A Doubly Accelerated Inexact Proximal Point Method for Nonconvex Composite Optimization Problem

Jiaming Liang *

*H. Milton Stewart School of Industrial and Systems Engineering,
Georgia Institute of Technology

November 27, 2018

Joint work with Renato D.C. Monteiro

Table of Contents

Outline

Assumptions

We are interested in the composite nonconvex optimization (CNO) problem

$$\phi_* := \min \{\phi(z) := f(z) + h(z) : z \in \mathbb{R}^n\} \tag{1}$$

where the following conditions are assumed to hold:

(A1) $h \in \overline{\mathrm{Conv}}\,(\mathbb{R}^n)$;

(A2) f is a differentiable function on $\mathrm{dom}\, h$ and there exist scalars $M \geq m > 0$ such that

$$f(u) \geq \ell_f(u; z) - \frac{m}{2}\|u - z\|^2 \quad \forall z, u \in \mathrm{dom}\, h. \tag{2}$$

holds and ∇f is M-Lipschitz continuous on $\mathrm{dom}\, h$, i.e.,

$$\|\nabla f(u) - \nabla f(z)\| \leq M\|u - z\| \quad \forall u, z \in \mathrm{dom}\, h;$$

(A3) the diameter D of $\mathrm{dom}\, h$ is finite.

Approximate solutions

- $\hat{\rho}$-approximate solution:
  if $(\hat{z}, \hat{v}) \in \mathbb{R}^n \times \mathbb{R}^n$ satisfies

$$\hat{v} \in \nabla f(\hat{z}) + \partial h(\hat{z}), \quad \|\hat{v}\| \leq \hat{\rho} \tag{3}$$

- $(\bar{\rho}, \bar{\varepsilon})$-prox-approximate solution:
  if $(\lambda, z^-, z, w, \varepsilon) \in \mathbb{R}_{++} \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+$ satisfies

$$w \in \partial_\varepsilon \left(\phi + \frac{1}{2\lambda} \| \cdot - z^- \|^2 \right)(z), \quad \left\| \frac{1}{\lambda}(z^- - z) \right\| \leq \bar{\rho}, \quad \varepsilon \leq \bar{\varepsilon}. \tag{4}$$

Refinement

The next proposition shows how an approximate solution as in (3) can be obtained from a prox-approximate solution by performing a composite gradient step.

Proposition

Let $h \in \overline{\mathrm{Conv}}\,(\mathbb{R}^n)$ *and* f *be a differentiable function on* $\mathrm{dom}\, h$ *whose gradient is* M*-Lipschitz continuous on* $\mathrm{dom}\, h$*. Let* $(\bar\rho, \bar\varepsilon) \in \mathbb{R}^2_{++}$ *and a* $(\bar\rho, \bar\varepsilon)$*-prox-approximate solution* $(\lambda, z^-, z, w, \varepsilon)$ *be given and define*

$$z_f := \underset{u}{\mathrm{argmin}} \left\{ \ell_f(u; z) + h(u) + \frac{M + \lambda^{-1}}{2}\|u - z\|^2 \right\}, \tag{5}$$

$$q_f := [M + \lambda^{-1}](z - z_f), \tag{6}$$

$$v_f := q_f + \nabla f(z_f) - \nabla f(z). \tag{7}$$

Then, (z_f, v_f) *satisfies*

$$v_f \in \nabla f(z_f) + \partial h(z_f), \quad \|v_f\| \le 2\|q_f\| \le 2\left[\bar\rho + \sqrt{2\bar\varepsilon(M + \lambda^{-1})}\right].$$

Literature review

- S. Ghadimi and G. Lan (2016) *Accelerated gradient methods for nonconvex nonlinear and stochastic programming* (**AG method**)
  - The first time that the convergence of the AG method has been established for solving nonconvex nonlinear programming
  - Small stepsize

- W. Kong, J.G. Melo and R.D.C. Monteiro (2018) *Complexity of a quadratic penalty accelerated inexact proximal point method for solving linearly constrained nonconvex composite programs* (**AIPP method**)
  - Apply an accelerated inexact proximal point method for solving approximately each prox-subproblem
  - Large stepsize

Literature review

- AG by Ghadimi and Lan

$$\mathcal{O}\left(\frac{MmD^2}{\hat{\rho}^2}+\left(\frac{Md_0}{\hat{\rho}}\right)^{2/3}\right)$$

- AIPP by Kong, Melo and Monteiro

$$\mathcal{O}\left(\frac{\sqrt{Mm}}{\hat{\rho}^2}\min\left\{\phi(z_0)-\phi_*, md_0^2\right\}+\sqrt{\frac{M}{m}}\log\left(\frac{M+m}{m}\right)\right)$$

- D-AIPP in this paper

$$\mathcal{O}\left(\frac{M^{1/2}m^{3/2}D^2}{\hat{\rho}^2}+\sqrt{\frac{M}{m}}\log\left(\frac{M+m}{m}\right)\right)$$

Literature review

- AG by Ghadimi and Lan

$$\mathcal{O}\left(\frac{MmD^2}{\hat{\rho}^2} + \left(\frac{Md_0}{\hat{\rho}}\right)^{2/3}\right)$$

- AIPP by Kong, Melo and Monteiro

$$\mathcal{O}\left(\frac{\sqrt{Mm}}{\hat{\rho}^2}\min\left\{\phi(z_0)-\phi_*, md_0^2\right\} + \sqrt{\frac{M}{m}}\log\left(\frac{M+m}{m}\right)\right)$$

- D-AIPP in this paper

$$\mathcal{O}\left(\frac{M^{1/2}m^{3/2}D^2}{\hat{\rho}^2} + \sqrt{\frac{M}{m}}\log\left(\frac{M+m}{m}\right)\right)$$

Literature review

- AG by Ghadimi and Lan

$$\mathcal{O}\left(\frac{MmD^2}{\hat{\rho}^2} + \left(\frac{Md_0}{\hat{\rho}}\right)^{2/3}\right)$$

- AIPP by Kong, Melo and Monteiro

$$\mathcal{O}\left(\frac{\sqrt{Mm}}{\hat{\rho}^2}\min\left\{\phi(z_0)-\phi_*, md_0^2\right\} + \sqrt{\frac{M}{m}}\log\left(\frac{M+m}{m}\right)\right)$$

- D-AIPP in this paper

$$\mathcal{O}\left(\frac{M^{1/2}m^{3/2}D^2}{\hat{\rho}^2} + \sqrt{\frac{M}{m}}\log\left(\frac{M+m}{m}\right)\right)$$

Outline

GAIPP framework

0. Let $x_0 = y_0 \in \operatorname{dom} h$, $0 < \theta < \alpha$, $\delta \geq 0$, $0 < \kappa < \min\{1, 1/\alpha\}$ be given, and set $k = 0$ and $A_0 = 0$;
1. compute

$$a_k = \frac{1 + \sqrt{1 + 4A_k}}{2}, \quad A_{k+1} = A_k + a_k, \quad \tilde{x}_k = \frac{A_k}{A_{k+1}} y_k + \frac{a_k}{A_{k+1}} x_k;$$

2. choose $\lambda_k > 0$ and find a triple $(y_{k+1}, \tilde{v}_{k+1}, \tilde{\varepsilon}_{k+1})$ satisfying

$$\tilde{v}_{k+1} \in \partial_{\tilde{\varepsilon}_{k+1}} \left(\lambda_k \phi(\cdot) + \frac{1}{2} \| \cdot - \tilde{x}_k \|^2 - \frac{\alpha}{2} \| \cdot - y_{k+1} \|^2 \right) (y_{k+1}),$$

$$\frac{1}{\alpha + \delta} \| \tilde{v}_{k+1} + \delta (y_{k+1} - \tilde{x}_k) \|^2 + 2 \tilde{\varepsilon}_{k+1} \leq (\kappa \alpha + \delta) \| y_{k+1} - \tilde{x}_k \|^2;$$

3. compute

$$x_{k+1} := \frac{-\tilde{v}_{k+1} + \alpha y_{k+1} + \delta x_k / a_k - (1 - 1/a_k) \theta y_k}{\alpha - \theta + (\theta + \delta)/a_k};$$

4. set $k \leftarrow k + 1$ and go to step 1.

FISTA for convex $\phi = f + h$

1. Let $x_0 = y_0 \in \operatorname{dom} h$, and set $A_0 = 0$, $k = 0$;
2. For $k \geq 0$ iterate:
   (1) compute
   $$a_k = \frac{1 + \sqrt{1 + 4MA_k}}{2M}, \quad A_{k+1} = A_k + a_k, \quad \tilde{x}_k = \frac{A_k}{A_{k+1}} y_k + \frac{a_k}{A_{k+1}} x_k;$$
   (2) compute
   $$y_{k+1} := \operatorname*{argmin}_u \{\tilde{\gamma}_k(u) + \frac{M}{2}\|u - \tilde{x}_k\|^2\},$$
   where $\tilde{\gamma}(u) = \ell_f(u; \tilde{x}_k) + h(u)$;
   (3) compute
   $$x_{k+1} := \operatorname*{argmin}_u \{a_k \gamma_k(u) + \frac{1}{2}\|u - x_k\|^2\} = x_k + a_k M(y_{k+1} - \tilde{x}_k),$$
   where $\gamma_k(u) = \tilde{\gamma}_k(y_{k+1}) + M\langle \tilde{x}_k - y_{k+1}, u - y_{k+1}\rangle$.

Convergence

Lemma

For $k \geq 0$, a_k and A_k defined as in FISTA, the following estimations hold

$$A_k \geq \frac{k^2}{4M}, \quad \sum_{i=0}^{k-1} A_{i+1} \geq \frac{k^3}{12M}, \quad \frac{\sum_{i=0}^{k-1} a_i}{\sum_{i=0}^{k-1} A_{i+1}} \leq \frac{4}{k}.$$

Theorem

$$\phi(y_k) - \phi^* \leq \frac{2M}{k^2}\|x_0 - x_*\|^2.$$

modified FISTA

0. Let $x_0 = y_0 \in \operatorname{dom} h$, a pair $(m, M) \in \mathbb{R}^2_{++}$ satisfying (A2), a tolerance $\bar{\rho} \in \mathbb{R}_{++}$ be given, and set $k = 0$ and $A_0 = 0$; also, choose positive parameters $0 < \theta < \alpha < 1$ and $\delta \geq 0$;
1. compute

$$a_k = \frac{1 + \sqrt{1 + 4A_k}}{2}, \quad A_{k+1} = A_k + a_k, \quad \tilde{x}_k = \frac{A_k}{A_{k+1}} y_k + \frac{a_k}{A_{k+1}} x_k;$$

2. choose $0 < \lambda < \min\{\alpha/M, (1-\alpha)/m\}$ and set y_{k+1} to be

$$y_{k+1} := \operatorname{argmin}\left\{ \ell_f(\cdot; \tilde{x}_k) + h + \frac{1}{2\lambda}\|\cdot - \tilde{x}_k\|^2 \right\};$$

3. compute

$$x_{k+1} := \frac{-\lambda[\nabla f(y_{k+1}) - \nabla f(\tilde{x}_k)] + \alpha y_{k+1} + \delta x_k / a_k - (1 - 1/a_k)\theta y_k}{\alpha - \theta + (\theta + \delta)/a_k};$$

4. set $k \leftarrow k + 1$ and go to step 1.

Results – basics

Lemma

Letting

$$\tilde{\phi}_k := \lambda_k \phi + \frac{1}{2}\|\cdot - \tilde{x}_k\|^2,$$

$$\gamma_k := \tilde{\phi}_k(y_{k+1}) + \langle \tilde{v}_{k+1}, \cdot - y_{k+1}\rangle + \frac{\alpha}{2}\|\cdot - y_{k+1}\|^2 - \frac{\theta}{2}\|\cdot - \tilde{x}_k\|^2 - \tilde{\varepsilon}_{k+1},$$

then, for every $k \geq 0$, the following statements hold:

(a) γ_k *is an $(\alpha - \theta)$-strongly convex quadratic function and* $\tilde{\phi}_k \geq \gamma_k + \theta\|\cdot - \tilde{x}_k\|^2/2$;

(b) $x_{k+1} = \operatorname{argmin}\left\{a_k\gamma_k(u) + (\theta + \delta)\|u - x_k\|^2/2 : u \in \mathbb{R}^n\right\}$;

(c) $\min\left\{\gamma_k(u) + (\theta + \delta)\|u - \tilde{x}_k\|^2/2 : u \in \mathbb{R}^n\right\} \geq \lambda_k\phi(y_{k+1}) + (1 - \kappa\alpha)\|y_{k+1} - \tilde{x}_k\|^2/2$;

(d) $\gamma_k(u) - \lambda_k\phi(u) \leq (1 - \theta)\|u - \tilde{x}_k\|^2/2$ *for every* $u \in \operatorname{dom} h$.

Results – boundedness

Lemma

Define

$$\beta := 3 + \frac{4(\theta+\delta)}{\alpha-\theta}, \quad \tau_0 := \frac{\sqrt{\kappa\alpha+\delta}}{\sqrt{\alpha+\delta}}. \tag{8}$$

where α, θ, δ and κ are the parameters as in step 0 of the GAIPP framework. Then, $\tau_0 < 1$ and, for every $\bar{x} \in \operatorname{dom} h$, we have

$$\|x_k - \bar{x}\| \le \tau_0^k \|x_0 - \bar{x}\| + \frac{\beta}{1-\tau_0} D \quad \forall k \ge 1.$$

where D is as in (A3). As a consequence, $\{x_k\}$ is bounded.

Results – convergence

Proposition

For every $k \geq 0$,

$$\frac{1-\kappa\alpha}{2}\sum_{i=0}^{k-1} A_{i+1}\|\tilde{x}_i - y_{i+1}\|^2 \leq \left[\frac{\theta+\delta}{2} + (1-\theta)\frac{2\beta^2 k}{(1-\tau_0)^2} + (1-\theta)\sum_{i=0}^{k-1} a_i\right] D^2.$$

As a consequence,

$$\min_{0\leq i\leq k-1} \frac{\|\tilde{x}_i - y_{i+1}\|^2}{\lambda_i^2} \leq \frac{\left[\theta+\delta+c_0 k + 2(1-\theta)\sum_{i=0}^{k-1} a_i\right] D^2}{(1-\kappa\alpha)\sum_{i=0}^{k-1} A_{i+1}\lambda_i^2}$$

where

$$c_0 := \frac{4(1-\theta)\beta^2}{(1-\tau_0)^2} = \mathcal{O}(\delta^4)$$

and β *and* τ_0 *are as in* (8).

Results – convergence

Corollary

If, for some $\underline{\lambda} > 0$*, we have* $\lambda_i \geq \underline{\lambda}$ *for every* $i = 0, \cdots, k-1$*, then*

$$\min_{0 \leq i \leq k-1} \frac{\|\tilde{x}_i - y_{i+1}\|^2}{\lambda_i^2} \leq \frac{D^2}{(1-\kappa\alpha)\underline{\lambda}^2} \left[\frac{12(\theta+\delta)}{k^3} + \frac{12c_0}{k^2} + \frac{8(1-\theta)}{k} \right].$$

Consequently,

$$\min_{0 \leq i \leq k-1} \frac{\|\tilde{x}_i - y_{i+1}\|^2}{\lambda_i^2} = \mathcal{O}\left(\frac{D^2}{\underline{\lambda}^2 k} \right).$$

Outline

Subproblem

Recall that, in the GAIPP framework, we solve a subproblem

$$\tilde{v}_{k+1} \in \partial_{\tilde{\varepsilon}_{k+1}} \left(\lambda_k \phi(\cdot) + \frac{1}{2}\| \cdot - \tilde{x}_k\|^2 - \frac{\alpha}{2}\| \cdot - y_{k+1}\|^2 \right) (y_{k+1}),$$

in each outer iteration.
In fact, when the objective function in the parentheses are strongly convex, we solve

$$\min\{\psi(z) := \psi_s(z) + \psi_n(z) : z \in \mathbb{R}^n\} \tag{9}$$

where the following conditions hold:

(B1) $\psi_n : \mathbb{R}^n \to (-\infty, +\infty]$ is a proper, closed and μ-strongly convex function with $\mu \geq 0$;

(B2) ψ_s is a convex differentiable function whose gradient is L-Lipschitz continuous on the domain of ψ_n.

Accelerated Composite Gradient (ACG) Method

0. Let a pair of functions (ψ_s, ψ_n) as in (9) and initial point $z_0 \in \operatorname{dom} \psi_n$ be given, and set $y_0 = z_0$, $B_0 = 0$, $\Gamma_0 \equiv 0$ and $j = 0$;
1. compute

$$B_{j+1} = B_j + \frac{\mu B_j + 1 + \sqrt{(\mu B_j + 1)^2 + 4L(\mu B_j + 1)B_j}}{2L},$$

$$\tilde{z}_j = \frac{B_j}{B_{j+1}} z_j + \frac{B_{j+1} - B_j}{B_{j+1}} y_j, \quad \Gamma_{j+1} = \frac{B_j}{B_{j+1}} \Gamma_j + \frac{B_{j+1} - B_j}{B_{j+1}} l_{\psi_s}(\cdot, \tilde{z}_j),$$

$$y_{j+1} = \operatorname*{argmin}_y \left\{ \Gamma_{j+1}(y) + \psi_n(y) + \frac{1}{2B_{j+1}} \|y - y_0\|^2 \right\},$$

$$z_{j+1} = \frac{B_j}{B_{j+1}} z_j + \frac{B_{j+1} - B_j}{B_{j+1}} y_{j+1},$$

2. compute

$$u_{j+1} = \frac{y_0 - y_{j+1}}{B_{j+1}},$$

$$\eta_{j+1} = \psi(z_{j+1}) - \Gamma_{j+1}(y_{j+1}) - \psi_n(y_{j+1}) - \langle u_{j+1}, z_{j+1} - y_{j+1} \rangle;$$

3. set $j \leftarrow j + 1$ and go to step 1.

ACG method

Proposition

Let positive constants α, δ *and* κ *be given and consider the sequence* $\{(B_j, \Gamma_j, z_j, u_j, \eta_j)\}$ *generated by the ACG method applied to* (9) *where* (ψ_s, ψ_n) *is a given pair of data functions satisfying (B1) and (B2) with* $\mu \geq 0$. *The ACG method obtains a triple* $(z, u, \eta) = (z_j, u_j, \eta_j)$ *satisfying*

$$u \in \partial_\eta(\psi_s + \psi_n)(z) \quad \frac{1}{\alpha + \delta}\|u_j + \delta(z_j - z_0)\|^2 + 2\eta_j \leq (\kappa\alpha + \delta)\|z_j - z_0\|^2$$

in at most

$$\left\lceil 2\sqrt{\frac{L(\kappa+1)}{\kappa\alpha + (\kappa+1)\delta}} \right\rceil$$

iterations.

D-AIPP method

0. Let $x_0 = y_0 \in \operatorname{dom} h$, a pair $(m, M) \in \mathbb{R}^2_{++}$ satisfying (A2), a stepsize $0 < \lambda \leq 1/(2m)$, and a tolerance pair $(\bar{\rho}, \bar{\varepsilon}) \in \mathbb{R}^2_{++}$ be given, and set $k = 0$, $A_0 = 0$ and $\xi = 1 - \lambda m$; also, choose parameters $0 < \theta < \xi/2$, $\delta \geq 0$;
1. compute

$$a_k = \frac{1 + \sqrt{1 + 4A_k}}{2}, \quad A_{k+1} = A_k + a_k, \quad \tilde{x}_k = \frac{A_k}{A_{k+1}} y_k + \frac{a_k}{A_{k+1}} x_k;$$

and perform at least $\left\lceil 6\sqrt{2\lambda M + 1} \right\rceil$ iterations of the ACG method started from $\tilde{x}_k$ and with

$$\psi_s = \psi_s^k := \lambda f + \frac{1}{4}\| \cdot - \tilde{x}_k \|^2, \quad \psi_n = \psi_n^k := \lambda h + \frac{1}{4}\| \cdot - \tilde{x}_k \|^2$$

to obtain a triple (z, u, η) satisfying

$$u \in \partial_\eta \left(\lambda \phi(\cdot) + \frac{1}{2}\| \cdot - \tilde{x}_k \|^2 \right)(z), \tag{10}$$

$$\frac{1}{\xi/2 + \delta} \|u + \delta(z - \tilde{x}_k)\|^2 + 2\eta \leq (\xi/4 + \delta)\|z - \tilde{x}_k\|^2; \tag{11}$$

D-AIPP method

2. if

$$\|z - \tilde{x}_k\| \leq \frac{\lambda \bar{\rho}}{2}$$

then go to step 3; otherwise, set $(y_{k+1}, \tilde{v}_{k+1}, \tilde{\varepsilon}_{k+1}) = (z, u, 2\eta)$,

$$x_{k+1} := \frac{-\tilde{v}_{k+1} + \xi y_{k+1}/2 + \delta x_k/a_k - (1 - 1/a_k)\theta y_k}{\xi/2 - \theta + (\theta + \delta)/a_k};$$

and $k \leftarrow k + 1$, and go to step 1;

3. restart the previous call to the ACG method in step 1 to find an iterate $(\tilde{z}, \tilde{u}, \tilde{\eta})$ satisfying (10), (11) with (z, u, η) replaced by $(\tilde{z}, \tilde{u}, \tilde{\eta})$ and the extra condition

$$\tilde{\eta} \leq \lambda \bar{\varepsilon}$$

and set $(y_{k+1}, \tilde{v}_{k+1}, \tilde{\varepsilon}_{k+1}) = (\tilde{z}, \tilde{u}, 2\tilde{\eta})$; finally, output $(\lambda, y^-, y, v, \varepsilon)$ where

$$(y^-, y, v, \varepsilon) = (\tilde{x}_k, y_{k+1}, \tilde{v}_{k+1}/\lambda, \tilde{\varepsilon}_{k+1}/(2\lambda)).$$

Technical result

Lemma

Assume that $\psi \in \overline{\mathrm{Conv}}\left(\mathbb{R}^n\right)$ *is a* ξ*-strongly convex function and let* $(y, \eta) \in \mathbb{R}^n \times \mathbb{R}$ *be such that* $0 \in \partial_\eta \psi(y)$. *Then,*

$$0 \in \partial_{2\eta}\left(\psi - \frac{\xi}{4}\| \cdot - y\|^2\right)(y).$$

Lemma

The following statements hold about the algorithm D-AIPP:

(a) *it is a special implementation of the GAIPP with* $\alpha = \xi/2$, *and* $\kappa = 1/2$;

(b) *the number of outer of iterations performed by the D-AIPP is bounded by*

$$\mathcal{O}\left(\frac{D^2}{\lambda^2 \bar{\rho}^2}\right);$$

(c) *at every outer iteration, the numer of calls to the ACG method in step 2 finds a triple* (z, u, η) *satisfying* (10) *and* (11) *is at most*

$$\mathcal{O}\left(\sqrt{\lambda M + 1}\right);$$

(d) *at the last outer iteration, say the* K*-th one, the triple* $(\tilde{z}, \tilde{u}, \tilde{\eta})$ *satisfies* $\|\tilde{x}_K - \tilde{z}\| \leq \lambda\bar{\rho}, \tilde{\eta} \leq \lambda\bar{\varepsilon}$ *and the extra number of ACG iterations is bounded by*

$$\mathcal{O}\left(\sqrt{\lambda M + 1}\log_1^+\left(\frac{\bar{\rho}\sqrt{\lambda^2 M + \lambda}}{\sqrt{\bar{\varepsilon}}}\right)\right).$$

Main results

Theorem

The D-AIPP method terminates with a $(\bar{\rho}, \bar{\varepsilon})$*-prox-solution* $(\lambda, y^-, y, v, \varepsilon)$ *by performing a total number of inner iterations bounded by*

$$\mathcal{O}\left\{\sqrt{\lambda M + 1}\left[\frac{D^2}{\lambda^2 \bar{\rho}^2} + \log_1^+\left(\frac{\bar{\rho}\sqrt{\lambda^2 M + \lambda}}{\sqrt{\bar{\varepsilon}}}\right)\right]\right\}.$$

As a consequence, if $\lambda = \Theta(1/m)$*, the above inner-iteration complexity reduces to*

$$\mathcal{O}\left(\frac{M^{1/2} m^{3/2} D^2}{\bar{\rho}^2} + \sqrt{\frac{M}{m}} \log_1^+\left(\frac{\bar{\rho}\sqrt{M}}{m\sqrt{\bar{\varepsilon}}}\right)\right).$$

Main results

Corollary

Let a tolerance $\hat{\rho} > 0$ *be given and let* $(\lambda, y^-, y, v, \varepsilon)$ *be the output obtained by the D-AIPP method with inputs* $\lambda = 1/(2m)$ *and* $(\bar{\rho}, \bar{\varepsilon})$ *defined as*

$$\bar{\rho} := \frac{\hat{\rho}}{4} \quad \text{and} \quad \bar{\varepsilon} := \frac{\hat{\rho}^2}{32(M + 2m)}.$$

Then the following statements hold:

(a) *the number of inner iterations for D-AIPP method to terminate is at most*

$$\mathcal{O}\left(\frac{M^{1/2} m^{3/2} D^2}{\hat{\rho}^2} + \sqrt{\frac{M}{m}} \log_1^+ \left(\frac{M}{m} \right) \right)$$

(b) *if* ∇g *is* M*-Lipschitz continuous, then the pair* $(\hat{z}, \hat{v}) = (z_g, v_g)$ *computed according to* (5) *and* (7) *is a* $\hat{\rho}$*-approximate solution of* (1)*, i.e.,* (3) *holds.*

Outline

Problems

Several instances of the quadratic programming problem

$$\min \left\{ g(z) := -\frac{\alpha_1}{2}\|DBz\|^2 + \frac{\alpha_2}{2}\|Az - b\|^2 : z \in \Delta_n \right\}$$

are considered for comparison among AG, AIPP and D-AIPP.
All three methods use the centroid of the set Δ_n as the initial point z_0 and are run until a pair (z, v) is generated satisfying the condition

$$v \in \nabla g(z) + N_{\Delta_n}(z), \qquad \frac{\|v\|}{\|\nabla g(z_0)\| + 1} \leq \bar{\rho}$$

for a given tolerance $\bar{\rho} > 0$.

Preliminary results

This is the case when $\bar{\rho} = 10^{-7}$.

Size		Iteration Count		
M	m	AG	AIPP	D-AIPP
16777216	16777216	**374**	14822	1841
16777216	1048576	4429	6711	**1246**
16777216	65536	22087	24129	**4920**
16777216	4096	26053	5706	**5585**
16777216	256	20371	**1625**	2883
16777216	16	20761	**2308**	3656

Table: Numerical results of AIPP with $\sigma = 0.3$ and $\lambda = 0.9/m$, D-AIPP with $\lambda = 0.9/m$, $\theta = 0.49\xi$ and $\theta' := \theta + \delta = 0.9(M/m)^{1/7}$

Thank you!

Manuscript is available at Optimization Online.