteratively remove the most crowded Pareto front point until the scale of remaining Pareto set is no more than μ ($\mu = 5$ in this work).

As shown in Fig.2(c), each of λ simulation budget is assigned to one of the promising regions by a softmax probability based on the HVC and constraint violation (CV) of the center points

$$P_i = softmax(HVC_i - CV_i) = \frac{\exp(HVC_i - CV_i)}{\sum_j \exp(HVC_j - CV_j)}, \quad (9)$$

where HVC_i and $CV_i = \sum_j \max(con_{pi}^{(j)}, 0)$ denote the hypervolume contribution and total constraint violation of the i-th promising region center. Equation (9) encourages allocating more simulation budget around promising regions with better HVC and lower CV.

Adaptive subspace sampling. Once a simulation budget is designated, we sample random candidates within a gradient-based subspace, subject to a truncated Gaussian distribution as detailed in Section 3.2. The Gaussian distributions are parametrized by mean, deviation, and covariance parameters. The promising region center $\boldsymbol{x}_p$ serves as the mean, and the deviation σ and covariance $\boldsymbol{C}$ can be found in the dictionary.

Subspace optimization. Random points in the subspace are chosen by BO to enhance the simulation efficiency and constraint compatiblity. We employ a lightweight BO based on approximate constrained EHVI (a-CEHVI) as detailed in Section 3.3. The selected promising candidates are then simulated in parallel.

Subspace update. Taking the terminology of evolutionary algorithms, we refer to the previous subspace as the parent subspace. For all newly simulated points, approximate gradients are calculated based on directional derivatives and locally weighted sparse regression without additional simulations, detailed in Section 3.4.

Then we incrementally update their covariance matrix from that of their parents as (11a) to integrate new gradients and success step. A step is considered as successful when the simulation point is within the current Pareto set. Failure steps are not incorporated

Algorithm 2: Pareto front pruning

Input: Pareto set $\mathcal{S}$, target # Pareto n_{pf}, # objectives m.
Output: Pruned Pareto set $\mathcal{S}_{pr}$.

while $|\mathcal{S}| > n_{pf}$ **do**
  For $\boldsymbol{x}_i \in \mathcal{S}$: Calculate average Euclidean distance $\bar{r}_i$ of nearest $(m+1)$ neighbours of $\boldsymbol{F}(\boldsymbol{x}_i)$ in objective space;
  Remove the most crowded Pareto front point from $\mathcal{S}$;
end
return pruned Pareto set $\mathcal{S}_{pr} = \mathcal{S}$.

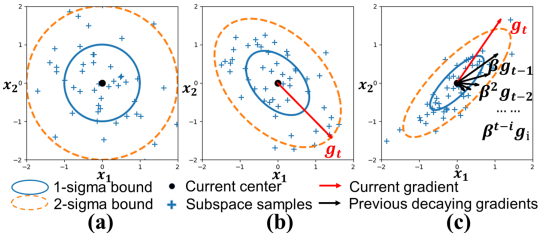

Figure 3: The illustration of adaptive gradient-based subspace sampling: (a) initial full dimension exploration, (b) adaptive subspace incorporating gradient, and (c) effective low-dimensional subspace.

into the covariance matrix. If a step is successful, the deviation parameter σ would expand with a ratio $\eta > 1$ for both this subspace and its parent. Otherwise, σ would shrinkage with a ratio $1/\eta$ for both. In this work, we set $\eta = 1.02$.

3.2 Adaptive Gradient-Based Subspace Sampling

We innovatively employ gradient information to address high-dimensional challenges in MOO problems through an adaptive gradient-based subspace sampling method based on truncated Gaussian distributions. Similar to (7), a marginal effective subspace for MOO problems could be spanned by gradients of objectives $\boldsymbol{g}^{(i)}$ and the success step $\boldsymbol{d}$.

$$subS^{(m)} = \text{span}\{\boldsymbol{g}^{(1)}, ..., \boldsymbol{g}^{(m)}, \boldsymbol{d}\} \tag{10}$$

To ensure robustness and effectiveness, we introduce more prior features from parent subspaces. Contrary to using a subspace with equally significant features from different iterations like (8), we attempt to priortize features from recent iterations to enhance their importance, allowing previous features to decay. Employing a decaying weight $\beta < 1$ for previous features from parent subspace covariance $\boldsymbol{C}_{t-1}$, we can construct the current covariance matrix $\boldsymbol{C}_t$ with the current gradients $\boldsymbol{G}_t$ and success step $\boldsymbol{d}_t$ using incremental update form as (11a) or equivalent accumulative form as (11b)

$$\boldsymbol{C}_t = \beta \boldsymbol{C}_{t-1} + \frac{\boldsymbol{G}_t \boldsymbol{G}_t^T + \boldsymbol{d}_t \boldsymbol{d}_t^T}{m+1}, \tag{11a}$$

$$= \beta^t \boldsymbol{C}_0 + \sum_{i=1}^{t} \frac{\beta^{t-i}}{m+1} (\boldsymbol{G}_i \boldsymbol{G}_i^T + \boldsymbol{d}_i \boldsymbol{d}_i^T), \tag{11b}$$

where $\boldsymbol{G}$ is the matrix containing normalized column approximate gradients of objectives and constraint violation calculated by (20). Specifically, $\boldsymbol{G}$ contains only one gradient of constraint violation, when the center point is infeasible, and $\boldsymbol{d}$ would be a zero vector if the last step is unsuccessful. $\boldsymbol{C}_0$ is the intial identity covariance matrix. We set $\beta = 0.9$ in this work, enabling previous features to decay by half of their significance after 7 generations. When close to convergence, the effective dimensionality is from $(m+1)$ to 20 depending on linear correlations of $\boldsymbol{G}$s and $\boldsymbol{d}$s from different iterations.

As illustrated in Fig.3, the orientation of subspace samples is mainly affected by the current gradient, with previous gradient features exponentially decaying over iterations with weight β. Consequently, samples are following eigenvectors of the covariance matrix, with distribution ranges corresponding to eigenvalues. The eigenvectors with small eigenvalues are effectively dropped, reducing the high-dimensional problem into an effective low-dimensional search. Given a promising region center $\boldsymbol{x}_p$, the covariance matrix $\boldsymbol{C}_p$ and deviation σ_p, we draw subspace samples subject to Gaussian distribution truncated by design space

$$\boldsymbol{X}_p^{(s)} = \boldsymbol{x}_p + \sigma_p \boldsymbol{W}_p \text{diag}(\boldsymbol{\Lambda_p}) Z \sim \mathcal{N}(\boldsymbol{x}_p, \sigma_p^2 \boldsymbol{C}_p), \tag{12}$$

where $\boldsymbol{C}_p$ is eigen decomposed as $\boldsymbol{W}_p \text{diag}(\boldsymbol{\Lambda}_p^2) \boldsymbol{W}_p^T$, and $Z \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ denotes samples from standard d-variate normal distribution.

3.3 Lightweight Bayesian Optimization

To improve the efficiency and constraint compatibility, we employ a lightweight BO to select promising candidates from subspace samples $\boldsymbol{X}_p^{(s)}$. The traditional Bayesian MOO methods encounter huge computation burdens due to $O(N^3)$ complexity of Gaussian process (GP) model in relation to the number of training samples N, and the $O(m^2 N_p^m)$ cost of hypervolume computation with respect to the number of Pareto front points N_p and the number of objectives m. The quantities for N and N_p can easily exceed thousands in high-dimensional MOO problems, resulting in scalability issues.

We reduce the aforementioned computational cost by employing a fixed number of training samples for GP models training and Pareto front points for EHVI computation. When training GP models, we exclusively consider the $(2 \times d + 1)$ nearest neighbors of the current region center $\boldsymbol{x}_p$ in the dataset. For EHVI approximation, we prune the crowded Pareto front by Algo.2 to obtain a non-crowded Pareto front with at most $25 \times m$ Pareto front points, calculating the approximate EHVI (a-EHVI) as

$$\alpha_{a-EHVI}(\boldsymbol{x}) = \mathbb{E}[HVI(\boldsymbol{Y}_{pr}, \boldsymbol{x}, \boldsymbol{r})], \tag{13}$$

where $\boldsymbol{Y}_{pr}$ denotes the objectives of the pruned non-crowded Pareto front. Concerning tight constraints, we use adaptive probability of feasibility (APF) as

$$\alpha_{APF}(\boldsymbol{x}) = \prod_{i=1}^{n} P(con_x^{(i)} \le \max(0, con_p^{(i)})), \tag{14}$$

where $P(\cdot)$ denotes the probability given by GP posterior prediction, and $con_p^{(i)}$ is the i-th constraint of the current region center. The APF prevents the extremely small value due to the tight constraints. The constrained approximate EHVI (a-CEHVI) acquisition function used in the lightweight BO is derived as

$$\alpha_{a-CEHVI}(\boldsymbol{x}) = \begin{cases} \alpha_{APF}(\boldsymbol{x}), & \boldsymbol{x}_p \text{ is infeasible} \\ \alpha_{APF}(\boldsymbol{x}) \cdot \alpha_{a-EHVI}(\boldsymbol{x}), & \text{else} \end{cases} \tag{15}$$

3.4 Gradient Approximation

We introduce a novel gradient approximation approach based on directional derivatives and locally weighted regression, requiring no additional simulations. A directional derivative of function f

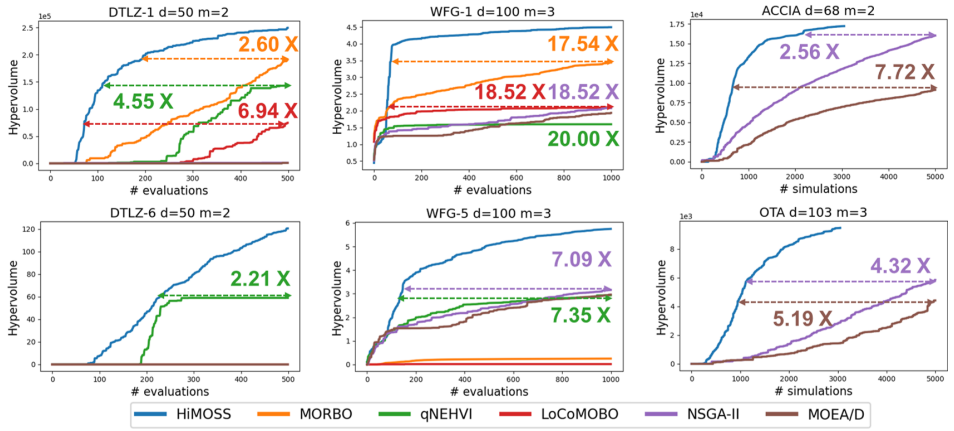

Figure 4: The hypervolume convergence curves on DTLZ and WFG synthetic benchmarks and ACCIA, OTA analog circuits. The optimization dimensionality and number of objectives are denoted as d and m in the sub-figure titles. The reference points are $(1000, 1000)$, $(20, 20)$, $(3, 3, 3)$, and $(3, 3, 3)$ for DTLZ-1, DTLZ-6, WFG-1, and WFG-5, respectively.

with respect to a unit vector $\boldsymbol{u}$ can be defined as

$$f'_{\boldsymbol{u}}(\boldsymbol{x}) = \lim_{\rho \to 0} \frac{f(\boldsymbol{x} + \rho \boldsymbol{u}) - f(\boldsymbol{x})}{\rho} = \boldsymbol{u}^T \boldsymbol{g}(\boldsymbol{x}), \tag{16}$$

where $\boldsymbol{g}(\boldsymbol{x})$ represents the gradient at $\boldsymbol{x}$. We modified (16) with a neighbor observation $(\boldsymbol{x}', f(\boldsymbol{x}'))$ as

$$f(\boldsymbol{x}') - f(\boldsymbol{x}) \approx (\boldsymbol{x}' - \boldsymbol{x})^T \boldsymbol{g}(\boldsymbol{x}), \tag{17}$$

Collecting the nearest $d+1$ observed neighbors of a center point $\boldsymbol{x}$ within the d-dimensional space, an equation set is established as

$$f(\boldsymbol{x} + \boldsymbol{\Psi}) - f(\boldsymbol{x}) = \Delta \boldsymbol{y} \approx \boldsymbol{\Psi}^T \boldsymbol{g}(\boldsymbol{x}), \tag{18}$$

where $\boldsymbol{\Psi}$ is a matrix with column difference vectors. We approximate the gradient by solving the linear problem (18), reformulating it with sparse ridge regression and incorporating weights based on distance for robustness.

$$\tilde{\boldsymbol{g}}(\boldsymbol{x}) := \underset{\boldsymbol{g}}{\text{minimize}} \quad |\boldsymbol{\omega}(\Delta \boldsymbol{y} - \boldsymbol{\Psi}^T \boldsymbol{g})|_2^2 + \alpha |\boldsymbol{g}|_2^2, \tag{19}$$

where α is the weight of ridge regression. $\boldsymbol{\omega}$ is the diagonal matrix with $\omega_{ii} = \exp(-|\boldsymbol{\Psi}_i|_2 / 2l)$ being the sample weight based on the distance of difference vector, where $l = 0.2$ is an emperical setting for normalized $\boldsymbol{\Psi}$. The explicit solution of (19) is derived as

$$\tilde{\boldsymbol{g}}(\boldsymbol{x}) = (\boldsymbol{\Psi}^T \boldsymbol{\omega}^2 \boldsymbol{\Psi} + \alpha \boldsymbol{I})^{-1} \boldsymbol{\Psi}^T \boldsymbol{\omega}^2 \Delta \boldsymbol{y} \tag{20}$$

4 EXPERIMENTS

We implement the proposed HiMOSS method with PyTorch. The performances are tested on synthetic benchmarks and analog circuit designs. The proposed method is compared to non-dominance sorting-based method NSGA-II [5], decomposition-based method MOEA/D [6], indicator-based Bayesian method qNEHVI [8], and state-of-the-art high-dimensional multi-objective Bayesian methods MORBO [15] and LoCoMOBO [17]. The number of trust regions is set to 5. The number of random initial points is set to 50, and the initialization time is counted in total runtime. For robustness, we run experiments 10 times independently. Simulations are invoked parallelly with batch size 5 for all methods. Experiments are conducted on a Linux server with Intel Xeon 6230 CPUs.

4.1 Synthetic benchmarks

We utilize four multi-objective benchmarks from DTLZ and WFG benchmarks, with 50 and 100 dimensions, to test the performance

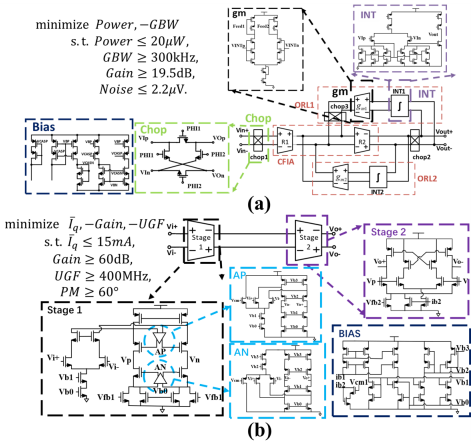

Figure 5: The schematic and multi-objective optimization formulation of (a) the ACCIA circuit and (b) the OTA circuit.

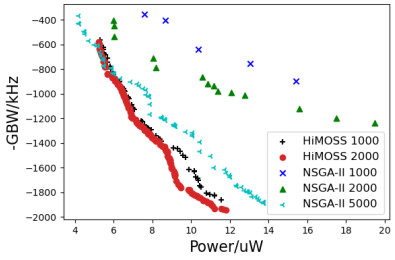

Figure 6: The Pareto fronts of ACCIA circuit. The red dots, representing HiMOSS with 2000 simulation budget, exhibit superior Pareto front quality.

of the proposed HiMOSS method. The maximum evaluation budget is set to $10 \times d$, corresponding to the dimensionality d. HiMOSS achieves the overall best and robust performances in all benchmarks, as shown in Fig.4. Notably, NSGA-II and MOEA/D fail to find feasible solutions in DTLZ-1 and DTLZ-6 functions, while MORBO and LoCoMOBO fail in DTLZ-6 and WFG-5 functions.

In terms of evaluation efficiency, HiMOSS demonstrates significant speedups compared to successful runs of other algorithms. Specifically, it achieves 2.60 ∼ 17.54× speedups compared to MORBO,

Table 1: Optimization Results of Real-World Analog Circuits

Circuit	Algo.	Hypervolume	# simu	# simu speedup	runtime	runtime speedup	# success
ACCIA	**HiMOSS**	**16133 ± 3070**	**1953**	**2.56×**	**118.16h**	**2.48×**	**10/10**
	NSGA-II	16034 ± 3067	5000	1.0×	293.05h	1.0×	10/10
	MOEA/D	9138 ± 2858	5000	1.0×	293.15h	0.99×	10/10
	MORBO	fail	3000	-	249.02h	-	0/10
	qNEHVI	fail	3000	-	475.41h	-	0/10
	LoCoMOBO	fail	3000	-	191.64h	-	0/10
OTA	**HiMOSS**	**5892 ± 3422**	**1157**	**4.32×**	**4.84h**	**3.39×**	**10/10**
	NSGA-II	5844 ± 4240	5000	1.0×	16.39h	1.0×	10/10
	MOEA/D	4418 ± 3811	5000	1.0×	16.22h	1.01×	10/10
	MORBO	fail	3000	-	97.02h	-	0/10
	qNEHVI	fail	3000	-	44.02h	-	0/10
	LoCoMOBO	fail	3000	-	64.95h	-	0/10

2.21 ∼ 20.00× speedups compared to qNEHVI, 6.94 ∼ 18.52× speedups compared to LoCoMOBO, and 7.09 ∼ 18.52× speedups compared to NSGA-II. In terms of algorithm complexity, with 500 evaluation budget, HiMOSS takes a total average runtime of 856.3s, whereas MORBO, qNEHVI, and LoCoMOBO require 17793.5s, 4505.8s, and 6258.4 respectively. NSGA-II and MOEA/D have total runtimes of 4.3s and 5.7s. Compared to Bayesian baselines, HiMOSS achieves over 5.26× algorithm complexity reduction.

4.2 Real-world Analog Circuits

In this subsection, we present experiments on two analog circuits: AC-coupled instrumentation amplifier (ACCIA) and operational trans-impedance amplifier (OTA). The optimized circuit parameters include the size of transistors, values of resistors and capacitors. Symmetry parameters (e.g., of differential pairs and current mirrors) are merged. The ACCIA, with 68 independent design variables, 2 objectives, and 4 main constraints as shown in Fig.5(a), takes average 18 minutes for one simulation. The OTA, with 103 independent design variables, 3 objectives, and 4 main constraints as shown in Fig.5(b), takes about 1 minute for one simulation. Other constraints like large signal behavior and operating points are considered. The reference points serve as constraints on objectives. These circuit performance requirements are practical and relatively strict.

The MORBO, qNEHVI, LoCoMOBO fail to find feasible designs in all circuit experiments. The hypervolume convergence curves is illustrated in Fig.4 and the Pareto fronts of ACCIA are depicted in Fig.6. Compared to NSGA-II, HiMOSS achieves a 2.56× simulation number speedup and a 2.48× total runtime speed up reaching similar hypervolume for ACCIA circuit, and a 4.32× simulation number speedup and a 3.39× runtime speedup for OTA circuit. Compard to MOEA/D, the proposed method achieves 7.72× and 5.19× simulation number speedups for ACCIA and OTA circuits.

5 CONCLUSION

We introduce HiMOSS, a novel high-dimensional multi-objective optimization method via adaptive gradient-based subspace sampling. The approach involves promising region selection from a non-crowded Pareto front, and a lightweight BO based on approximate constrained EHVI for MOO problems with tight constraints. To address high-dimensional challenges, HiMOSS leverages adaptive gradient-based subspace sampling based on truncated Gaussian distributions. The gradients are approximated by sparse regression without additional simulations, and the lightweight BO is computational efficient with constant complexity regarding the sample scale. The results on synthetic benchmarks and analog circuits demonstrate the effectiveness and advantages of the HiMOSS method.

ACKNOWLEDGMENTS

This research is supported partially by National Key R&D Program of China 2020YFA0711900, 2020YFA0711901, partly by National Natural Science Foundation of China (NSFC) research projects 62141407, 62304052, 92373207, 62090025, partly by Beijing Natural Science Foundation (Grant NO. Z220003), partly by the Hong Kong GRC under Grant PolyU 153054/20P and Grant PolyU 153066/21P.

REFERENCES

[1] R. A. Rutenbar, G. G. E Gielen, and J. Roychowdhury. Hierarchical modeling, optimization, and synthesis for system-level analog and rf designs. *Proceedings of the IEEE*, 95(3):p.640–669, 2007.

[2] R. Phelps, M. Krasnicki, R.A. Rutenbar, L.R. Carley, and J.R. Hellums. Anaconda: simulation-based synthesis of analog circuits via stochastic pattern search. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 19(6):703–717, 2000.

[3] Bo Liu, Hadi Aliakbarian, Soheil Radiom, Guy A. E. Vandenbosch, and Georges Gielen. Efficient multi-objective synthesis for microwave components based on computational intelligence techniques. In *2012 Design Automation Conference (DAC)*, pages 542–548, 2012.

[4] Wenlong Lyu, Fan Yang, Changhao Yan, Dian Zhou, and Xuan Zeng. Multi-objective bayesian optimization for analog/rf circuit synthesis. In *2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC)*, pages 1–6, 2018.

[5] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

[6] Qingfu Zhang and Hui Li. Moea/d: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, 11(6):712–731, 2007.

[7] Lyndon While, Lucas Bradstreet, and Luigi Barone. A fast way of calculating exact hypervolumes. *IEEE Transactions on Evolutionary Computation*, 16(1):86–95, 2012.

[8] Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. Differentiable expected hypervolume improvement for parallel multi-objective bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 33, pages 9851–9864. Curran Associates, Inc., 2020.

[9] David Eriksson, Michael Pearce, Jacob Gardner, Ryan D Turner, and Matthias Poloczek. Scalable global optimization via local bayesian optimization. *Advances in neural information processing systems*, 32, 2019.

[10] Konstantinos Touloupas, Nikos Chouridis, and Paul P Sotiriadis. Local bayesian optimization for analog circuit sizing. In *2021 58th ACM/IEEE Design Automation Conference (DAC)*, pages 1237–1242. IEEE, 2021.

[11] Aidong Zhao, Xianan Wang, Zixiao Lin, Zhaori Bi, Xudong Li, Changhao Yan, Fan Yang, Li Shang, Dian Zhou, and Xuan Zeng. cvts: A constrained voronoi tree search method for high dimensional analog circuit synthesis. In *2023 60th ACM/IEEE Design Automation Conference (DAC)*, pages 1–6, 2023.

[12] Hanbin Hu, Peng Li, and Jianhua Z Huang. Enabling high-dimensional bayesian optimization for efficient failure detection of analog and mixed-signal circuits. In *2019 56th ACM/IEEE Design Automation Conference (DAC)*, pages 1–6. IEEE, 2019.

[13] Tuotian Liao and Lihong Zhang. High-dimensional many-objective bayesian optimization for lde-aware analog ic sizing. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 30(1):15–28, 2022.

[14] Tianchen Gu, Wangzhen Li, Aidong Zhao, Zhaori Bi, Xudong Li, Fan Yang, Changhao Yan, Wenchuang Hu, Dian Zhou, Tao Cui, Xin Liu, Zaikun Zhang, and Xuan Zeng. Bbgp-sdfo: Batch bayesian and gaussian process enhanced subspace derivative free optimization for high-dimensional analog circuit synthesis. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 43(2):417–430, 2024.

[15] Samuel Daulton, David Eriksson, Maximilian Balandat, and Eytan Bakshy. Multi-objective bayesian optimization over high-dimensional search spaces. In *Proceedings of 38th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 507–517, 2022.

[16] Su Zheng, Hao Geng, Chen Bai, Bei Yu, and Martin D. F. Wong. Boosting vlsi design flow parameter tuning with random embedding and multi-objective trust-region bayesian optimization. *ACM Trans. Des. Autom. Electron. Syst.*, 28(5), sep 2023.

[17] Konstantinos Touloupas and Paul P. Sotiriadis. Locomobo: A local constrained multiobjective bayesian optimization for analog circuit sizing. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 41(9):2780–2793, 2022.

[18] Ya-xiang Yuan and J. Stoer. A subspace study on conjugate gradient algorithms. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 75(1):69–77, 1995.

[19] Ya-xiang Yuan. A review on subspace methods for nonlinear optimization. In *International Congress of Mathematicians*, volume 4, pages 807–827, 2014.