Linearized proximal algorithms with adaptive stepsizes for convex composite optimization with applications

Linglingzhi Zhu

School of Mathematical Sciences
Zhejiang University

Joint work with Yaohua Hu, Chong Li, Jinhua Wang, and Xiaoqi Yang

Outline

1. Convex Composite Optimization
2. Linearized Proximal Algorithm
   - LPA
   - Inexact LPA
   - Globalized LPA
3. Applications
   - Convex Inclusion Problem
   - Nonnegative Inverse Eigenvalue Problem
   - Sensor Network Localization Problem

Convex composite optimization problem

$$\min_{x\in\mathbb{R}^n} f(x) := h(F(x)),$$

- the outer function $h : \mathbb{R}^m \to \mathbb{R}$ is convex,
- the inner function $F : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable,
- $C := \arg\min_{y\in\mathbb{R}^m} h(y) \quad \text{and} \quad X^* := \arg\min_{x\in\mathbb{R}^n} h(F(x))$.

It provides a unified framework for

- convex inclusions,
- nonsmooth and nonconvex optimization,
- penalty methods for nonlinear programming,
- regularized minimization problems.

Convex composite optimization problem

$$\min_{x\in\mathbb{R}^n} f(x) := h(F(x)),$$

- the outer function $h : \mathbb{R}^m \to \mathbb{R}$ is convex,
- the inner function $F : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable,
- $C := \arg\min_{y\in\mathbb{R}^m} h(y) \quad \text{and} \quad X^* := \arg\min_{x\in\mathbb{R}^n} h(F(x))$.

It provides a unified framework for

- convex inclusions,
- nonsmooth and nonconvex optimization,
- penalty methods for nonlinear programming,
- regularized minimization problems.

Algorithm (Gauss-Newton method)

Initializing: $\rho \geq 1$, $\Delta \in (0, +\infty]$ *and* $x_0 \in \mathbb{R}^n$.
Iteration of $x_k \to x_{k+1}$:

- *If* $h(F(x_k)) = \min\{h(F(x_k) + F'(x_k)d) : \|d\| \leq \Delta\}$, *then* *stop*;
- *otherwise, we denote* $D_\Delta(x_k) := \arg\min_{\|d\|\leq\Delta}\{h(F(x_k) + F'(x_k)d)\}$, *choose* $d_k \in D_\Delta(x_k)$ *to satisfy* $\|d_k\| \leq \rho\,\mathrm{dist}(0, D_\Delta(x_k))$, *and set* $x_{k+1} = x_k + d_k$.

Convergence study of GNM:

- [R. S. Womersley, Math. Program. 1985]: quadratically converges to a local minima under the assumption of strong uniqueness.
- [J. V. Burke and M. C. Ferris, Math. Program. 1995]: quadratically converges to a global minima under the assumptions of weak sharp minima and regularity condition.
- [C. Li and X. Wang, Math. Program. 2002]: quadratically converges to a global minima under the assumption of regularity condition.
- [C. Li and K. F. Ng, SIAM J. Optim. 2007]: semilocal linear/quadratic convergence with quasi-regular initial points.

- From the practical view, the GNM is inconvenient to implement, because the search direction d_k, which is the vector in $D_\Delta(x_k)$ with minimal norm, is difficult to be found for many applications, especially for the large scale problems.
- Based on the idea of the proximal point algorithm due to [R. T. Rockafellar, SIAM J. Control Optim. 1976], [A. S. Lewis and S. J. Wright, arXiv:math.OC/0812.0423v1 2008] proposed an linearized proximal algorithm called ProxDescent to solve CCO and investigated the properties of local solutions of the subproblem.
- [Y. Hu, C. Li and X. Yang, SIAM J. Optim. 2016] proposed the following general stepsizes of linearized proximal algorithm (LPA) for solving CCO.

- From the practical view, the GNM is inconvenient to implement, because the search direction d_k, which is the vector in $D_\Delta(x_k)$ with minimal norm, is difficult to be found for many applications, especially for the large scale problems.
- Based on the idea of the proximal point algorithm due to [R. T. Rockafellar, SIAM J. Control Optim. 1976], [A. S. Lewis and S. J. Wright, arXiv:math.OC/0812.0423v1 2008] proposed an linearized proximal algorithm called ProxDescent to solve CCO and investigated the properties of local solutions of the subproblem.
- [Y. Hu, C. Li and X. Yang, SIAM J. Optim. 2016] proposed the following general stepsizes of linearized proximal algorithm (LPA) for solving CCO.

- From the practical view, the GNM is inconvenient to implement, because the search direction d_k, which is the vector in $D_\Delta(x_k)$ with minimal norm, is difficult to be found for many applications, especially for the large scale problems.
- Based on the idea of the proximal point algorithm due to [R. T. Rockafellar, SIAM J. Control Optim. 1976], [A. S. Lewis and S. J. Wright, arXiv:math.OC/0812.0423v1 2008] proposed an linearized proximal algorithm called ProxDescent to solve CCO and investigated the properties of local solutions of the subproblem.
- [Y. Hu, C. Li and X. Yang, SIAM J. Optim. 2016] proposed the following general stepsizes of linearized proximal algorithm (LPA) for solving CCO.

Algorithm (LPA)

Initializing: an initial point $x_0 \in \mathbb{R}^n$ and a stepsize $v_0 > 0$.
Iteration of $x_k \to x_{k+1}$*: (by setting* v_k*)*

- If $h(F(x_k)) = \min_{d\in\mathbb{R}^n} \left\{ h(F(x_k) + F'(x_k)d) + v_k\|d\|^2 \right\}$, then stop;
- otherwise, set

$$
\begin{aligned}
d_k &= \arg\min_{d\in\mathbb{R}^n} \left\{ h(F(x_k) + F'(x_k)d) + v_k\|d\|^2 \right\}, \\
x_{k+1} &= x_k + d_k.
\end{aligned}
$$

[Y. Hu, C. Li and X. Yang, SIAM J. Optim. 2016] studied convergence results for the stepsize satisfying:

$$v_k = \frac{1}{2v}$$

where $v > 0$ is a constant for each $k > 0$, or more general, the stepsize $\{v_k\}$ satisfying:

$$0 < \underline{v} \leq v_k \leq \bar{v} < +\infty$$

for each $k > 0$. We call this algorithm CLPA.

The notion of the weak sharp minimizer (of order 1) is proposed by [J. V. Burke and M. C. Ferris, SIAM J. Control Optim. 1993]. [M. Studniarski and D. E. Ward, SIAM J. Control Optim. 1999] extended the concept to the weak sharp minimizer of $p (p \geq 1)$ for a function $g : \mathbb{R}^n \to \mathbb{R}$ as follows:

Definition (Local weak sharp minimizer of order p)

Let $\bar{x} \in \mathbb{R}^n$. $\bar{x}$ is a local weak sharp minimizer of order p for g if $\bar{x} \in \arg\min g$ and there exist $r > 0$ and $\eta_r > 0$ such that

$$g(x) - g(\bar{x}) \geq \eta_r \operatorname{dist}^p(x, \arg\min g) \quad \text{for each } x \in \mathbf{B}(\bar{x}, r).$$

Let $D(x) := \{d \in \mathbb{R}^n : F(x) + F'(x)d \in C\}$ and $\bar{x} \in \mathbb{R}^n$. We say that $\bar{x}$ is

- a regular point (proposed by [J. V. Burke and M. C. Ferris, Math. Program. 1995]) of inclusion $F(x) \in C$ if

$$\ker(F'(\bar{x})^T) \cap (C - F(\bar{x}))^{\ominus} = \{0\};$$

- a quasi-regular point (proposed by [C. Li and K. F. Ng, SIAM J. Optim. 2007]) of inclusion $F(x) \in C$ if there exist $r > 0$ and $\beta_r > 0$ such that

$$\beta_r \mathrm{dist}(0, D(x)) \leq \mathrm{dist}(F(x), C) \quad \text{for each } x \in \mathbf{B}(\bar{x}, r).$$

Drawbacks/limitations of the CLPA:

- The CLPA does not work very efficiently in the case when $p > 2$.
- In the case when $p = 2$, the convergence performance of the CLPA is sensitive to the choice of the stepsize (related to weak sharp minima modulus and the quasi-regular modulus).

In our recent paper, assume that the optimal value $h_{\min}$ of the function h is known and propose an algorithm called ALPA by using an adaptive stepsize:

$$v_k = \min\{\theta w_k^{\alpha}, v\},$$

where $w_k := h(F(x_k)) - h_{\min}$, and $0 < \theta < 1$, $\alpha > 0$, $v > 0$ are constants.

We assume the following blanket assumptions denoted by (H):

- (A1): $F(\bar{x}) \in C$, F' is locally Lipschitz around $\bar{x}$,
- (A2): $\bar{x}$ is a quasi-regular point of inclusion $F(x) \in C$,
- (A3): $F(\bar{x})$ is a local weak sharp minimizer of order p for h.

Theorem

Assumptions:

- *CLPA: $p \in [1,2)$ or the stepsize $v_k < \frac{\eta(\bar{x})\beta(\bar{x})^2}{4}$*[a] *(if $p = 2$).*[b]
- *ALPA: $p \geq 1$, $\alpha > p - 2$.*

Conclusion:

$\{x_k\}$ converges locally to a solution x^* satisfying $F(x^*) \in C$ at a rate of

- *CLPA: $\frac{2}{p}$,*[5]
- *ALPA: $\min\{2, \frac{2+\alpha}{p}\}$. In particular, $\{x_k\}$ converges quadratically if $\alpha \geq 2p - 2$.*

---

[a] $\eta(\bar{x})$ is the local weak sharp minimizer constant of order 2, and $\beta(\bar{x})$ is the quasi-regularity constant.

[b] [Y. Hu, C. Li and X. Yang, SIAM J. Optim. 2016].

If we further assume that

$$\text{(A4): } (h - h_{\min})^{\frac{1}{s}} \text{ is locally Lipschitz at } F(\bar{x}) \text{ with } s \geq 1.$$

Theorem

Assumptions[a]*:*

- $p \geq 1$, $s > \frac{1+\sqrt{[1+2p(p-2)]_+}}{2}$[b], $\alpha \in (\frac{p-2}{s}, \frac{2s-2}{p}]$

Conclusion:

$\{x_k\}$ generated by the ALPA converges locally at a rate of $\min\{\frac{2s}{p}, \frac{2+\alpha s}{p}\}$ *to a solution x^* satisfying $F(x^*) \in C$. In particular, $\{x_k\}$ converges quadratically if $s = p$ and $\alpha = 2 - \frac{2}{p}$.*

[a] (A2)+(A3) could be weakened to the following:
(A5): $\bar{x}$ is a local weak sharp minimizer of order p for $h \circ F$.

[b] (A4)+(A5) imply that $s \leq p$.

Comparison between the CLPA and the ALPA:

- The restriction $p \leq 2$ for CLPA is dropped and estimates of constant $v_k < \frac{\eta(\bar{x})\beta(\bar{x})^2}{4}$[1] are avoided when $p = 2$.
- The convergence rate is improved by choosing suitable α, and even in the case when $p \geq 2$, the convergence could be quadratic.

[1] $\eta(\bar{x})$ is the local weak sharp minimizer constant of order 2, and $\beta(\bar{x})$ is the quasi-regularity constant.

Algorithm (Inexact CLPA [Y. Hu et al., SIAM J. Optim. 2016])

Initializing: $\theta > 0$, $\rho > 0$, *an initial point* $x_0 \in \mathbb{R}^n$, $d_{-1} \in \mathbb{R}^n$ *and a stepsize* $v_0 > 0$.

Iteration of $x_k \to x_{k+1}$ *(by setting* $\epsilon_k = \theta\|d_{k-1}\|^\rho$*):*

- *If* $h(F(x_k)) = \min_{d\in\mathbb{R}^n}\left\{h(F(x_k) + F'(x_k)d) + v_k\|d\|^2\right\}$, *then stop;*
- *else if* $h(F(x_k)) \le \min_{d\in\mathbb{R}^n}\left\{h(F(x_k) + F'(x_k)d) + v_k\|d\|^2\right\} + \epsilon_k$, *then we set*

$$\begin{aligned} d_k &= \|d_{k-1}\|^{\rho-1} d_{k-1}, \\ x_{k+1} &= x_k + d_k. \end{aligned}$$

- *otherwise, we set*

$$\begin{aligned} d_k &= \epsilon_k\text{-}\arg\min_{d\in\mathbb{R}^n}\left\{h(F(x_k) + F'(x_k)d) + v_k\|d\|^2\right\}, \\ x_{k+1} &= x_k + d_k. \end{aligned}$$

Algorithm (Inexact ALPA)

Initializing: $0 < \theta < 1$, $\alpha > 0$, $\rho > 0$, $v > 0$ *and an initial point* $x_0 \in \mathbb{R}^n$.
Iteration of $x_k \to x_{k+1}$ *(by setting* $\epsilon_k \leq \theta w_k^\rho$*):*

- *Set* $v_k = \min\{\theta w_k^\alpha, v\}$.
- *If* $h(F(x_k)) = \min_{d\in\mathbb{R}^n} \left\{ h(F(x_k) + F'(x_k)d) + v_k\|d\|^2 \right\}$, *then stop;*
- *otherwise, we set* $\epsilon_k = \theta\epsilon_k$ *until*

$$h(F(x_k)) > \min_{d\in\mathbb{R}^n} \left\{ h(F(x_k) + F'(x_k)d) + v_k\|d\|^2 \right\} + \epsilon_k$$

and then set

$$\begin{aligned} d_k &= \epsilon_k\text{-}\arg\min_{d\in\mathbb{R}^n} \left\{ h(F(x_k) + F'(x_k)d) + v_k\|d\|^2 \right\}, \\ x_{k+1} &= x_k + d_k. \end{aligned}$$

Theorem

Assumptions:

- *CLPA*: $p \in [1, 2)$ *or the stepsize* $v_k < \frac{\eta(\bar{x})\beta(\bar{x})^2}{32}$ *(if* $p = 2$*);* $\rho > 2$.[a]
- *ALPA*: $p \geq 1$, $\alpha > p - 2$, $\rho \geq \alpha + 2$.

Conclusion:

$\{x_k\}$ *converges locally to a solution* x^* *satisfying* $F(x^*) \in C$ *at a rate of*

- *CLPA*: $\min\left\{\frac{\rho}{2}, \frac{2}{p}\right\}$,[a]
- *ALPA*: $\min\{2, \frac{2+\alpha}{p}\}$. *In particular,* $\{x_k\}$ *converges quadratically if* $\alpha \geq 2p - 2$.

[a] [Y. Hu, C. Li and X. Yang, SIAM J. Optim. 2016].

- For the ALPA, since larger ρ requires more cost for solving the subproblem, we can apply $\rho = \alpha + 2$ instead of $\rho \geq \alpha + 2$ but with same convergence rate.

Theorem

Assumptions:

- *CLPA*: $p \in [1,2)$ *or the stepsize* $v_k < \frac{\eta(\bar{x})\beta(\bar{x})^2}{32}$ *(if* $p = 2$*);* $\rho > 2$.[a]
- *ALPA*: $p \geq 1$, $\alpha > p - 2$, $\rho \geq \alpha + 2$.

Conclusion:

$\{x_k\}$ *converges locally to a solution* x^* *satisfying* $F(x^*) \in C$ *at a rate of*

- *CLPA*: $\min\left\{\frac{\rho}{2}, \frac{2}{p}\right\}$,[a]
- *ALPA*: $\min\{2, \frac{2+\alpha}{p}\}$. *In particular,* $\{x_k\}$ *converges quadratically if* $\alpha \geq 2p - 2$.

[a] [Y. Hu, C. Li and X. Yang, SIAM J. Optim. 2016].

- For the ALPA, since larger ρ requires more cost for solving the subproblem, we can apply $\rho = \alpha + 2$ instead of $\rho \geq \alpha + 2$ but with same convergence rate.

If we further assume that

$$\text{(A4): } (h - h_{\min})^{\frac{1}{s}} \text{ is locally Lipschitz at } F(\bar{x}) \text{ with } s \geq 1.$$

Theorem

Assumptions:

- $p \geq 1$, $s > \frac{1+\sqrt{[1+2p(p-2)]_+}}{2}$, $\alpha \in (\frac{p-2}{s}, \frac{2s-2}{p}]$, $\rho \geq \alpha + 2$.

Conclusion:

$\{x_k\}$ generated by the ALPA converges locally at a rate of $\min\{\frac{2s}{p}, \frac{2+\alpha s}{p}\}$ to a solution x^* satisfying $F(x^*) \in C$. In particular, $\{x_k\}$ converges quadratically if $s = p$ and $\alpha = 2 - \frac{2}{p}$.

In the case when assumptions (A1)-(A4) of theorems hold, then the inexact ALPA owns the following convergence rate

$$q = \begin{cases} \min\{2, \frac{2+\alpha}{p}\}, & \textit{if } (p, \alpha, \rho) \textit{ satisfies } (\mathrm{C1}); \\ \min\{\frac{2s}{p}, \frac{2+\alpha s}{p}\}, & \textit{if } (p, s, \alpha, \rho) \textit{ satisfies } (\mathrm{C2}). \end{cases}$$

$$(\mathrm{C1}): \; p \geq 1, \; \alpha > p - 2, \; \rho \geq \alpha + 2,$$

$$(\mathrm{C2}): \; p \geq 1, \; s > \frac{1+\sqrt{[1+2p(p-2)]_+}}{2}, \; \alpha \in \left(\frac{p-2}{s}, \frac{2s-2}{p}\right], \; \rho \geq \alpha + 2.$$

Algorithm (Globalized CLPA [Y. Hu et al., SIAM J. Optim. 2016])

Initializing: $\theta > 0$, $\rho > 0$, $c \in (0,1)$, $\gamma \in (0,1)$, $x_0 \in \mathbb{R}^n$ *and* $v_0 > 0$.
Iteration of $x_k \to x_{k+1}$*: (by setting* $\epsilon_k = \theta \|d_{k-1}\|^\rho$*)*

- *Generate* d_k *by exact CLPA,*
- *Set*

$$x_{k+1} = x_k + t_k d_k,$$

where t_k *is the maximum value of* γ^i *for* $i = 0, 1, \ldots,$ *such that*

$$\begin{aligned} h(F(x_k + \gamma^i d_k)) - h(F(x_k)) \leq\ & c\gamma^i (h(F(x_k) + F'(x_k) d_k) \\ & + v_k \|d_k\|^2 - h(F(x_k))). \end{aligned}$$

Algorithm (Globalized ALPA)

Initializing: $0 < \theta < 1$, $\rho > 0$, $c \in (0,1)$, $\gamma \in (0,1)$, $x_0 \in \mathbb{R}^n$ *and* $v_0 > 0$.

Iteration of $x_k \to x_{k+1}$*: (by setting* $\epsilon_k \leq \theta w_k^\rho$*)*

- *Generate* d_k *by inexact ALPA,*
- *Set*

$$x_{k+1} \;=\; x_k + t_k d_k,$$

where t_k *is the maximum value of* γ^i *for* $i = 0, 1, \ldots,$ *such that*

$$\begin{aligned} h(F(x_k + \gamma^i d_k)) - h(F(x_k)) \;& \leq c\gamma^i (h(F(x_k) + F'(x_k)d_k) \\ & \quad + v_k \|d_k\|^2 - h(F(x_k))). \end{aligned}$$

We assume for the following theorems that:

- $\{x_k\}$ is the sequence generated by the Globalized LPA and $\bar{x}$ is a cluster point of this sequence such that the blanket assumptions (H) holds.

Theorem

Assumptions:

- *Globalized exact CLPA*: $1 \leq p < 2$.[a]
- *Globalized inexact ALPA*: $1 \leq p < 2$, $\rho \geq \alpha + 2$.

Conclusion:

$\{x_k\}$ *converges to* $\bar{x}$ *at a rate of*

- *Globalized exact CLPA*: $\frac{2}{p}$,[a]
- *Globalized inexact ALPA*: $\min\{2, \frac{2+\alpha}{p}\}$.

[a] [Y. Hu, C. Li and X. Yang, SIAM J. Optim. 2016].

If we further assume that

$$\text{(A4)}: \ (h - h_{\min})^{\frac{1}{s}} \text{ is locally Lipschitz at } F(\bar{x}) \text{ with } s \geq 1.$$

Theorem

Assumptions:

- $p \geq 1$, $s > \frac{1+\sqrt{[1+2p(p-2)]_+}}{2}$, $\alpha \in (\frac{p-2}{s}, \frac{2s-2}{p}]$, $\rho \geq \alpha + 2$.

Conclusion:

$\{x_k\}$ *converges to* $\bar{x}$ *at a rate of* $\min\{\frac{2s}{p}, \frac{2+\alpha s}{p}\}$.

- Convex inclusion problem is at the core of the modeling of many problems in various areas of mathematics and physical sciences:

$$\text{Find } x \text{ such that } F(x) \in Q,$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable and $Q \subseteq \mathbb{R}^m$ is a closed convex set.

- Convex composite optimization reformulation:

$$\min_{x \in \mathbb{R}^n} h(F(x)), \text{ where } h(\cdot) := \frac{1}{p}\mathrm{dist}^p(\cdot, Q).$$

- The first-order optimality condition of the LPA's subproblem ($p > 1$):

$$G_{x,u}(d) := \mathrm{d}^{p-2}(F(x) + F'(x)d, Q)F'(x)^\top(\mathbb{I} - P_Q)(F(x) + F'(x)d) \\ +2ud = 0$$

Algorithm (ALPA for $p > 1$)

Initializing: $0 < \theta < 1$, $\alpha > 0$, $\rho > 0$, $x_0 \in \mathbb{R}^n$.

Iteration of $x_k \to x_{k+1}$ *(by setting* $\epsilon_k \leq \theta w_k^{\rho}$*):*

- *Set* $v_k = \min\{\theta w_k^{\alpha}, v\}$.
- *If* $G_{x_k, v_k}(0) = 0$, *then stop;*
- *otherwise, we set* $\epsilon_k = \theta \epsilon_k$ *until*

$$\|G_{x_k, v_k}(0)\| > \sqrt{2 v_k \epsilon_k},$$

and then generate d_k *by solving the nonlinear equations*

$$G_{x_k, v_k}(d) = 0$$

such that $\|G_{x_k, v_k}(d_k)\| \leq \sqrt{2 v_k \epsilon_k}$, *and set* $x_{k+1} = x_k + d_k$.

We assume the following assumptions for the remainder:

- $\bar{x} \in X^*$ [2],
- F' is locally Lipschitz around $\bar{x}$,
- $\mathrm{im} F'(\bar{x}) - Q = \mathbb{R}^m$.

[2] X^* is the solution set of the convex inclusion problem.

Theorem

Assumptions:

- *CLPA*: $p = 2$, $v_k < \frac{1}{64\bar{\beta}^2}$, $\rho > 1$, *where* $\bar{\beta} := \sup_{\|y\| \le 1} \inf_{F'(\bar{x})d \in y + \mathbb{R}^m_-} \|d\|$.[a]
- *ALPA*: $(\mathrm{C1})$ $p > 1$, $\alpha > \max\{p-2, 2-\frac{2}{p}\}$, $\rho \ge \alpha + 2$, *or*
  $(\mathrm{C2})$ $p > 1$, $\alpha \in \left(1-\frac{2}{p}, 2-\frac{2}{p}\right]$, $\rho \ge \alpha + 2$.

Conclusion: $\{x_k\}$ *converges locally to a solution* x^* *satisfying* $x^* \in X^*$

- *CLPA*: *linearly*,[a]
- *ALPA*: *at a rate of*

$$q = \begin{cases} \min\{2, \frac{2+\alpha}{p}\}, & \textit{if } (p, \alpha, \rho) \textit{ satisfies } (\mathrm{C1}), \\ \min\{2, \frac{2+\alpha p}{p}\}, & \textit{if } (p, \alpha, \rho) \textit{ satisfies } (\mathrm{C2}). \end{cases}$$

[a] [Y. Hu, C. Li and X. Yang, SIAM J. Optim. 2016].

Algorithm (Globalized ALPA for $p > 1$)

Initializing: $c \in (0,1)$, $\gamma \in (0,1)$ *and* $x_0 \in \mathbb{R}^n$.
Iteration of $x_k \to x_{k+1}$*:*

- *Generate* d_k *by the ALPA for* $p > 1$,
- *Set*

$$x_{k+1} = x_k + t_k d_k,$$

where t_k *is the maximum value of* γ^i *for* $i = 0, 1, \dots,$ *such that*

$$\begin{aligned} &\frac{1}{p}\|F(x_k + \gamma^i d_k)_+\|^p - \frac{1}{p}\|F(x_k)_+\|^p \\ &\le c\gamma^i\left(\frac{1}{p}\|(F(x_k) + F'(x_k)d_k)_+\|^p + v_k\|d_k\|^2 - \frac{1}{p}\|F(x_k)_+\|^p\right). \end{aligned}$$

Theorem

Assumptions:

- $\{x_k\}$ *is the sequence generated by the* *Globalized ALPA* *for* $p > 1$ *and* $\bar{x}$ *is a cluster point of this sequence,*
- $(\mathrm{C1})$ $p \in (1, 2)$, $\alpha > 2 - \frac{2}{p}$, $\rho \geq \alpha + 2$, *or*
  $(\mathrm{C2})$ $p > 1$, $\alpha \in \left(1 - \frac{2}{p}, 2 - \frac{2}{p}\right]$, $\rho \geq \alpha + 2$.

Conclusion:

$\{x_k\}$ *converges to* $\bar{x} \in X^*$ *at a rate of*

$$q = \begin{cases} \min\{2, \frac{2+\alpha}{p}\}, & \text{if } (p, \alpha, \rho) \text{ satisfies } (\mathrm{C1}), \\ \min\{2, \frac{2+\alpha p}{p}\}, & \text{if } (p, \alpha, \rho) \text{ satisfies } (\mathrm{C2}). \end{cases}$$

Nonnegative Inverse Eigenvalue Problem (NIEP)

Given an n-tuple $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ which is a spectrum for nonnegative matrices, find $X \in \mathbb{R}_+^{n\times n}$ whose eigenvalues are $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$.

- Define the block diagonal matrix
  $\Lambda := \mathrm{blkdiag}(\lambda_1^{[2]}, \cdots, \lambda_s^{[2]}, \lambda_{2s+1}, \cdots, \lambda_n)$, where

  $$\lambda_i^{[2]} := \begin{bmatrix} a_i & b_i \\ -b_i & a_i \end{bmatrix}, \; a_i, b_i \in \mathbb{R} \text{ with } b_i \neq 0 \text{ for all } i = 1, \cdots, s$$

  and $\lambda_i \in \mathbb{R},\; i = 2s+1, \cdots, n$.
- The set of all isospectral matrices (by Schur decomposition):

  $$\mathcal{S}(\Lambda) := \{X \in \mathbb{R}^{n\times n} : X = U(\Lambda + V)U^T,\; U \in \mathcal{O}(n),\; V \in \mathcal{V}\}$$

  where $\mathcal{O}(n) := \{U \in \mathbb{R}^{n\times n} : U^T U = \mathbb{I}_{n\times n}\}$,
  $\mathcal{V} := \{V \in \mathbb{R}^{n\times n} : V_{ij} = 0 \text{ for all } (i,j) \in \mathcal{I}\}$ and
  $\mathcal{I} := \{(i,j) : i \geq j \text{ or } \Lambda_{ij} \neq 0\}$.

Let the mapping $F : \mathbb{R}^{n\times n} \times \mathcal{V} \to \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$

$$F(U, V) := \left(U(\Lambda + V)U^\top, UU^\top - \mathbb{I}_{n\times n}\right),$$

The NIEP has a solution if and only if there exists $(U, V) \in \mathcal{O}(n) \times \mathcal{V}$ such that $F(U, V) \in Q := \mathbb{R}_+^{n\times n} \times \{0\}$.

Now the NIEP can be solved by the CLPA and ALPA by letting the outer function $h : \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \to \mathbb{R}$

$$h(\cdot) := \frac{1}{p}\mathrm{d}^p(\cdot, \mathbb{R}_+^{n\times n} \times \{0\}).$$

- The accuracy of algorithms are evaluated by

$$\mathrm{RES} := \sqrt{\|[U_*(\Lambda + V_*)U_*^\top]_-\|_F^2 + \|U_* U_*^\top - \mathbb{I}_{n\times n}\|_F^2},$$

where U^* and V^* forming Schur decomposition of estimated results.

- The stopping criterion of the CLPA and ALPA type algorithms:
Outer iteration: the number of iterations is greater than 100 or RES $<$ 1e-4.
Inner iteration: the number of iterations is greater than 50 or
CLPA: $G_{x_k,v}(d) < \max\{\theta\|d_{k-1}\|^\rho, 10^{-(2p+4)}\}$;
ALPA: $G_{x_k,v_k}(d) < \max\{\sqrt{2\theta v_k w_k^\rho}, 10^{-(2p+4)}\}$.
Except extra assumptions, set $v = 0.005$, $c = \gamma = 0.9$, $\theta = 0.5$, $\alpha = 1$, $\rho = 2$.

- The accuracy of algorithms are evaluated by

$$\mathrm{RES} := \sqrt{\|[U_*(\Lambda + V_*)U_*^\top]_-\|_F^2 + \|U_* U_*^\top - \mathbb{I}_{n\times n}\|_F^2},$$

where U^* and V^* forming Schur decomposition of estimated results.
- The stopping criterion of the CLPA and ALPA type algorithms:
Outer iteration: the number of iterations is greater than 100
or RES $<$ 1e-4.
Inner iteration: the number of iterations is greater than 50 or
CLPA: $G_{x_k,v}(d) < \max\{\theta\|d_{k-1}\|^\rho, 10^{-(2p+4)}\}$;
ALPA: $G_{x_k,v_k}(d) < \max\{\sqrt{2\theta v_k w_k^\rho}, 10^{-(2p+4)}\}$.
Except extra assumptions, set $v = 0.005$, $c = \gamma = 0.9$, $\theta = 0.5$, $\alpha = 1$, $\rho = 2$.

Table 1: The result of the NIEP solved by the CLPA and ALPA when $p = 2$.

Algorithm	CLPA		ALPA	
n	CPU time	RES	CPU time	RES
10	0.0849 s	6.5e-05	0.0639 s	3.8e-05
50	1.7763 s	2.9e-05	1.3076 s	4.7e-06
100	17.422 s	1.1e-05	8.6562 s	8.0e-05
150	70.795 s	7.8e-05	43.408 s	1.1e-06
200	163.46 s	4.3e-05	88.427 s	5.9e-07

Table 2: The result of the NIEP solved by the CLPA and ALPA when $p = 4$.

Algorithm	CLPA		ALPA	
n	CPU time	RES	CPU time	RES
10	N/A[3]		0.1601 s	4.2e-05
50	N/A		27.741 s	2.5e-05
100	N/A		148.77 s	8.5e-05
150	N/A		459.96 s	6.3e-05
200	N/A		1260.5 s	1.1e-05

[3] It means the algorithm cannot reach the stopping criterion in tenfold CPU time of the ALPA for corresponding case.

Table 3: The result of the RINC and ALPA ($p = 2$) for the NIEP (dense matrices).

	dense matrices			
Algorithm	RINC[4]		ALPA	
n	CPU time	RES	CPU time	RES
10	0.01 s	9.5e-05	0.05 s	8.6e-06
20	0.03 s	5.9e-06	0.12 s	6.0e-06
50	0.21 s	4.5e-05	1.05 s	1.2e-07
80	0.52 s	1.4e-06	4.71 s	5.3e-07
100	1.02 s	2.4e-05	8.15 s	2.1e-07

[4] Riemannian inexact Newton-CG method [Z. Zhao, Z. Bai and X. Jin, Numer. Math., 2018].

Table 4: The result of the RINC and ALPA ($p = 2$) for the NIEP (1% sparse matrices).

	1% sparse matrices			
Algorithm	RINC		ALPA	
n	CPU time	RES	CPU time	RES
10	1.22 s	9.9e-05	0.21 s	9.5e-07
20	21.7 s	9.7e-05	2.73 s	3.7e-07
50	N/A		5.76 s	7.7e-06
80	N/A		11.7 s	3.1e-05
100	N/A		18.4 s	1.6e-07

Sensor Network Localization Problem

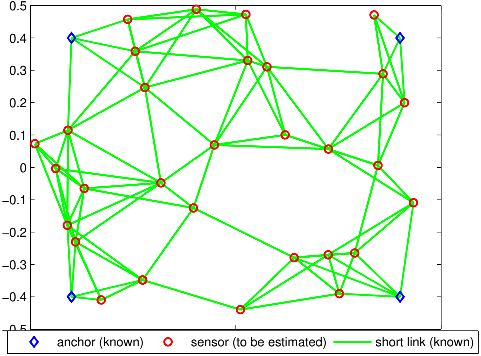


Figure 1: The sensor network localization problem is to estimate the positions of the sensors in a network by using the given incomplete pairwise short distance measurements. The sensors can only detect each other when their distance is within the radio range (depends on the quality of the sensors).

- The sensor network localization problem is to find $\{x_1, \ldots, x_n\}$ satisfying:

$$\|x_i - x_j\|^2 = d_{ij}^2, \quad \|a_k - x_j\|^2 = \bar{d}_{kj}^2, \quad (i,j) \in N_e, \ (k,j) \in M_e,$$
$$\|x_i - x_j\|^2 > R^2, \quad \|a_k - x_j\|^2 > R^2, \quad (i,j) \notin N_e, \ (k,j) \notin M_e.$$

- x_i: the position of sensor (variable), a_k: the position of anchor.
- d_{ij}: the short distance between sensors,
  $\bar{d}_{kj}$: the short distance between anchor and sensor,
  R: the radio range.
- N_e: the sets of sensor-sensor edges, whose length is less or equal to the radio range,
  M_e: the sets of sensor-anchor edges, whose length is less or equal to the radio range.

Let

$$F(x) := (g(x), \bar{g}(x)) \quad \text{for each } x \in \mathbb{R}^{2\times n},$$

where

$$g(x) := ((g_{i,j,1}(x))_{(i,j)\notin N_e}, (g_{i,j,2}(x))_{(i,j)\notin M_e}),$$

$$\bar{g}(x) := ((\bar{g}_{i,j,1}(x))_{(i,j)\in N_e}, (\bar{g}_{i,j,2}(x))_{(i,j)\in M_e}),$$

and

$$\begin{aligned}
g_{i,j,1}(x) &:= R^2 - \|x_i - x_j\|^2, \quad (i,j) \notin N_e,\\
g_{i,j,2}(x) &:= R^2 - \|a_i - x_j\|^2, \quad (i,j) \notin M_e,\\
\bar{g}_{i,j,1}(x) &:= \|x_i - x_j\|^2 - d_{ij}^2, \quad (i,j) \in N_e,\\
\bar{g}_{i,j,2}(x) &:= \|a_i - x_j\|^2 - \bar{d}_{ij}^2, \quad (i,j) \in M_e.
\end{aligned}$$

Let

$$Q := \mathbb{R}_{-}^{\frac{1}{2}n(n-1)+mn-|N_e|-|M_e|} \times \{0\} \subseteq \mathbb{R}^{\frac{1}{2}n(n-1)+mn},$$

where $|\cdot|$ denotes the cardinality of a set.

Now, the CLPA and ALPA can solve the sensor network localization problem as a convex inclusion problem $F(x) \in Q$.

- Neglecting all inequality constraints, many works concentrate on the following relaxation model

$$\begin{aligned} \|x_i - x_j\|^2 &= d_{ij}^2, && (i,j) \in N_e, \\ \|a_k - x_j\|^2 &= \bar{d}_{kj}^2, && (k,j) \in M_e. \end{aligned}$$

- The MDS (Multidimensional Scaling) can be used to solve the incorporate distance measurement localization problem as above model [X. Ji and H. Zha, IEEE INFOCOM 2004].
- The SDR (Semi-Definite Relaxation) [P. Biswas et al., IEEE Trans. Automat. Sci. Engrg. 2006], [Z.-Q. Luo et al., IEEE Signal Proc. Mag. 2010] is also a popular technique to solve the above relaxation model.

- Neglecting all inequality constraints, many works concentrate on the following relaxation model

$$\begin{aligned} \|x_i - x_j\|^2 &= d_{ij}^2, \qquad (i,j) \in N_e, \\ \|a_k - x_j\|^2 &= \bar{d}_{kj}^2, \qquad (k,j) \in M_e. \end{aligned}$$

- The MDS (Multidimensional Scaling) can be used to solve the incorporate distance measurement localization problem as above model [X. Ji and H. Zha, IEEE INFOCOM 2004].
- The SDR (Semi-Definite Relaxation) [P. Biswas et al., IEEE Trans. Automat. Sci. Engrg. 2006], [Z.-Q. Luo et al., IEEE Signal Proc. Mag. 2010] is also a popular technique to solve the above relaxation model.

- Neglecting all inequality constraints, many works concentrate on the following relaxation model

$$\begin{aligned} \|x_i - x_j\|^2 &= d_{ij}^2, && (i,j) \in N_e, \\ \|a_k - x_j\|^2 &= \bar{d}_{kj}^2, && (k,j) \in M_e. \end{aligned}$$

- The MDS (Multidimensional Scaling) can be used to solve the incorporate distance measurement localization problem as above model [X. Ji and H. Zha, IEEE INFOCOM 2004].
- The SDR (Semi-Definite Relaxation) [P. Biswas et al., IEEE Trans. Automat. Sci. Engrg. 2006], [Z.-Q. Luo et al., IEEE Signal Proc. Mag. 2010] is also a popular technique to solve the above relaxation model.

Table 5: List of the algorithms for solving the sensor network localization problem.

Abbreviations	Algorithms
MDS	**M**ulti**D**imensional **S**caling method (relaxed problem).
SDR	**S**emi**D**efinite **R**elaxation method (relaxed problem).
CLPA	CLPA (full problem).
ALPA	ALPA (full problem).
CLPA-R	CLPA (relaxed problem).
ALPA-R	ALPA-R (relaxed problem).
MDS-CLPA	CLPA with initial points by MDS (full problem).
MDS-ALPA	ALPA with initial points by MDS (full problem).
MDS-CLPA-R	CLPA with initial points by MDS (relaxed problem).
MDS-ALPA-R	ALPA with initial points by MDS (relaxed problem).

- The root mean square distance (RMSD) is a key criterion to characterize the accuracy of the estimation:

$$\mathrm{RMSD} = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} \|s_i - x_i\|^2 \right)^{\frac{1}{2}},$$

where s_i is the true position of the sensor, and x_i is the estimated position of the sensor.

- The stopping criterion of the CLPA and ALPA type algorithms:
Outer iteration: the number of iterations is greater than 100 or RMSD $<$ 1e-10.
Inner iteration: the number of iterations is greater than 50 or
CLPA: $G_{x_k,v}(d) < \theta \|d_{k-1}\|^{\rho}$;
ALPA: $G_{x_k,v_k}(d) < \sqrt{2\theta v_k w_k^{\rho}}$.
Except extra assumptions, set $p = 2$, $v = 0.005$, $c = \gamma = 0.9$, $\theta = 0.5$, $\alpha = 1$, $\rho = 2$.

- The root mean square distance (RMSD) is a key criterion to characterize the accuracy of the estimation:

$$\mathrm{RMSD} = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} \|s_i - x_i\|^2 \right)^{\frac{1}{2}},$$

where s_i is the true position of the sensor, and x_i is the estimated position of the sensor.

- The stopping criterion of the CLPA and ALPA type algorithms:
Outer iteration: the number of iterations is greater than 100
or RMSD $<$ 1e-10.
Inner iteration: the number of iterations is greater than 50 or
CLPA: $G_{x_k,v}(d) < \theta \|d_{k-1}\|^{\rho}$;
ALPA: $G_{x_k,v_k}(d) < \sqrt{2\theta v_k w_k^{\rho}}$.
Except extra assumptions, set $p = 2$, $v = 0.005$, $c = \gamma = 0.9$, $\theta = 0.5$, $\alpha = 1$, $\rho = 2$.

Table 6: The numerical results for a WSN localization problem (200 sensors, 20 anchors, radio range=0.3 and initial points: random).

Algorithm	MDS	SDR	CLPA	ALPA	CLPA-R	ALPA-R
RMSD	1.0e-3	2.8e-8	1.8e-11	2.2e-13	1.6e-11	7.8e-13
CPU time	0.3 s	38.6 s	1.7 s	1.5 s	0.6 s	0.5 s
3s-S rate[5]	0%	0%	99%	99%	67%	68%
2s-S rate	0%	0%	64%	64%	67%	68%

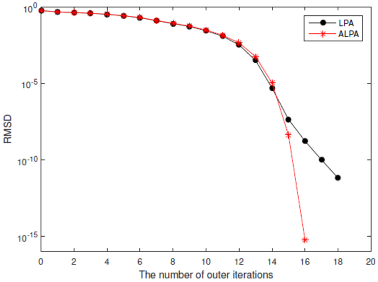
[5] The estimation is regarded as "*t*s-successful" if the estimated RMSD is less than 1e-5 within *t* seconds. "*t*s-S rate" denotes the ratio of "*t*s-successful" estimating in 100 random trials.

Table 7: The numerical results for a WSN localization problem (200 sensors, 20 anchors, radio range=0.3 and initial points: given by the MDS).

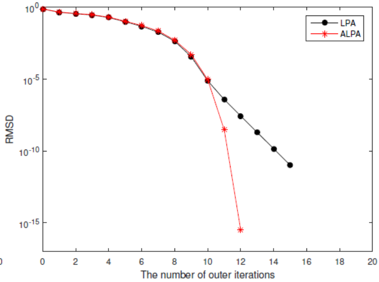
Algorithm	CLPA	ALPA	CLPA-R	ALPA-R
RMSD	1.7e-11	2.2e-13	1.4e-11	5.6e-13
CPU time	0.6 s	0.3 s	0.2 s	0.1 s
CPU time (+MDS)	0.9 s	0.6 s	0.5 s	0.4 s
1s-S rate	99%	99%	99%	99%

Obesrvations:

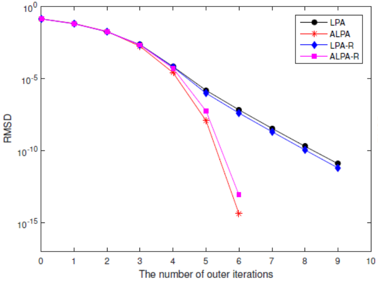
- The CLPA and ALPA achieve a more precise solution within less CPU time than the SDR.
- The CLPA and ALPA consume more CPU time than the CLPA-R and ALPA-R, because the CLPA and ALPA are designed to solve the full version problem whose number of constraints is more than double that of relaxation problem solved by the CLPA-R and ALPA-R.
- When random initial points are used, the CLPA and ALPA own more robust 3s-successful rate than the CLPA-R and ALPA-R as well as the MDS and SDR, which is benefited from more constraints information.
- When good initial points are given by the MDS, the CLPA/CLPA-R and ALPA/ALPA-R will be faster with high successful rate; particularly the ALPA/ALPA-R are much faster than the CLPA/CLPA-R.



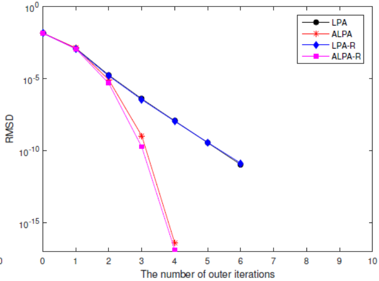
(a) random initial points



(b) $sensor + 0.5 * randn(2, n)$



(c) $sensor + 0.1 * randn(2, n)$

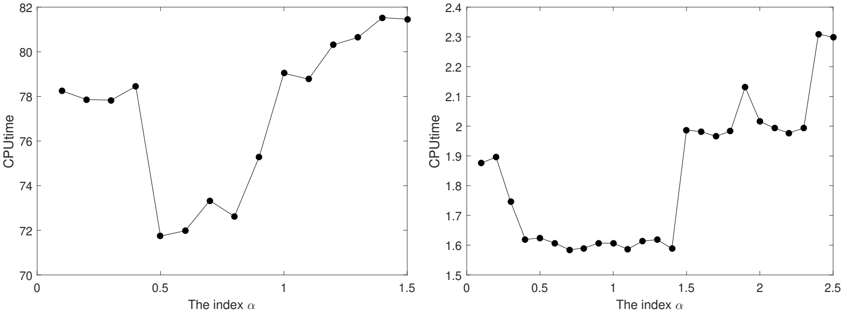


(d) $sensor + 0.01 * randn(2, n)$

Table 8: The CPU time of CLPA/ALPA with different initial points for the WSN localization problem (200 sensors, 20 anchors, radio range=0.3).

Algorithm	CLPA	ALPA	CLPA-R	ALPA-R
random initial points	1.7 s	1.5 s	0.6 s	0.5 s
sensor+0.5∗randn(2,n)	1.5 s	1.2 s	0.6 s	0.5 s
sensor+0.1∗randn(2,n)	1.0 s	0.6 s	0.4 s	0.2 s
sensor+0.01∗randn(2,n)	0.6 s	0.3 s	0.3 s	0.1 s

Results of the ALPA with different α



(a) NIEP (b) Sensor Network Localization

Figure 3: The CPU time of ALPA along with α for the NIEP ($p = 2$ and $n = 200$)/Sensor Network Localization Problem (200 sensors, 20 anchors, radio range = 0.3, and initial points: random.)

Y. H. Hu, C. Li and X. Q. Yang, *On Convergence Rates of Linearized Proximal Algorithms for Convex Composite Optimization with Applications*, SIAM J. Optim, 26 (2016), pp. 1207–1235.

P. Biswas, T.-C. Liang, K.-C. Toh, Y. Ye, and T.-C. Wang, *Semidefinite programming approaches for sensor network localization with noisy distance measurements*, IEEE Trans. Autom. Sci. Eng., 3 (2006), pp. 360–371.

J. V. Burke and M. C. Ferris, *A Gauss-Newton method for convex composite optimization*, Math. Program., 71 (1995), pp. 179–194.

J. V. Burke and M. C. Ferris, *Weak sharp minima in mathematical programming*, SIAM J. Control Optim., 31 (1993), pp. 1340–1359.

A. S. Lewis and S. J. Wright, *A proximal method for composite minimization*, tech. report, 2008.

C. Li and X. Wang, *On convergence of the Gauss-Newton method for convex composite optimization*, Math. Program., 91 (2002), pp. 349–356.

C. Li and K. F. Ng, *Majorizing functions and convergence of the Gauss-Newton method for convex composite optimization*, SIAM J. Optim., 18 (2007), pp. 613–642.

Z.-Q. Luo, W.-K. Ma, A.M.-C. So, Y. Ye, and S. Zhang, *Semidefinite relaxation of quadratic optimization problems*, IEEE Signal Proc. Mag., 27 (2010), pp. 20–34.

L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Program., 58 (1993), pp. 353–367.

R. S. Womersley, *Local properties of algorithms for minimizing nonsmooth composite functions*, Math. Program., 32 (1985), pp. 69–89.

X. Y. Zheng and X. Q. Yang, *Weak sharp minima for semi-infinite optimization problems with applications*, SIAM J. Optim., 18 (2007), pp. 573–588.

X. Ji and H. Y. Zha, *Sensor positioning in wireless ad-hoc sensor networks using multidimensional scaling*, IEEE INFOCOM, 4 (2004), pp. 2652–2661.

Thank You for Your Attention.