Accuracy Certificates for Convex Optimization at Accelerated Rates via Primal-Dual Averaging

PI: Jiaming Liang, University of Rochester

Funded Project: Modern Developments of Optimization Methods via Inexact Proximal Point Framework

Award Number: FA9550-25-1-0182

Funding Period: August 1, 2025-July 31, 2028

September 9, 2026

AFOSR Annual Review, Arlington, VA

Motivation

Primal-Dual Correspondences with Averaging Dynamics

- Averaging is a fundamental operation in convex optimization (primal and dual)
- Correspondences provide alternative convergence/certificate proofs (Bach, 2015), (Tibshirani, 2017), (Lu and Freund, 2021), novel algorithm variants (Tibshirani, 2017), (Liang, 2025)

Q1: How do primal-dual correspondences change as we increase the number of average sequences?

$\star$ MDA$\leftrightarrow$GCG, AggGCG$\leftrightarrow$AggGCG, TAA$\leftrightarrow$GEM

Computable Certificates from Primal-Dual Averaging

- (Nemirovski, Onn, and Rothblum, 2010) popularized "Accuracy certificates" to certify ε-suboptimality.
- Applied to Subgradient Ellipsoid (Rodomanov and Nesterov, 2023), Cutting Planes (Gladin, Gasnikov, and Dvurechensk, 2025), PEP proofs (Grimmer and Wang, 2026)
- Accelerated inexact proximal point frameworks (Monteiro and Svaiter, 2013), (Burns and Liang, 2026)

Q2: Can we use "inner" subproblem certificates to construct "outer" certificates?

$\star$ PD-AIPP

Problem of Interest

Problem: Convex Smooth Composite Optimization

$$\phi_* = \min_{x \in \mathbb{R}^n} \{\phi(x) := f(x) + h(x)\}$$

- f is closed, proper, convex and L_f-smooth on $\mathbb{R}^n$, i.e.,

  $$\|\nabla f(x) - \nabla f(y)\| \leq L_f \|x - y\|$$

  for all $x, y \in \mathbb{R}^n$.
- h is closed, proper, convex and sufficiently "simple" (to be elaborated on).

Goal: ε-primal-dual gap: find $(x, s) \in \operatorname{dom} h \times \mathbb{R}^n$ satisfying

$$\phi(x) + \psi(s) \leq \varepsilon.$$

where $\psi(s) = f^*(s) + h^*(-s)$ is the negative Fenchel-Rockafellar dual.

Accuracy Certificates

$\phi(x) + \psi(s)$ may not be easily computable

Proposition

Let $\gamma : \mathbb{R}^n \to (-\infty, \infty]$ *be closed, convex, and differentiable with* $\gamma(\cdot) \leq f(\cdot)$. *Define* $x_\gamma = \operatorname{argmin}_{x \in \mathbb{R}^n} \{\gamma(x) + h(x)\}$ *and* $s_\gamma = \nabla \gamma(x_\gamma)$. *Then*

$$\psi(s_\gamma) \leq - \min_{x \in \mathbb{R}^n} \{\gamma(x) + h(x)\}$$

Definition (Accuracy Certificate)

We call the pair $(y, \gamma) \in \mathbb{R}^n \times \left(\overline{\text{Conv}}(\mathbb{R}^n) \cap C^1(\mathbb{R}^n)\right)$ an ε-accuracy certificate for a CSCO problem with objective ϕ if

$$\phi(y) - \min_{x \in \mathbb{R}^n} \{\gamma(x) + h(x)\} \leq \varepsilon. \tag{1}$$

An ε-certificate implies an ε-primal-dual gap $\phi(y) + \psi(s_\gamma) \leq \varepsilon$.

Regularized Notation

Assume $w : \mathbb{R}^n \to [0, \infty]$ is differentiable, 1-strongly convex on $\operatorname{dom} h$, and the generalized linear minimization oracle

$$x^+ = \operatorname*{argmin}_{x \in \mathbb{R}^n} \{\langle s, x \rangle + h^\alpha(x)\}$$

is solvable for any $s \in \mathbb{R}^n$ and $\alpha > 0$, where

$$h^\alpha(\cdot) := h(\cdot) + \alpha w(\cdot).$$

Define $\phi^\alpha(\cdot) = f(\cdot) + h^\alpha(\cdot)$ with dual $\psi^\alpha(\cdot) = f^*(\cdot) + (h^\alpha)^*(-\cdot)$.

Lemma

Suppose $M := \max_{x \in \operatorname{dom} h} w(x) < 0$. Then, if $\alpha \leq \varepsilon/(2M)$ and (y, γ) is an $(\varepsilon/2)$-certificate for $\phi^\alpha(\cdot)$, then (y, γ) is an ε-certificate for $\phi(\cdot)$.

Outline

Outline

One Average: MDA and GCG

Algorithm Modified Dual Averaging

Initialize: given $y_0 \in \operatorname{dom} h$, $\alpha > 0$, set $\eta = \alpha/(L_f + \alpha)$, $s_1 = \nabla f(y_0)$, and compute

$$x_1 = \operatorname*{argmin}_{x \in \mathbb{R}^n} \{\langle s_1, x\rangle + h^\alpha(x)\}.$$

for $k \geq 1$ **do**

Compute

$$x_{k+1} = \operatorname*{argmin}_{x \in \mathbb{R}^n} \{\langle s_{k+1}, x\rangle + h^\alpha(x)\},$$

where $s_{k+1} = (1-\eta)s_k + \eta \nabla f(x_k)$.

end for

Algorithm Generalized Conditional Gradient

Initialize: $z_1 \in \operatorname{dom} f^*$, set $\eta = \alpha/(L_f + \alpha)$.

for $k \geq 1$ **do**

Compute

$$\bar{z}_k = \operatorname*{argmin}_{v \in \mathbb{R}^n} \{\langle -\nabla (h^\alpha)^*(-z_k), v\rangle + f^*(v)\},$$

$$z_{k+1} = \eta \bar{z}_k + (1-\eta) z_k.$$

end for

$\tilde{\mathcal{O}}(L_f/\alpha)$-certificate complexity for ϕ^α, which implies $\tilde{\mathcal{O}}(L_f M/\varepsilon)$-certificate complexity for ϕ.

MDA and GCG Correspondence

Also observed in (Liang, 2025) in the context of proximal bundle methods.

Proposition

Fix $y_0 \in \operatorname{dom} h$ and set $z_1 = \nabla f(y_0)$. Then, on every iteration k of Algorithm 1 and Algorithm 2, we have the following correspondence

$$\nabla f(x_k) = \bar{z}_k, \quad x_k = \nabla (h^\alpha)^*(-z_k), \quad s_k = z_k. \tag{2}$$

$$y_0 \longrightarrow (s_1 = z_1) \longrightarrow \left(x_k \overset{\textbf{(a)}}{=} \nabla (h^\alpha)^*(-z_k)\right) \longrightarrow \left(\nabla f(x_k) \overset{\textbf{(b)}}{=} \bar{z}_k\right) \longrightarrow \left(s_{k+1} \overset{\textbf{(c)}}{=} z_{k+1}\right).$$

Two Averages: Aggregated GCG (Zhao and Zhu, 2023)

Algorithm AggGCG (Primal)

Initialize: $y_0 \in \operatorname{dom} h$, $s_0 \in \mathbb{R}^n$, $\eta = \alpha/(L_f + \alpha)$

for $k \geq 0$ **do**

1. Compute
$$x_{k+1} = \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ \langle s_k, x \rangle + h^\alpha(x) \right\}.$$

2. Compute
$$y_{k+1} = (1 - \eta) y_k + \eta x_{k+1},$$
$$s_{k+1} = (1 - \eta) s_k + \eta \nabla f(y_k).$$

end for

Algorithm AggGCG (Dual)

Initialize: $z_0 \in \operatorname{dom} f^*$, $v_0 \in \mathbb{R}^n$, $\eta = \alpha/(L_f + \alpha)$

for $k \geq 0$ **do**

1. Compute
$$\bar{z}_{k+1} = \operatorname*{argmin}_{z \in \mathbb{R}^n} \left\{ -\langle v_k, z \rangle + f^*(z) \right\}. \tag{3}$$

2. Compute
$$z_{k+1} = (1 - \eta) z_k + \eta \bar{z}_{k+1},$$
$$v_{k+1} = (1 - \eta) v_k + \eta \nabla (h^\alpha)^*(-z_k).$$

end for

$\tilde{\mathcal{O}}(L_f/\alpha)$ complexity to obtain $\phi^\alpha(y_k) + \psi^\alpha(s_k) \leq \varepsilon$, but *without* a natural certificate.

Self-dual algorithm

Outline

Three Average Acceleration

One-average (MDA$\leftrightarrow$GCG): $\tilde{\mathcal{O}}(L_f/\alpha)$ certificate guarantees
Two-average (AggGCG): Symmetric updates but loses certificate guarantees
Three-average (TAA): Regain certificate guarantees with a new primal average

Algorithm Three Average Acceleration

Initialize: given $y_{-1} \in \mathbb{R}^n$, $L_f > 0$, $\alpha > 0$, set $s_0 = \nabla f(y_{-1})$ and compute

$$y_0 = x_0 = \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ \langle s_0, x \rangle + h^\alpha(x) \right\}, \quad \beta = \frac{2\alpha}{\alpha + \sqrt{\alpha^2 + 4L_f \alpha}}. \tag{4}$$

for $k \geq 0$ **do**

1. Compute

$$\tilde{x}_{k+1} = (1-\beta) y_k + \beta x_k, \quad s_{k+1} = (1-\beta) s_k + \beta \nabla f(\tilde{x}_{k+1}),$$

$$x_{k+1} = \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ \langle s_{k+1}, x \rangle + h^\alpha(x) \right\},$$

$$y_{k+1} = (1-\beta) y_k + \beta x_{k+1}.$$

end for

Accelerated Guarantees

Theorem

For all iterations $k \geq 0$ *of TAA,*

$$\phi^\alpha(y_k) - \min_{x \in \mathbb{R}^n} \{\Gamma_k(x) + h^\alpha(\cdot)\} \leq \frac{\Delta}{\left(1 + \frac{\sqrt{\alpha}}{2\sqrt{L_f}}\right)^{2k}}, \tag{5}$$

where $\Delta = \phi^\alpha(y_0) - \min_{x \in \mathbb{R}^n} \{\Gamma_0(x) + h^\alpha(x)\}$ *and* $\Gamma_k(\cdot)$ *is the average convex model defined as*

$$\Gamma_{k+1}(\cdot) = (1 - \beta)\Gamma_k(\cdot) + \beta \ell_f(\cdot; \tilde{x}_{k+1}), \tag{6}$$

with $\Gamma_0(\cdot) = \ell_f(\cdot; y_{-1})$.

$\tilde{\mathcal{O}}(\sqrt{L_f/\alpha})$ for regularized problem, implies $\tilde{\mathcal{O}}(\sqrt{ML_f}\varepsilon^{-1/2})$ complexity for original problem.
Optimal in ε, explicit parameter dependence on M

Gradient Extrapolation Method (Lan and Zhou, 2018) in the Dual

Suppose ν^* is a 1-s.c. function and L_f-smooth relative to f^* (e.g., $\nu^* = L_f f^*$ or $\nu^* = \|\cdot\|_*^2/2$)

Algorithm GEM

Initialize: given $\alpha > 0$ and $g_0 \in \operatorname{dom} f^*$, set $z_0 = g_0$, $v_{-1} = v_0 = \nabla(h^\alpha)^*(-g_0)$, $\tau_0 = \alpha/L_f$, $A_0 = 1$, and $a_{-1} = 0$

for $k \geq 0$ **do**

1. Compute

$$\tau_{k+1} = \tau_k + \frac{\alpha a_k}{L_f}, \quad a_k = \frac{\tau_k + \sqrt{\tau_k^2 + 4\tau_k A_k}}{2}, \tag{7}$$

$$A_{k+1} = A_k + a_k, \quad \hat{v}_k = v_k + \frac{a_{k-1}}{a_k}(v_k - v_{k-1}). \tag{8}$$

2. Compute

$$g_{k+1} = \operatorname*{argmin}_{g \in \mathbb{R}^n} \left\{ a_k[\langle -\hat{v}_k, g\rangle + f^*(g)] + \frac{\tau_k}{\alpha} \mathrm{d}_{\nu^*}(g \| g_k) \right\}, \tag{9}$$

$$z_{k+1} = \frac{A_k}{A_{k+1}} z_k + \frac{a_k}{A_{k+1}} g_{k+1}, \quad v_{k+1} = \nabla(h^\alpha)^*(-z_{k+1}). \tag{10}$$

end for

Primal-Dual Correspondence

Proposition

Fix $y_{-1} \in \operatorname{dom} h$ and suppose that we choose $\nu^* = L_f f^*$ in GEM. Then, setting $s_0 = z_0 = g_0 = \nabla f(y_{-1})$ and $\tilde{x}_0 = y_{-1}$, GEM with objective $\psi^\alpha(\cdot)$ is equivalent to TAA with objective $\phi^\alpha(\cdot)$ under the following correspondence [a]

$$\nabla f(\tilde{x}_k) = g_k, \quad s_k = z_k, \quad x_k = v_k. \tag{11}$$

[a] f^* may not necessarily differentiable.

$$\begin{tikzcd}
 & & & \overset{\textbf{(a)}}{\begin{pmatrix} \tilde{x}_{k+1} \\ \hat{v}_k \end{pmatrix}} \arrow[r] & \overset{\textbf{(b)}}{(\nabla f(\tilde{x}_{k+1}) = g_{k+1})} \arrow[d] \\
y_{-1} \arrow[r] & (s_0/z_0 = \nabla f(y_{-1})) \arrow[r] & (y_0/x_0/v_0 = \nabla(h^\alpha)^*(-z_0)) \arrow[ur] & y_{k+1}\ \textbf{(e)} \arrow[u] & (s_{k+1} = z_{k+1})\ \textbf{(c)} \arrow[dl] \\
 & & & \overset{\textbf{(d)}}{(x_{k+1}/v_{k+1} = \nabla(h^\alpha)^*(-z_{k+1}))} \arrow[u] &
\end{tikzcd}$$

Outline

TAA for Proximal Subproblems

Consider the specific choice $w(\cdot) = \frac{1}{2}\| \cdot - y_{-1}\|^2$ and

$$\phi^\alpha(\cdot) := f(\cdot) + h(\cdot) + \frac{\alpha}{2}\| \cdot - y_{-1}\|^2. \tag{12}$$

Then, TAA offers the following relative error guarantee

Proposition

Fix $\sigma > 0$ and choose α such that $4L_f(\alpha + L_f)^2 \geq \sigma\alpha^3$. Suppose the number of inner iterations k taken by TAA satisfies

$$k \geq (1 + \sqrt{\kappa}) \log\left(\frac{4\kappa(1+\kappa)^2}{\sigma}\right), \tag{13}$$

where $\kappa := L_f/\alpha$. Then, we have the following guarantee

$$\phi^\alpha(y_k) - \min_{x \in \mathbb{R}^n}\{\Gamma_k(x) + h^\alpha(x)\} \leq \frac{\alpha\sigma}{2}\|y_k - y_{-1}\|^2. \tag{14}$$

PD-AIPP Framework

Algorithm Primal-Dual Accelerated Inexact Proximal Point

Initialize: $y_0 \in \operatorname{dom} h$, $\lambda > 0$, $\sigma > 0$, set $x_0 = y_0$, $s_0 = 0^n$, and $A_0 = 0$.

for $k \geq 0$ **do**

Step 1) Compute

$$a_{k+1} = \frac{\lambda + \sqrt{\lambda^2 + 4\lambda A_k}}{2}, \quad A_{k+1} = A_k + a_{k+1}, \tag{15}$$

and define the linear map $T_k(x) = A_k y_k / A_{k+1} + a_{k+1} x / A_{k+1}$.

Step 2) Define $F(\cdot) = (A_{k+1} f(T_k(\cdot)) - A_k f(y_k))/a_{k+1}$, and find $(\tilde{y}_{k+1}, x_{k+1}, \Gamma^F_{k+1})$ such that an affine function $\Gamma^F_{k+1}(\cdot) \leq F(\cdot)$ and the IPP conditions are satisfied.

Step 3) Update

$$y_{k+1} = \frac{A_k}{A_{k+1}} y_k + \frac{a_{k+1}}{A_{k+1}} \tilde{y}_{k+1}, \quad s_{k+1} = \frac{A_k}{A_{k+1}} s_k + \frac{a_{k+1}}{A_{k+1}} \nabla \Gamma^F_{k+1}. \tag{16}$$

end for

IPP Conditions

IPP Conditions: $(\tilde{y}_{k+1}, x_{k+1}, \Gamma^F_{k+1})$ such that

$$\Gamma^F_{k+1}(\cdot) \le F(\cdot), \quad x_{k+1} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \Gamma^F_{k+1}(x) + h(x) + \frac{1}{2a_{k+1}} \|x - x_k\|^2 \right\},$$

$$(F + h)(\tilde{y}_{k+1}) + \frac{1}{2a_{k+1}} \|\tilde{y}_{k+1} - x_k\|^2 - (\Gamma^F_{k+1} + h)(x_{k+1}) - \frac{1}{2a_{k+1}} \|x_{k+1} - x_k\|^2$$
$$\le \frac{\sigma}{2a_{k+1}} \|\tilde{y}_{k+1} - x_k\|^2 + \delta.$$

Relative Error Term: $\sigma \|\tilde{y}_{k+1} - x_k\|^2 / (2a_{k+1})$ satisfied by TAA for solving the proximal subproblem
Absolute Error Term: δ

Comparison with AT Rule

Monteiro-Svaiter acceleration generalizes FISTA. Similarly, PD-AIPP generalizes AT-acceleration.

AT Rule (Auslender and Teboulle, 2006)

$$\tilde{x}_{k+1} = T_k(x_k), \quad x_{k+1} = \operatorname{argmin} \left\{ \ell_f(x; \tilde{x}_{k+1}) + h(x) + \frac{1}{2a_{k+1}} \|x - x_k\|^2 \right\}, \quad y_{k+1} = T_k(x_{k+1}) \tag{17}$$

Observations: $\nabla f(T_k(x)) = \nabla F(x)$, so $F(\cdot)$ is $a_{k+1} L_f / A_{k+1}$-smooth.

Proposition

For $\sigma \in (0,1)$, the AT rule satisfies the IPP conditions with

$$\Gamma^F_{k+1}(\cdot) = \ell_F(\cdot; x_k), \quad \sigma = \sigma, \quad \lambda = \frac{\sigma}{L_f}, \quad \tilde{y}_{k+1} = x_{k+1}, \quad x_k = x_k, \quad \delta = 0. \tag{18}$$

The proximal subproblem is solved by one proximal mapping of h.

PD-AIPP+TAA Guarantees

Achieves optimal complexity $\tilde{\mathcal{O}}(\sqrt{L_f \varepsilon^{-1}})$ for the primal gap, however an accuracy certificate requires a localization set Q.

Theorem

Let $Q \subset \mathbb{R}^n$ be a compact and convex set such that $\{x_k\}_{k\geq 0} \cup \{\tilde{y}_{k+1}\}_{k\geq 0} \cup \{y_k\}_{k\geq 0} \subset Q$ for all $k \geq 1$. For every $k \geq 1$, we have the following inequality

$$f(y_k) + \hat{h}(y_k) - \min_{u\in\mathbb{R}^n} \left\{ \bar{\Gamma}_k(x) + \hat{h}(x) \right\} \leq \frac{R_Q^2}{2A_k} + \delta,$$

where $R_Q = \max_{x\in Q} \|x - x_0\|$ and $\hat{h}(\cdot) = h(\cdot) + I_Q(\cdot)$, and $\bar{\Gamma}_k(\cdot)$ is

$$\bar{\Gamma}_k(\cdot) = A_k^{-1} \sum_{i=0}^{k-1} a_{i+1} \Gamma_{i+1}^F(\cdot). \tag{19}$$

Localization Set: Intuition

Consider $h(\cdot) = \|\cdot\|_1$, with $h^*(\cdot) = I_{B_\infty}(\cdot)$
$s_k \notin B_\infty$ implies $\psi(s_k) = \infty$, so $\phi(x) + \psi(s_k) = \infty$ for any $x \in \operatorname{dom} h$. Localization gives

$$\hat{h}^*(s_k) = \max_{x \in \mathbb{R}^n} \{\langle s_k, x\rangle - \|x\|_1 - I_Q(x)\}, \tag{20}$$

which is finite for any $s_k \in \mathbb{R}^n$.
Choice 1: $\operatorname{dom} h$ is bounded, so $Q = \operatorname{dom} h$ (so $\hat{h} = h$).
Choice 2: Ball centered at x_0 with radius dependent on $\|x_0 - x_*\|$.

Lemma

For all $k \geq 0$, we have

$$\max\{\|x_k - x_*\|, \|\tilde{y}_{k+1} - x_*\|, \|y_{k+1} - x_*\|\} \leq \frac{2}{\sqrt{1-\sigma}}\|x_0 - x_*\|$$

for $x_* = \operatorname{argmin}\{\|x_0 - x\| : \phi(x) = \phi_*\}$.

Conclusions and Future Work

Contributions

- **TAA Method**
  - Accelerated method via primal/dual averaging
  - Dual to GEM
  - Produces certificates for regularized problems with optimal complexity
- **PD-AIPP Framework**
  - Generic framework for acceleration generalizing AT rule
  - Produces a (localized) certificate from inner proximal certificates
  - Paired with TAA inner $\Rightarrow$ achieves optimal complexity

Future Directions

- Parameter-free
- Localization-free
- Application to high-order methods

References

- Auslender and Teboulle. Interior Gradient and Proximal Methods for Convex and Conic Optimization. SIAM Journal on Optimization, 2006.
- Bach. Duality between subgradient and conditional gradient methods. SIAM Journal on Optimization, 2015.
- Burns and Liang. Improved analysis of restarted accelerated gradient and augmented Lagrangian methods via inexact proximal point frameworks. Preprint, 2026.
- Grimmer and Wang. Implicit Primal-Dual Guarantees in Unconstrained First-Order Minimization. Preprint, 2026.
- Gladin, Gasnikov, and Dvurechensky. Accuracy certificates for convex minimization with inexact oracle. Journal of Optimization Theory and Applications, 2025.
- Lan and Monteiro. Iteration-complexity of first-order augmented Lagrangian methods for convex programming. Mathematical Programming, 2016.
- Lan and Zhou. Random gradient extrapolation for distributed and stochastic optimization. SIAM Journal on Optimization, 2018.
- Liang. Primal-dual proximal bundle and conditional gradient methods for convex problems. Mathematical Programming, 2025.

References

- Lu and Freund. Generalized stochastic Frank-Wolfe algorithm with stochastic "substitute" gradient for structured convex optimization. Mathematical Programming, 2021.
- Monteiro and Svaiter. An accelerated hybrid proximal extragradient method for convex optimization and its implications to second-order methods. SIAM Journal on Optimization, 2013.
- Nemirovski, Onn, and Rothblum. Accuracy certificates for computational problems with convex structure. Mathematics of Operations Research, 2010.
- Rodomanov and Nesterov. Subgradient ellipsoid method for nonsmooth convex problems. Mathematical Programming, 2023.
- Tibshirani. Dykstra's algorithm, ADMM, and coordinate descent: Connections, insights, and extensions. NeurIPS, 2017.
- Zhao and Zhu. A generalized Frank-Wolfe method with "dual averaging" for strongly convex composite optimization. Optimization Letters, 2023.

Thank you!