A Linear Convergence Result for the Jacobi-Proximal Alternating Direction Method of Multipliers

Hyelin Choi
Joint work with Prof. Woocheol Choi

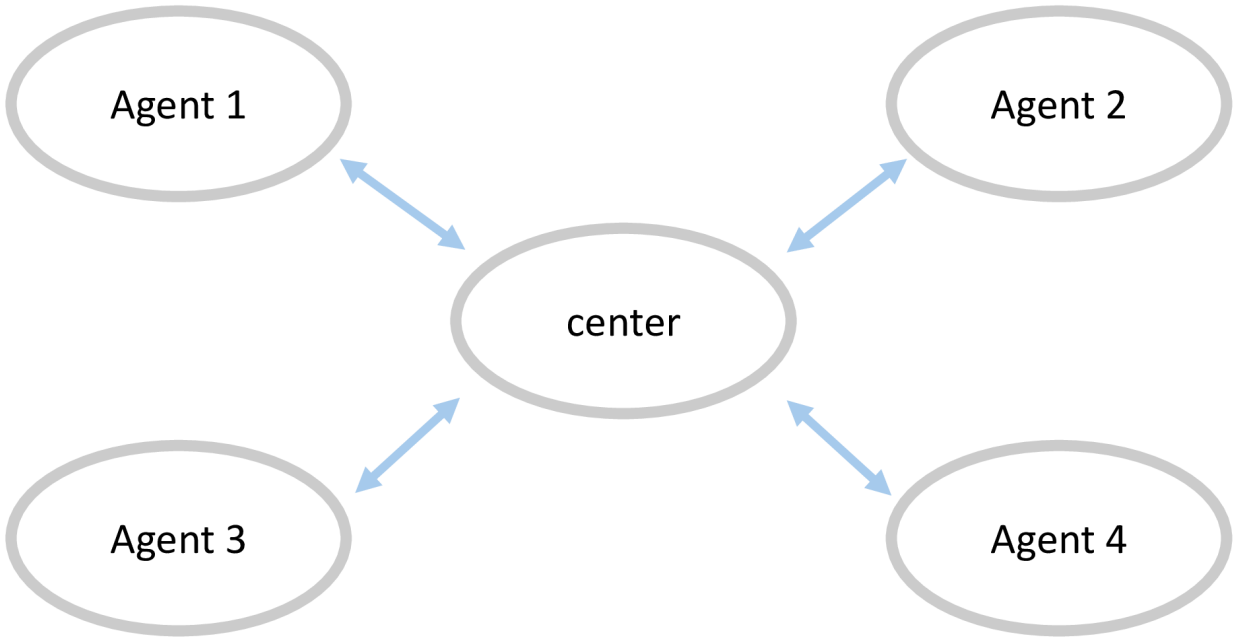
Department of Mathematics
Sungkyunkwan University

April 9, 2025

Outline

Distributed Optimization

Distributed optimization is a method of solving optimization problems by dividing computations across multiple agents, allowing them to work collaboratively.



Classical Form of an Optimization Problem

The classical form of an optimization problem is as follows:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_j(x) = 0, \quad j = 1, \ldots, p \end{aligned}$$

where

- $f(x)$: Objective function (to minimize)
- $g_i(x)$: Inequality constraints
- $h_j(x)$: Equality constraints
- x: Decision variable (Primal variable)

Big O Notation

Definition

Let $f(x)$ and $g(x)$ be real-valued functions. We say that

$$f(x) = \mathcal{O}(g(x)) \quad \text{as } x \to \infty$$

if there exists a positive real number M and a real number x_0 such that

$$|f(x)| \leq M|g(x)| \quad \text{for all } x \geq x_0.$$

Example Suppose there exists a constant $0 < \sigma < 1$ such that

$$\|x^{k+1} - x^*\| \leq \sigma \|x^k - x^*\| \quad \text{for all } k \geq 0.$$

Then we have

$$\|x^k - x^*\| \leq \sigma^k \|x^0 - x^*\|$$

which implies that

$$\|x^k - x^*\| = \mathcal{O}(\sigma^k).$$

$\rightarrow$ Linear Converge

Problem Setting

We consider the following convex optimization problem with N-block variables involving a linear constraint given as follows:

$$\begin{aligned} &\min_{x_1,\cdots,x_N} \sum_{i=1}^{N} \tilde{f}_i(x_i) \\ &\text{subject to } \sum_{i=1}^{N} A_i x_i = c, \\ &\qquad x_i \in \mathcal{X}_i, \qquad i = 1, \cdots, N. \end{aligned} \tag{2.1}$$

where $N \geq 2$, $x_i \in \mathbb{R}^{n_i}$, $A_i \in \mathbb{R}^{m \times n_i}$, $c \in \mathbb{R}^m$ and $\mathcal{X}_i \subset \mathbb{R}^{n_i}$ are closed convex sets and $\tilde{f}_i : \mathbb{R}^{n_i} \to (-\infty, +\infty]$ $(i = 1, 2, \cdots, N)$ are closed proper convex functions. The solution set of (2.1) is assumed to be nonempty.

Problem Reformulation

Note that the convex constraints $x_i \in \mathcal{X}_i$ can be incorporated into the objective $\tilde{f}_i$ using the indicator function:

$$I_{\mathcal{X}_i}(x_i) = \begin{cases} 0 & \text{if } x_i \in \mathcal{X}_i, \\ +\infty & \text{otherwise.} \end{cases} \tag{2.2}$$

Then the problem (2.1) can be rewritten as

$$\begin{aligned} &\min_{x_1,\cdots,x_N} \sum_{i=1}^{N} f_i(x_i) \\ &\text{subject to } \sum_{i=1}^{N} A_i x_i = c, \end{aligned} \tag{2.3}$$

where $f_i(x_i) = \tilde{f}_i(x_i) + I_{\mathcal{X}_i}(x_i)$.

Lagrangian Function

Consider the Lagrangian for problem (2.1):

$$\mathcal{L}(x_1, \cdots, x_N, \lambda) = \sum_{i=1}^{N} f_i(x_i) - \left\langle \lambda, \sum_{i=1}^{N} A_i x_i - c \right\rangle \tag{2.4}$$

where $\lambda \in \mathbb{R}^m$ is the dual variable.

Dual Decomposition

A simple distributed algorithm for solving (2.3) is dual decomposition. The update rules for dual decomposition are given as follows: for $k \geq 1$,

$$\begin{cases} \left(x_1^{k+1}, x_2^{k+1}, \cdots, x_N^{k+1}\right) = \arg\min_{\{x_i\}} \mathcal{L}(x_1, \cdots, x_N, \lambda^k), \\ \lambda^{k+1} = \lambda^k - \alpha_k \left(\sum_{i=1}^N A_i x_i^{k+1} - c\right), \end{cases}$$

where $\alpha_k > 0$ is a step-size. With a suitable choice of α_k and certain assumptions, dual decomposition is guaranteed to converge to an optimal solution [5]. However, dual decomposition exhibits slow convergence in practice, with a convergence rate of $\mathcal{O}(1/\sqrt{k})$ for general convex problems.

[5] Shor, N. Z. (2012). Minimization methods for non-differentiable functions (Vol. 3). Springer Science & Business Media.

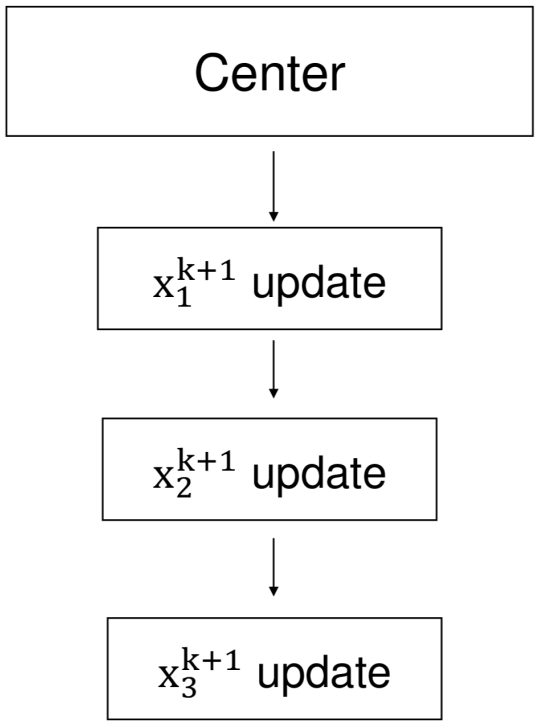
Augmented Lagrangian Function

An effective distributed algorithm for (2.3) can be designed based on the alternating direction method of multipliers (ADMM) using the following augmented Lagrangian:

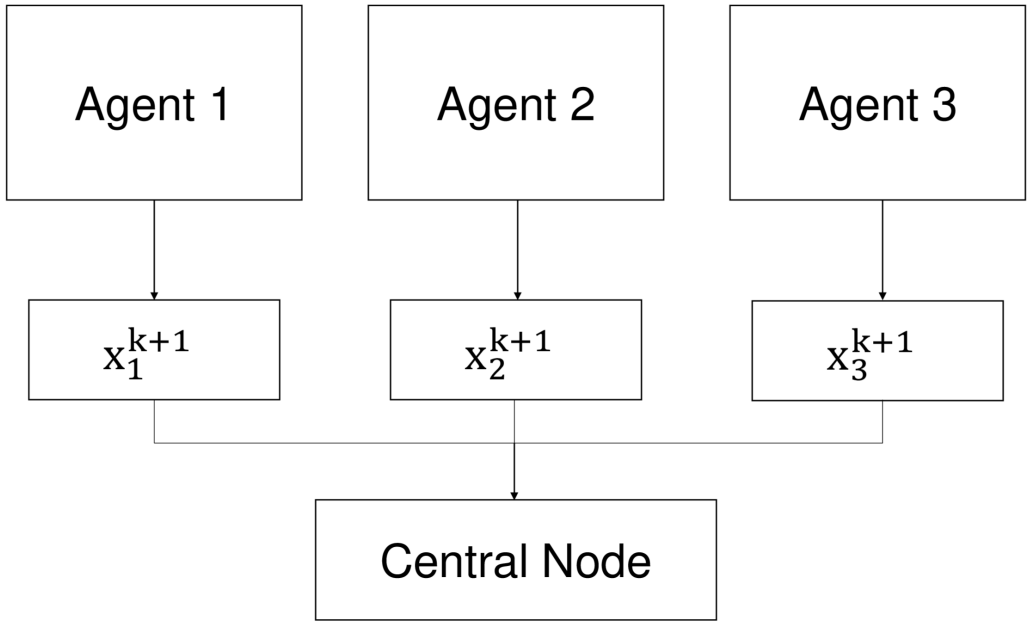
$$\mathcal{L}_\rho(x_1, \cdots, x_N; \lambda) := \sum_{i=1}^{N} f_i(x_i) - \left\langle \lambda, \sum_{i=1}^{N} A_i x_i - c \right\rangle + \frac{\rho}{2} \left\| \sum_{i=1}^{N} A_i x_i - c \right\|^2,$$

where $\rho > 0$ is a penalty parameter.

Two Types of ADMM Updates



(a) Gauss-Seidel-type



(b) Jacobi-type

Two Types of ADMM Updates: Gauss-Seidel vs Jacobi

Gauss-Seidel-type

$$\begin{cases} x_1^{k+1} := \arg\min_{x_1} \mathcal{L}_\rho(x_1, x_2^k, \cdots, x_N^k; \lambda^k) \\ x_2^{k+1} := \arg\min_{x_2} \mathcal{L}_\rho(x_1^{k+1}, x_2, \cdots, x_N^k; \lambda^k) \\ \vdots \\ x_N^{k+1} := \arg\min_{x_N} \mathcal{L}_\rho(x_1^{k+1}, \cdots, x_{N-1}^{k+1}, x_N; \lambda^k) \\ \lambda^{k+1} := \lambda^k - \rho\left(\sum_{i=1}^N A_i x_i^{k+1} - b\right) \end{cases}$$

Jacobi-type

$$\begin{cases} x_1^{k+1} := \arg\min_{x_1} \mathcal{L}_\rho(x_1, x_2^k, \cdots, x_N^k; \lambda^k) \\ x_2^{k+1} := \arg\min_{x_2} \mathcal{L}_\rho(x_1^k, x_2, \cdots, x_N^k; \lambda^k) \\ \vdots \\ x_N^{k+1} := \arg\min_{x_N} \mathcal{L}_\rho(x_1^k, \cdots, x_{N-1}^k, x_N; \lambda^k) \\ \lambda^{k+1} := \lambda^k - \rho\left(\sum_{i=1}^N A_i x_i^{k+1} - b\right) \end{cases}$$

Gauss-Seidel-type vs. Jacobi-type

	Gauss-Seidel type	Jacobi type
Update Method	sequential	parallel (simultaneous)
Update Source	Latest values	Previous iteration values
Convergence Speed	Generally faster	Generally slower
Parallelism	not parallelizable	parallelizable
Convergence	stable	Can diverge, especially when $N = 2$ [2].

[2] He, B., Hou, L., & Yuan, X. (2015). *On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming*. SIAM Journal on Optimization, 25(4), 2274–2312.

Jacobi-Proximal ADMM

To address this issue, Deng et al. [1] introduced the Jacobi-Proximal ADMM, which incorporates the following modifications:

1. A **proximal term** $\frac{1}{2}\|x_i - x_i^k\|_{P_i}^2$ for each x_i-subproblem, where P_i is a positive semi-definite matrix.
2. A **damping parameter** $\gamma > 0$ for updating λ in order to guarantee convergence of the method.

[1] Deng, W., Lai, M.-J., Peng, Z., & Yin, W. (2017). *Parallel multi-block ADMM with $O(1/k)$ convergence*. Journal of Scientific Computing, 71, 712–736.

Jacobi-type ADMM vs Jacobi-Proximal ADMM

Jacobi-type ADMM

$$\begin{cases} x_1^{k+1} := \arg\min_{x_1} \mathcal{L}_\rho(x_1, x_2^k, \cdots, x_N^k; \lambda^k) \\ x_2^{k+1} := \arg\min_{x_2} \mathcal{L}_\rho(x_1^k, x_2, \cdots, x_N^k; \lambda^k) \\ \vdots \\ x_N^{k+1} := \arg\min_{x_N} \mathcal{L}_\rho(x_1^k, \cdots, x_{N-1}^k, x_N; \lambda^k) \\ \lambda^{k+1} := \lambda^k - \rho\left(\sum_{i=1}^N A_i x_i^{k+1} - b\right) \end{cases}$$

Jacobi-Proximal ADMM

$$\begin{cases} x_1^{k+1} := \arg\min_{x_1} \mathcal{L}_\rho(x_1, x_2^k, \cdots, x_N^k; \lambda^k) + \frac{1}{2}\|x_1 - x_1^k\|_{P_1}^2 \\ x_2^{k+1} := \arg\min_{x_2} \mathcal{L}_\rho(x_1^k, x_2, \cdots, x_N^k; \lambda^k) + \frac{1}{2}\|x_2 - x_2^k\|_{P_2}^2 \\ \vdots \\ x_N^{k+1} := \arg\min_{x_N} \mathcal{L}_\rho(x_1^k, \cdots, x_{N-1}^k, x_N; \lambda^k) + \frac{1}{2}\|x_N - x_N^k\|_{P_N}^2 \\ \lambda^{k+1} := \lambda^k - \gamma\rho\left(\sum_{i=1}^N A_i x_i^{k+1} - b\right) \end{cases}$$

Jacobi-Proximal ADMM

Algorithm 1: Jacobi-Proximal ADMM

Input: x_i^0 $(i = 1, 2, \cdots, N)$ and λ^0;

for $k = 0, 1, \cdots$ **do**

Update x_i for $i = 1, \cdots, N$ in parallel by:

$$x_i^{k+1} = \arg\min_{x_i} \left\{ f_i(x_i) + \frac{\rho}{2} \left\| A_i x_i + \sum_{j \neq i} A_j x_j^k - c - \frac{\lambda^k}{\rho} \right\|_2^2 + \frac{1}{2} \|x_i - x_i^k\|_{P_i}^2 \right\}; \quad (3.1)$$

$$\text{Update } \lambda^{k+1} = \lambda^k - \gamma\rho \left(\sum_{i=1}^{N} A_i x_i^{k+1} - c \right). \quad (3.2)$$

Linear Convergence Results for ADMM

	Ref.	Convexity	Regularity	Rank Conditions	Rate
Gauss-Seidel ADMM for $N \geq 3$	[4]	$f_2, \ldots, f_N$: SC	f_N: L-smooth	A_N: full row rank	linear
		$f_1, \ldots, f_N$: SC	$f_1, \ldots, f_N$: L-smooth	—	
		$f_2, \ldots, f_N$: SC	$f_1, \ldots, f_N$: L-smooth	A_1: full column rank	
Gauss-Sediel ADMM for $N \geq 3$	[3]	$f_i(x_i) = g_i(E_i x_i) + h_i(x_i)$ g_i: strictly C, h_i: C	g_i: L-smooth	E_i: full column rank	linear
Jacobi-Proximal ADMM for $N \geq 3$	[1]	$f_1, \ldots, f_N$: C	—	A_i: full column rank	sub-linear
Jacobi-Proximal ADMM for $N \geq 3$	This work	$f_1, \ldots, f_N$: SC	$f_1, \ldots, f_N$: L-smooth	$[A_1^T; A_2^T; \cdots; A_N^T]$ full column rank	linear

[4] Lin, T., Ma, S., & Zhang, S. (2015). On the global linear convergence of the ADMM with multiblock variables. SIAM Journal on Optimization, 25(3), 1478–1497.
[3] Hong, M., & Luo, Z.-Q. (2017). *On the linear convergence of the alternating direction method of multipliers*. Mathematical Programming, 162(1), 165–199.
[1] Deng, W., Lai, M.-J., Peng, Z., & Yin, W. (2017). *Parallel multi-block ADMM with $O(1/k)$ convergence*. Journal of Scientific Computing, 71, 712–736.

Assumptions

Assumption

The functions $f_i : \mathbb{R}^{n_i} \to (-\infty, +\infty]$ *are closed proper convex for* $1 \leq i \leq N$.

Assumption

We assume that $x^* = (x_1^*, \cdots, x_N^*) \in \mathbb{R}^n$ *is a minimizer of problem* (2.3) *companied with a multiplier* $\lambda^* \in \mathbb{R}^m$ *satisfying the following KKT condition*

$$\nabla f_i(x_i^*) = A_i^T \lambda^* \quad for\ i = 1, \cdots, N, \tag{4.1}$$

$$Ax^* = \sum_{i=1}^{N} A_i x_i^* = c. \tag{4.2}$$

Assumptions

Assumption

There are positive constants L_i such that for all $x, y \in \mathbf{R}^{n_i}$,

$$\|\nabla f_i(x) - \nabla f_i(y)\| \leq L_i \|x - y\| \quad for\ 1 \leq i \leq N.$$

Assumption

The function $f_i : \mathbb{R}^{n_i} \to (-\infty, +\infty]$ is α-strongly convex for some $\alpha > 0$, i.e.,

$$f_i(y) \geq f_i(x) + \langle y - x, \nabla f_i(x) \rangle + \alpha \|x - y\|^2$$

for all $x, y \in \mathbb{R}^{n_i}$.

Assumption

We assume that the matrix $\left[A_1^T;\ A_2^T;\ \cdots;\ A_N^T\right] \in \mathbb{R}^{n \times m}$ with $n = n_1 + \cdots + n_N$ has full column rank.

Main Result

Theorem

Suppose Assumptions 4.1-4.5 hold. Choose any $s > 0$ such that

$$s\left[\rho^2 D\|A_i\| + \frac{L}{N}\right] < \frac{\alpha}{2N}, \qquad \forall 1 \le i \le N,$$

and suppose that $0 < \gamma < 2$, $\rho > 0$ are chosen so that there exist values $\xi_i > 0$ such that

$$\begin{cases} \rho A_i^T A_i + P_i - 8s(\rho A_i^T A_i + P_i)^T(\rho A_i^T A_i + P_i) - \frac{\rho}{\xi_i} A_i^T A_i \succ 0, \\ \sum_{i=1}^N \xi_i < 2 - \gamma. \end{cases} \tag{4.3}$$

Let $\sigma = \max\left\{1 - 2\gamma\rho s c_A^2, \mu_s\right\}$. Here $c_A > 0$ and $\mu_s \in (0,1)$ are constants satisfying

$$\begin{cases} c_A^2 I_m \preceq \sum_{i=1}^N A_i A_i^T, \\ (\rho + 4Ns\rho^2 D)A_i^T A_i + P_i \preceq \mu_s(\rho A_i^T A_i + P_i + 2(\alpha - Ls)I_{n_i}, \end{cases}$$

Then we have

$$\phi(x^{k+1}, \lambda^{k+1}) \le \sigma\phi(x^k, \lambda^k), \tag{4.4}$$

and hence

$$\phi(x^k, \lambda^k) = O(\sigma^k).$$

Lemmas

To prove the main theorem, we prepare several lemmas.

Lemma

Suppose Assumption 4.5 holds. Consider matrices $A_i \in \mathbb{R}^{m \times n_i}$ for $1 \leq i \leq N$. Then there exists $c_A > 0$ such that

$$\left(\sum_{i=1}^{N} \|A_i^T \lambda\|^2 \right)^{\frac{1}{2}} \geq c_A \|\lambda\| \tag{4.5}$$

for all $\lambda \in \mathbb{R}^m$.

(Assumption 4.5: The matrix $\left[A_1^T;\ A_2^T;\ \cdots;\ A_N^T\right] \in \mathbb{R}^{n \times m}$ with $n = n_1 + \cdots + n_N$ has full column rank.)

Lemmas

By the strongly convexity of the objective function, we have the following lemma.

Lemma

Suppose Assumptions 4.1, 4.2, and 4.4 hold. Then it holds that

$$\begin{aligned}
&\frac{1}{2\gamma\rho}\left\|\lambda^k-\lambda^*\right\|^2+\frac{1}{2}\sum_{i=1}^{N}\left\|x_i^k-x_i^*\right\|^2_{\rho A_i^T A_i+P_i}\\
&\geq \alpha\sum_{i=1}^{N}\left\|x_i^{k+1}-x_i^*\right\|^2+\frac{1}{2\gamma\rho}\left\|\lambda^{k+1}-\lambda^*\right\|^2 \qquad (4.6)\\
&+\frac{1}{2}\sum_{i=1}^{N}\left\|x_i^{k+1}-x_i^*\right\|^2_{\rho A_i^T A_i+P_i}+H_k
\end{aligned}$$

where $H_k := \frac{1}{2}\sum_{i=1}^{N}\left\|x_i^{k+1}-x_i^k\right\|^2_{\rho A_i^T A_i+P_i}+\frac{2-\gamma}{2\gamma^2\rho}\left\|\lambda^{k+1}-\lambda^k\right\|^2+\frac{1}{\gamma}\left\langle\lambda^k-\lambda^{k+1},A\left(x^k-x^{k+1}\right)\right\rangle.$

We notice that the term $\alpha\sum_{i=1}^{N}\left\|x_i^{k+1}-x_i^*\right\|^2$ in the inequality (4.6) is a gain for the strongly convexity of the costs.

Lemmas

Lemma

Suppose Assumption 4.1-4.5 hold. Let $L = \max_{1 \le i \le N} L_i^2$ *and* $D = \sum_{i=1}^{N} \left\| A_i^T \right\|^2$. *Then it holds that*

$$
\begin{aligned}
& c_A^2 \left\| \lambda^k - \lambda^* \right\|^2 \\
& \le 2L \sum_{i=1}^{N} \left\| x_i^{k+1} - x_i^* \right\|^2 + 4\rho^2 D \left\| \sum_{j=1}^{N} A_j \left(x_j^k - x_j^* \right) \right\|^2 \\
& \quad + 4 \sum_{i=1}^{N} \left\| \left(\rho A_i^T A_i + P_i \right) \left(x_i^{k+1} - x_i^k \right) \right\|^2 .
\end{aligned}
\tag{4.7}
$$

Propositions

We derive the following result using Lemma 4.8 and Lemma 4.9.

Proposition

Suppose Assumptions 4.1-4.5 hold. For $s > 0$, we have the following inequality

$$\begin{aligned}&\left(\frac{1}{2\gamma\rho} - sc_A^2\right)\left\|\lambda^k - \lambda^*\right\|^2 + \frac{1}{2}\sum_{i=1}^{N}\left\|x_i^k - x_i^*\right\|^2_{(\rho+4Ns\rho^2D)A_i^TA_i+P_i} \\ &\geq \frac{1}{2\gamma\rho}\left\|\lambda^{k+1} - \lambda^*\right\|^2 + \frac{1}{2}\sum_{i=1}^{N}\left\|x_i^{k+1} - x_i^*\right\|^2_{\rho A_i^TA_i+P_i+2(\alpha-2Ls)I_{n_i}} + G_k,\end{aligned} \tag{4.8}$$

where
$H_k := \frac{1}{2}\sum_{i=1}^{N}\left\|x_i^{k+1} - x_i^k\right\|^2_{\rho A_i^TA_i+P_i} + \frac{2-\gamma}{2\gamma^2\rho}\left\|\lambda^{k+1} - \lambda^k\right\|^2 + \frac{1}{\gamma}\left\langle\lambda^k - \lambda^{k+1}, A\left(x^k - x^{k+1}\right)\right\rangle$
and $G_k := H_k - 4s\sum_{i=1}^{N}\left\|\left(\rho A_i^TA_i + P_i\right)\left(x_i^{k+1} - x_i^k\right)\right\|^2$.

Lemmas

Lemma

If $s > 0$ satisfies

$$s\left[\rho^2 D\|A_i\|^2 + \frac{L}{N}\right] < \frac{\alpha}{2N}, \tag{4.9}$$

there exists $\mu_s \in (0,1)$ such that

$$\|x\|^2_{(\rho+4Ns\rho^2 D)A_i^T A_i + P_i} \leq \mu_s \|x\|^2_{\rho A_i^T A_i + P_i + 2(\alpha - 2Ls) I_{n_i}} \quad \forall\, x \in \mathbb{R}^{n_i}. \tag{4.10}$$

Lemma

If one chooses some $\xi_i > 0$ such that $0 < \gamma < 2$, $\rho > 0$, and $s > 0$ satisfies the following conditions:

$$\begin{cases} \rho A_i^T A_i + P_i - 8s(\rho A_i^T A_i + P_i)^T(\rho A_i^T A_i + P_i) - \frac{\rho}{\xi_i} A_i^T A_i \succ 0, \\ \sum_{i=1}^N \xi_i < 2 - \gamma, \end{cases}$$

then we have

$$\frac{1}{2}\sum_{i=1}^N \|x_i^{k+1} - x_i^k\|^2_{\rho A_i^T A_i + P_i} - 4s\sum_{i=1}^N \left\| \left(\rho A_i^T A_i + P_i\right)\left(x_i^{k+1} - x_i^k\right)\right\|^2 + \frac{2-\gamma}{2\gamma^2\rho}\|\lambda^{k+1} - \lambda^k\|^2 + \frac{1}{\gamma}\left\langle \lambda^k - \lambda^{k+1}, A\left(x^k - x^{k+1}\right)\right\rangle \geq 0.$$

Main Result

Theorem

Suppose Assumptions 4.1-4.5 hold. Choose any $s > 0$ such that

$$s\left[\rho^2 D\|A_i\| + \frac{L}{N}\right] < \frac{\alpha}{2N}, \qquad \forall 1 \le i \le N,$$

and suppose that $0 < \gamma < 2$, $\rho > 0$ are chosen so that there exist values $\xi_i > 0$ such that

$$\begin{cases} \rho A_i^T A_i + P_i - 8s(\rho A_i^T A_i + P_i)^T(\rho A_i^T A_i + P_i) - \frac{\rho}{\xi_i} A_i^T A_i \succ 0, \\ \sum_{i=1}^N \xi_i < 2 - \gamma. \end{cases} \tag{4.11}$$

Let $\sigma = \max\left\{1 - 2\gamma\rho s c_A^2, \mu_s\right\}$. Here $c_A > 0$ and $\mu_s \in (0,1)$ are constants satisfying

$$\begin{cases} c_A^2 I_m \preceq \sum_{i=1}^N A_i A_i^T, \\ (\rho + 4Ns\rho^2 D)A_i^T A_i + P_i \preceq \mu_s(\rho A_i^T A_i + P_i + 2(\alpha - Ls)I_{n_i}, \end{cases}$$

Then we have

$$\phi(x^{k+1}, \lambda^{k+1}) \le \sigma\phi(x^k, \lambda^k), \tag{4.12}$$

and hence

$$\phi(x^k, \lambda^k) = O(\sigma^k).$$

Main Theorem

We are now ready to show our main convergence and rate results.

Proof of Theorem 4.6.

By applying Lemma 4.11 and Lemma 4.12 to the inequality of Proposition 4.10, we get

$$\left(\frac{1}{2\gamma\rho} - sc_A^2\right)\left\|\lambda^k - \lambda^*\right\|^2 + \frac{\mu_s}{2}\sum_{i=1}^{N}\left\|x_i^k - x_i^*\right\|^2_{\rho A_i^T A_i + P_i + 2(\alpha - 2Ls)I_{n_i}}$$

$$\geq \frac{1}{2\gamma\rho}\left\|\lambda^{k+1} - \lambda^*\right\|^2 + \frac{1}{2}\sum_{i=1}^{N}\left\|x_i^{k+1} - x_i^*\right\|^2_{\rho A_i^T A_i + P_i + 2(\alpha - 2Ls)I_{n_i}}.$$

This gives the following inequality

$$\sigma\left(\frac{1}{2\gamma\rho}\left\|\lambda^k - \lambda^*\right\|^2 + \frac{1}{2}\sum_{i=1}^{N}\left\|x_i^k - x_i^*\right\|^2_{\rho A_i^T A_i + P_i + 2(\alpha - 2Ls)I_{n_i}}\right)$$

$$\geq \frac{1}{2\gamma\rho}\left\|\lambda^{k+1} - \lambda^*\right\|^2 + \frac{1}{2}\sum_{i=1}^{N}\left\|x_i^{k+1} - x_i^*\right\|^2_{\rho A_i^T A_i + P_i + 2(\alpha - 2Ls)I_{n_i}},$$

where $\sigma = \max\left\{1 - 2\gamma\rho sc_A^2, \mu_s\right\}$. The proof is done. □

Numerical Experiments: The LCQP Model

We present numerical tests of Algorithm 1 supporting the linear convergence property proved in Theorem 4.6. Here we consider the following LCQP model:

$$\begin{aligned} &\min \sum_{i=1}^{N} f_i(x_i) \\ &\text{s.t.} \sum_{i=1}^{N} A_i x_i = c \end{aligned} \tag{5.1}$$

where $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$ is given by $f_i(x_i) = \frac{1}{2} x_i^T H_i x_i + x_i^T q_i$ with $H_i \in \mathbb{S}_+^{n_i}$, $q_i \in \mathbb{R}^{n_i}$, $A_i \in \mathbb{R}^{m \times n_i}$ and $c \in \mathbb{R}^m$.

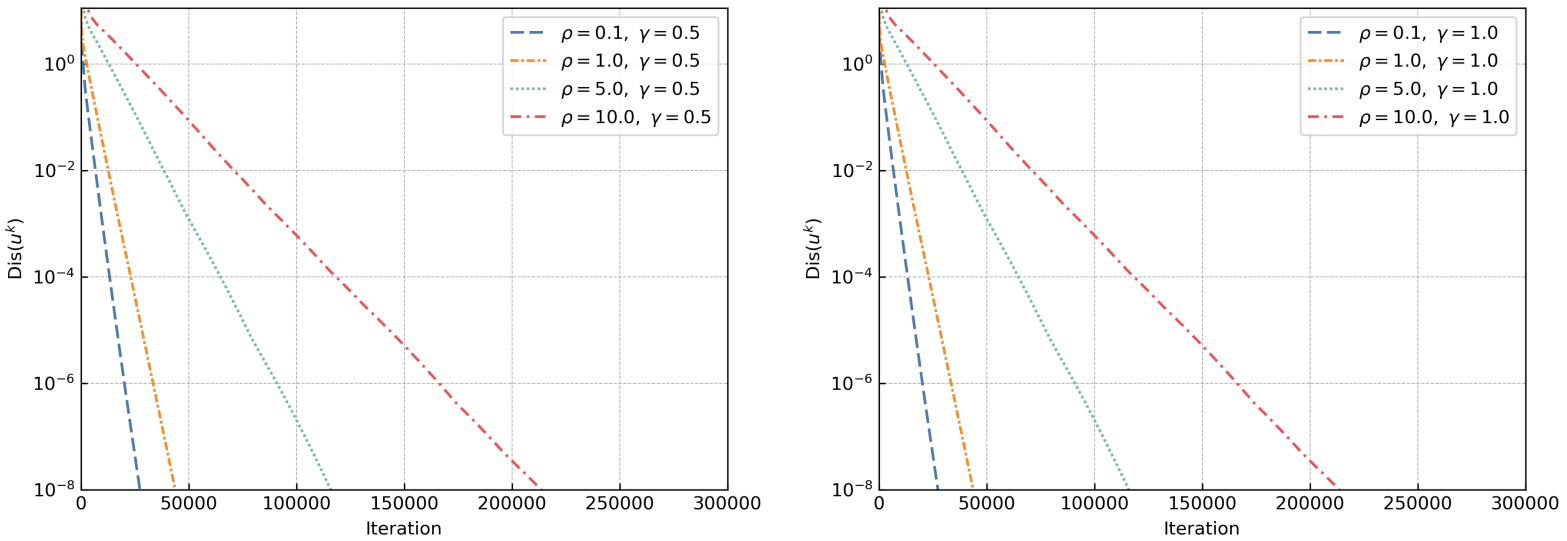
Numerical Experiments: The LCQP Model

We conduct experiments with the damping parameter $\gamma \in \{0.5, 1.0, 1.5, 1.9\}$, and the penalty parameter $\rho \in \{0.1, 1, 5, 10\}$ in Algorithm 1. We define the error metric at each k-th iteration for $u^k = \left(x_1^k, \ldots, x_N^k; \lambda^k\right)$ as follows:

$$\text{dis}(u^k) := \max\left\{\left\|x_1^k - x_1^*\right\|, \ldots, \left\|x_N^k - x_N^*\right\|, \left\|\lambda^k - \lambda^*\right\|\right\}. \tag{5.2}$$

Numerical Experiments: The LCQP Model

We find that the value $\text{dis}(u^k)$ decreases to zero linearly as expected in Theorem 4.6.



(a) $\rho = \{0.1, 1.0, 5.0, 10.0\}$, $\gamma = 0.5$

(b) $\rho = \{0.1, 1.0, 5.0, 10.0\}$, $\gamma = 1.0$

Figure: Experimental results on 10-block LCQP model for fixed γ under varying ρ.

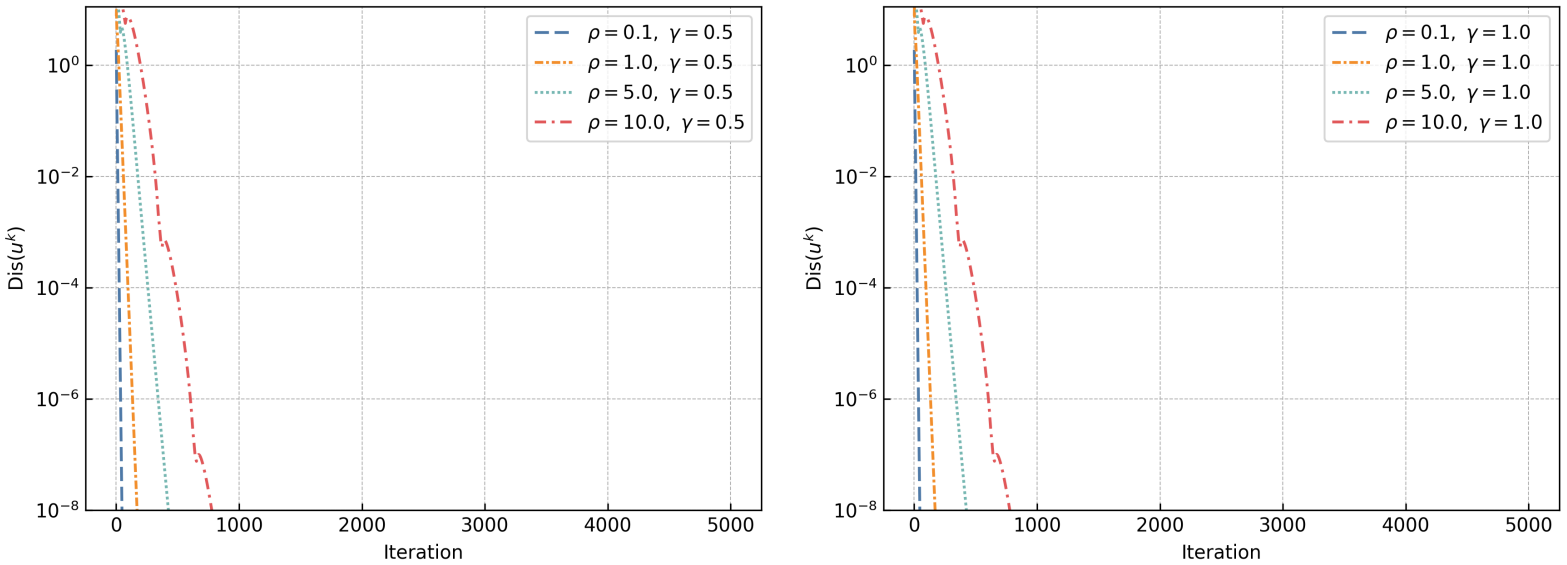
Numerical Experiments: The Optimal Resource Allocation Problem

In this example, we consider the following case:

$$\min \sum_{i=1}^{N} f_i(x_i)$$
$$\text{s.t.} \sum_{i=1}^{N} x_i = 0$$

where $f_i : \mathbb{R} \to \mathbb{R}$ is given as $f_i(x_i) = (1/2)a_i(x_i - c_i)^2 + \log[1 + exp(b_i(x_i - d_i))]$ with the coefficients a_i, b_i, c_i, d_i generated randomly with uniform distributions on $[0, 2]$, $[-2, 2]$, $[-10, 10]$, $[-10, 10]$ respectively.

Numerical Experiments: The Optimal Resource Allocation Problem



(a) $\rho = \{0.1, 1.0, 5.0, 10.0\}$, $\gamma = 0.5$

(b) $\rho = \{0.1, 1.0, 5.0, 10.0\}$, $\gamma = 1.0$

Figure: Experimental results on 6-block problem for the optimal resource allocation model with a fixed γ under varying ρ.


Wei Deng, Ming-Jun Lai, Zhimin Peng, and Wotao Yin.
Parallel multi-block admm with o (1/k) convergence.
Journal of Scientific Computing, 71:712–736, 2017.


Bingsheng He, Liusheng Hou, and Xiaoming Yuan.
On full jacobian decomposition of the augmented lagrangian method for separable convex programming.
SIAM Journal on Optimization, 25(4):2274–2312, 2015.


Mingyi Hong and Zhi-Quan Luo.
On the linear convergence of the alternating direction method of multipliers.
Mathematical Programming, 162(1):165–199, 2017.


Tianyi Lin, Shiqian Ma, and Shuzhong Zhang.
On the global linear convergence of the admm with multiblock variables.
SIAM Journal on Optimization, 25(3):1478–1497, 2015.


Naum Zuselevich Shor.
Minimization methods for non-differentiable functions, volume 3.
Springer Science & Business Media, 2012.

Thank You