MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

PROXICBO: A CONSENSUS-BASED METHOD FOR COMPOSITE OPTIMIZATION

Zhang, Haoyu; Ma, Yanting; Kitichotkul, Ruangrawee; Rapp, Joshua; Boufounos, Petros T.

TR2026-041 March 28, 2026

Abstract

This paper presents an interacting-particle optimization method for composite optimization problems. The proposed approach combines ideas from consensus-based optimization (CBO) with proximal gradient descent. We establish theoretical convergence guarantees for the continuous-time finite-particle dynamics and introduce an alternating update scheme for practical implementation. Finally, we validate the effectiveness of our method across several signal processing problems, demonstrating its advantages over proximal gradient de- scent, CBO, and their variants.

IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2026

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

PROXICBO: A CONSENSUS-BASED METHOD FOR COMPOSITE OPTIMIZATION

Haoyu Zhang[⋆]*, Yanting Ma*[†]*, Ruangrawee Kitichotkul*[‡]*, Joshua Rapp*[†]*, Petros T Boufounos*[†]

[⋆]Department of Mathematics, University of California San Diego
[†] Mitsubishi Electric Research Laboratories (MERL)
[‡] Department of Electrical and Computer Engineering, Boston University

ABSTRACT

This paper presents an interacting-particle optimization method for composite optimization problems. The proposed approach combines ideas from consensus-based optimization (CBO) with proximal gradient descent. We establish theoretical convergence guarantees for the continuous-time finite-particle dynamics and introduce an alternating update scheme for practical implementation. Finally, we validate the effectiveness of our method across several signal processing problems, demonstrating its advantages over proximal gradient descent, CBO, and their variants.

***Index Terms*—** Composite optimization, interacting particles, consensus-based optimization, proximal mapping

1. INTRODUCTION

In this paper, we propose an interacting-particle method for solving *composite optimization* problems of the form

$$\min_{v\in\mathbb{R}^d} \{\mathcal{E}(v) := f(v) + g(v)\}, \tag{1}$$

where $f(v)$ is differentiable but possibly non-convex, and $g(v)$ is convex but possibly non-differentiable.

Composite optimization provides a unifying framework for many inverse problems in signal processing applications, where $f(v)$ is the data fidelity term that promotes measurement consistency and $g(v)$ is the regularization term that encodes prior knowledge about the underlying signal to be reconstructed. For example, in a single-photon lidar problem, we are interested in estimating the reflectivity and depth of a target given photon detection times. We can solve it by maximizing the (non-convex) log-likelihood function under positivity constraint for the parameters [1, 2]. In this case, f is the negative log-likelihood of a Poisson process and g is the indicator function of the constraint set. Another example is PDE-constrained optimization such as diffractive imaging [3, 4], where we have $f(v) = 1/2\|\mathcal{A}(v) - y\|_2^2$ with $\mathcal{A}$ being the map from dielectric permittivity to electric field, and g is usually chosen to be the image total variation semi-norm [5] to promote piecewise constant structure in the reconstructed image. When multiple scattering exists, $\mathcal{A}$ is nonlinear and f is non-convex.

Classical approaches such as proximal gradient descent [6] and its accelerated variant [7] have been widely used for composite optimization. However, it is well known that purely gradient-based methods are susceptible to local minima. Recently, consensus-based optimization (CBO) [8,9], a class of multi-agent metaheuristic methods, has emerged as a promising alternative. Through particle interactions, CBO and its variants have successfully tackled diverse non-convex problems. For instance, anisotropic diffusion has been introduced to handle high-dimensional problems in neural network training [10], while other works apply CBO to constrained problems such as phase retrieval [11–14], federated learning [15], and sampling [16]. Mirror-map extensions have shown success in sparse neural network training and constrained optimization [17]. A key advantage of CBO over other metaheuristics is its amenability to theoretical convergence analysis. Prior studies have analyzed the convergence of CBO in the mean-field limit, either under strong assumptions on particle initialization [9] or through refined frameworks that exploit loss landscape properties [18]. Other works have investigated the finite-particle to mean-field discrepancy [19, 20].

Related Works. A key feature of composite optimization is that the regularization term g encodes prior knowledge about the solution, such as ℓ_1 regularization for sparsity or total variation for piecewise-constant images. Several works within the CBO framework have sought to incorporate such prior information into the particle dynamics, beyond the vanilla variants [8–10].

When g is an indicator function, [12] introduces projection operators as a separate step, while [14] embeds the projection directly into the dynamics. Similarly, [11] and [13] incorporate indicator constraints as regularization either through the objective function or within the drift term. Another line of work employs mirror maps, extending the vanilla CBO framework in analogy to mirror descent [17]. However, these approaches either do not explicitly exploit the composite structure, namely, the interplay between the gradient of f and the proximal operator of g, or cannot be extended to more sophisticated choices of g. In contrast, [21] relies solely on gradient information of $\mathcal{E}$, and therefore cannot be applied when g is non-differentiable, as in the general composite objective $f + g$. Overall, these methods are designed for broader purposes rather than being tailored specifically for composite optimization.

Contributions. In this work, we introduce *ProxiCBO*, a novel CBO variant tailored for composite optimization, and provide theoretical guarantees and empirical validation. Our main contributions are:

- We propose ProxiCBO, a CBO variant that integrates proximal gradient flow to exploit the composite structure, thereby empirically improving convergence and particle efficiency.
- We present theoretical global convergence guarantees for the proposed continuous-time finite-particle dynamics.
- We empirically demonstrate that ProxiCBO outperforms proximal gradient descent, vanilla CBO, and their variants in convergence accuracy and particle efficiency on signal processing tasks. It is worth noting that we implemented our algorithm in JAX [22], and for a 200-dimensional problem, it requires only about 0.1 seconds with 100 particles and about 1.7 seconds with 10,000 particles over 5,000 iterations on a single RTX 4090 GPU.

This paper was completed while H. Z. and R. K. were interns at MERL.

2. METHODOLOGY

In this section, we introduce *ProxiCBO*, our proposed method for solving composite optimization problems of the form (1). The approach combines the interacting-particle framework of vanilla CBO dynamics [8, 10] with proximal gradient techniques to effectively handle the composite structure. We begin with the continuous-time formulation (2) and then provide its discretized counterpart in Algorithm 1 for practical implementation.

2.1. Continuous-Time Dynamics

At the initialization stage of the dynamics, we take N particles $V_0^1, \dots, V_0^N \in \mathbb{R}^d$, which are independently sampled from a common initial probability distribution ρ_0. We use V_t^i to denote the position of the i-th particle at time t, and $\widehat{\rho}_t^N(v) := \frac{1}{N}\sum_{i=1}^N \delta_{V_t^i}(v)$ to denote the empirical measure at time t, where δ_v is the Dirac measure concentrated at v. The continuous-time dynamics of the i-th particle follow the below stochastic differential equation (SDE),

$$\begin{aligned} dV_t^i = \quad & \underbrace{-\lambda_1 \left(V_t^i - v_\alpha(\widehat{\rho}_t^N)\right) dt}_{T_1} \\ & \underbrace{-\lambda_2 \left(\nabla f(V_t^i) + \nabla M_{\mu g}\left(V_t^i - \mu \nabla f(V_t^i)\right)\right) dt}_{T_2} \\ & \underbrace{+\sigma_1 D\left(V_t^i - v_\alpha(\widehat{\rho}_t^N)\right) dB_t^{i,1}}_{T_3} \\ & \underbrace{+\sigma_2 D\left(\nabla f(V_t^i) + \nabla M_{\mu g}\left(V_t^i - \mu \nabla f(V_t^i)\right)\right) dB_t^{i,2}}_{T_4}, \end{aligned} \tag{2}$$

where $v_\alpha(\widehat{\rho}_t^N) = \int v \cdot \frac{\omega_\alpha(v)}{\|\omega_\alpha\|_{L^1(\widehat{\rho}_t^N)}} d\widehat{\rho}_t^N$, with the weight $\omega_\alpha(v) = \exp\left(-\alpha \mathcal{E}(x)\right)$ and $\|\omega_\alpha\|_{L^1(\widehat{\rho}_t^N)} = \frac{1}{N}\sum_{i=1}^N \omega_\alpha(V_t^i)$, $M_{\mu g}$ is the Moreau envelope of g with parameter $\mu > 0$, defined as $M_{\mu g}(v) := \inf_{u \in \mathbb{R}^d}\left\{g(u) + \frac{1}{2\mu}\|v-u\|_2^2\right\}$, $B_t^{i,1}$, $B_t^{i,2}$ are two independent d-dimensional Wiener processes, and $D : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is a map we shall introduce later. In the following, we explain each term in (2).

Term T_1. This drift term is inherited from vanilla CBO methods [8]. It exploits the current information owned by particles, guiding all particles toward the *consensus point* $v_\alpha(\widehat{\rho}_t^N)$. Motivated by the Laplace principle [23], the consensus point $v_\alpha(\widehat{\rho}_t^N)$ smoothly approximates the particle with the lowest objective function value at the current iteration. Consequently, the particles are encouraged to gather around a location where the objective value is small and $\lambda_1 > 0$ controls the magnitude of this move.

Term T_2. Inspired by [21], this drift term exploits first-order information of the objective function, in the spirit of proximal gradient descent. It is based on the observation in [24] that, under proper assumptions, (1) can be solved using the proximal gradient flow,

$$\dot{v}(t) = -\mu\left(\nabla f(v) + \nabla M_{\mu g}\left(v - \mu \nabla f(v)\right)\right). \tag{3}$$

One can notice that the standard proximal gradient descent can be obtained via an explicit forward Euler discretization of (3) with step size one. For each particle, a drift in the direction of $\nabla f(v) + \nabla M_{\mu g}\left(v - \mu \nabla f(v)\right)$ provides extra first-order information of the objective landscape, thereby augmenting the dynamics. Notably, when all the particles are concentrated at the minimizer of (1), T_2 vanishes. Here, $\lambda_2 > 0$ controls the magnitude of this force.

Terms T_3 and T_4. The two diffusion terms facilitate exploration. The above two drift terms T_1 and T_2 are based on the current information obtained by particles. However, this incurs the danger of biased information. For example, if all initial particles are around a local minimizer, then particles will fail to explore the remaining landscape and concentrate at the local minimizer. The two diffusion terms help prevent this undesired situation. The matrix-valued function D determines the way of exploration. The isotropic exploration [8] can be done by choosing $D(v) = \|v\|_2 I_d$, where I_d is the $d \times d$ identity matrix, while choosing $D(v) = \mathrm{diag}(v)$ gives the anisotropic exploration [10]. $\sigma_1 > 0$ and $\sigma_2 > 0$ are parameters that determine of willingness of exploration.

2.2. Numerical Implementation

The practical algorithm is obtained by discretizing (2) using a standard Euler–Maruyama scheme. To accommodate the composite structure, we adopt an alternating update: a consensus step that discretizes terms T_1, T_3, and T_4, followed by a proximal step corresponding to T_2. This formulation arises by choosing λ_2 such that $\lambda_2 \Delta t = \mu$, so that the contribution of T_2 is carried out as a proximal update rather than appearing explicitly in the particle drift.

The gradient of the Moreau envelope is computed using the proximal operator of g via the following relation $\nabla M_{\mu g}(v) = \frac{1}{\mu}(v - \mathrm{prox}_{\mu g}(v))$ [25, Proposition 12.30], where the proximal operator is defined as $\mathrm{prox}_{\mu g}(v) = \mathrm{argmin}_{u\in\mathbb{R}^d}\left\{g(u) + \frac{1}{2\mu}\|v-u\|_2^2\right\}$. One can either record the historical best location (the one that achieves the best objective value throughout the run) or record the best location at the last iteration as the output. The pseudocode of *ProxiCBO* is summarized in Algorithm 1.

Algorithm 1 *ProxiCBO*

1: **Input:** $\{V^i\}_{i=1}^N \overset{\text{i.i.d.}}{\sim} \rho_0$
2: **while** not STOP **do**
3: **Compute consensus point:**

$$v_\alpha \leftarrow \frac{1}{\sum_{i=1}^N \exp(-\alpha\,\mathcal{E}(V^i))} \sum_{i=1}^N \exp\Big(-\alpha\,\mathcal{E}(V^i)\Big)\, V^i.$$

4: **Update particles:** for i=1,…,N

$$V^i \leftarrow V^i - \lambda_1\left(V^i - v_\alpha\right)\Delta t + \sigma_1 D\Big(V^i - v_\alpha\Big) z^{i,1}\sqrt{\Delta t}$$
$$+ \sigma_2 D\left(\frac{1}{\mu}\left[V^i - \mathrm{prox}_{\mu g}\left(V^i - \mu\nabla f(V^i)\right)\right]\right) z^{i,2}\sqrt{\Delta t},$$

where $\{z^{i,1}\}_{i=1}^N$ and $\{z^{i,2}\}_{i=1}^N$ are independent d-dimensional standard Gaussian vectors.
5: **Apply proximal map:** for i=1,…,N

$$V^i \leftarrow \mathrm{prox}_{\mu g}\Big(V^i - \mu\nabla f(V^i)\Big)$$

6: **end while**
7: **Output:** Historical or current best particle location

A key advantage of this alternating update is feasibility preservation. Specifically, if g includes the indicator function $\iota_{\mathcal{S}}$ of a constraint set $\mathcal{S}$, then $\mathrm{prox}_{\mu g}$ contains projection onto $\mathcal{S}$, ensuring that all particles remain in $\mathcal{S}$ after every iteration.

3. CONVERGENCE ANALYSIS

In this section, we state our convergence result (Theorem 1) for the continuous-time dynamics (2); the proof is deferred to an extended version of this paper. While the theoretical results for other CBO variants [11–16] focus on infinite-particle (mean-field) analysis and omit convergence rates in terms of the number of particles N, our result provides an explicit convergence bound that depends on the number of particles N.

3.1. Notation

We use $\|\cdot\|_2$ and $\|\cdot\|_\infty$ to denote the ℓ_2 and ℓ_∞ norms of vectors, respectively. For $x \in \mathbb{R}^d$ and $r > 0$, let $B_r^\infty(x)$ denote the closed ℓ_∞-ball centered at x with radius r. For probability measures μ and ν, $\mathcal{W}_2(\mu,\nu)$ denotes the Wasserstein-2 distance between them. δ_v denotes the Dirac measure supported at v.

3.2. Assumptions

We make the following assumptions.

Assumption 1. *(1) The objective function $\mathcal{E}$ with f differentiable and g convex is bounded from below with minimum being $\underline{\mathcal{E}}$ achieved by a unique global minimizer $v^* \in \mathbb{R}^d$.*
(2) There exist $s, L_\mathcal{E} > 0$ such that for all $u, v \in \mathbb{R}^d$, $|\mathcal{E}(u) - \mathcal{E}(v)| \le L_\mathcal{E}(1 + \|u\|_2 + \|v\|_2)^s \|u - v\|_2$.
(3) ∇f is Lipschitz with Lipschitz constant $L_f > 0$.
(4) There exist $l \ge 0$ and $c_l, C_l, c_u, C_u > 0$ such that for all $v \in \mathbb{R}^d$, $\mathcal{E}(v) - \underline{\mathcal{E}} \le c_u\|v\|_2^l + C_u$ and $\mathcal{E}(v) - \underline{\mathcal{E}} \ge c_l\|v\|_2^l - C_l$.
(5) There exist $R_0, \mathcal{E}_\infty > 0$ such that for all $v \in (B_{R_0}^\infty(v^*))^c$, $\mathcal{E}(v) - \mathcal{E}(v^*) > \mathcal{E}_\infty$.
(6) There exist $\eta, \nu > 0$ such that for all $v \in B_{R_0}^\infty(v^*)$, $\|v - v^*\|_\infty \le \frac{1}{\eta}(\mathcal{E}(v) - \mathcal{E}(v^*))^\nu$.

Assumptions (1)–(4) ensure that the SDE dynamics in (2) are well-posed and that their behavior can be approximated by the mean-field limit, with an error that can be quantified in terms of the number of particles N. Assumptions (5) and (6) guarantee that the unique minimizer v^* lies in a well-defined valley, making it identifiable, and that the mean-field dynamics are capable of converging to it.

3.3. Main Theorem

In this subsection, we present our main global convergence result.

Theorem 1. *Let Assumption 1 hold with $0 < l \le s+1$ or $l = s = 0$, and suppose the initial distribution ρ_0 satisfies $\rho_0(B_r^\infty(v^*)) > 0$ for all $r > 0$. Let $\widehat{\rho}_t^N$ be the empirical distribution generated by (2) with $D(\cdot) = diag(\cdot)$ and $\{V_0^i\}_{i=1}^N$ sampled from ρ_0 independently. Choose algorithm parameters such that $\sigma_1, \sigma_2 > 0$ and*

$$2\lambda_1 - \sigma_1^2 - \lambda_2\Big(2L_f + \tfrac{1}{\mu}\Big) - \sigma_2^2\Big(2L_f + \tfrac{1}{\mu}\Big)^2 > 0.$$

Then, for any error tolerance $\delta > 0$, if $\mathcal{W}_2^2(\rho_0, \delta_{v^*}) > \delta$, there exists a choice of inverse temperature $\alpha > 0$ such that

$$\min_{t\in[0,T^*]} \mathcal{W}_2^2(\widehat{\rho}_t^N, \delta_{v^*}) \;\le\; CN^{-1} + \frac{\delta}{2}, \tag{4}$$

where C is a constant independent of N and

$$T^* = \frac{2\log\big(4\mathcal{W}_2^2(\rho_0,\delta_{v^*})/\delta\big)}{2\lambda_1 - \sigma_1^2 - \lambda_2(2L_f + \frac{1}{\mu}) - \sigma_2^2(2L_f + \frac{1}{\mu})^2}.$$

Theorem 1 guarantees global convergence for possibly non-convex functions $\mathcal{E}$: for any prescribed accuracy $\delta > 0$, if the initial distribution assigns positive probability mass arbitrarily close to the global minimizer v^*, then with sufficiently many particles and appropriate parameter choices, the empirical distribution $\widehat{\rho}_t^N$ approaches δ_{v^*} within tolerance δ before a finite time T^*.

The requirement on the initialization is mild, as it only demands that ρ_0 is not identically zero in a neighborhood of v^*. The error bound in (4) has two components: the first term CN^{-1} arises from approximating the finite-particle system with its infinite-particle mean-field limit, while the second term $\frac{\delta}{2}$ reflects the convergence of the mean-field dynamics to the global minimizer.

4. NUMERICAL EXAMPLES

We now compare the empirical performance of our algorithm with those of proximal gradient (PG) [6], accelerated proximal gradient (APG) [7] and existing CBO-type algorithms [10, 12] for two signal reconstruction examples. All algorithms use the same initial particles and the final result for each algorithm is chosen to be the particle that achieves the lowest objective function value. Note that PG and APG can be seen as particle systems without interactions. Hyper-parameters of the algorithms are tuned empirically and all algorithms use the same stepsize. The resampling trick [10] is applied to the vanilla CBO to further improve its performance, as we have noticed that its performance remains limited on numerical examples considered in this paper even after careful parameter tuning. Our first metric for quantifying performance is the success rate of achieving global minimum, which is important for an optimization algorithm. For each trial, we estimate the global minimum by running PG initialized at the ground truth signal. Let v^* be the estimated global minimizer and let $\widehat{v}$ be the reconstructed signal, then a trial is successful if

$$\frac{\mathcal{E}(\widehat{v}) - \mathcal{E}(v^*)}{\mathcal{E}(v^*)} < 10^{-4}. \tag{5}$$

Our second metric is related to mean squared error with respect to the ground truth signal, which is important for signal processing applications. The definitions will be provided below in each example.

4.1. Example 1: One-Bit Signal Quantization

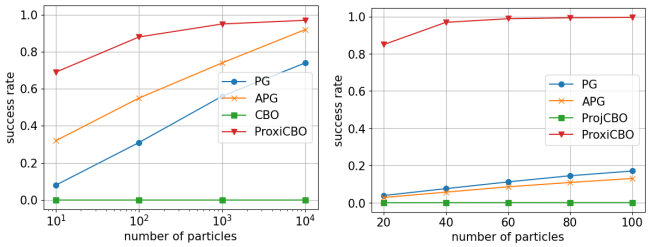
Our first example is the non-monotonic quantization problem first proposed and analyzed in [26]. Consider the measurement model

$$y = \operatorname{sign}\left(\sin\left(\frac{\pi}{\Delta}\left(Ax_0 + u\right)\right)\right), \tag{6}$$

where the unknown signal $x_0 \in \mathbb{R}^d$ is sparse, the measurement matrix $A \in \mathbb{R}^{m\times d}$ has i.i.d. 0-mean Gaussian entries with variance $1/d$, Δ is the quantization bin-size, $\sin(\cdot)$ and $\operatorname{sign}(\cdot)$ are applied element-wise to the arguments, and the vector $u \in \mathbb{R}^m$ is a known i.i.d. uniform dither. Consistent reconstruction is a combinatorial not differentiable problem that may be infeasible in the presence of measurement errors or noise. We relax it, replacing $\operatorname{sign}(\cdot)$ consistency with an ℓ_2 penalty. Since x_0 is sparse, it is estimated by solving

$$\min_{x\in\mathbb{R}^d}\left\{\mathcal{E}(x) := \frac{1}{2}\left\|y - \sin\left(\frac{\pi}{\Delta}\left(Ax + u\right)\right)\right\|_2^2 + \lambda\|x\|_1\right\}.$$

Even with this relaxation, the problem has a very difficult optimization landscape, providing a good test case for our approach. To date, there are no good solutions known, unless a good estimate of the signal already exists (e.g., solving a hierarchy of problems [27]).



(a) One-bit signal quantization. (b) Single-photon lidar.

Fig. 1: Success rate by objective function value defined in (5).

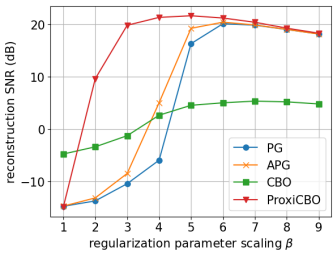


Fig. 2: Reconstruction signal to noise ratio (SNR) for one-bit signal quantization reconstruction with different regularization parameters λ. Specifically, $\lambda = \beta \times 0.05\|y\|_2^2$. The number of particles is 100.

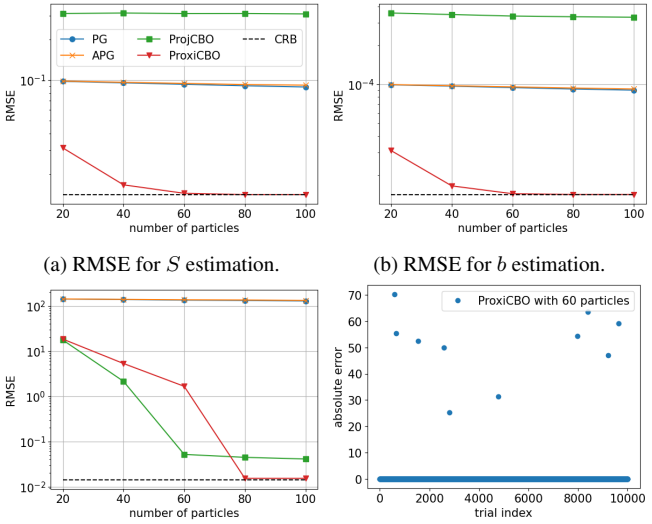
In our simulations, x_0 has dimension $d = 200$ and sparsity $s = 10$. We use $m = 800$ measurements and $\Delta = \pi/14$. The dither u has i.i.d. entries, uniform in $[-\Delta/2, \Delta/2]$. All initial particles have i.i.d. standard normal entries.

Fig. 1a compares the success rate (5) for PG, APG, CBO [10], and the proposed ProxiCBO and Fig. 2 compares the reconstruction signal to noise ratio (SNR), which is defined as $10\log_{10}(\|x_0\|_2^2/\|\widehat{x}-x_0\|_2^2)$. The presented results are computed from 500 trials, and in each trial, A, x_0, u are independently sampled according to their distributions. In Fig. 1a, we fix the regularization parameter $\lambda = 0.025\|y\|_2^2$ and vary the number of particles. We can see that ProxiCBO has achieved a success rate close to 90% with only 100 particles and that ProxiCBO with 1000 particles outperforms other methods with 10,000 particles, showing ProxiCBO's superior particle-efficiency. In Fig. 2, we fix the number of particles to be 100 and change the regularization parameter λ. Specifically, $\lambda = \beta \times 0.05\|y\|_2^2$, where $\beta = 1, 2, \ldots, 9$ is the regularization parameter scaling factor. A larger λ value can improve the optimization landscape and make the optimization problem easier to solve. However, it may also lead to a larger measurement mismatch, driving the optimizer of the optimization problem away from the ground truth signal. Fig. 2 shows that the performance of ProxiCBO is more robust to the choice of λ than other methods.

4.2. Example 2: Single-Photon Lidar

In a typical single-photon lidar setup, a target is illuminated by a pulsed laser, the reflected light is detected by a single-photon detector, and the detection times relative to the laser pulse times are recorded by a timing system. Those detection times contain information about the reflectivity S and distance z of the target. Suppose the pulse shape of the laser is defined by $h(t)$, which is normalized such that $\int_{-\infty}^{\infty} h(t)\,dt = 1$. Let $\{t_k\}_{k=1}^K$ be the timestamps when the laser pulses are sent out, then the photon detection process is a time-inhomogeneous Poisson process [28] with intensity function

$$\lambda(t) = S\sum_{k=1}^{K} h\left(t - \tau - t_k\right) + b,$$



(a) RMSE for S estimation. (b) RMSE for b estimation.

(c) RMSE for τ estimation. (d) Error for τ in each trial.

Fig. 3: Estimation of S (Fig. 3a), b (Fig. 3b) and τ (Fig. 3c) for single-photon lidar. The relatively high root mean squared error (RMSE) of ProxiCBO with 60 particles for τ estimation in Fig. 3c is due to a few (10 out of 10,000) outliers as we can see in Fig. 3d, which also explains the high success rate in achieving global minimum as reported in Fig. 1b

where b is the background intensity, which is assumed to be constant, and $\tau = 2z/c$ with c being the speed of light is the photon roundtrip time. The log-likelihood function of parameters (S, b, τ) given a set of detection times $\mathcal{T}$ is defined as [28]

$$\mathcal{L} = -SK - bt_a + \sum_{t\in\mathcal{T}} \log\left(S\sum_{k=1}^{K} h\left(t - \tau - t_k\right) + b\right),$$

where t_a is the total acquisition time. Hence, given detection times $\mathcal{T}$, we can estimate S, b and τ (thus z) by solving [1, 2]

$$\min_{S,b,\tau} \left\{\mathcal{E}(S, b, \tau) := -\mathcal{L}(\mathcal{T}; S, b, \tau) + \iota_{\mathcal{C}}(S, b, \tau)\right\}, \qquad (7)$$

where $\mathcal{C}$ is the feasible set for (S, b, τ) and $\iota_{\mathcal{C}}$ is the indicator function for $\mathcal{C}$. In our simulations, $K = 500$, $S_0 = 0.1$, $b_0 = 10^{-4}$, $\tau_0 = 234$ ns, $t_a = 5 \times 10^5$ ns, thus the signal to background ratio (SBR) is $(K \cdot S)/(b \cdot t_a) = 1$. The pulse shape $h(t)$ is the probability density function of the Gaussian distribution with mean zero and standard deviation 0.1. The feasible set $\mathcal{C} = [10^{-8}, 10] \times [10^{-8}, 10] \times [0, \infty)$. The initial particles are i.i.d. uniform in $[0, 1] \times [0, 1] \times [0, 500]$.

Fig. 1b compares the success rate (5) for PG, APG, projected CBO [12], and the proposed ProxiCBO. Fig. 3 compares the root mean squared error (RMSE) for each of the parameters S, b, and τ, where the RMSE for S is defined as $\sqrt{\frac{1}{M}\sum_{i=1}^{M}(\widehat{S}_i - S_0)^2}$ with M being the number of trials and the definition of RMSE for other parameters is similar. We also include the Cramér-Rao lower bound (CRB) in the plots showing the best achievable RMSE for any unbiased estimators [29]. The presented results are computed from $M = 10,000$ independent trials. The results show that ProxiCBO has better particle-efficiency than all comparison methods. Moreover, with sufficient particles, ProxiCBO can accurately solve the maximum likelihood problem (7) and achieve the CRB.

5. REFERENCES

[1] Dongeek Shin, Ahmed Kirmani, Vivek K Goyal, and Jeffrey H Shapiro, "Photon-efficient computational 3-d and reflectivity imaging with single-photon detectors," *IEEE Transactions on Computational Imaging*, vol. 1, no. 2, pp. 112–125, 2015.

[2] Ruangrawee Kitichotkul, Joshua Rapp, and Vivek K Goyal, "The role of detection times in reflectivity estimation with single-photon lidar," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 30, no. 1: Single-Photon Technologies and Applications, pp. 1–14, 2023.

[3] Emmanuel Soubies, Thanh-An Pham, and Michael Unser, "Efficient inversion of multiple-scattering model for optical diffraction tomography," *Optics express*, vol. 25, no. 18, pp. 21786–21800, 2017.

[4] Yanting Ma, Hassan Mansour, Dehong Liu, Petros T Boufounos, and Ulugbek S Kamilov, "Accelerated image reconstruction for nonlinear diffractive imaging," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 6473–6477.

[5] Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, Thomas Pock, et al., "An introduction to total variation for image analysis," *Theoretical foundations and numerical methods for sparse recovery*, vol. 9, no. 263-340, pp. 227, 2010.

[6] Neal Parikh and Stephen Boyd, "Proximal algorithms," *Foundations and Trends® in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[7] Amir Beck and Marc Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[8] René Pinnau, Claudia Totzeck, Oliver Tse, and Stephan Martin, "A consensus-based model for global optimization and its mean-field limit," *Mathematical Models and Methods in Applied Sciences*, vol. 27, no. 01, pp. 183–204, 2017.

[9] José A. Carrillo, Young-Pil Choi, Claudia Totzeck, and Oliver Tse, "An analytical framework for consensus-based global optimization method," *Mathematical Models and Methods in Applied Sciences*, vol. 28, no. 06, pp. 1037–1066, 2018.

[10] Carrillo, José A., Jin, Shi, Li, Lei, and Zhu, Yuhua, "A consensus-based global optimization method for high dimensional machine learning problems," *ESAIM: COCV*, vol. 27, pp. S5, 2021.

[11] Giacomo Borghi, Michael Herty, and Lorenzo Pareschi, "Constrained consensus-based optimization," *SIAM Journal on Optimization*, vol. 33, no. 1, pp. 211–236, 2023.

[12] Hyeong-Ohk Bae, Seung-Yeal Ha, Myeongju Kang, Hyuncheul Lim, Chanho Min, and Jane Yoo, "A constrained consensus based optimization algorithm and its application to finance," *Applied Mathematics and Computation*, vol. 416, pp. 126726, 2022.

[13] José A Carrillo, Shi Jin, Haoyu Zhang, and Yuhua Zhu, "An interacting particle consensus method for constrained global optimization," *arXiv preprint arXiv:2405.00891*, 2024.

[14] Massimo Fornasier, Lorenzo Pareschi, Hui Huang, and Philippe Sünnen, "Consensus-based optimization on the sphere: Convergence to global minimizers and machine learning," *Journal of Machine Learning Research*, vol. 22, no. 237, pp. 1–55, 2021.

[15] José A. Carrillo, Nicolás García Trillos, Sixu Li, and Yuhua Zhu, "Fedcbo: Reaching group consensus in clustered federated learning through consensus-based optimization," *Journal of Machine Learning Research*, vol. 25, no. 214, pp. 1–51, 2024.

[16] José A Carrillo, Franca Hoffmann, Andrew M Stuart, and Urbain Vaes, "Consensus-based sampling," *Studies in Applied Mathematics*, vol. 148, no. 3, pp. 1069–1140, 2022.

[17] Leon Bungert, Franca Hoffmann, Doh Yeon Kim, and Tim Roith, "Mirrorcbo: A consensus-based optimization method in the spirit of mirror descent," *arXiv preprint arXiv:2501.12189*, 2025.

[18] Massimo Fornasier, Timo Klock, and Konstantin Riedl, "Consensus-based optimization methods converge globally," *SIAM Journal on Optimization*, vol. 34, no. 3, pp. 2973–3004, 2024.

[19] Nicolai Jurek Gerber, Franca Hoffmann, and Urbain Vaes, "Mean-field limits for consensus-based optimization and sampling," *arXiv preprint arXiv:2312.07373*, 2023.

[20] Nicolai Gerber, Franca Hoffmann, Dohyeon Kim, and Urbain Vaes, "Uniform-in-time propagation of chaos for consensus-based optimization," *arXiv preprint arXiv:2505.08669*, 2025.

[21] Konstantin Riedl, "Leveraging memory effects and gradient information in consensus-based optimisation: On global convergence in mean-field law," *European Journal of Applied Mathematics*, vol. 35, no. 4, pp. 483–514, 2024.

[22] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang, "JAX: composable transformations of Python+NumPy programs," 2018.

[23] P. D. Miller, *Applied asymptotic analysis*, vol. 75, American Mathematical Soc., 2006.

[24] Sepideh Hassan-Moghaddam and Mihailo R. Jovanović, "Proximal gradient flow and douglas–rachford splitting dynamics: Global exponential stability via integral quadratic constraints," *Automatica*, vol. 123, pp. 109311, 2021.

[25] Heinz H Bauschke and Patrick L Combettes, "Convex analysis and monotone operator theory in hilbert spaces. 2nd edition," in *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2017.

[26] Petros T Boufounos, "Universal rate-efficient scalar quantization," *IEEE transactions on information theory*, vol. 58, no. 3, pp. 1861–1872, 2011.

[27] P. T. Boufounos, "Hierarchical distributed scalar quantization," in *Proc. Int. Conf. Sampling Theory and Applications (SampTA)*, Singapore, May 2-6 2011.

[28] Donald L Snyder and Michael I Miller, *Random point processes in time and space*, Springer Science & Business Media, 2012.

[29] Steven M Kay, *Fundamentals of statistical signal processing: estimation theory*, Prentice-Hall, Inc., 1993.