# An Approximated ADMM based Algorithm for $\ell_1 - \ell_2$ Optimization Problem

Rui Lin* and Kazunori Hayashi†
* Graduate School of Informatics, Kyoto University, Kyoto, Japan
E-mail: ruilin@acs.i.kyoto-u.ac.jp Tel/Fax: +81-75-753-9690
† Graduate School of Informatics, Kyoto University, Kyoto, Japan
E-mail: hayashi.kazunori.4w@kyoto-u.ac.jp Tel/Fax: +81-75-753-9690

***Abstract*—Compressed sensing is a technique to recover a sparse vector from its underdetermined linear measurements. Since a naive $\ell_0$ optimization approach is hard to tackle due to the discreteness and the non-convexity of $\ell_0$ norm, a relaxed problem of the $\ell_1-\ell_2$ optimization is often employed for the reconstruction of the sparse vector especially when the measurement noise is not negligible. FISTA (fast iterative shrinkage-thresholding algorithm) is one of popular algorithms for the $\ell_1 - \ell_2$ optimization, and is known to achieve optimal convergence rate among the first order methods. Recently, the employment of optical circuits for various signal processing including deep neural networks has been considered intensively, but it is difficult to implement FISTA with the optical circuit, because it requires operations of divisions with a dynamic value in the algorithm. In this paper, assuming the implementation with the optical circuit, we propose an ADMM (alternating direction method of multipliers) based algorithm for the $\ell_1 - \ell_2$ optimization. It is true that an ADMM based algorithm for the $\ell_1-\ell_2$ optimization has been already proposed in the literature, but the proposed algorithm is derived with the different formulation from the existing method, and unlike the existing ADMM based algorithm, the proposed algorithm does not include the calculation of the inverse of a matrix. Computer simulation results demonstrate that the proposed algorithm can achieve comparable performance as FISTA or existing ADMM based algorithm while requiring no division operations and no matrix inversions.**

## I. Introduction

In order to improve the processing speed and to reduce the power consumption, the analog signal processing with optical circuits instead of conventional digital electric circuits has been drawing much attention recently [1]. In particular, since the product-sum operation, in other words, the product between a matrix and a vector, can be realized with the optical circuits, it is expected to speed up various signal processing algorithms or machine learning techniques with the implementation by the optical circuit [2].

Compressed sensing is a mathematical framework for reconstructing an unknown vector from its underdetermined linear measurements taking advantage of the sparsity of the unknown vector. When the linear measurement contains the observation noise, a method that minimizes the weighted sum of the $\ell_2$ norm data fidelity term and the $\ell_1$ sparse regularization term of the unknown vector is often used. Such a problem is called the $\ell_1 - \ell_2$ optimization problem, and ISTA (iterative shrinkage-thresholding algorithm) and its high-speed version FISTA (fast ISTA) are widely used to solve the optimization problem. In particular, FISTA is known to achieve the optimal convergence rate among the first order methods, which use only the value of the cost function and its derivative [3], [4], However, FISTA requires an operation of division with a dynamic value in the algorithm, which is difficult to implement with the optical circuit.

In this paper, assuming the implementation with the optical circuit, we propose an ADMM (alternating direction method of multipliers) based algorithm for the $\ell_1-\ell_2$ optimization. So far, an ADMM based algorithm for the $\ell_1 - \ell_2$ optimization problem has been already shown in [5], but this algorithm requires matrix inversion, which may difficult to realize with the optical circuit. In the proposed approach, we utilize a different formulation from [5], and the resultant proposed algorithm requires neither the operation of division with a dynamic value nor the matrix inversion, which is suited for the implementation with the optical circuit. Computer simulation results demonstrate that the proposed algorithm can achieve comparable performance as FISTA or existing ADMM based algorithm while requiring no division operations and no matrix inversions.

## II. $\ell_1 - \ell_2$ Optimization Problem and Existing Algorithms

In this paper, we consider an underdetermined linear measurement model for an unknown sparse vector $\boldsymbol{x} \in \mathbb{R}^N$ as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}, \tag{1}$$

where $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ is a known measurement matrix and $\boldsymbol{w} \in \mathbb{R}^M$ is the measurement noise. We also assume that $N > M$. In compressed sensing, the following $\ell_1 - \ell_2$ optimization problem is often utilized in order to reconstruct the sparse vector $\boldsymbol{x}$ from the information of $\boldsymbol{y}$ and $\boldsymbol{A}$:

$$\hat{\boldsymbol{x}}_{\ell_1\text{-}\ell_2} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \left(\frac{1}{2}||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}||_2^2 + \lambda||\boldsymbol{x}||_1\right). \tag{2}$$

Here, $\lambda > 0$ is a parameter that controls the balance between the data fidelity term $||\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}||_2^2$ and the sparse regularization term $||\boldsymbol{x}||_1$.

In order to solve the $\ell_1-\ell_2$ optimization problem, algorithms using the proximal operator are often employed. Here, the

**Algorithm 1** ISTA

**Fix:** $\boldsymbol{x}[0]$, $\gamma > 0$.
**Output:** Estimate of $\boldsymbol{x}$
**for** $k = 0, 1, 2, \dots$ **:**
$\boldsymbol{x}[k+1] = S_{\gamma\lambda}(\boldsymbol{x}[k] - \gamma \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}[k] - \boldsymbol{y}))$
**end for**
**return** $\boldsymbol{x}[k+1]$

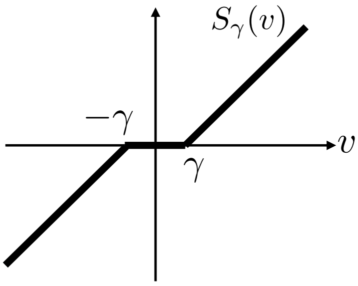


Fig. 1. Soft thresholding operator

**Algorithm 2** FISTA

**Fix:** $\boldsymbol{x}[0]$, $\boldsymbol{z}[0]$, $t[0] > 0$, $\gamma > 0$.
**Output:** Estimate of $\boldsymbol{x}$
**for** $k = 0, 1, 2, \dots$ **:**
$\boldsymbol{x}[k+1] = S_{\gamma\lambda}(\boldsymbol{z}[k] - \gamma \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{z}[k] - \boldsymbol{y}))$
$t[k+1] = \dfrac{1 + \sqrt{1 + 4t[k]^2}}{2}$
$\boldsymbol{z}[k+1] = \boldsymbol{x}[k+1] + \dfrac{t[k] - 1}{t[k+1]}(\boldsymbol{x}[k+1] - \boldsymbol{x}[k])$
**end for**
**return** $\boldsymbol{x}[k+1]$

**Algorithm 3** ADMM

**Fix:** $\boldsymbol{z}_0$,$\boldsymbol{v}_0$, $\gamma > 0$.
**Output:** $\boldsymbol{s}$ and $\boldsymbol{z}$
**for** $k = 0, 1, 2, \dots$ **:**
$\boldsymbol{s}_{k+1} = \arg\min_{\boldsymbol{s} \in \mathbb{R}^L} \left\{ \phi_1(\boldsymbol{s}) + \dfrac{1}{2\gamma} ||\boldsymbol{B}\boldsymbol{s} - \boldsymbol{z}_k + \boldsymbol{v}_k||_2^2 \right\}$
$\boldsymbol{z}_{k+1} = \mathrm{prox}_{\gamma\phi_2}(\boldsymbol{B}\boldsymbol{s}_{k+1} + \boldsymbol{v}_k)$
$\boldsymbol{v}_{k+1} = \boldsymbol{v}_k + \boldsymbol{B}\boldsymbol{s}_{k+1} - \boldsymbol{z}_{k+1}$
**end for**
**return** $\boldsymbol{s}[k+1]$, $\boldsymbol{z}[k+1]$

proximal operator with the parameter $\gamma > 0$ is defined as

$$\mathrm{prox}_{\gamma f}(\boldsymbol{z}) = \arg\min_{\boldsymbol{u} \in \mathrm{dom}(f)} \left\{ f(\boldsymbol{u}) + \frac{1}{2\gamma} ||\boldsymbol{u} - \boldsymbol{z}||_2^2 \right\}, \quad (3)$$

where $f : \mathbb{R}^N \to \mathbb{R} \cup \{\infty\}$ is a proper closed-convex function, that is, $\{(\boldsymbol{x}, t) \in \mathbb{R}^N \times \mathbb{R} : f(\boldsymbol{x}) \leq t\}$ is a non-empty closed-convex set, and $\mathrm{dom}(f)$ is the effective domain of the function $f$, namely, $\mathrm{dom}(f) = \{\boldsymbol{x} \in \mathbb{R}^N : f(\boldsymbol{x}) < +\infty\}$.

ISTA (iterative shrinkage-thresholding algorithm) is one of the simplest and most popular algorithms that use proximal operator. The details of ISTA is shown in Algorithm 1, where $\mathcal{S}_\gamma(v)$ is the proximal operator of $\ell_1$ norm and is called a soft thresholding operator (see Fig. 1).

One of major concerns with ISTA is its slow rate of convergence, while its simplicity is attractive. Therefore, FISTA, which is a faster version of ISTA, is widely used when faster convergence is required. Detailed steps of FISTA are summarized in Algorithm 2. It should be noted that FISTA achieves faster convergence by the employment of the update equation using not only the estimate in the previous iteration but also the estimate of one more previous iteration, while the update equation in ISTA uses only the estimate in the previous iteration. It is known that FISTA can achieve the optimum convergence rate among the first order methods that use the information of the gradient and the function value of the cost function [3], [4]. However, it is difficult to directly implement FISTA in the optical circuit, because it includes the operation of division with a dynamic value in the update formula of $\boldsymbol{z}[k+1]$, which is difficult to realize with the optical circuit.

Yet, another famous algorithm that uses proximal operator is ADMM. In ADMM, we consider an optimization problem of the form

$$\min_{\boldsymbol{s} \in \mathbb{R}^L, \boldsymbol{z} \in \mathbb{R}^K} \{\phi_1(\boldsymbol{s}) + \phi_2(\boldsymbol{z})\} \text{ subject to } \boldsymbol{z} = \boldsymbol{B}\boldsymbol{s}, \quad (4)$$

where $\phi_1(\cdot)$ and $\phi_2(\cdot)$ are proper closed-convex functions. The ADMM algorithm for this canonical form of the problem is given as shown in Algorithm 3.

The $\ell_1 - \ell_2$ optimization problem can be solved using ADMM. Specifically, the original optimization problem (2) is equivalent to the optimization problem of

$$\min_{\boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^N} \left( \frac{1}{2} ||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}||_2^2 + \lambda ||\boldsymbol{z}||_1 \right), \ \boldsymbol{x} = \boldsymbol{z}, \quad (5)$$

and by setting

$$\phi_1(\boldsymbol{x}) = \frac{1}{2} ||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}||_2^2 \quad (6)$$

$$\phi_2(\boldsymbol{z}) = \lambda ||\boldsymbol{z}||_1 \quad (7)$$

$$\boldsymbol{B} = \boldsymbol{I}, \quad (8)$$

where $\boldsymbol{I}$ is an identity matrix, it turns out that the $\ell_1 - \ell_2$ optimization problem is reduced to the canonical form problem (4) of ADMM [5]. From this formulation, the ADMM based algorithm for the $\ell_1 - \ell_2$ optimization problem is obtained as shown in Algorithm 4.

## III. Proposed Algorithm

The previous section has described existing algorithms for $\ell_1$–$\ell_2$ optimization problem based on ISTA, FISTA, and ADMM. Here, we propose a novel algorithm for the optimization problem based on ADMM by using a different formulation from the existing ADMM based algorithm.

In the proposed method, we rewrite $\ell_1$–$\ell_2$ optimization problem (2) as an equivalent optimization problem:

$$\min_{\boldsymbol{x} \in \mathbb{R}^N, \boldsymbol{z} \in \mathbb{R}^M} \left( \frac{1}{2} ||\boldsymbol{z} - \boldsymbol{y}||_2^2 + \lambda ||\boldsymbol{x}||_1 \right), \ \boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}. \quad (9)$$

**Algorithm 4** ADMM for $\ell_1 - \ell_2$ (existing))

**Fix:** $\boldsymbol{z}[0]$, $\boldsymbol{v}[0]$, $\gamma > 0$.
**Output:** Estimate of $\boldsymbol{x}$
**for** $k = 0, 1, 2, \dots$ **:**

$$\boldsymbol{x}[k+1] = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \frac{1}{\gamma}\boldsymbol{I}\right)^{-1} \cdot \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} + \frac{1}{\gamma}(\boldsymbol{z}[k] - \boldsymbol{v}[k])\right)$$

$$\boldsymbol{z}[k+1] = S_{\gamma\lambda}(\boldsymbol{x}[k+1] + \boldsymbol{v}[k])$$

$$\boldsymbol{v}[k+1] = \boldsymbol{v}[k] + \boldsymbol{x}[k+1] - \boldsymbol{z}[k+1]$$

**end for**
**return** $\boldsymbol{x}[k+1]$

Here, by setting

$$\phi_1(\boldsymbol{x}) = \lambda||\boldsymbol{x}||_1, \tag{10}$$

$$\phi_2(\boldsymbol{z}) = \frac{1}{2}||\boldsymbol{z} - \boldsymbol{y}||_2^2, \tag{11}$$

and $\boldsymbol{B} = \boldsymbol{A}$ in (4), $\ell_1$–$\ell_2$ optimization problem can be reduced to the canonical form of ADMM. To derive the proposed algorithm, we then obtain the augmented Lagrangian function of problem (9) as

$$\begin{aligned} L_\rho(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{v}) =& \frac{1}{2}||\boldsymbol{z} - \boldsymbol{y}||_2^2 + \lambda||\boldsymbol{x}||_1 + \boldsymbol{v}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{z}) \\ &+ \frac{\rho}{2}||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{z}||_2^2 \\ =& \frac{1}{2}||\boldsymbol{z} - \boldsymbol{y}||_2^2 + \lambda||\boldsymbol{x}||_1 + \frac{\rho}{2}\left|\left|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{z} + \frac{\boldsymbol{v}}{\rho}\right|\right|_2^2 \\ &- \frac{1}{2\rho}||\boldsymbol{v}||_2^2 \end{aligned} \tag{12}$$

In this form, the variables corresponding to the two iterative directions of the ADMM algorithm are completely independent.

The first step of the ADMM update is given by

$$\begin{aligned} \boldsymbol{x}[k+1] = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \Bigg(& \frac{\lambda}{\rho}||\boldsymbol{x}||_1 \\ &+ \frac{1}{2}\left|\left|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{z}[k] + \frac{1}{\rho}\boldsymbol{v}[k]\right|\right|_2^2\Bigg), \end{aligned} \tag{13}$$

which is also an $\ell_1$–$\ell_2$ optimization problem. Since it is difficult to directly obtain the closed form solution of the problem, we consider to utilize ISTA here to obtain the the solution. Note that, in ISTA, a sufficient number of iterations are required to ensure the convergence of the solution. However, we set the number of iterations $J$ in ISTA for the first step of ADMM as a design parameter, and evaluate the impact of $J$ on the performance via computer simulations later. Also, this is the reason why we call the proposed algorithm as the 'approximated' ADMM based algorithm.

**Algorithm 5** ADMM for $\ell_1 - \ell_2$ (proposed))

**Fix:** $\boldsymbol{x}[0]$, $\boldsymbol{z}[0]$, $\boldsymbol{v}[0]$, $\gamma > 0$, $\rho > 0$.
**Output:** Estimate of $\boldsymbol{x}$
**for** $k = 0, 1, 2, \dots$ **:**
$\bar{\boldsymbol{x}}[0] = \boldsymbol{x}[k]$
**for** $j = 0, 1, 2, \dots J-1$ **:**

$$\bar{\boldsymbol{x}}[j+1] = S_{\gamma\lambda/\rho}\left(\bar{\boldsymbol{x}}[j] - \gamma\boldsymbol{A}^{\mathrm{T}}\left(\boldsymbol{A}\bar{\boldsymbol{x}}[j] - \boldsymbol{z}[k] + \frac{1}{\rho}\boldsymbol{v}[k]\right)\right)$$

**end for**
$\boldsymbol{x}[k+1] = \bar{\boldsymbol{x}}[J]$

$$\boldsymbol{z}[k+1] = \frac{1}{1+\rho}\left(\boldsymbol{y} + \rho\left(\boldsymbol{A}\boldsymbol{x}[k+1] + \frac{1}{\rho}\boldsymbol{v}[k]\right)\right)$$

$$\boldsymbol{v}[k+1] = \boldsymbol{v}[k] + \rho(\boldsymbol{A}\boldsymbol{x}[k+1] - \boldsymbol{z}[k+1])$$

**end for**
**return** $\boldsymbol{x}[k+1]$

On the other hand, the second step of the ADMM update

$$\begin{aligned} \boldsymbol{z}[k+1] = \arg\min_{\boldsymbol{z}\in\mathbb{R}^M} \Bigg(& \frac{1}{2}||\boldsymbol{z} - \boldsymbol{y}||_2^2 \\ &+ \frac{\rho}{2}\left|\left|\boldsymbol{A}\boldsymbol{x}[k+1] - \boldsymbol{z} + \frac{1}{\rho}\boldsymbol{v}[k]\right|\right|_2^2\Bigg) \end{aligned} \tag{14}$$

is rather simple problem, and thus we can obtain closed form update equation for this step as

$$\boldsymbol{z}[k+1] = \frac{1}{1+\rho}\left(\boldsymbol{y} + \rho\left(\boldsymbol{A}\boldsymbol{x}[k+1] + \frac{1}{\rho}\boldsymbol{v}[k]\right)\right). \tag{15}$$

From above formulations, the proposed algorithm for the $\ell_1$–$\ell_2$ optimization problem based on ADMM is obtained as shown in Algorithm 5. Note that the proposed algorithm does not include division by dynamically changing values and matrix inversion unlike FISTA or existing ADMM based algorithm, so it could be easily implemented by the optical circuit.

## IV. Numerical Experiment

We have conducted computer simulations to evaluate the performance of the proposed ADMM based algorithm. In the simulations, we have generated measurement matrix $\boldsymbol{A}$ as a submatrix of unitary DFT (discrete Fourier transform) matrix by randomly choosing appropriate number of rows of the DFT matrix. Also, we have set $\lambda = 5$ in the $\ell_1$–$\ell_2$ optimization problem, each element of the observation noise is assumed to be white Gaussian with mean 0 and variance 0.05, and the variance of the nonzero element of the unknown signal to be 1.

Firstly, in order to evaluate the impact of the number of inner loops $J$ of the proposed algorithm, we show the numerical results of MSE (mean squared error) versus number of outer loops of the proposed algorithm in Fig. 2, where $J$ is set to be 1, 2, 3, 4, 5, and 100. Also, we have set $N = 512$, $M = 256$, and $K = 32$ in the simulation. From the figure, we can see that

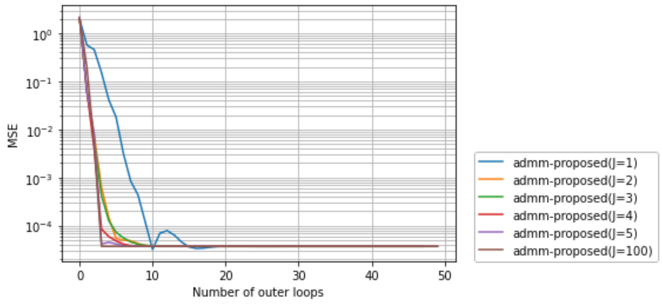


Fig. 2. MSE of Proposed Method vs Number of Outer Loops ($N = 512,\ M = 256,\ K = 32$)

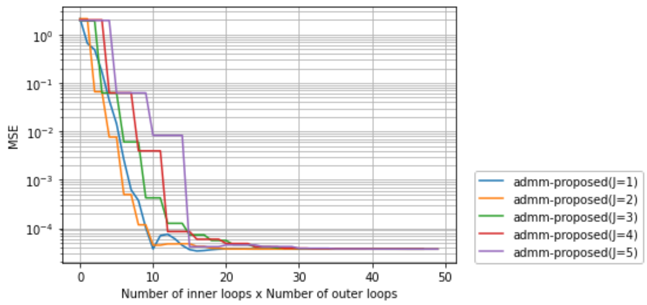


Fig. 3. MSE of Proposed Method vs Number of Inner and Outer Loops ($N = 512,\ M = 256,\ K = 32$)

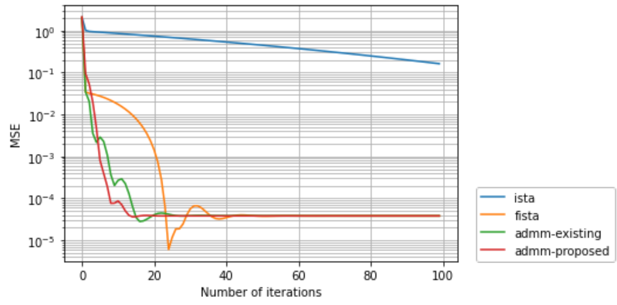


Fig. 4. Comparison of Learning Curves ($N = 512,\ M = 256,\ K = 32$)

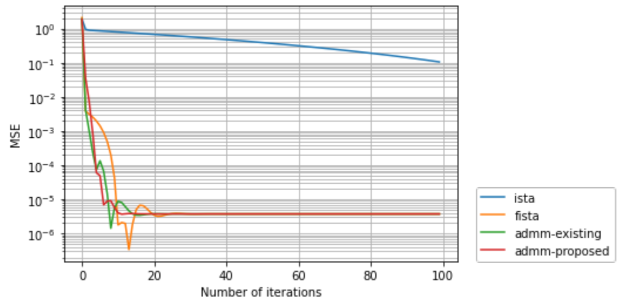


Fig. 5. Comparison of Learning Curves ($N = 512,\ M = 256,\ K = 4$)

the MSE performance improves as the number of inner loops $J$ increases. However, the comparison of the number of outer loops will not be fair for the proposed algorithm with different value of $J$, because the computational complexity and the processing delay largely depend on the number of inner loops $J$. Therefore, we have also evaluated the MSE performance versus the number of inner loops times the number of outer loops, which is the total number of iterations of the inner loops of the proposed algorithm, in Fig. 3. From the figure, we can see that $J = 1$ or 2 will be the best choice for the proposed algorithm in terms of the convergence rate of the MSE. Thus, we employ $J = 1$ in the subsequent simulations, which means that the proposed algorithm is no more double loop algorithm.

In order to compare the performance of the proposed algorithm with that of existing algorithms for the $\ell_1$–$\ell_2$ optimization problem, we show the learning curves (MSE performance) of the proposed ADMM, ISTA, FISTA and existing ADMM in Figs. 4-7 with different system parameters. We have set $N = 512,\ M = 256,\ K = 32$ in Fig.4, $N = 512,\ M = 256,\ K = 4$ in Fig. 5, $N = 64,\ M = 32,\ K = 4$ in Fig. 6, and $N = 64, M = 32, K = 1$ in Fig. 7. In all figures, the convergence rate of ISTA is significantly worse than other algorithms. Also, we can see that the convergence performance of the proposed ADMM, FISTA and existing ADMM is almost identical when the number of nonzero elements of the unknown vector is small ($K = 4$) or the system size is small ($N = 64,\ M = 32$). When the system size is large and the unknown vector is not that sparse, the proposed ADMM can achieve comparable or even better performance than existing ADMM and FISTA.

## V. Conclusions

In this paper, we have proposed a novel algorithm for the $\ell_1 - \ell_2$ optimization problem based on ADMM, aiming at the implementation with the optical circuit. The proposed algorithm is free from the operation of division by a dynamic value and also from the matrix inversion, and thus suited for the implementation with the optical circuit. Although the proposed algorithm has a double loop configuration, we have verified that the number of iterations of the inner loop can be set to be one from the numerical results. Also, we have demonstrated via computer simulations that the proposed algorithm can achieve comparable performance to FISTA or existing ADMM based algorithm.

Future work includes theoretical analysis of the convergence characteristics and detailed optimization of the parameters in the algorithm.

## Acknowledgment

This work has been supported in part by JST CREST JPMJCR21C3.

## References

[1] K. Nozaki, S. Matsuo, T. Fujii, K. Takeda, A. Shinya, E. Kuramochi and M. Notomi, "Femtofarad optoelectronic integration demonstrating energy-saving signal conversion and nonlinear functions," *Nature Photonics*, Vol. 13, No. 7, pp. 454–459, 2019

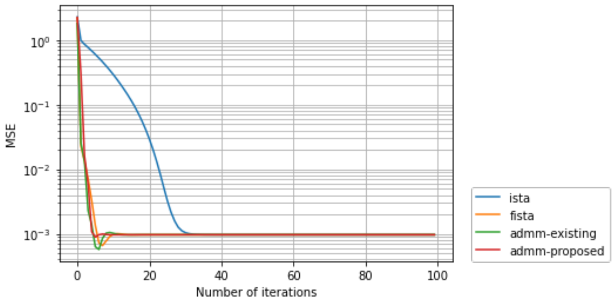


Fig. 6. Comparison of Learning Curves $(N = 64,\ M = 32,\ K = 4)$

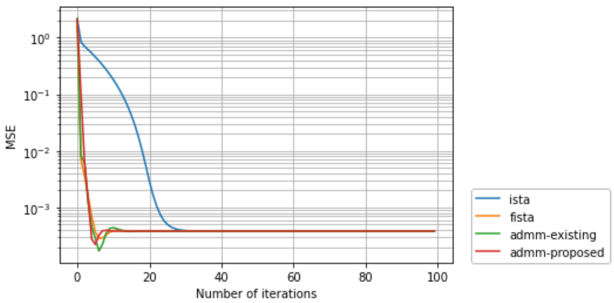


Fig. 7. Comparison of Learning Curves $(N = 64, M = 32, K = 1)$

[2] H. Zhou, J. Dong, J. Cheng, W. Dong, C. Huang, Y. Shen, Q. Zhang, M. Gu, C. Qian, H. Chen, Z. Ruan and X. Zhang, "Photonic matrix multiplication lights up photonic accelerator and beyond," *Light Sci Appl*, Vol. 11, No. 30, 2022
[3] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, Springer, 2004.
[4] Y. Nesterov, "Gradient methods for minimizing composite functions," *Mathematical Programming*, Vol. 140, No. 1, pp. 125-161, 2013.
[5] M. Nagahara, *Sparsity methods for systems and control*, now Publishers, 2020.