An Alternating Direction Method of Multipliers for Inverse Lithography Problem

Junqing Chen and Haibo Liu

Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China.

Received 21 September 2022; Accepted (in revised version) 11 April 2023

Abstract. We propose an alternating direction method of multipliers (ADMM) to solve an optimization problem stemming from inverse lithography. The objective functional of the optimization problem includes three terms: the misfit between the imaging on wafer and the target pattern, the penalty term which ensures the mask is binary and the total variation regularization term. By variable splitting, we introduce an augmented Lagrangian for the original objective functional. In the framework of ADMM method, the optimization problem is divided into several subproblems. Each of the subproblems can be solved efficiently. We give the convergence analysis of the proposed method. Specially, instead of solving the subproblem concerning sigmoid, we solve directly the threshold truncation imaging function which can be solved analytically. We also provide many numerical examples to illustrate the effectiveness of the method.

AMS subject classifications: 78A46, 78M50

Key words: Inverse lithography techniques, ADMM framework, total variation regularization.

1. Introduction

Optical lithography system plays a critical role in semiconductor industry. The optical lithography system can print mask patterns onto the wafer through an optical system. As shrinkage of integrated circuit device size and subject to the resolution limit of the optical system, the diffraction of mask makes the patterns on wafer distorted very much. To remedy that, many resolution enhancement techniques are proposed and used extensively in optical lithography. Optical proximity correction (OPC) is the most popular resolution enhancement method. With optical proximity correction, people adjust the mask layout such that the output pattern approximates the desired one. OPC can be divided into two categories: rule based OPC and model based OPC. Rule
based method modifies the mask layout and compensates for the warping in local features empirically, which is easy to implement. Model based method relies on the physical and mathematical theories of optical imaging. With model based OPC, the mask pattern is represented by a pixel-wise function or a level-set function, then the OPC process is modeled as an inverse problem. Usually, the inverse problem is formulated as a non-convex optimization problem about the mask pattern. So the model based OPC is usually called inverse lithography technology (ILT) [26]. The objective functional of this inverse problem is the misfit between the image on wafer and the target pattern. The image on wafer is modeled by threshold and truncation of aerial image projected by the optical system and the threshold is used to mimic the exposure process of photoresist on wafer. By solving the inverse problem, ILT can adjust the shape of mask such that the mask gets close to the optimal one. It is known that the aerial image of mask is approximated by Abbe formula which can be regarded as convolution [14]. Besides, the threshold function is discontinuous, then the objective functional of inverse lithography is non-smooth and non-convex, and traditional optimization algorithms are difficult to achieve good results.

Since the importance of the inverse lithography problem, tremendous efforts have been devoted to solve the problem in the past decades. Based on the pixel-wise mask, the inverse lithography problem was stated as nonlinear, constrained minimization problems, and linear, quadratic, and nonlinear formulations of the objective function were considered in [15]. The steepest descent methods were proposed in [20, 27]. A conjugate gradient method was suggested to accelerate the computation in [20]. And there are many other works related to the gradient-based method such as [3, 4, 18, 29, 31]. The level set method is another effective way to solve inverse lithography problem. With this method, the target pattern, mask pattern, and wafer pattern are represented by level set functions, then the inverse lithography problem is regarded as a shape and topology optimization problem [32, 33]. There are many other miscellaneous methods in inverse lithography, such as the genetic algorithm [11], the route of particle swarm optimizer combined with the adaptive nonlinear control strategy [35], deep convolution neural network methods [40, 41]. A fuller review of inverse lithography techniques could be found in [25].

It is known that the inverse lithography problems are ill-posed. Regularization method is an efficient way to remedy the ill-posedness [10]. The regularization framework was proposed to control the tone and complexity of the synthesized masks [30], $L^1$-regularization, wavelet and total-variation (TV) regularizations are discussed in [18]. Sparsity and low-rank regularization terms are used to constrain the solution space and reduce the mask complexity. The split Bregman algorithm is used to solve the inverse optimization problem with sparsity ($L^1$ norm) and low rank (nuclear norm) regularizations [21]. Specially, the TV regularization is much attractive due to its good edge-preserving property [34]. An objective functional that consists of several total-variation regularization terms is introduced in [9], where the total variations about mask and image on wafer are both considered. But the TV regularization is not differentiable and it is difficult to implement. Some methods are proposed to deal with the
TV-regularization such as smoothness method [8], dual methods [2,5,9] and Bregman iteration method [24].

While various computational schemes have been proposed solving the inverse lithography problems, the computing process of ILT is still complicated. Especially when there are several regularization terms in the loss function [9]. Furthermore, since the optimization problem in optical lithography is not convex, the convergence analysis of the solution algorithms is lacking in the literature. The ADMM method [12] has been found to work well in solving non-smooth and non-convex problems, such as phase retrieval problem [37], compressive sensing [7], noisy color image restoration [16]. Besides, the convergence analysis of ADMM method has made a lot of progress in recent years [36].

In the present work, we focus on inverse lithography problem with TV regularization and propose an ADMM method to deal with the regularization. There are several advantages of our method. Firstly, comparing with [9], our method can avoid the tedious computation of total variation of image on wafer and focuses on the TV regularization of mask. Secondly, in the framework of ADMM, we split the original optimization problem to three subproblems to avoid the expensive computation of the gradient of a composite function about sigmoid and convolution. Thirdly, we establish the convergence of proposed ADMM method under certain conditions. Finally, when dealing with the subproblem about the sigmoid function, we use its original nonsmooth form to accelerate the computation.

The paper is organized as follows. In Section 2, we introduce the inverse lithography problem and prove the stability of the solution to inverse problem with total variation regularization. We develop the ADMM method for inverse lithography problem and prove the convergence of the proposed algorithm in Section 3. Some numerical examples are provided in Section 4 to show the effectiveness of our algorithm. We draw some conclusion in Section 5.

2. Inverse lithography problem

2.1. Lithography system and imaging process

The Fig. 1 describes the most important elements inside a photo-lithography system. The light emitted by the source \( \{ S \} \) goes through the condenser \( \{ C \} \), becomes parallel light and then projects the mask pattern \( \{ M \} \) by the projection lens \( \{ P \} \) onto the wafer \( \{ W \} \). With the help of photochemical reaction in photo-resist on wafer, this optical lithography process can transfer the mask pattern to wafer pattern for next stage use. Under the assumption that the mask is thin, the projection processing can be described by Fourier Optics [14]. Using the Kirchhoff approximation, with the help of Abbe formulation, the aerial image on the wafer with partially coherent illumination can be approximately described as

\[
I^a(U)(r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(r_1)U^*(r_2)\gamma(r_1 - r_2)H(r - r_1)H^*(r - r_2)dr_1dr_2,
\]
where \( r = (x, y) \), \( r_1 = (x_1, y_1) \), and \( r_2 = (x_2, y_2) \). \( U(r) \) is the mask pattern, \( \gamma(r_1 - r_2) \) is the complex degree of coherence, and \( H(r) \) represents the amplitude impulse response of the optical system. The complex degree of coherence \( \gamma(r_1 - r_2) \) is generally a complex number, whose magnitude represents the extent of optical interaction between two spatial locations \( r_1 \) and \( r_2 \) of the light source [38]. Specially, for completely coherent system, namely, the light is emitted by a point source, \( \gamma(r_1 - r_2) = 1 \) and the aerial image can be simplified as

\[
I_{\text{aerial}}(U)(r) = |H(r) * U(r)|^2,
\]

where “\(*\)” denote the convolution operation. As for the point spread function \( H(r) \) which describes the response of an imaging system to a point source,

\[
H(x, y) = F^{-1}[h(f, g)],
\]

where \( F \) is the Fourier transform and \( h(f, g) \) is the projection lens transfer function.

In Fourier optics, the projection lens acts as a low pass filter for Fourier transform of the mask. The frequencies which are lower than the projection lens cutoff frequency can pass through the lens and frequencies larger than the cutoff frequency are lost. The cutoff frequency \( f_{\text{cut}} \) is

\[
f_{\text{cut}} = \frac{NA}{\lambda},
\]

where \( \lambda \) is the wavelength of the optic wave and \( NA \) is the numerical aperture of the projection lens [19, 39]. The effect of the projection lens is described as a circular transfer function

\[
h(f, g) = \begin{cases} 
1, & \text{if } \sqrt{f^2 + g^2} \leq \frac{NA}{\lambda}, \\
0, & \text{otherwise}.
\end{cases}
\]

Then the corresponding point spread function reads as

\[
H(x, y) = \frac{J_1(2\pi r NA/\lambda)}{2\pi r NA/\lambda},
\]
and is shown in Fig. 2, where $J_1(x)$ is the Bessel functions of order 1. In a real optical lithography system, defocus is an important property that should be considered. It corresponds to a vertical shift of the substrate relative to the plane of best focus. Defocus can be treated similarly with an aberration which describes deviation from the ideal wave front [14], even though it is actually not an aberration. Let $W(f, g)$ be an aberration function, the according transfer function for the pupil function is written as

$$h(f, g) = \begin{cases} e^{-i \frac{2\pi}{\lambda} W(f, g)}, & \text{if } |(f, g)| \leq \frac{NA}{\lambda}, \\ 0, & \text{otherwise}. \end{cases}$$

Where the defocus aberration function is

$$W(f, g) = \delta \sqrt{1 - (f^2 + g^2)\lambda^2}, \quad (2.1)$$

and $\delta$ represents the defocus.

The last step in the simulation of lithography process is the exposure of the photoresist. If we regard the output pattern on wafer is binary, the resist effects could be modeled as a hard threshold operation $T(x)$

$$T(x) = \begin{cases} 1, & x \geq tr, \\ 0, & x < tr, \end{cases} \quad (2.2)$$

and $tr$ is the threshold. Then the exposure pattern on wafer is

$$I_t(U)(x, y) = T(I_{aerial}(U)(x, y)).$$

Clearly, $T(x)$ is not a differentiable function. For the optimization method based on gradient, this threshold model is not suitable. Usually, $T$ is approximated with a sigmoid.
function

$$\text{Sig}_a(x) = \frac{1}{1 + \exp\left(-a(x-t\tau)\right)}.$$  

For this smoothness, the function value tends to 1 when $x$ is larger than $t\tau$, tends to 0 otherwise. The coefficient $a$ controls the steepness. Fig. 3 is the description of a sigmoid function.

Finally, we have the approximate smooth exposure image on wafer

$$I(U)(x, y) = \text{Sig}_a\left(\left| (H \ast U)(x, y) \right|^2 \right), \quad (2.3)$$

which is the image model we used in the following sections.

2.2. BV space and total variation

Let $\Omega$ be a 2-D bounded Lipschitz domain. In the followings, we assume that the mask belongs to the BV-space

$$BV(\Omega) = \left\{ U \in L^1(\Omega) : \sup_{\phi \in C^1_c(\Omega)} \int_{\Omega} U(x) \text{div}\phi(x) \, dx < \infty, |\phi| \leq 1 \right\},$$

which is equipped with the norm

$$\|U\|_{BV} = \|U\|_{L^1(\Omega)} + \|U\|_{TV},$$

where

$$\|U\|_{TV} = \int_{\Omega} |DU| \, dx \equiv \sup_{\phi \in C^1_c(\Omega), |\phi| \leq 1} \int_{\Omega} U \text{div}\phi \, dx$$

is the total variation of $U$. It is known that the BV-space is compactly embedding to $L^1(\Omega)$, which is crucial for the existence and stability of solutions to the inverse problem. The total variation measures the total image gradient and possesses stronger constraining effects. The regularization based on total variation has the capability of denoising, deblurring, and yielding sharp edges and is applied in many fields as we have mentioned in Section 1.
2.3. The inverse lithography problem

Within the confines of geometrical optics, the image of a point is sharp. But the actual image is smeared because of diffraction. For an imaging system with a circular aperture, the resolution is proportional to $\lambda/(NA)$, which is always bigger than the critical size of the target pattern on wafer. So the image on the wafer is blurred. To remedy the blurring and improve the resolution, the resolution enhancement techniques (RETS) have been proposed to minimize mask distortions as they are projected onto wafers. There are many kinds of techniques included in RETs. For more details of RETs technique, we recommend [19, 39]. Among the RETs, inverse lithography technology (ILT) is a promising method and used widely. ILT determines the mask $U(x, y)$ that produces the desired on-wafer results $I(x, y) = I(U)$ by formulating the resolution enhancement process as an inverse problem. This process could be modeled as an optimization problem

$$\min_{U} \|I(U) - I\|_{L^2(\Omega)}$$

where $U$ is the unknown mask pattern to be found, $I(U)$ represents the computed field (2.3), $I$ is desired lithography pattern.

Over the past years, some of the ideas have been innovatively introduced into ILT and significantly enhance the quality of reconstructed mask [9, 17, 21, 28]. Basically, all these methods are based on constructing a cost function comprising one quadratic $L^2$ norm error term and some regularization terms. The regularization terms are used to deal with the ill-posedness of the inverse problem. Then the cost function is expressed as

$$E(U) = \|I(U) - I\|_{L^2(\Omega)}^2 + \alpha R(U),$$

where $\alpha$ is the regularization parameter, and $R(U)$ represents the regularization. We assume that the mask is a BV image and binary valued [6], which means that the mask $U \in K_0 = \{U(x)|U(x) \in BV(\Omega), U(x) = \{0, 1\}\}$. Then with (2.3) and (2.4) in hand, by considering the total variation regularization, the original ILT problem could be modeled as the following constraint optimization problem:

$$\min_{U} \|\text{Sig}_\alpha(|H * U|^2) - I\|_{L^2(\Omega)}^2 + \alpha \|U\|_{TV}$$

s.t. $U \in K_0$. (2.5)

Since $U(x)$ is binary-valued in $\{0, 1\}$, problem (2.5) will cause a nonlinear integer programming problem which is NP-hard. Besides, $K_0$ is a non-convex set. So we relax the binary constrain $U(x) = \{0, 1\}$ with $0 \leq U \leq 1, U(x) \in BV(\Omega)$, add a penalty term $\|U(1-U)||_1$. Then let $K = \{U(x)|U(x) \in BV(\Omega), 0 \leq U(x) \leq 1\}$, we have

$$\min_{U} \|\text{Sig}_\alpha(|H * U|^2) - I\|_{L^2(\Omega)}^2 + \beta_1 \|U\|_{TV} + \beta_2 \|U(1-U)||_{L^1(\Omega)}$$

s.t. $U(x) \in K$. (2.6)
It should be mentioned that the cost function (2.6) is different with the one proposed in [9]. Since there is only one regularization term, we can avoid complicated computation of the total variation of aerial image. Besides, there are less parameters needed to be adjusted, so the corresponding algorithm is easy to implement.

2.4. Well-posedness analysis

To consider the well-posedness analysis of the solution, we rewrite the objective function as

\[ E(U; H) = \| \text{Sig}_a(\| H * U \|^2) - I \|^2_{L^2(\Omega)} + \beta_1 \| U \|_{TV} + \beta_2 \| U(1 - U) \|_{L^1(\Omega)}. \]

Because the regularization terms of problem (2.6) is different with [9]. We give the existence of (2.6) for the completeness.

**Theorem 2.1.** Assume that \( I \in L^\infty(\Omega) \), then for any given \( H \in L^2(\Omega) \) and \( \beta_1, \beta_2 > 0 \), there exists a minimizer of \( E(U; H) \) in \( K \).

**Proof.** Since

\[ E(0; H) = \| I \|^2_{L^2(\Omega)} \leq \| I \|_{L^\infty(\Omega)} \| I \|^2_{L^2(\Omega)} < + \infty, \]

we know \( \inf_{U \in K} E(U; H) \) is finite. Suppose

\[ \lim_{n \to \infty} E(U_n; H) = \inf_{U \in K} E(U; H), \]

we can conclude that \( \{U_n\} \) is uniformly bounded in \( BV(\Omega) \). Since \( BV(\Omega) \) is compactly embedded in \( L^1(\Omega) \), there exists a subsequence (also denoted by \( \{U_n\} \)) and a function \( U^* \in L^1(\Omega) \) such that

\[ U_n \to U^* \text{ strongly in } dL^1(\Omega). \tag{2.7} \]

From [1, Corollary 2.17], there is a subsequence (also denoted by \( \{U_n\} \), s.t.

\[ U_n \to U^* \text{ a.e. } x \in \Omega. \]

Then

\[ \text{Sig}_a(\| H * U_n \|^2) - I \to \text{Sig}_a(\| H * U^* \|^2) - I \text{ a.e. } x \in \Omega, \]

\[ U_n(1 - U_n) \to U^*(1 - U^*) \text{ a.e. } x \in \Omega. \]

By Fatou’s Lemma, we have

\[ \| \text{Sig}_a(\| H * U^* \|^2) - I \|^2_{L^2(\Omega)} + \beta_2 \| U^*(1 - U^*) \|_{L^1(\Omega)} \]

\[ \leq \liminf_{n \to \infty} \| \text{Sig}_a(\| H * U_n \|^2) - I \|^2_{L^2(\Omega)} + \beta_2 \| U_n(1 - U_n) \|_{L^1(\Omega)}. \tag{2.8} \]

From (2.7), we know that [8]

\[ \int_\Omega |DU^*| dx \leq \liminf_n \int_\Omega |DU_n| dx. \tag{2.9} \]
By (2.8) and (2.9), we have
\[ E(U^*; H) \leq \lim \inf_{n} E(U_n, H) = \inf_{U \in K} E(U; H), \]
so \( U^* \) is a minimizer of \( E(U; H) \).

The following result shows the stability of solution.

**Theorem 2.2.** Assume that \( I \in L^\infty(\Omega) \). \( H_n \to H \) in \( L^2(\Omega) \) as \( n \to \infty \), suppose \( U_n \) is a minimizer of \( E(U; H_n) \). Then there exists a subsequence of \( \{ U_n \} \) which converges to a minimizer of \( E(U, H) \) strongly in \( L^1(\Omega) \).

**Proof.** It is easy to find that
\[
|H * U|^2(x) - |H_n * U|^2(x)
= \int_\Omega (H(y) - H_n(y))U(x - y)dy \cdot \int_\Omega H(y)U(x - y)dy
+ \int_\Omega H_n(y)U(x - y)dy \cdot \int_\Omega (H(y) - H_n(y))U(x - y)dy.
\]
Since \( H_n \to H \) in \( L^2(\Omega) \), there exists a constant \( C \) such that
\[
\|H_n\|_{L^2(\Omega)} \leq C, \quad \|H - H_n\|_{L^2(\Omega)} \leq C.
\]
Be aware of that \( 0 \leq U \leq 1 \), we have
\[
\left| |H * U|^2(x) - |H_n * U|^2(x) \right| \leq C \|H - H_n\|_{L^2(\Omega)}, \quad \forall x \in \Omega. \tag{2.10}
\]
By the definition of \( E(U, H) \), we have
\[
|E(U; H) - E(U; H_n)| = \int_\Omega \left| \Sigma_a \left( |H * U|^2 \right) - \Sigma_a (|H_n * U|^2) \right|
\leq C \|H - H_n\|_{L^2(\Omega)}.
\]
The last inequality follows from the Lipschitz property of \( \Sigma_a \) and (2.10), so for any \( U \in K \)
\[
\lim_{n \to \infty} E(U; H_n) = E(U; H), \tag{2.11}
\]
and \( E(U; H_n) \) is bounded. By Theorem 2.1 and the definition of \( E(U; H_n) \), since \( E(U_n; H_n) = \inf_{U \in K} E(U; H_n) \), we have
\[
\beta_1 \|U_n\|_{TV} \leq E(U_n, H_n) \leq E(U, H_n),
\]
and \( \|U_n\|_{TV} \) is bounded. From \( 0 \leq U \leq 1 \),
\[
\|U_n\|_{L^1(\Omega)} = \int_\Omega |U_n|dx \leq |\Omega|.
\]
We can conclude that \( \{U_n\} \) is uniformly bounded in \( BV(\Omega) \). Since \( BV(\Omega) \) is compactly embedded in \( L^1(\Omega) \), there exist a subsequence (also denoted by \( \{U_n\} \)) and a function \( U^* \in L^1(\Omega) \) such that
\[
U_n \rightarrow U^* \quad \text{strongly in } L^1(\Omega). \tag{2.12}
\]

From [1, Corollary 2.17], there is a subsequence (also denoted by \( \{U_n\} \)), s.t.
\[
U_n \rightarrow U^* \quad \text{a.e. } x \in \Omega,
\]
and a subsequence of \( \{H_n\} \) (also denoted by \( \{H_n\} \)), s.t.
\[
H_n \rightarrow H \quad \text{a.e. } x \in \Omega.
\]

Then
\[
\begin{align*}
\Sig_a(|H_n * U_n|^2) - I & \rightarrow \Sig_a(|H * U^*|^2) - I \quad \text{a.e. } x \in \Omega, \\
U_n(1 - U_n) & \rightarrow U^*(1 - U^*) \quad \text{a.e. } x \in \Omega.
\end{align*}
\]

By Fatou’s Lemma, we have
\[
\begin{align*}
\| \Sig_a(|H * U^*|^2) - I \|_{L^2(\Omega)} & + \beta_2 \| U^*(1 - U^*) \|_{L^1(\Omega)} \\
& \leq \liminf_{n \to \infty} \| \Sig_a(|H_n * U_n|^2) - I \|_{L^2(\Omega)} + \beta_2 \| U_n(1 - U_n) \|_{L^1(\Omega)}.
\end{align*} \tag{2.13}
\]

From (2.12), we know that [8]
\[
\int_{\Omega} |DU^*|dx \leq \liminf_{n} \int_{\Omega} |DU_n|dx. \tag{2.14}
\]

By (2.13) and (2.14), we have
\[
E(U^*; H) \leq \liminf_{n} E(U_n, H_n).
\]

From the definition of \( U_n \)
\[
E(U_n; H_n) \leq E(U; H_n), \quad \forall U \in K.
\]

Then
\[
E(U^*; H) \leq \liminf_{n} E(U_n, H_n) \leq \liminf_{n} E(U, H_n) = E(U; H), \quad \forall U \in K,
\]
hence, \( U^* \) is a minimizer of \( E(U; H) \).

We remark that the above result shows the stability with respect to the convolution kernel \( H \). Actually, as we have mentioned before, \( H \) represents the point spread function of the photo-lithography system which is very complex and many factors can affect \( H \). So this stability is important in reality.
3. The ADMM method for the inverse lithography problem

In this section, we will give an ADMM framework to solve the inverse lithography problem and analyze the convergence, then propose an efficient ADMM method with threshold truncation.

In the beginning, we give some notations similar to [19]. Let the mask domain $\Omega$ be rectangle. We discretize the mask into an $N \times N$ Cartesian grid and let the mask be valued on grid points. In this case, the mask is pixel-based and can be represented by a matrix $U \in \mathbb{R}^{N \times N}$. For the sake of convenience, we use the vector representation of the matrix $U$ and also denote it by $U \in \mathbb{R}^{N^2}$. Similarly, we use $I \in \mathbb{R}^{N^2}$ to write the desired binary output pattern. A convolution matrix $H \in \mathbb{R}^{N^2 \times N^2}$ is the discrete form of the two-dimensional point spread function. Let $G_x, G_y \in \mathbb{R}^{N^2 \times N^2}$ be the first-order finite difference matrices in the horizontal and vertical directions, respectively. Here and in what follows, we let $G_U = (G_x U, G_y U)$.

With these notations, we define

$$\|U\|_{TV} = \|G_U\|_1 = \|G_x U\|_1 + \|G_y U\|_1.$$

By variable-splitting, problem (2.6) can be recast to

$$\begin{align*}
\min_{U,V} & \|\operatorname{Sig}_a(|V|^2) - I\|_2^2 + \beta_1 \|G_U\|_1 + \beta_2 \|U \odot (1 - U)\|_1 \\
\text{s.t.} & \quad 0 \leq U \leq 1, \quad V = HU.
\end{align*}
$$

(3.1)

Here “$\odot$” denotes the element-wise product of matrices. Then the augmented Lagrangian for problem (3.1) is

$$L_a(U, V, P) = \|\operatorname{Sig}_a(|V|^2) - I\|_2^2 + \beta_1 \|G_U\|_1 + \beta_2 \|U \odot (1 - U)\|_1$$

$$+ \langle P, V - HU \rangle + \frac{\rho}{2} \|V - HU\|_2^2,$$

(3.2)

where $\rho$ is a sufficient large penalty parameter. With this Lagrangian, we have the following alternating direction method of multipliers for given initial $U^0, V^0, P^0$:

$$\begin{align*}
U^{k+1} &= \arg \min_U L_a(U, V^k, P^k), \\
V^{k+1} &= \arg \min_V L_a(U^{k+1}, V, P^k), \\
P^{k+1} &= P^k + \rho(V^{k+1} - HU^{k+1}).
\end{align*}
$$

(3.3)

3.1. Convergence analysis

The convergence analysis of (3.3) is inspired by the global convergence result of [36]. In order to make the proof more concise, we first give some notations. We define the following functions:

$$h_a(V) = \|\operatorname{Sig}_a(|V|^2) - I\|_2^2,$$
then the derivative with respect to $V$

$$\nabla h_a(V) = 2a V \odot (\text{Sig}_a(|V|^2) - I) \odot (1 - \text{Sig}_a(|V|^2)) \odot \text{Sig}_a(|V|^2).$$

Clearly, $\nabla h_a(V)$ has the same size with $V$, the Hessian of $h_a(V)$ is a matrix. According to the definition of $h_a(V)$, each entry of $\nabla h_a(V)$ is a single value function, then the Hessian of $h_a(V)$ is diagonal. In Fig. 4, we plot the $i$-th entry of $\nabla h_a(V)$ and the according diagonal entry of Hessian with $a = 20, \text{tr} = 0.3$. It is easy to know that $\nabla h_a(V)$ is Lipschitz continuous for any given $a$. We suppose the Lipschitz constant is $l_h$. We define

$$f(U) = \beta_1 \|GU\|_1 + \beta_2 \|U \odot (1 - U)\|_1.$$

In such situation, the Lagrangian function can be recast as

$$L_a(U, V, P) = f(U) + h_a(V) + \langle P, V - HU \rangle + \frac{\rho}{2} \|V - HU\|_2^2. \quad (3.4)$$

For convenience, we omit the subscript $a$ in the followings. Then

$$U^{k+1} = \arg \min_U L(U, V^k, P^k),$$

$$V^{k+1} = \arg \min_V L(U^{k+1}, V, P^k),$$

$$P^{k+1} = P^k + \rho(V^{k+1} - HU^{k+1}). \quad (3.5)$$

Now we will state and prove the main theorem about the convergence.

**Theorem 3.1.** Suppose sequence $(U^k, V^k, P^k)$ is generated by (3.5). If $\rho/2 - l_h / \rho - l_h > 0$, then there is a subsequence which converges to a stationary point of $L(U, V, P)$.

We begin our analysis by showing a sufficient decrease property of the Lagrange function.
Lemma 3.1. If $\rho/2 - l_h/\rho - l_h > 0$, where $l_h$ is the Lipschitz constant of $\nabla h$, then there is a constant $C = C(\rho, l_h) > 0$ such that for sufficiently large $k$, we have
\begin{equation}
L(U^k, V^k, P^k) - L(U^{k+1}, V^{k+1}, P^{k+1}) \geq C\|V^{k+1} - V^k\|^2. 
\end{equation}

Proof. With the optimality condition of $V^k$
\[\nabla h(V^k) + P^{k-1} + \rho(V^k - HU^k) = 0,\]
and
\[P^k = P^{k-1} + \rho(V^k - HU^k),\]
we have
\[P^k = -\nabla h(V^k).\]
Then by the Lipschitz continuity of $\nabla h$,
\[\|P^{k+1} - P^k\|_2 = \|\nabla h(V^k) - \nabla h(V^{k+1})\|_2 \leq l_h\|V^k - V^{k+1}\|_2.\]
With the definition of $U^{k+1}$, it follows that:
\begin{equation}
L(U^k, V^k, P^k) - L(U^{k+1}, V^k, P^k) \geq 0, 
\end{equation}
and
\begin{align*}
L(U^{k+1}, V^k, P^k) - L(U^{k+1}, V^{k+1}, P^{k+1}) & = h(V^k) - h(V^{k+1}) - \langle P^{k+1}, V^{k+1} - HU^{k+1} \rangle \\
& + \langle P^k, V^k - HU^{k+1} \rangle + \frac{\rho}{2}\|V^k - HU^{k+1}\|^2 - \frac{\rho}{2}\|V^{k+1} - HU^{k+1}\|^2 \\
& = h(V^k) - h(V^{k+1}) + \langle P^k - P^{k+1}, V^k - HU^{k+1} \rangle \\
& - \langle P^{k+1}, V^{k+1} - V^k \rangle + \frac{\rho}{2}\|V^k - V^{k+1}\|^2 \\
& + \rho\langle V^k - V^{k+1}, V^{k+1} - HU^{k+1} \rangle. \\
\end{align*}
Then combining (3.7) and (3.8) and using Lipschitz property of $\nabla h(V)$, we have
\begin{align*}
L(U^{k+1}, V^k, P^k) - L(U^{k+1}, V^{k+1}, P^{k+1}) & = h(V^k) - h(V^{k+1}) - \langle \nabla h(V^{k+1}), V^k - V^{k+1} \rangle \\
& + \frac{\rho}{2}\|V^{k+1} - V^k\|^2 - \frac{1}{\rho}\|P^{k+1} - P^k\|^2 \\
& \geq -l_h\|V^{k+1} - V^k\|^2 + \frac{\rho}{2}\|V^{k+1} - V^k\|^2 - \frac{l_h}{\rho}\|V^{k+1} - V^k\|^2 \\
& \geq \left(\frac{\rho}{2} - \frac{l_h}{\rho} - l_h\right)\|V^{k+1} - V^k\|^2. \\
\end{align*}
Finally, with the help of (3.9) and the above inequality, we obtain the result (3.6). \qed
The next lemma tells us that the Lagrangian is bounded below and the sequence \((U^k, V^k, V^k)\) is bounded.

**Lemma 3.2.** Under the same condition as Lemma 3.1, \(L(U^k, V^k, P^k)\) is lower bounded and \((U^k, V^k, P^k)\) is bounded.

**Proof.** Since \(\rho > 2l_h\), by the definition of \(L_a(U, V, P)\), we have
\[
L(U^k, V^k, P^k) = f(U^k) + h(V^k) + \langle P^k, V^k - HU^k \rangle + \frac{\rho}{2} \|V^k - HU^k\|^2
\]
\[
= f(U^k) + h(V^k) - \langle \nabla h(V^k), V^k - HU^k \rangle + \frac{\rho}{2} \|V^k - HU^k\|^2
\]
\[
\geq f(U^k) + h(HU^k) + \left(\frac{\rho}{2} - l_h\right) \|V^k - HU^k\|^2
\]
\[
\geq \frac{l_h}{\rho} \|V^k - HU^k\|^2 \geq 0.
\]

Lemma 3.1 tells us that \(L(U^k, V^k, P^k)\) is upper bounded, then \(\|V^k - HU^k\|\) is bounded. We know that \(0 \leq U^k \leq 1\), then \(HU^k\) is bounded. So \(V^k\) is bounded and \(P^k = -\nabla h(V^k)\) is bounded, too.

**Lemma 3.3.** The sequence \(L(U^k, V^k, P^k)\) is convergent and
\[
\lim_{k \to \infty} \|V^k - V^{k+1}\|_2 \to 0, \quad \lim_{k \to \infty} \|P^k - P^{k+1}\|_2 \to 0.
\]

**Proof.** Based on Lemmas 3.1 and 3.2, we know that the sequence \(\{L(U^k, V^k, P^k)\}\) is convergent, then by Lemma 3.1, we have \(\|V^k - V^{k+1}\|_2 \to 0 (k \to 0)\). Since \(P^k = -\nabla h(V^k)\), we have
\[
P^k - P^{k+1} = -\nabla h(V^k) + \nabla h(V^{k+1}).
\]

Because \(\nabla h(V)\) is Lipschitz continuous, we can get
\[
\lim_{k \to \infty} \|P^k - P^{k+1}\|_2 \to 0.
\]

The proof is complete.

We now establish the following two convergence results.

**Lemma 3.4.** For any \(k\), there is a \(\bar{d}^k \in \partial_U L(U^{k+1}, V^{k+1}, P^{k+1})\) such that
\[
\|\bar{d}^k\|_2 \to 0 \quad \text{as} \quad k \to \infty.
\]

**Proof.** By (3.4), we have
\[
\partial_U L(U^{k+1}, V^{k+1}, P^{k+1}) = \partial f(U^{k+1}) - H^T P^{k+1} - \rho H^T (V^{k+1} - HU^{k+1}).
\]

Let
\[
\Psi(U) = h(V^k) + \langle P^k, V^k - HU^k \rangle + \frac{\rho}{2} \|V^k - HU^k\|^2.
\]
From the optimality condition of $U^{k+1}$ and [22, Section 3.1.7]

$$d^k = -\nabla_U \Psi(U^{k+1}) = H^T P^k + \rho H^T (V^k - HU^{k+1}) \in \partial f(U^{k+1}).$$

Then

$$\bar{d}^k = d^k - H^T P^{k+1} - \rho H^T (V^{k+1} - HU^{k+1})$$
$$= H^T (P^k - P^{k+1}) + \rho H^T (V^k - V^{k+1}) \in \partial_U L(U^{k+1}, V^{k+1}, P^{k+1}),$$

by virtue of Lemma 3.1 and (3.2),

$$\|\bar{d}^k\|_2 \leq C^s (\|P^k - P^{k+1}\|_2 + \|V^k - V^{k+1}\|_2) \leq C\|V^k - V^{k+1}\|_2 \to 0.$$

The proof is complete.

**Lemma 3.5.** If $(U^*, V^*, P^*)$ is the limit point of a subsequence $(U^{k_s}, V^{k_s}, P^{k_s})$, then $L(U^*, V^*, P^*) = \lim_{s \to \infty} L(U^{k_s}, V^{k_s}, P^{k_s})$.

**Proof.** We first note that $L(U^{k_s}, V^{k_s}, P^{k_s})$ is monotonic nonincreasing, which implies the convergence of $L(U^{k_s}, V^{k_s}, P^{k_s})$. Since $L$ is lower semicontinuous, we have

$$\lim_{s \to \infty} L(U^{k_s}, V^{k_s}, P^{k_s}) \geq L(U^*, V^*, P^*). \quad (3.10)$$

Because the only potentially discontinuous terms in $L$ is $f(U)$, we have

$$\lim_{s \to \infty} L(U^{k_s}, V^{k_s}, P^{k_s}) - L(U^*, V^*, P^*) \leq \limsup_{s \to \infty} f(U^{k_s}) - f(U^*). \quad (3.11)$$

However, because $U^{k_s}$ is the optimal solution for the subproblem

$$\min_U L(U, V^{k_s-1}, P^{k_s-1}),$$

we know $L(U^{k_s}, V^{k_s-1}, P^{k_s-1}) \leq L(U^*, V^{k_s-1}, P^{k_s-1})$. Since the only discontinuous part of $L(U, V, P)$ is $f(U)$, we have

$$\limsup_{s \to \infty} L(U^{k_s}, V^{k_s-1}, P^{k_s-1}) - L(U^*, V^{k_s-1}, P^{k_s-1}) = \limsup_{s \to \infty} f(U^{k_s}) - f(U^*),$$

then $\limsup_{s \to \infty} f(U^{k_s}) - f(U^*) \leq 0$. Hence, by virtue of (3.10) and (3.11),

$$\lim_{s \to \infty} L(U^{k_s}, V^{k_s}, P^{k_s}) - L(U^*, V^*, P^*) = 0.$$

The proof is complete.

Finally, we establish the proof of Theorem 3.1.

**Proof.** By Lemma 3.2, the sequence $(U^k, V^k, P^k)$ is bounded, so there exists a convergent subsequence denoted by $(U^{k_s}, V^{k_s}, P^{k_s})$ and a limit point $(U^*, V^*, P^*)$ such that $(U^{k_s}, V^{k_s}, P^{k_s}) \to (U^*, V^*, P^*)$ as $s \to +\infty$. By Lemmas 3.1 and 3.2, $L(U^{k_s}, V^{k_s}, P^{k_s})$
is monotonically nonincreasing and lower bounded, and therefore \( \|V^{k_s} - V^{k_{s+1}}\| \to 0 \) as \( s \to 0 \). Based on Lemma 3.4, there exists \( \bar{d} \in \partial U(U^{k_s}, V^{k_s}, P^{k_s}) \) such that \( \|\bar{d}\| \to 0 \). Based on Lemma 3.5, \( L(U^*, V^*, P^*) = \lim_{s \to \infty} L(U^{k_s}, V^{k_s}, P^{k_s}) \). By definition of general subgradient, we have

\[
L(U, V^{k_s}, P^{k_s}) - L(U^{k_s}, V^{k_s}, P^{k_s}) - \langle \bar{d}^s, U - U^{k_s} \rangle \geq 0, \quad \forall U.
\]

Taking limit of \( s \), since \( L(U, V, P) \) is continuous with respect to \( V \) and \( P \), we have

\[
L(U, V^*, P^*) - L(U^*, V^*, P^*) \geq 0, \quad \forall U.
\]

Therefore, \( 0 \in \partial U(U^*, V^*, P^*) \). By the optimality of \( V^{k_s} \), we know that

\[
0 = \partial V L(U^{k_s}, V^{k_s}, P^{k_s-1}) = \nabla h(V^{k_s}) + P^{k_s-1} + \rho(V^{k_s} - HU^{k_s}).
\]

Clearly, \( \partial V L(U, V, P) \) is continuous with respect to \( U, V, P \), then by Lemma 3.3, we have

\[
\partial V L(U^*, V^*, P^*) = \nabla h(V^*) + P^* + \rho(V^* - HU^*) = \lim_{k \to \infty} \nabla h(V^{k_s}) + P^{k_s-1} + \rho(V^{k_s} - HU^{k_s}) = 0.
\]

Still by the optimality of \( V^{k_s} \),

\[
\rho(V^{k_s} - HU^{k_s}) = -\nabla h(V^{k_s}) - P^{k_s-1} = \nabla h(V^{k_s-1}) - \nabla h(V^{k_s}),
\]

then we have

\[
\partial P L(U^*, V^*, P^*) = V^* - HU^* = \lim_{s \to \infty} V^{k_s} - HU^{k_s} = 0.
\]

Here we use the Lipschitz property of \( \nabla h \) and Lemma 3.3. So \( 0 \in \partial L(U^*, V^*, P^*) \). \( \square \)

### 3.2. The ADMM algorithm with threshold truncation

#### 3.2.1. U-subproblem

For given \( V^k, P^k \), the U-update step in (3.3) reduces to solving

\[
U^{k+1} = \arg \min_U \|HU - W\|_2^2 + \beta_1 \|G_U\|_1 + \beta_2 \|U \odot (1 - U)\|_1,
\]

subject to \( 0 \leq U \leq 1 \), where

\[
W = V^k + \frac{1}{\rho} P^k.
\]

Let

\[
\Phi(U) = (\beta_1 G_U, \beta_2 U \odot (1 - U)),
\]

\[
0 \in \partial \Phi(U).
\]
we can solve problem (3.12) using split Bregman iteration method [13]. Namely, given $b^0, d^0$, for $s = 0, 1, \ldots$, we solve (3.12) with the following iterative procedure:

\[
\begin{align*}
\hat{U}^{s+1} &= \arg\min_U \|HU - W\|^2_2 + \frac{\gamma}{2}\|d^s - \Phi(U) - b^s\|^2_2, \\
d^{s+1} &= \arg\min_d \|d\|_1 + \frac{\gamma}{2}\|d - \Phi(\hat{U}^{s+1}) - b^s\|^2_2, \\
b^{s+1} &= b^s + \Phi(\hat{U}^{s+1}) - d^{s+1}.
\end{align*}
\] (3.14)

Denote

\[
F(U) = \|HU - W\|^2_2 + \frac{\gamma}{2}\|d^s - \Phi(U) - b^s\|^2_2
\]

\[
= \|HU - W\|^2_2 + \frac{\gamma}{2}(\|d_1^s - \beta_1 GU - b_1^s\|^2_2 + \|d_2^s - \beta_2 U \odot (1 - U) - b_2^s\|^2_2),
\]

then by a direct calculation, the gradient of $F(U)$ is

\[
\nabla F(U) = 2\text{Re}\{H^*(HU - W)\} - \gamma\beta_1 G^T(d_1^s - \beta_1 GU - b_1^s) + \gamma\beta_2 (d_2^s - \beta_2 U \odot (1 - U) - b_2^s) \odot (2U - 1).
\] (3.15)

The first problem is solved with gradient decent method. The initial value is $\tilde{U}^0 = U^k$, the iteration is

\[\tilde{U}^{m+1} = \tilde{U}^m - \eta_m \nabla F(\tilde{U}^m), \quad m = 0, 1, \ldots.\] (3.16)

Here the coefficient $\eta_m$ is given by the Armijo line search method. The Armijo line search method is an inexact line search method. It does not find the exact minimum but guarantees sufficient decrease. Armijo rule allows for the step-size to be computed using only local properties of the objective, as opposed to other approaches that use global quantities like the gradient’s Lipschitz constant. The selection process of step size works as following. Fix parameters $0 < \alpha < 0.5, 0 < \beta < 1$, then start with an initial step size $t = t_0$, when

\[
F(\tilde{U}^m - t\nabla F(\tilde{U}^m)) > F(\tilde{U}^m) - \alpha t\|\nabla F(\tilde{U}^m)\|^2,
\] (3.17)

we update step size as $t = \beta t$. We repeat this process until (3.17) fails, then choose $t$ as the step size $\eta_m$. More details of Armijo method could be find in [23, Chapter 3]. When the convergence is achieved at $m = M$, we let $\hat{U}^{s+1} = \tilde{U}^M$. To make the mask $U$ satisfy $0 \leq U \leq 1$, we project $\hat{U}^{s+1}$ onto $[0, 1]$ by

\[
\hat{U}^{s+1} = \min \left\{ \max \left\{ 0, \hat{U}^{s+1} \right\}, 1 \right\}.
\]

Finally, we update $d$ using the shrinkage method [13]. Namely, for each element of $d$, we write

\[
d^{s+1}_j = \text{shrink}(\Phi(\hat{U}^{s+1})_j + b^s_j, 1/\gamma), \quad j = 1, 2, \ldots, N^2,
\] (3.18)

with the shrink function defined by

\[
\text{shrink}(x, \lambda) = \frac{x}{|x|} \max(|x| - \lambda, 0).
\]
We stop the Bregman iteration (3.14) when \( n \) is equal to a given number \( S \) and let
\[
U^{k+1} = \hat{U}^{S+1}
\]
be the solution of (3.12).

### 3.2.2. V-subproblem

For given \( U^{k+1}, P^k \), after some calculations, the V-update step becomes to the following problem:
\[
V^{k+1} = \arg \min_V \| \text{Sig}_a(|V|^2) - I \|^2_2 + \frac{\rho}{2} \| V - W^k \|^2_2,
\]
(3.19)
where
\[
W^k = HU^{k+1} - \frac{1}{\rho}P^k.
\]

Fortunately, the V-subproblem could be decomposed to a series of single variable optimization problem
\[
V^{k+1}_i = \arg \min_{V_i} (\text{Sig}_a(|V_i|^2) - I_i)^2 + \frac{\rho}{2} (V_i - W^{k}_i)^2.
\]
(3.20)
As we have mentioned before, \( \text{Sig}_a \) is a smoothness of the truncation function. Then, to avoid the nonlinear iteration to solve (3.20), we solve the following problem directly:
\[
V^{k+1}_i = \arg \min_{V_i} (T(|V_i|^2) - I_i)^2 + \frac{\rho}{2} (V_i - W^{k}_i)^2.
\]
(3.21)

Where \( T \) is the operator defined in (2.2). Problem (3.21) has a closed-form solution
\[
V^{k+1}_i = \begin{cases}
W^k_i, & |W^k_i| < \sqrt{\text{tr}} \quad \text{and} \quad I_i = 0, \\
W^k_i, & |W^k_i| < \sqrt{\text{tr}} \quad \text{and} \quad I_i = 1 \quad \text{and} \quad \frac{\rho}{2} |W^k_i - \sqrt{\text{tr}}|^2 > 1, \\
W^k_i, & |W^k_i| > \sqrt{\text{tr}} \quad \text{and} \quad I_i = 0 \quad \text{and} \quad \frac{\rho}{2} |W^k_i - \sqrt{\text{tr}}|^2 > 1, \\
\sqrt{\text{tr}} \cdot \text{sgn}(W^k_i), & \text{otherwise}.
\end{cases}
\]
(3.22)

Fig. 5 shows the curves of objective function in (3.21) with \( W^k_i = 0.2, \rho = 1, \text{tr} = 0.3 \) and different \( I_i \). We can see that \( V^{k+1}_i = \sqrt{\text{tr}} \cdot \text{sgn}(W^k_i) = \sqrt{0.3} \) and \( V^{k+1}_i = W^k_i \) take the minimums for \( I_i = 1 \) and \( I_i = 0 \) respectively, which confirms the formula (3.22).

Now, as summary, we give the detailed Algorithm 3.1 for minimization problem (3.1).
Algorithm 3.1 The ADMM with threshold function approximation.

Input: Desired image $I$.
Output: Optimized mask $U$.
Parameters: $\beta_1, \beta_2, \rho, \gamma, \eta$.
Initialize: $U = I, V = H \ast U, P = 1$.

while error $>$ tol do
    Compute $W = V + P/\rho$.
    Initialize $d = \Phi(U), b = 0$.
    while error $>$ tol do
        Update $U$ with gradient decent method (3.16).
        Project $U$ onto $[0, 1]$:
        $$U = \min \{ \max \{0, U\}, 1\}.$$  
        Update $d = \text{shrink}(b + \Phi(U), 1/\gamma)$.
        Update $b = b + \Phi(U) - d$.
        Update $V$ with (3.22).
        Update $P$ with $P = P + \rho(V - HU)$.
    return $U$

4. Numerical examples

In this section, we will give some numerical examples to show the performance of our algorithm. The parameters in all our experiments are chosen as follows: the wavelength $\lambda = 193 \text{ nm}$, $\text{NA} = 0.85$, pixel size $= 5 \text{ nm}$, and the size of the point spread function $H$ is $100 \times 100$. We choose $tr = 0.3$ in the threshold function (2.2). The algorithm are implemented with MATLAB and all the numerical experiments are done on a laptop with 1.8 GHz CPU.
The edge placement error (EPE) is defined as a picture

$$EPE = |I(U) - I|,$$

which shows that the error occurs at the edge of the image and is an important property to evaluate the optimality of inversion. Then we can define the error

$$error = \|EPE\|_2$$

to measure the distance between the output image with the target image, here $\| \cdot \|_2$ represents the $l^2$ norm for vectorized image as mentioned in Section 3. We give numerical examples at the best focus $D = 0$ nm and with defocus $D = 50$ nm, respectively in (2.1). When the input is the target pattern, the output image is blurred. After optimization, the synthesized mask will have some small and isolated assist features, then the optimal mask achieves satisfactory output patterns where the corners of the image are well preserved.

4.1. Choosing parameters

In this example, we will discuss the influence of the parameters $\rho, \beta_1, \beta_2$ in (3.2) and $\gamma$ in (3.14). Fig. 6 shows the target image in this example. The size of the image is $144 \times 144$ and the blue part is valued 0, and the yellow represents value 1. Fig. 7 shows error decreasing in the optimization process under different $\rho$. We can see that with the increase of $\rho$, the decline is more consistent, but more gently. So $\rho = 10$ is a good choice. Fig. 8 gives the error decreasing for different $\gamma$. We can see that $\gamma = 30$ is a suitable choice. Then in the followings, we fix $\rho = 10$ and $\gamma = 30$. Fig. 9 shows the optimized masks with different $\beta_1$. Fig. 10 shows the optimized images with different $\beta_2$. These numerical examples suggest us to choose $\beta_1 = 0.01, \beta_2 = 0.015$ in the following examples. Here we remark that the choices are quite heuristic because it is quite hard to find the optimal parameters.

Figure 6: The target pattern.
4.2. Numerical experiments

Figs. 11-13 show the comparative results at the best focus $D = 0$ nm and under defocus $D = 50$ nm.

Fig. 11(a) shows the target pattern consisting of 10 squares. Fig. 11(b) display the
output patterns with target pattern as input. Fig. 11(c) is the edge placement error (EPE). Compared with the desired target pattern, we observe that the error is visually noticeable, especially for the regions located at the edge of each square. Fig. 11(d) shows the optimal mask generated by our proposed method at the best focus, Fig. 11(d) give the output pattern on wafer and Fig. 11(f) is the corresponding EPE. The last two rows displays the corresponding results under defocus $D = 50$ nm. As can be seen, some assist features near the squares are automatically generated in the minimization

Figure 11: Left column: mask, Middle column: output pattern, Right column: EPE.
process, which would help to improve the overall pattern fidelity. In addition, the corner-rounding effect at the end of each square is reduced compared to initial pattern, and the optimal mask gives the satisfactory contours in the output patterns.

Fig. 12 shows the example in which the target pattern contains long strips. From the first and third rows, we can find that the corners are round for the direct imaging with target pattern as input. By optimization with our proposed algorithm, the corners become sharp and the EPEs are significantly decreased, which can be found from the second and forth rows.
Fig. 13 shows the example in which the target pattern contains small squares and long strips. For this kind of mixed pattern, the algorithm still works well. From the second and forth rows, the small details can be imaged with optimized mask patterns, even for imaging system with defocus.

5. Conclusion

We have developed an ADMM method to solve the inverse lithography problem. The TV regularization is introduced to deal with ill-posedness of the inverse problem.
and sharpen the optimized mask pattern. To deal with the TV regularization term, we propose the ADMM framework. To avoid the complicated computation of gradient of sigmoid function in gradient descent method, we use the threshold function in the ADMM subproblem directly instead of sigmoid. With this replacement, we find that the optimal solution to the subproblem is explicit. Furthermore, we prove the convergence of our proposed ADMM algorithm. The parameters of the algorithm is chosen in a heuristic way. The numerical experiments illustrate the efficiency of the proposed algorithm.

Acknowledgments

The authors are very grateful to the referees for the constructive and helpful comments which lead to great improvement of the paper.

This research is supported partly by the National Key R&D Program of China (Grant Nos. 2019YFA0709600, 2019YFA0709602) and by the NSFC (Grant No. 11871300).

References

[1] R. A. Adams and J. F. Fournier, Sobolev Spaces, Elsevier, 2003.
[2] A. Chambolle, An algorithm for total variation minimization and applications, J. Math. Imaging Vision 20 (2004), 89–97.
[3] S. H. Chan and E. Y. Lam, Inverse image problem of designing phase shifting masks in optical lithography, in: Proceedings of the 15th IEEE International Conference on Image Processing, IEEE, (2008), 1832–1835.
[4] S. H. Chan, A. K. Wong, and E. Y. Lam, Initialization for robust inverse synthesis of phase-shifting masks in optical projection lithography, Opt. Express 16 (2008), 14746–14760.
[5] T. F. Chan, G. H. Golub, and P. Mulet, A nonlinear primal-dual method for total variation-based image restoration, SIAM J. Comput. 20 (1999), 1964–1977.
[6] T. F. Chan and J. Shen, Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods, SIAM, 2005.
[7] R. Chartrand and B. Wohlberg, A nonconvex ADMM algorithm for group sparsity with sparse groups, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, (2013), 6009–6013.
[8] Z. Chen and J. Zou, An augmented Lagrangian method for identifying discontinuous parameters in elliptic systems, SIAM J. Control Optim. 37(3) (1999), 892–910.
[9] S. K. Choy, N. Jia, C. S. Tong, M. L. Tang, and E. Y. Lam, A robust computational algorithm for inverse photomask synthesis in optical projection lithography, SIAM J. Imaging Sci. 5 (2012), 625–651.
[10] H. W. Engl, M. Hanke, and A. Neubauer, Regularization of Inverse Problems, Kluwer Academic Publishers, 1996.
[11] A. Ermann, T. Fuehner, T. Schnattinger, and B. Tollkuehn, Toward automatic mask and source optimization for optical lithography, in: Proceedings of the SPIE Optical Microlithography XVII, Vol. 5377 (2004), 646–657.
[12] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximation, Comput. Math. with Appl. 2(1) (1976), 17–40.
[13] T. Goldstein and S. Osher, The split Bregman method for L1-regularized problems, SIAM J. Imaging Sci. 2(2) (2009), 323–343.
[14] J. W. Goodman, Introduction to Fourier Optics, Roberts & Co. Publishers, 2005.
[15] Y. Granik, Fast pixel-based mask optimization for inverse lithography, J. Micro Nano-lithogr. MEMS MOEMS 5(4) (2006), 043002.
[16] R. Lai and S. Osher, A splitting method for orthogonality constrained problems, J. Sci. Comput. 58 (2014), 431–449.
[17] J. Lih, S. Liu, and E. Y. Lam, Efficient source and mask optimization with augmented Lagrangian methods in optical lithography, Opt. Express, 21(7) (2013), 8076–8090.
[18] X. Ma and G. R. Arce, Generalized inverse lithography methods for phase-shifting mask design, Opt. Express 19 (2011), 2165–2180.
[19] X. Ma, Z. Wang, J. Zhu, S. Zhang, G. R. Arce, and S. Zhao, Nonlinear compressive inverse lithography aided by low-rank regularization, Opt. Express 27 (2019), 29992–30008.
[20] Y. Nesterov, Lectures on Convex Optimization, in: Springer Optimization and Its Applications, Vol. 137, Springer, 2018.
[21] J. Nocedal and S. J. Wright, Numerical Optimization, in: Springer Series in Operations Research and Financial Engineering, Springer, 2006.
[22] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, An iterative regularization method for total variation-based image restoration, Multiscale Model. Simul. 4 (2005), 460–489.
[23] L. Pang, Inverse lithography technology: 30 years from concept to practical, full-chip reality, Journal of Micro/Nanopatterning, Materials, and Metrology 20 (2021), 030901.
[24] L. Pang, Y. Liu, and D. Abrams, Inverse lithography technology (ILT): What is the impact to the photomask industry?, in: Proceedings of SPIE Photomask and Next-Generation Lithography Mask Technology XIII, Vol. 6283, 2006.
[25] Y. Peng, J. Zhang, Y. Wang, and Z. Yu, Gradient-based source and mask optimization in optical lithography, IEEE Trans. Image Process. 20 (2011), 2856–2864.
[26] A. Poonawala and P. Milanfar, OPC and PSM design using inverse lithography: A nonlinear optimization approach, Optical Microlithography XIX, 6154 (2006), 1159–1172.
[27] A. Poonawala and P. Milanfar, Double-exposure mask synthesis using inverse lithography, J. Micro Nanolithogr. MEMS MOEMS 6 (2007), 043001.
[28] A. Poonawala and P. Milanfar, Mask design for optical microlithography – an inverse imaging problem, IEEE Trans. Image Process. 16(3) (2007), 774–788.
[29] A. Poonawala, B. Painter, and J. Mayhew, Model-based assist feature placement: An inverse imaging approach, in: Proceedings of SPIE Photomask Technology 2008, Vol. 7122 (2008), 277–286.
[30] Y. Shen, F. Peng, and Z. Zhang, Semi-implicit level set formulation for lithographic source and mask optimization, Opt. Express 27 (2019), 29659–29668.
[31] Y. Shen, N. Wong, and E. Y. Lam, Level-set-based inverse lithography for photomask synthesis, Opt. Express 17 (2009), 23690–23701.
[32] D. Strong and T. Chan, Edge-preserving and scale-dependent properties of total variation regularization, Inverse Problems 19 (2003), S165.
[33] H. Sun, J. Du, C. Jin, J. Feng, J. Wang, S. Hu, and J. Liu, Global source optimization based on adaptive nonlinear particle swarm optimization algorithm for inverse lithography,
IEEE Photon. J. 13 (2021), 1–7.

[36] Y. WANG, W. YIN, AND J. ZENG, Global convergence of ADMM in nonconvex nonsmooth optimization, J. Sci. Comput. 78 (2019), 29–63.

[37] Z. WEN, C. YANG, X. LIU, AND S. MARCHESINI, Alternating direction methods for classical and ptychographic phase retrieval, Inverse Problems, 28 (2012), 115010.

[38] A. K.-K. WONG, Resolution enhancement techniques in optical lithography, in: SPIE Tutorial Texts in Optical Engineering, Vol. 47, SPIE Press, 2001.

[39] A. K.-K. WONG, Optical imaging in projection microlithography, in: SPIE Tutorial Texts in Optical Engineering, Vol. 66, SPIE Press, 2005.

[40] H. YANG, S. LI, Z. DENG, Y. MA, B. YU, AND E. F. YOUNG, GAN-OPC: Mask optimization with lithography-guided generative adversarial nets, IEEE T. Comput. Aid. D. 39(10) (2020), 2822–2834.

[41] H. YANG, Z. LI, K. SASTRY, S. MUKHOPADHYAY, M. KILGARD, A. ANANDKUMAR, B. KHALANY, V. SINGH, AND H. REN, Generic lithography modeling with dual-band optics-inspired neural networks, arXiv:2203.08616 (2022).