Zicong XIA, Yang LIU, Wenlian LU, Weihua GUI, 2023. Matrix-valued distributed stochastic optimization with constraints. *Frontiers of Information Technology & Electronic Engineering,* 24(9):1239-1252. https://doi.org/10.1631/FITEE.2200381

Matrix-valued distributed stochastic optimization with constraints

Key words: Distributed optimization; Matrix-valued optimization; Stochastic optimization; Penalty method; Gossip model

Corresponding author: Yang LIU
E-mail: liuyang@zjnu.edu.cn
ORCID: https://orcid.org/0000-0003-3761-0104

Motivation

1. As an optimization technique with enhanced safety and parallel computation capability, distributed optimization is of great significance.

2. Due to environmental interference or technical errors, stochastic optimization has always been a hot topic in optimization theories. The studies on distributed stochastic optimization techniques are also open.

3. Existing works on distributed optimization are vector-variable systems. Hence, the computation time relies on the dimension of the state in the optimization problem. However, if the dimension is high, the methods converge slowly. In many key areas (e.g., image processing), matrix-valued optimization models can overcome this difficulty.

Contributions

1. An auxiliary function is proposed to analyze several properties of the matrix-valued functions. Many common properties for vector-valued optimization methods are proposed in a matrix-valued fashion.
2. A selection principle of the penalty functions and the penalty gains is derived. Based on the selection principle, an exact penalty method is proposed for transforming a matrix-valued optimization problem with inequality and equality constraints into a problem without inequality or equality constraints.
3. A distributed optimization algorithm based on a gossip model is developed for solving the matrix-valued distributed stochastic optimization, and its convergence is analyzed. Two numerical examples are provided to illustrate the efficiency of the proposed algorithm for solving matrix-valued distributed stochastic optimization problems.

Matrix-valued optimization

To extend the results to matrix-valued domains, several common properties of vector-valued optimization methods are reformulated in a matrix-valued manner.

Definition 1 (L-Lipschitz continuity) $f: \mathbb{R}^{n\times m} \to \mathbb{R}$ is said to be L-Lipschitz continuous if $\forall \boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{n\times m}$, $\exists L > 0$, such that $|f(\boldsymbol{X}) - f(\boldsymbol{Y})| \le L\|\boldsymbol{X} - \boldsymbol{Y}\|_{\mathrm{F}}$, where L is a Lipschitz constant.

Definition 2 (l-smoothness) $f: \mathbb{R}^{n\times m} \to \mathbb{R}$ is said to be l-smooth if $\forall \boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{n\times m}$, $\exists l > 0$, such that $\|\nabla f(\boldsymbol{X}) - \nabla f(\boldsymbol{Y})\|_{\mathrm{F}} \le l\|\boldsymbol{X} - \boldsymbol{Y}\|_{\mathrm{F}}$.

Definition 3 (μ-strong convexity) $f: \mathbb{R}^{n\times m} \to \mathbb{R}$ is said to be μ-strongly convex if $\forall \boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{n\times m}$, $\exists \mu > 0$, such that $f(\boldsymbol{Y}) \ge f(\boldsymbol{X}) + \mathrm{tr}((\nabla f(\boldsymbol{X}))^{\mathrm{T}}(\boldsymbol{Y} - \boldsymbol{X})) + \mu\|\boldsymbol{Y} - \boldsymbol{X}\|_{\mathrm{F}}^2/2$.

Note that Definition 3 defines the convexity of f if $\mu = 0$.

Definition 4 For any convex function $f: \mathbb{R}^{n\times m} \to \mathbb{R}$, the subdifferential of f with respect to $\boldsymbol{X}$ is defined by

$$\partial f(\boldsymbol{X}) = \left\{ \boldsymbol{G} | f(\boldsymbol{Y}) \ge f(\boldsymbol{X}) + \mathrm{tr}\left(\boldsymbol{G}^{\mathrm{T}}(\boldsymbol{Y} - \boldsymbol{X})\right) \right\}.$$

In addition, $\boldsymbol{G} \in \partial f(\boldsymbol{X})$ is called a subgradient of f at $\boldsymbol{X}$.

Lemma 1 Assume $f: \mathbb{R}^{n\times m} \to \mathbb{R}$ and $\alpha: \mathbb{R}^{nm} \to \mathbb{R}$ with $\mathrm{vec}(\boldsymbol{X}) = \boldsymbol{x}$. $\forall \boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{n\times m}$, we have the following statements:

(1) $\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}) = (\mathrm{vec}(\boldsymbol{X}))^{\mathrm{T}}\mathrm{vec}(\boldsymbol{Y})$;

(2) $\|\boldsymbol{X}\|_{\mathrm{F}} = \|\boldsymbol{x}\|$;

(3) $\forall \eta > 0$, $\|\boldsymbol{X} + \boldsymbol{Y}\|_{\mathrm{F}} \le (1+\eta)\|\boldsymbol{X}\|_{\mathrm{F}} + (1 + \eta^{-1})\|\boldsymbol{Y}\|_{\mathrm{F}}$;

(4) $f(\boldsymbol{X})$ is l-Lipschitz continuous if and only if $\alpha(\boldsymbol{x})$ is l-Lipschitz continuous;

(5) $f(\boldsymbol{X})$ is l-smooth if and only if $\alpha(\boldsymbol{x})$ is l-smooth;

(6) $f(\boldsymbol{X})$ is μ-strongly convex if and only if $\alpha(\boldsymbol{x})$ is μ-strongly convex.

Lemma 2 If $f(\boldsymbol{X})$ is μ-strongly convex and bounded with M', and $\nabla f(\boldsymbol{X})$ is bounded with M'', then $f(\boldsymbol{X})$ is $2M'\mu/M''$-Lipschitz continuous.

Lemma 3 If $l\|\boldsymbol{X}\|_{\mathrm{F}}^2/2 - f(\boldsymbol{X})$ is convex, then $f(\boldsymbol{X})$ is l-pseudo smooth.

Stochastic optimization models and penalty function methods

Consider the following optimization model:

Stochastic optimization model

$$\min \sum_{i=1}^{N} f_i(X) = \sum_{i=1}^{N} \mathbb{E}_{\xi_i \in \mathcal{D}_i} F_i(X, \xi_i)$$
$$\text{s.t. } g(X) \le 0,\ h(X) = 0,$$

Stochastic optimization model in a distributed settings

$$\min \sum_{i=1}^{N} f_i(X_i) = \sum_{i=1}^{N} \mathbb{E}_{\xi_i \in \mathcal{D}_i} F_i(X_i, \xi_i)$$
$$\text{s.t. } \begin{cases} X_i = X_j, & i, j \in \mathcal{I}_N, \\ g(X_i) \le 0, & i \in \mathcal{I}_N, \\ h(X_i) = 0, & i \in \mathcal{I}_N. \end{cases}$$

To deal with the constraints, a penalty function method is applied, and a selection principle of penalty functions and gains is proposed as follows:

Selection principle 1 Penalty gain c ($c > 0$) and penalty function $\tau_S(X) : \mathbb{R}^{n\times m} \to \mathbb{R}$ satisfy the following conditions:

(1) $\forall X \in \mathbb{R}^{n\times m}$, $f(X) + c\tau_S(X) \ge f(P_S(X))$;
(2) $\forall X \in \mathbb{R}^{n\times m}$, $\tau_S(X) \ge 0$;
(3) $\forall X \in S$, $\tau_S(X) = 0$.

Model via penalty function methods

$$\min \sum_{i=1}^{N} \tilde{f}_i(X_i) \quad (3)$$
$$\text{s.t. } X_i = X_j,\ i, j \in \mathcal{I}_N,$$

where $\tilde{f}_i(X_i) = f_i(X_i) + P_{g_i}\mathcal{A}_i(X_i) + P_{h_i}\mathcal{B}_i(X_i)$ with $\mathcal{A}_i(X_i) = (g(X_i) + |g(X_i)|)^2$ and $\mathcal{B}_i(X_i) = |h(X_i)|^2$ for $i \in \mathcal{I}_N$.

Table 1 Comparison of existing works with this study

Reference	Matrix-valued	Distributed	Stochastic	Constraint
Huang et al. (2021)	✓	×	×	Bound constraints
Zhang et al. (2022)	✓	×	×	Equality constraints
Zhou et al. (2019), Xia ZC et al. (2021)	×	✓	×	Bound constraints
Koloskova et al. (2019)	×	✓	✓	None
This study	✓	✓	✓	Equality constraints; inequality constraints

Distributed stochastic optimization algorithm

A distributed stochastic optimization algorithm is established based on the gossip model, and it converges to an optimal solution with suitable step sizes.

Gossip model

$$X_i(t+1) = X_i(t) + \kappa \sum_{j \in \mathcal{N}_i} A(i,j)(X_j(k) - X_i(k))$$

Convergence analysis

Theorem 3 Under Assumption 2, for $p > 0$, Algorithm 1 for $\zeta(k) = 4/(\chi_1(a+k))$ with $a \geq 5/p$ converges at the rate

$$\sum_{i=1}^{N} f_i\left(X_{\text{avg}}(K)\right) - \sum_{i=1}^{N} f_i(X^*)$$
$$\leq \frac{\chi_1 a^3}{8S(K)} \|\bar{X}(0) - X^*\|_{\text{F}}^2 + \frac{K(K+2a)}{\chi_1 S(K)} \left(\frac{\bar{\sigma}^2}{N} + \chi_2\right)$$
$$+ \frac{320 K \chi_3 Q}{\chi_1^2 S(K) p^2},$$

where $X_{\text{avg}}(K) = \sum_{k=0}^{K-1} \omega(k)\bar{X}(k)/S(K)$ for $\omega(k) = (a+k)^2$, and $S(K) = \sum_{k=0}^{K-1} \omega(k) \geq K^3/3$.

Algorithm 1 Distributed stochastic gradient descent algorithm

1: **Initialization**
2: **Input**: $X_i(0)$, time-varying step $\zeta(k)$, total number of iterations K, and A
3: **For** $k = 1, 2, ..., K$
4: Sample $\xi_i(k)$, and calculate $\nabla F_i(X_i(k), \xi_i(k))$
5: Choose P_{g_i} and P_{h_i} satisfying

$$L_{g_i}(X_i(k))P_{g_i} + L_{h_i}(X_i(k))P_{h_i} \geq \frac{2m\mu}{M}$$

6: Choose $\mathcal{H}_i(k) \in P_{g_i}\partial\mathcal{A}(X_i(k)) + P_{h_i}\partial\mathcal{B}(X_i(k))$
7: Calculate

$$Y_i(k) = X_i(k) - \zeta(k)(\nabla F_i(X_i(k), \xi_i(k)) + \mathcal{H}_i(k)) \quad (5)$$

8: Calculate

$$X_i(k+1) = Y_i(k) + \kappa \sum_{j \in \mathcal{N}_i} A(i,j)(Y_j(k) - Y_i(k)) \quad (6)$$

9: **End for**
10: **Output**: $X_{\text{avg}}(K)$

Major results

The penalty function method deals with the equality and inequality constraints well.

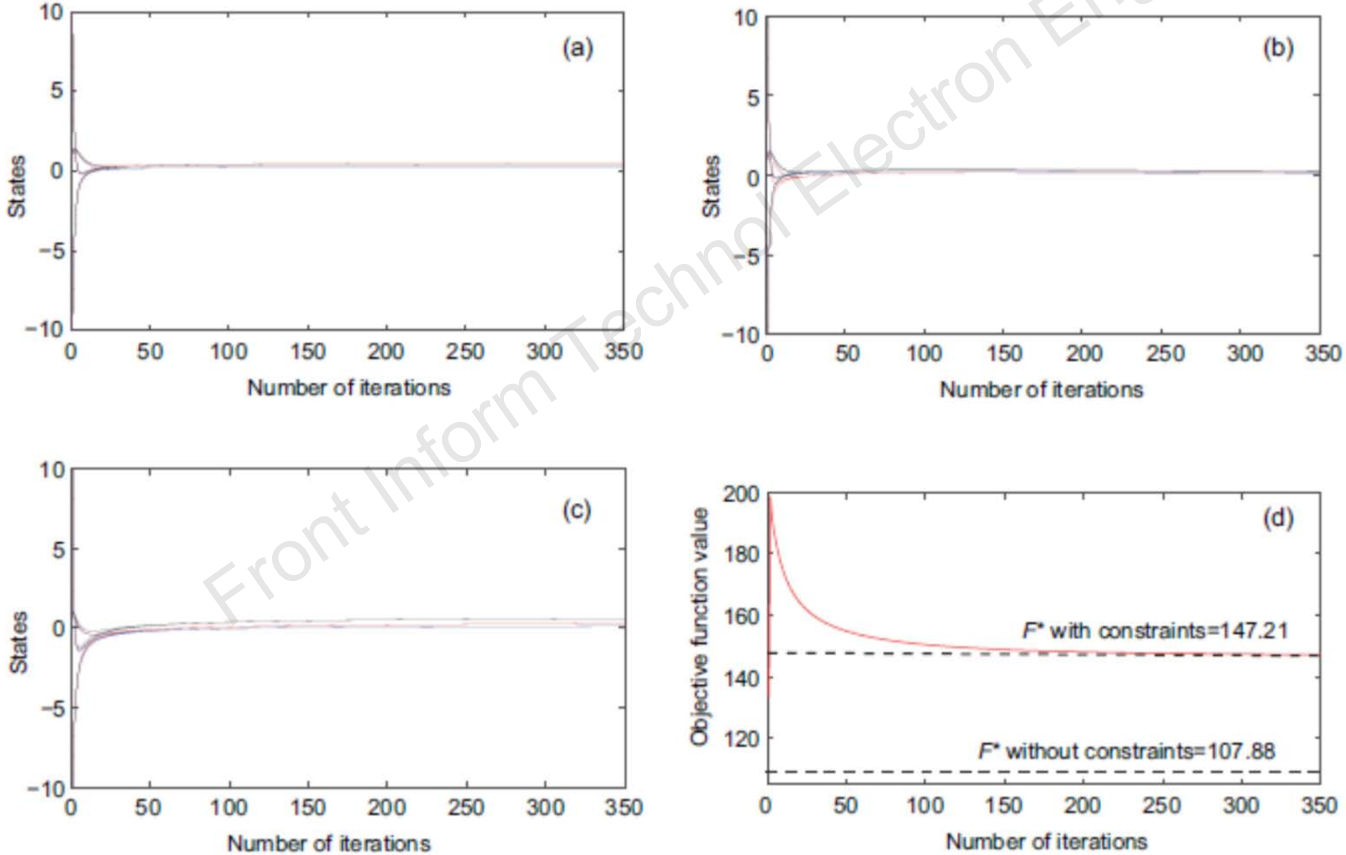


Fig. 1 Transient states of $X_k(1,i)$ (a), $X_k(2,i)$ (b), $X_k(3,i)$ (c), and the transient values of the objective function (d) in Example 1 ($k,\ i \in \{1,2,3\}$)

Major results (Cont'd)

The distributed stochastic optimization algorithm converges to an optimal solution with suitable step sizes, and it can deal with the stochastic variables well.

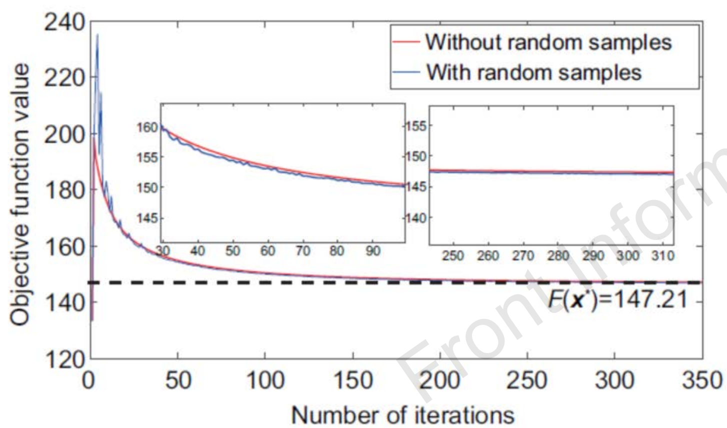


Fig. 2 Transient values of the objective function in Example 1

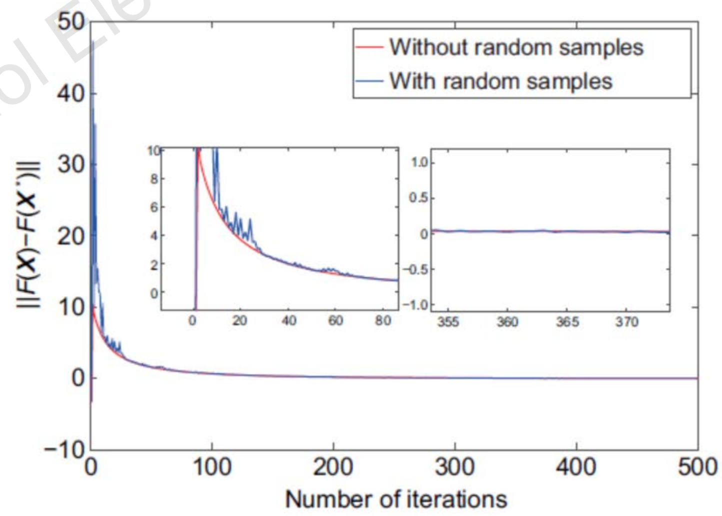
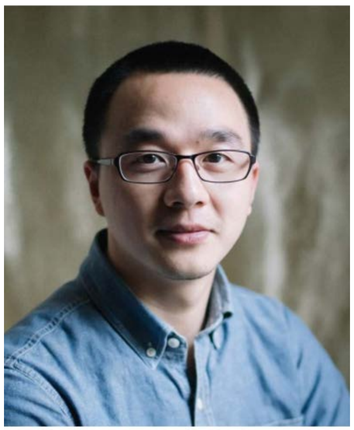


Fig. 3 Errors between the transient values of the objective function obtained by Algorithm 1 and the optimal values of the objective function in Example 2

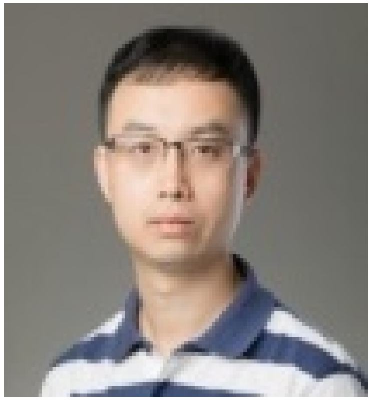
Conclusions

1. We have focused on a special constrained optimization called matrix-valued distributed stochastic optimization subject to inequality and equality constraints.
2. We have adopted an exact penalty for the handling of the constraints. Based on a gossip model, we have developed a distributed stochastic gradient descent algorithm and analyzed its stability.
3. Two illustrative examples have been provided to explain the validity of the exact penalty method and the optimization method.

Zicong XIA received the BS degree in mathematics and MS degree from Zhejiang Normal University, Jinhua, China, in 2019 and 2023, respectively. He is currently pursuing the PhD degree with the College of Mathematics, Southeast University, Nanjing, China. His current research interests include distributed optimization, multi-agent systems, and optimal control.



Yang LIU received the BS degree in mathematics from Zhejiang Normal University, Jinhua, China, in 2003, and the PhD degree from Tongji University, Shanghai, in 2008. He is currently a vice dean of Jinhua Intelligent Manufacturing Research Institute. He is also a Professor with the School of Mathematical Sciences and with Key Laboratory of Intelligent Education Technology and Application of Zhejiang Province, Zhejiang Normal University. He has authored over 100 publications and three books. He is an Associate Editor of *Neural Processing Letters* (Springer). He was recognized by Elsevier as a Most Cited Chinese Researcher in 2020–2022, and by Clarivate Analytics as a Highly Cited Researcher in 2019–2022. He was a supervisor of ICCM Best Thesis Award in 2016 and 2022. His research interests include logical systems, hybrid systems, and distributed optimization.



Wenlian LU received the BS degree in mathematics and the PhD degree in applied mathematics from Fudan University, Shanghai, China, in 2000 and 2005, respectively. He was a Postdoctoral Fellow with the Max Planck Institute for Mathematics in the Science, Leipzig, Germany, from 2005 to 2007. He was also a Marie-Curie International Incoming Research Fellow with the Department of Computer Sciences, University of Warwick, Coventry, UK, from 2012 to 2014. He is currently a Professor with the School of Mathematical Sciences and the Institute for Science and Technology of Brain-Inspired AI, Fudan University. His current research interests include neural networks, cyber security dynamics, computational systems biology, nonlinear dynamical systems, and complex systems. He served as an Associate Editor for the *IEEE Transactions on Neural Networks and Learning Systems*, from 2013 to 2019 and *Neurocomputing*, from 2010 to 2015.

Weihua GUI received the BE degree in electrical engineering and the MS degree in automatic control engineering from Central South University, Changsha, China, in 1976 and 1981, respectively. From 1986 to 1988, he was a Visiting Scholar with the University Duisburg–Essen, Duisburg, Germany. Since 1991, he has been a Full Professor with the School of Automation, Central South University. Since 2013, he has been also an Academician of the Chinese Academy of Engineering. His main research interests include the modeling and optimal control of CIPs, fault diagnoses, and distributed robust control.