# RECURSIVE LINEARIZATION METHOD FOR INVERSE MEDIUM SCATTERING PROBLEMS WITH COMPLEX MIXTURE GAUSSIAN ERROR LEARNING 

JUNXIONG JIA, BANGYU WU, JIGEN PENG, AND JINGHUAI GAO


#### Abstract

This paper is concerned with the numerical errors appeared in the calculation of inverse medium scattering problems (IMSP). Optimization based iterative methods are widely employed to solve IMSP, which are computationally intensive due to a series of Helmholtz equations need to be solved numerically. Hence, rough approximations of Helmholtz equations can significantly speed up the iterative procedure. However, rough approximations will lead to instability and inaccurate estimation. Inspired by mixture Gaussian error construction used widely in the machine learning community, we model numerical errors brought by rough forward solver as some complex mixture Gaussian (CMG) random variables. Based on this assumption, a new nonlinear optimization problem is derived by using infinite-dimensional Bayes' inverse method. Then, we generalize the real valued expectation-maximization (EM) algorithm to our complex valued case to learn parameters in the CMG distribution. Next, we generalize the recursive linearization method (RLM) to a new iterative method named as mixture Gaussian recursive linearization method (MGRLM) which consists of two stages: 1) learn CMG; 2) solve IMSP. Through the learning stage, numerical errors and some prior knowledge of the true scatterer have been incorporated into the proposed optimization problem. Hence, both the convergence speed and the resolution of the obtained result can be enhanced in the second stage. Finally, we provide two numerical examples to illustrate the effectiveness of the proposed method.


## 1. Introduction

Scattering theory has played a central role in the field of mathematical physics, which is concerned with the effect that an inhomogeneous medium has on an incident particle or wave [16. Usually, the total field is regarded as the sum of an incident field and a scattered field. Then, the inverse scattering problems focus on determining the nature of the inhomogeneity from knowledge of the scattered field [9, 15], which have played important roles in diverse scientific areas such as radar and sonar, geophysical exploration, medical imaging and nano-optics.

Deterministic computational methods for inverse scattering problems can be classified into two categories: nonlinear optimization based iterative methods [4, 33, 34 ] and imaging based direct methods [11, 37]. Direct methods are called qualitative methods which need no direct solvers and visualize the scatterer by highlighting its boundary with designed imaging functions. Iterative methods are usually called

[^0]quantitative methods, which aim at providing some functions to represent the scatterer. Because a sequence of direct and adjoint scattering problems need to be solved, the quantitative methods are computationally intensive.

This paper is concerned with the nonlinear optimization based iterative methods, especially focus on the recursive linearization method (RLM) for inverse medium scattering problems (IMSP) reviewed in 4. In order to obtain global minima in optimization, RLM was proposed in [13, 14, which requires high computational resources. Then, new and efficient RLMs have been proposed in a series of papers [2, 3, 4, 5, 6, 7, which used the differential equation formulation and highly reduced the computational cost. However, a series of forward solver need to be solved and the accuracy of the forward solver is critical for obtaining an acceptable computational results, particularly for applications in seismic exploration [22] and medical imaging [30. A lot of efficient forward solvers based on finite difference method, finite element methods and spectral methods have been proposed [38, 40]. Here, we will not propose a new forward solver to reduce the computational cost, but attempt to introduce the ideas originated from the machine learning community. Specifically speaking, our idea comes from mixture Gaussian based error construction method which is an active research direction in the field of machine learning [8, 32, 42 , 43, 44. By combining mixture Gaussian error learning process with RLM, we expect to obtain high resolution results when rough approximate forward solvers are employed.

To provide a clear illustration, let us provide the abstract formulation of IMSP in the following. Denote $X$ to be some separable Banach space, then the forward problem usually modeled as follows

$$
\begin{equation*}
d=\mathcal{F}(m)+\epsilon, \tag{1.1}
\end{equation*}
$$

where $d \in \mathbb{C}^{N_{d}}\left(N_{d} \in \mathbb{N}^{+}\right)$stands for the measured data, $m \in X$ represents the interested parameter and $\epsilon$ denotes noise. For inverse scattering problems, $m$ is just the scatterer, $\mathcal{F}$ represents the solution operator of a Helmholtz equation combined with some measurement operator. The nonlinear optimization based iterative methods just formulate inverse problem as follows

$$
\begin{equation*}
\min _{m \in X}\left\{\frac{1}{2}\|d-\mathcal{F}(m)\|_{2}^{2}+\mathcal{R}(m)\right\} \tag{1.2}
\end{equation*}
$$

where $\mathcal{R}(\cdot)$ stands for some regularization operator and $\|\cdot\|_{2}$ represents the $\ell^{2}$-norm.
In real world applications, we would like to use a fast forward solver (limited accuracy) to obtain an estimation as accurately as possible. Hence, the noise is usually not only brought by inaccurate measurements but also induced by a rough forward solver and inaccurate physical assumptions [12, 30]. Following [29, 30, let us denote $\mathcal{F}_{a}(\cdot)$ to be the forward operator related to some rough forward solver, then (1.1) can be rewritten as follows

$$
\begin{equation*}
d=\mathcal{F}_{a}(m)+\left(\mathcal{F}(m)-\mathcal{F}_{a}(m)\right)+\epsilon \tag{1.3}
\end{equation*}
$$

Denoting $\xi:=\left(\mathcal{F}(m)-\mathcal{F}_{a}(m)\right)$, we find that

$$
\begin{equation*}
d=\mathcal{F}_{a}(m)+\xi+\epsilon \tag{1.4}
\end{equation*}
$$

Following the mixture Gaussian based error learning, we adopt Bayes' inference method and the maximum a posteriori (MAP) estimate to deduce our nonlinear
optimization problem. Hence, we model $\xi$ as a random variable which obviously has the following two important features

- $\xi$ depend on the unknown function $m$;
- $\xi$ may distribute according to a complicated probability measure.

For the first feature, we relax this tough problem to assume that $\xi$ is independent of $m$, but the probability distribution of $\xi$ and the prior probability measure of $m$ are related with each other [31]. For the second feature, to the best of our knowledge, only Gaussian assumption has been studied in the field of inverse problems with partial differential equations [27, 30]. Inspired by the mixture Gaussian error learning, we provide a more realistic assumption about the random variable $\xi$.

Before going further, let us introduce the density function of mixture Gaussian distribution as follow

$$
\begin{equation*}
\sum_{k=1}^{K} \pi_{k} \mathcal{N}\left(\cdot \mid \zeta_{k}, \Sigma_{k}\right) \tag{1.5}
\end{equation*}
$$

where $\mathcal{N}\left(\cdot \mid \zeta_{k}, \Sigma_{k}\right)$ stands for a Gaussian probability density function with mean value $\zeta_{k}$ and covariance matrix $\Sigma_{k}$ and for every $k, \pi_{k} \in(0,1)$ satisfy $\sum_{k=1}^{K} \pi_{k}=$ 1. Employing the ideas originated from the mixture Gaussian error learning, we model $\xi$ as a random variable obeying a mixture Gaussian distribution which can approximate any probability distribution in some sense [8. Then, considering the infinite-dimensional nature of our problem (usually finite-dimensional for machine learning problems), we generalize Bayes' inverse method developed in [17, 18, 26, 29, 39 to our case. Using general theory of the MAP estimate developed for white Gaussian noise in [10, 19, 21, we prove the validity of the MAP estimate in our case, which naturally leads to the following nonlinear optimization problem

$$
\begin{equation*}
\min _{m \in X}\left\{-\ln \left(\sum_{k=1}^{K} \pi_{k} \mathcal{N}\left(d-\mathcal{F}_{a}(m) \mid \zeta_{k}, \Sigma_{k}+\nu I\right)\right)+\mathcal{R}(m)\right\} \tag{1.6}
\end{equation*}
$$

where we assumed that the measurement noise $\epsilon$ is a Gaussian random variable with mean zero and covariance matrix $\nu I\left(\nu \in \mathbb{R}^{+}\right.$and $I$ is an identity matrix). Through introducing the responsibilities appeared in some learning algorithms, we can use the main ideas shown in Section 2.5 of [4] to deduce the adjoint problem.

In the field of machine learning, there are usually a lot of sampling data and the forward problems are not computationally intensive compared with the IMSP. Hence, they use alternative iterative methods to find the optimal solution and estimate the modeling error simultaneously 43. However, considering the lack of learning data and the high computational cost of our forward problems, we cannot trivially generalize their alternative iterative methods to our case. Actually, the proposed method named as mixture Gaussian recursive linearization method (MGRLM) includes two stages: 1) learn mixture Gaussian distribution; 2) solve IMSP.

The first stage is based on an essential assumption that is we have some learning examples obtained from historical data or constructed by some prior knowledge. Here, we would like to clarify three confusing points. Firstly, we assume that the learning examples are generated from a probability measure which can reflect some key features of the true scatterer. However, the probability measure may not capture the full features of the true scatterer, that means the true scatterer may not be a sample from the probability measure. Secondly, this stage is highly depending
on the real problems we faced, so this stage should be performed for different application areas. But the learning stage only performs once for a certain application area, hence, the learning stage will not consume computational resources when we solve IMSP for a specific problem. Thirdly, since the learning data contains some information about the true scatterer, some prior knowledge will also be incorporated into the mixture Gaussian distribution, which accelerates the convergence speed in stage 2 .

The second stage consists of a modified RLM deduced from optimization problem (1.6). With the parameters $\left\{\pi_{k}, \zeta_{k}, \Sigma_{k}\right\}$ learned in the first stage for a certain application area, both the convergence speed and the resolution of the computational results can be enhanced.

In the end, we briefly review the main contributions as follows:

- Inspired by the studies on error learning, we firstly introduce mixture Gaussian based error learning to enhance the performance of RLM for IMSP.
- We derive nonlinear-optimization problem (1.6) rigorously from infinite dimensional Bayes' inverse method.
- Relying on the relations between real valued Gaussian distribution and complex valued Gaussian distribution, we rigorously deduce the mean and covariance estimation formulas in EM algorithm for complex valued variables. From nonlinear-optimization problem (1.6), we derive a modified RLM.
The outline of this paper is as follows. In Section 2, new nonlinear optimization problems are obtained by using infinite-dimensional Bayes' inverse method. During the deduction, well-posedness of the posterior measure and MAP estimate with mixture Gaussian noise are established. Then, the general theory is applied to the inverse medium scattering problem under some appropriate conditions. In Section 3, we firstly generalize the real valued expectation-maximization (EM) algorithm to the complex Gaussian case. Secondly, we provide the adjoint equation and derive the Fréchet derivative of the objective functional. In Section 4, two typical numerical examples are given, which illustrate the effectiveness of the proposed method.


## 2. Model Derivation

In this section, we deduce the optimization problem (1.6) from infinite-dimensional Bayes' inverse method, which provides a foundation for constructing the learning algorithm and the modified RLM. The presentation has been divided into two subsections. Firstly, we provide a general theory based on the infinite-dimensional Bayes' framework [17, 18, 39, 19. Then, the general theory has been applied to IMSP under appropriate conditions.

Before diving into the main contents, let us provide a brief notation list which will be used in all of the following parts of this paper.

## Notations:

- For an integer $N$, denote $\mathbb{C}^{N}$ as $N$-dimensional complex vector space; $\mathbb{R}^{+}$ and $\mathbb{N}^{+}$represent positive real numbers and positive integers respectively;
- For a Banach space $X,\|\cdot\|_{X}$ stands for the norm defined on $X$ and, particularly, $\|\cdot\|_{2}$ represents the $\ell^{2}$-norm of $\ell^{2}$ space.
- For a matrix $\Sigma$, denote its determinant as $\operatorname{det}(\Sigma)$;
- Denote $B(m, R)$ as a ball with center $m$ and radius $R$. Particularly, denote $B_{R}:=B(0, R)$ when the ball is centered at the origin;
- Denote $X$ and $Y$ to be some Banach spaces; For an operator $F: X \rightarrow Y$, denote $F^{\prime}\left(x_{0}\right)$ as the Fréchet derivative of $F$ at $x_{0} \in X$.
- Denote $\operatorname{Re}(\xi), \operatorname{Imag}(\xi), \xi^{T}, \xi^{H}$ and $\bar{\xi}$ as the real part, imaginary part, transpose, conjugate transpose and complex conjugate of $\xi \in \mathbb{C}^{N}$ respectively;
- The notation $\eta \sim p(\eta)$ stands for a random variable $\eta$ obeys the probability distribution with density function $p(\cdot)$.
2.1. General theory. Let $\mathcal{N}_{c}(\eta \mid \zeta, \Sigma)$ represents the density function of $N_{d}$-dimensional complex valued Gaussian probability distribution 23] defined as follows

$$
\begin{equation*}
\mathcal{N}_{c}(\eta \mid \zeta, \Sigma):=\frac{1}{(\pi)^{N_{d}} \operatorname{det}(\Sigma)} \exp \left(-\|\eta-\zeta\|_{\Sigma}^{2}\right) \tag{2.1}
\end{equation*}
$$

where $\zeta$ is a $N_{d}$-dimensional complex valued vector, $\Sigma$ is a positive definite Hermitian matrix and $\|\cdot\|_{\Sigma}^{2}$ is defined as follow

$$
\begin{equation*}
\|\eta-\zeta\|_{\Sigma}^{2}:=(\eta-\zeta)^{H} \Sigma^{-1}(\eta-\zeta) \tag{2.2}
\end{equation*}
$$

with the superscript $H$ stands for conjugate transpose. Denote $\eta:=\xi+\epsilon$, then formula 1.4 can be written as follows

$$
\begin{equation*}
d=\mathcal{F}_{a}(m)+\eta \tag{2.3}
\end{equation*}
$$

where

$$
\begin{equation*}
d \in \mathbb{C}^{N_{d}}, \quad \eta \sim \sum_{k=1}^{K} \pi_{k} \mathcal{N}_{c}\left(\eta \mid \zeta_{k}, \Sigma_{k}+\nu I\right) \tag{2.4}
\end{equation*}
$$

with $N_{d}, K$ denote some positive integers and $\nu \in \mathbb{R}^{+}$.
Before going further, let us provide the following basic assumptions about the approximate forward operator $\mathcal{F}_{a}$.

## Assumption 1.

(1) for some $\epsilon>0$ there is $M=M(\epsilon) \in \mathbb{R}, C \in \mathbb{R}$ such that, for all $m \in X$,

$$
\left\|\mathcal{F}_{a}(m)\right\|_{2} \leq C \exp \left(\epsilon\|m\|_{X}^{2}+M\right)
$$

(2) for every $r>0$ there is $L=L(r)>0$ such that, for all $m \in X$ with $\|m\|_{X}<r$, we have

$$
\left\|\mathcal{F}_{a}^{\prime}(m)\right\|_{o p} \leq L
$$

where $\|\cdot\|_{o p}$ denotes the operator norm.
At this stage, we need to provide some basic notations of the Bayesian inverse method when $m$ in some infinite-dimensional space. Following the work [17, 39], let $\mu_{0}$ stands for the prior probability measure defined on a separable Banach space $X$ and denote $\mu^{d}$ to be the posterior probability measure. Then the Bayes' formula may be written as follows

$$
\begin{align*}
\frac{d \mu^{d}}{d \mu_{0}}(m) & =\frac{1}{Z(d)} \exp (\Phi(m ; d))  \tag{2.5}\\
Z(d) & =\int_{X} \exp (\Phi(m ; d)) \mu_{0}(d m) \tag{2.6}
\end{align*}
$$

where $\frac{d \mu^{d}}{d \mu_{0}}(\cdot)$ represents the Radon-Nikodym derivative and

$$
\Phi(m ; d):=\ln \left\{\sum_{k=1}^{K} \pi_{k} \frac{1}{\pi^{N_{d}} \operatorname{det}\left(\Sigma_{k}+\nu I\right)} \exp \left(-\left\|d-\mathcal{F}_{a}(m)-\zeta_{k}\right\|_{\Sigma_{k}+\nu I}^{2}\right)\right\}
$$

In the following, we provide a theorem that gives a rigorous formulation of 2.5 and 2.6.
Theorem 2.1. Let Assumption 1 hold for some $\epsilon, r, L$ and $M$. Assume that $X$ is some separable Banach space, $\mu_{0}(X)=1$ and that $\mu_{0}(X \cap B)>0$ for some bounded set $B$ in $X$. In addition, we assume $\int_{X} \exp \left(2 \epsilon\|m\|_{X}^{2}\right) \mu_{0}(d m)<\infty$. Then, for every $d \in \mathbb{C}^{N_{d}}, Z(d)$ given by (2.6) is positive and the probability measure $\mu^{d}$ given by (2.5) is well-defined. In addition, there is $C=C(r)>0$ such that, for all $d_{1}, d_{2} \in B(0, r)$

$$
d_{\mathrm{Hell}}\left(\mu^{d_{1}}, \mu^{d_{2}}\right) \leq C\left\|d_{1}-d_{2}\right\|_{2}
$$

where $d_{\text {Hell }}(\cdot, \cdot)$ denotes the Hellinger distance defined for two probability measures.
The proof of this theorem is postponed to Appendix.
Remark 2.2. The assumptions of the prior probability measure are rather general, which include Gaussian probability measure and TV-Gaussian probability measure [41] for certain space $X$.

In the last part of this subsection, we prove the validity of the MAP estimate which links Bayes' inverse method and regularization method. Firstly, let us assume that the prior probability measure $\mu_{0}$ is a Gaussian probability measure and define the following functional

$$
J(m)= \begin{cases}-\Phi(m ; d)+\frac{1}{2}\|m\|_{E}^{2} & \text { if } m \in E, \text { and }  \tag{2.7}\\ +\infty, & \text { else. }\end{cases}
$$

Here $\left(E,\|\cdot\|_{E}\right)$ denotes the Cameron-Martin space associated to $\mu_{0}$. In infinite dimensions, we adopt small ball approach constructed in 19. For $m \in E$, let $B(m, \delta) \in X$ be the open ball centred at $m \in X$ with radius $\delta$ in $X$. Then, we can prove the following theorem which encapsulates the idea that probability is maximized where $J(\cdot)$ is minimized.

Theorem 2.3. Let Assumption 1 hold and assume that $\mu_{0}(X)=1$. Then the function $J(\cdot)$ defined by (2.7) satisfies, for any $m_{1}, m_{2} \in E$,

$$
\lim _{\delta \rightarrow 0} \frac{\mu^{d}\left(B\left(m_{1}, \delta\right)\right)}{\mu^{d}\left(B\left(m_{2}, \delta\right)\right)}=\exp \left(J\left(m_{2}\right)-J\left(m_{1}\right)\right)
$$

The proof of this theorem is postponed to Appendix.
Now, if we assume $\mu_{0}$ is a TV-Gaussian probability measure, then we can define the following functional

$$
J(m)= \begin{cases}-\Phi(m ; d)+\lambda\|m\|_{\mathrm{TV}}+\frac{1}{2}\|m\|_{E}^{2} & \text { if } m \in E, \text { and }  \tag{2.8}\\ +\infty, & \text { else. }\end{cases}
$$

Using similar methods as for the Gaussian case and the above functional 2.8, we can prove a similar theorem to illustrate that the MAP estimate is also the minimal solution of $\min _{m \in X} J(m)$.
2.2. Applications to IMSP. Before applying the general theory to IMSP, let us provide some basic settings of the inverse scattering problem considered in this paper. In the following, we usually assume that the total field $u$ satisfies

$$
\begin{equation*}
\Delta u+\kappa^{2}(1+m) u=0 \quad \text { in } \mathbb{R}^{2} \tag{2.9}
\end{equation*}
$$

where $\kappa>0$ is the wavenumber, and $m(\cdot)$ is a real function known as the scatterer representing the inhomogeneous medium. As in 4], we assume that the scatterer is compactly supported and the support contained in the ball $B_{R}=\left\{\mathbf{r} \in \mathbb{R}^{2}\right.$ : $\left.\|\mathbf{r}\|_{2}<R\right\}$ with boundary $\partial B_{R}=\left\{\mathbf{r} \in \mathbb{R}^{2}:\|\mathbf{r}\|_{2}=R\right\}$, and satisfies $-1<$ $m_{\min } \leq m \leq m_{\max }<\infty$, where $m_{\min }$ and $m_{\max }$ are two constants. Denote $\mathbf{d}=(\cos \theta, \sin \theta) \in\left\{\mathbf{r} \in \mathbb{R}^{2}:\|\mathbf{r}\|_{2}=1\right\}$ as the incident direction with $\theta \in(0,2 \pi)$ is the incident angle, we assume that the scatterer is illuminated by a plane incident field

$$
\begin{equation*}
u^{\mathrm{inc}}(\mathbf{r})=e^{i \kappa \mathbf{r} \cdot \mathbf{d}} \tag{2.10}
\end{equation*}
$$

Apparently, the incident field satisfies

$$
\begin{equation*}
\Delta u^{\mathrm{inc}}+\kappa^{2} u^{\mathrm{inc}}=0 \quad \text { in } \mathbb{R}^{2} \tag{2.11}
\end{equation*}
$$

Combining the incident field $u^{\text {inc }}$ and the scattered field $u^{s}$, we obtain the total field $u$ as follow

$$
\begin{equation*}
u=u^{\mathrm{inc}}+u^{s} . \tag{2.12}
\end{equation*}
$$

It follows from $2.9,2.11$ and 2.12 that the scattered field satisfies

$$
\begin{equation*}
\Delta u^{s}+\kappa^{2}(1+m) u^{s}=-\kappa^{2} m u^{\mathrm{inc}} \quad \text { in } \mathbb{R}^{2} \tag{2.13}
\end{equation*}
$$

with the following Sommerfeld radiation condition

$$
\begin{equation*}
\lim _{\|\mathbf{r}\|_{2} \rightarrow \infty} r^{1 / 2}\left(\partial_{r} u^{s}-i \kappa u^{s}\right)=0 \tag{2.14}
\end{equation*}
$$

where $r=\|\mathbf{r}\|_{2}$. In the following, we suppose that the scatterer $m(\cdot)$ appeared in (2.9) has compact support and $\operatorname{supp}(m) \subset \Omega \subset B_{R}$ where $\Omega$ is a square region.

Because the scatterer $m(\cdot)$ is assumed to have compact support, the problem (2.13) and (2.14) defined on $\mathbb{R}^{2}$ can be reformulated to the following problem defined on bounded domain (4)

$$
\left\{\begin{array}{l}
\Delta u^{s}+\kappa^{2}(1+m) u^{s}=-\kappa^{2} m u^{\mathrm{inc}} \quad \text { in } B_{R}  \tag{2.15}\\
\partial_{\mathbf{n}} u^{s}=\mathcal{T} u^{s} \quad \text { on } \partial B_{R}
\end{array}\right.
$$

where $\mathcal{T}$ is the Dirichlet-to-Neumann ( DtN ) operator defined exactly as (2.8) in 4 .
For problem 2.15, we define the map $\mathcal{S}(m, \kappa) u^{\text {inc }}$ by $u^{s}=\mathcal{S}(m, \kappa) u^{\text {inc }}$. From [1, 16], we easily know that the following estimate holds for equations 2.15)

$$
\begin{equation*}
\left\|u^{s}\right\|_{H^{2}(\Omega)} \leq C\|m\|_{L^{\infty}(\Omega)}\left\|u^{\mathrm{inc}}\right\|_{L^{2}(B(0, R))} \tag{2.16}
\end{equation*}
$$

Considering Sobolev embedding theorem, we can define the following measurement operator

$$
\begin{equation*}
\mathcal{M}\left(\mathcal{S}(m, \kappa) u^{\mathrm{inc}}\right)(x)=\left(u^{s}\left(x_{1}\right), \ldots, u^{s}\left(x_{N_{d}}\right)\right)^{T} \tag{2.17}
\end{equation*}
$$

where $x_{i} \in \partial \Omega, i=1,2, \ldots, N_{d}$, are the points where the wave field $u^{s}$ is measured.
In practice, we employ a uniaxial PML technique to transform the problem defined on the whole domain to a problem defined on a bounded rectangular domain [1], as seen in Figure 1 Let $D$ be the rectangle which contain $\Omega=\left[x_{1}, x_{2}\right] \times\left[y_{1}, y_{2}\right]$ with $\operatorname{supp}(m) \subset \Omega$ and let $d_{1}$ and $d_{2}$ be the thickness of the PML layers along $x$


Figure 1. Geometry of the scattering problem with a uniaxial PML layer.
and $y$, respectively. Let $s_{1}(x)=1+i \sigma_{1}(x)$ and $s_{2}(y)=1+i \sigma_{2}(y)$ be the model medium property and usually we can simply take

$$
\sigma_{1}(x)= \begin{cases}\sigma_{0}\left(\frac{x-x_{2}}{d_{1}}\right)^{p} & \text { for } x_{2}<x<x_{2}+d_{1} \\ 0 & \text { for } x_{1} \leq x \leq x_{2} \\ \sigma_{0}\left(\frac{x_{1}-x}{d_{1}}\right)^{p} & \text { for } x_{1}-d_{1}<x<x_{1}\end{cases}
$$

and

$$
\sigma_{2}(y)= \begin{cases}\sigma_{0}\left(\frac{y-y_{2}}{d_{2}}\right)^{p} & \text { for } y_{2}<y<y_{2}+d_{2} \\ 0 & \text { for } y_{1} \leq y \leq y_{2} \\ \sigma_{0}\left(\frac{y_{1}-y}{d_{2}}\right)^{p} & \text { for } y_{1}-d_{2}<y<y_{1}\end{cases}
$$

where the constant $\sigma_{0}>1$ and the integer $p \geq 2$. Denote

$$
s=\operatorname{diag}\left(s_{2}(x) / s_{1}(x), s_{1}(x) / s_{2}(y)\right),
$$

then the truncated PML problem can be defined as follow

$$
\left\{\begin{array}{l}
\nabla \cdot\left(s \nabla u^{s}\right)+s_{1} s_{2} \kappa^{2}(1+m) u^{s}=-\kappa^{2} m u^{\mathrm{inc}} \quad \text { in } D  \tag{2.18}\\
u^{s}=0 \quad \text { on } \partial D
\end{array}\right.
$$

Similar to the physical problem 2.15), we introduce the map $\mathcal{S}_{a}(m, \kappa)$ defined by $u_{a}^{s}=\mathcal{S}_{a}(m, \kappa) u^{\mathrm{inc}}$ where $u_{a}^{s}$ stands for the solution of the truncated PML problem 2.18). Through similar methods for equations 2.15, we can prove that $u_{a}^{s}$ is a continuous function and satisfies

$$
\begin{equation*}
\left\|u_{a}^{s}\right\|_{L^{\infty}(D)} \leq C\|m\|_{L^{\infty}(D)}\left\|u^{\mathrm{inc}}\right\|_{L^{2}(D)} \tag{2.19}
\end{equation*}
$$

Now, we can define the measurement operator similar to (2.17) as follow

$$
\begin{equation*}
\mathcal{M}\left(\mathcal{S}_{a}(m, \kappa) u^{\mathrm{inc}}\right)(x)=\left(u_{a}^{s}\left(x_{1}\right), \ldots, u_{a}^{s}\left(x_{N_{d}}\right)\right)^{T} \tag{2.20}
\end{equation*}
$$

where $x_{i} \in \partial D, i=1,2, \ldots, N_{d}$.
In order to introduce appropriate Gaussian probability measures, we present the following assumptions proposed in [20].

Assumption 2. Denote $A$ to be an operator, densely defined on the Hilbert space $\mathcal{H}=L^{2}\left(D ; \mathbb{R}^{d}\right)$, satisfies the following properties:
(1) $A$ is positive-definite, self-adjoint and invertible;
(2) the eigenfunctions $\left\{\varphi_{j}\right\}_{j \in \mathbb{N}}$ of $A$, form an orthonormal basis for $\mathcal{H}$;
(3) the eigenvalues satisfy $\alpha_{j} \asymp j^{2 / d}$, for all $j \in \mathbb{N}$;
(4) there is $C>0$ such that

$$
\sup _{j \in \mathbb{N}}\left(\left\|\varphi_{j}\right\|_{L^{\infty}}+\frac{1}{j^{1 / d}} \operatorname{Lip}\left(\varphi_{j}\right)\right) \leq C
$$

where $\operatorname{Lip}\left(\varphi_{j}\right)$ represents the Lipschitz constant of the function $\varphi_{j}$.
At this moment, we can show well-posedness for inverse medium scattering problem with some Gaussian prior probability measures. For a constant $s>1$, we consider the prior probability measure to be a Gaussian measure $\mu_{0}:=\mathcal{N}\left(\bar{m}, A^{-s}\right)$ where $\bar{m}$ is the mean value and the operator $A$ satisfies Assumption 2. In addition, we take $X=C^{t}$ with $t<s-1$. Then we know that $\mu_{0}(X)=1$ by Theorem 12 shown in [20].

For the scattering problem, we can take $\mathcal{F}_{a}(m)=\mathcal{M}\left(\mathcal{S}_{a}(m, \kappa) u^{\text {inc }}\right)$ and let the noise $\eta$ obeys a mixture Gaussian distribution with density function

$$
\sum_{k=1}^{K} \pi_{k} \mathcal{N}_{c}\left(\eta \mid \zeta_{k}, \Sigma_{k}+\nu I\right)
$$

Then, the measured data $d \in \mathbb{C}^{N_{d}}$ are

$$
\begin{equation*}
d=\mathcal{F}_{a}(m)+\eta \tag{2.21}
\end{equation*}
$$

Theorem 2.4. For the two dimensional problem 2.18(problem 2.15), if we assume space $X, m \sim \mu_{0}$ and $\eta$ are specified as previous two paragraphs in this subsection. Then, the Bayesian inverse problems of recovering input $m \in X$ of problem 2.18)(problem (2.15)) from data $d$ given as in (2.21) is well formulated: the posterior $\mu^{d}$ is well defined in $X$ and it is absolutely continuous with respect to $\mu_{0}$, the Radon-Nikodym derivative is given by (2.5) and (2.6). Moreover, there is $C=C(r)$ such that, for all $d_{1}, d_{2} \in \mathbb{C}^{N_{d}}$ with $\left\|d_{1}\right\|,\left\|d_{2}\right\|_{2} \leq r$,

$$
\begin{equation*}
d_{H e l l}\left(\mu^{d_{1}}, \mu^{d_{2}}\right) \leq C\left\|d_{1}-d_{2}\right\|_{2} \tag{2.22}
\end{equation*}
$$

Proof. Relying on the general theory, we easily know that Theorem 2.4 holds when Assumption 1 is satisfied. According to the estimates 2.19) and 2.20, we find that

$$
\begin{equation*}
\left\|\mathcal{F}_{a}(m)\right\|_{2} \leq C\|m\|_{L^{\infty}(D)}, \tag{2.23}
\end{equation*}
$$

which indicates that statement (1) of Assumption 1 holds. In order to verify statement (2) of Assumption 1, we denote $u_{a}^{s}+\delta u=\mathcal{F}_{a}(m+\delta m)$. By simple calculations, we deduce that $\delta u$ satisfies

$$
\left\{\begin{array}{l}
\nabla \cdot(s \nabla \delta u)+s_{1} s_{2} \kappa^{2}(1+m) \delta u=-\kappa^{2} \delta m\left(u^{\mathrm{inc}}+s_{1} s_{2} u_{a}^{s}\right) \quad \text { in } D  \tag{2.24}\\
\delta u=0 \quad \text { on } \partial D .
\end{array}\right.
$$

Now, denote $\mathcal{F}_{a}^{\prime}(m)$ to be the Fréchet derivative of $\mathcal{F}_{a}(m)$, we find that

$$
\begin{equation*}
\mathcal{F}_{a}^{\prime}(m) \delta m=\mathcal{M}(\delta u) \tag{2.25}
\end{equation*}
$$

where $\delta u$ is the solution of equations 2.24 . Using similar estimates as for deriving (2.19) to equation 2.24), we obtain

$$
\begin{equation*}
\|\delta u\|_{L^{\infty}(D)} \leq C\|\delta m\|_{L^{\infty}(D)}\left(\left\|u^{\mathrm{inc}}\right\|_{L^{2}(D)}+\left\|u_{a}^{s}\right\|_{L^{2}(D)}\right) \tag{2.26}
\end{equation*}
$$

Taking estimate 2.19 into the above inequality 2.26 , we obtain

$$
\begin{equation*}
\left\|\mathcal{F}_{a}^{\prime}(m) \delta m\right\|_{2} \leq\|\delta u\|_{L^{\infty}(D)} \leq C\left(1+\|m\|_{L^{\infty}(\Omega)}\right)\|\delta m\|_{L^{\infty}(D)} \tag{2.27}
\end{equation*}
$$

where $C$ depends on $\kappa, D, s_{1}$ and $s_{2}$. Estimate (2.27) ensures that statement (2) of Assumption 1 holds, and the proof is completed by employing Theorem 2.1.

Remark 2.5. From the proof of Theorem 2.4 , we can see that Theorem 2.3 holds true for inverse medium scattering problem considered in this subsection. Hence, we can compute the MAP estimate by minimizing functional defined in 2.7 with the forward operator defined in (2.21).

Remark 2.6. If we assume $\mu_{0}$ is a TV-Gaussian probability measure, similar results can be obtained. The posterior probability measure is well-defined and the MAP estimate can be obtained by solving $\min _{m \in X} J(m)$ with $J$ defined in 2.8 . Since there are no new ingredients, we omit the details.

## 3. Algorithm construction

In Section 2, we illustrate the theoretical foundations for nonlinear optimization problems with functionals 2.7 ) and 2.8 . Here, we give a detailed description of the two stage algorithm stated in the introduction.
3.1. Learn parameters of complex mixture Gaussian distribution. How to estimate the parameters is one of the key steps for modeling noises by some complex mixture Gaussian distributions. This key step consists of two fundamental elements: learning examples and learning algorithms.

For the learning examples, they are the approximate errors $e:=\mathcal{F}(m)-\mathcal{F}_{a}(m)$ that is the difference of measured values for slow explicit forward solver and fast approximate forward solver. In order to obtain this error, we need to know the unknown function $m$ which is impossible. However, in practical problems, we usually know some prior knowledge of the unknown function $m$. Relying on the prior knowledge, we can construct some probability measures to generate functions which we believe to maintain some important properties as the real unknown function $m$. For this, we refer to a recent paper [24]. Since this procedure depends on specific application fields, we only provide details in Section 4 for concrete numerical examples.

For the learning algorithms, expectation-maximization (EM) algorithm is often employed in the machine learning community [8. Here, we need to notice that the variables are complex valued and the complex Gaussian distribution are used in our case. This leads some differences to the classical real variable situation.

In order to provide a clear explanation, let us recall some basic relationships between complex Gaussian distributions and real Gaussian distributions which are proved in [23]. Denote $e=\left(e_{1}, \ldots, e_{N_{d}}\right)^{T}$ is a $N_{d}$-tuple of complex Gaussian random variables. Let $\tau_{k}:=\operatorname{Re}\left(e_{k}\right)$ and $\varsigma_{k}:=\operatorname{Imag}\left(e_{k}\right)$ as the real and imaginary parts of $e_{k}$ with $k=1, \ldots, N_{d}$, then define

$$
\begin{equation*}
\tau=\left(\tau_{1}, \varsigma_{1}, \ldots, \tau_{N_{d}}, \varsigma_{N_{d}}\right) \tag{3.1}
\end{equation*}
$$

is $2 N_{d}$-tuple of random variables. From the basic theories of complex Gaussian distributions, we know that $\tau$ is $2 N_{d}$-variate Gaussian distributed. Denote the covariance matrix of $e$ by $\Sigma$ and the covariance matrix of $\tau$ by $\tilde{\Sigma}$. As usual, we assume $\Sigma$ is a positive definite Hermitian matrix, then $\tilde{\Sigma}$ is a positive definite symmetric matrix by Theorem 2.2 and Theorem 2.3 in [23]. In addition, we have the following lemma which is proved in [23].
Lemma 3.1. For complex Gaussian distributions, we have that the matrix $\Sigma$ is isomorphic to the matrix $2 \tilde{\Sigma}, e^{H} \Sigma e=\tau^{T} \tilde{\Sigma} \tau$ and $\operatorname{det}(\Sigma)^{2}=\operatorname{det}(\tilde{\Sigma})$.

Let $N_{s} \in \mathbb{N}^{+}$stand for the number of learning examples. Let $e_{n}=\left(e_{1}^{n}, \ldots, e_{N_{d}}^{n}\right)^{T}$ with $n=1, \ldots, N_{s}$ represent $N_{s}$ learning examples. Then, for some fixed $K \in$ $\mathbb{N}^{+}$, we need to solve the following optimization problem to obtain estimations of parameters

$$
\begin{equation*}
\min _{\left\{\pi_{k}, \zeta_{k}, \Sigma_{k}\right\}_{k=1}^{K}} J_{G}\left(\left\{\pi_{k}, \zeta_{k}, \Sigma_{k}\right\}_{k=1}^{K}\right) \tag{3.2}
\end{equation*}
$$

where

$$
\begin{equation*}
J_{G}\left(\left\{\pi_{k}, \zeta_{k}, \Sigma_{k}\right\}_{k=1}^{K}\right):=\sum_{n=1}^{N_{s}} \ln \left\{\sum_{k=1}^{K} \pi_{k} \mathcal{N}_{c}\left(e_{n} \mid \zeta_{k}, \Sigma_{k}\right)\right\} \tag{3.3}
\end{equation*}
$$

From the real valued case, we can easily infer the complex valued algorithm. However, we can not find a rigorous proof for the mean and covariance estimation formulas in the existing literatures. In the following, we provide a rigorous calculation.

Estimation of means: Setting the derivatives of $J_{G}\left(\left\{\pi_{k}, \zeta_{k}, \Sigma_{k}\right\}_{k=1}^{K}\right)$ with respect to $\zeta_{k}$ of the complex Gaussian components to zero and using Lemma 3.1, we obtain

$$
\begin{equation*}
0=-\sum_{n=1}^{N_{s}} \frac{\pi_{k} \mathcal{N}_{c}\left(e_{n} \mid \zeta_{k}, \Sigma_{k}\right)}{\sum_{j=1}^{K} \pi_{j} \mathcal{N}_{c}\left(e_{j} \mid \zeta_{j}, \Sigma_{j}\right)} \tilde{\Sigma}_{k}^{-1}\left(\tau_{n}-\tilde{\zeta}_{k}\right) \tag{3.4}
\end{equation*}
$$

where $\tau_{n}$ is defined as in (3.1) with $e$ replaced by $e_{n}, \tilde{\zeta}_{k}$ also defined as in 3.1 with $e$ replaced by $\zeta_{k}$ and $\tilde{\Sigma}_{k}$ is the covariance matrix corresponding to $\Sigma_{k}$. Hence, by some simple simplification, we find that

$$
\begin{equation*}
\zeta_{k}=\frac{1}{\tilde{N}_{k}} \sum_{n=1}^{N_{s}} \gamma_{n k} e_{n} \tag{3.5}
\end{equation*}
$$

where

$$
\begin{equation*}
\tilde{N}_{k}=\sum_{n=1}^{N_{s}} \gamma_{n k}, \quad \gamma_{n k}=\frac{\pi_{k} \mathcal{N}_{c}\left(e_{n} \mid \zeta_{k}, \Sigma_{k}\right)}{\sum_{j=1}^{K} \pi_{j} \mathcal{N}_{c}\left(e_{j} \mid \zeta_{j}, \Sigma_{j}\right)} \tag{3.6}
\end{equation*}
$$

In the above formula, $\tilde{N}_{k}$ usually interpret as the effective number of points assigned to cluster $k$ and $\gamma_{n k}$ usually is a variable depend on latent variables [8].

Estimation of covariances: For the covariances, we need to use latent variables to provide the following complete-data $\log$ likelihood function as formula (9.40) shown in [8]

$$
\begin{equation*}
\sum_{n=1}^{N_{s}} \sum_{k=1}^{K} \gamma_{n k}\left\{\ln \pi_{k}+\ln \mathcal{N}_{c}\left(e_{n} \mid \zeta_{k}, \Sigma_{k}\right)\right\} \tag{3.7}
\end{equation*}
$$

In the following lemma, we give the maximum likelihood estimation of the covariances, which is crucial for constructing the corresponding EM algorithm.

Lemma 3.2. Let $\left\{e_{n}\right\}_{n=1}^{N_{s}},\left\{\pi_{k}, \zeta_{k}, \Sigma_{k}\right\}_{k=1}^{K}$ and $\left\{\gamma_{n k}\right\}_{n, k=1}^{N_{s}, K}$ be specified as in (3.2), (3.3) and (3.6). Then the maximization problem

$$
\begin{equation*}
\max _{\left\{\Sigma_{k}\right\}_{k=1}^{K}}\left\{\sum_{n=1}^{N_{s}} \sum_{k=1}^{K} \gamma_{n k}\left(\ln \pi_{k}+\ln \mathcal{N}_{c}\left(e_{n} \mid \zeta_{k}, \Sigma_{k}\right)\right)\right\} \tag{3.8}
\end{equation*}
$$

possesses a solution with the following form

$$
\begin{equation*}
\Sigma_{k}:=\frac{1}{\tilde{N}_{k}} \sum_{n=1}^{N_{s}} \gamma_{n k}\left(e_{n}-\zeta_{k}\right)\left(e_{n}-\zeta_{k}\right)^{H}, \quad \text { for } k=1, \ldots, K \tag{3.9}
\end{equation*}
$$

Proof. Denote

$$
\begin{equation*}
L=\sum_{n=1}^{N_{s}} \sum_{k=1}^{K} \gamma_{n k}\left(\ln \pi_{k}+\ln \mathcal{N}_{c}\left(e_{n} \mid \zeta_{k}, \Sigma_{k}\right)\right) \tag{3.10}
\end{equation*}
$$

Let

$$
\begin{equation*}
B_{k}:=\frac{1}{\tilde{N}_{k}} \sum_{n=1}^{N_{s}} \gamma_{n k}\left(e_{n}-\zeta_{k}\right)\left(e_{n}-\zeta_{k}\right)^{H} \tag{3.11}
\end{equation*}
$$

and notice that

$$
\begin{aligned}
\sum_{n=1}^{N_{s}} \sum_{k=1}^{K} \gamma_{n k}\left(e_{n}-\zeta_{k}\right)^{H} \Sigma_{k}^{-1}\left(e_{n}-\zeta_{k}\right) & =\sum_{n=1}^{N_{s}} \sum_{k=1}^{K} \gamma_{n k} \operatorname{tr}\left(\Sigma_{k}^{-1}\left(e_{n}-\zeta_{k}\right)\left(e_{n}-\zeta_{k}\right)^{H}\right) \\
& =\sum_{k=1}^{K} \operatorname{tr}\left(\Sigma_{k}^{-1} \sum_{n=1}^{N_{s}} \gamma_{n k}\left(e_{n}-\zeta_{k}\right)\left(e_{n}-\zeta_{k}\right)^{H}\right) \\
& =\sum_{k=1}^{K} \tilde{N}_{k} \operatorname{tr}\left(\Sigma_{k}^{-1} B_{k}\right)
\end{aligned}
$$

where $\tilde{N}_{k}$ defined as in 3.6 . Then, using the explicit form of density function, we obtain

$$
\begin{equation*}
L=-\sum_{k=1}^{K} \tilde{N}_{k} \ln \operatorname{det}\left(\Sigma_{k}\right)-\sum_{k=1}^{K} \tilde{N}_{k} \operatorname{tr}\left(\Sigma_{k}^{-1} B_{k}\right)-N_{d} \ln \pi+\sum_{k=1}^{K} \tilde{N}_{k} \ln \pi_{k} \tag{3.12}
\end{equation*}
$$

Define $p(\xi, \Sigma):=\frac{1}{\pi^{N_{d}} \operatorname{det}(\Sigma)} \exp \left(-\xi^{H} \Sigma^{-1} \xi\right)$, then we have

$$
\begin{align*}
J= & \sum_{k=1}^{K} \tilde{N}_{k} \int_{\xi} p\left(\xi, \Sigma_{k}^{-1}\right) \ln \left(p\left(\xi, B_{k}^{-1}\right) / p\left(\xi, \Sigma_{k}^{-1}\right)\right) d \xi \\
= & \int_{\xi}\left\{\left(\ln \operatorname{det}\left(B_{k}\right)-\xi^{H} B_{k} \xi\right) p\left(\xi, \Sigma_{k}^{-1}\right)\right. \\
& \left.-\left(\ln \operatorname{det}\left(\Sigma_{k}\right)-\xi^{H} \Sigma_{k} \xi\right) p\left(\xi, \Sigma_{k}^{-1}\right)\right\} d \xi  \tag{3.13}\\
= & \sum_{k=1}^{K} \tilde{N}_{k} \ln \operatorname{det}\left(B_{k}\right)+\sum_{k=1}^{K} \tilde{N}_{k} \operatorname{tr}(I) \\
& -\sum_{k=1}^{K} \tilde{N}_{k} \operatorname{tr}\left(\Sigma_{k}^{-1} B_{k}\right)-\sum_{k=1}^{K} \tilde{N}_{k} \ln \operatorname{det}\left(\Sigma_{k}\right)
\end{align*}
$$

where Corollary 4.1 in 23 has been used for the last equality. Comparing the final result of (3.12) with 3.13 , we observe that any series Hermitian positive definite matrixes $\left\{\Sigma_{k}\right\}_{k=1}^{K}$ that maximize $L$ maximize $J$ and conversely. Now, $\ln u \leq u-1$ with equality holding if and only if $u=1$. Thus

$$
\begin{align*}
J & =\sum_{k=1}^{K} \tilde{N}_{k} \int_{\xi} p\left(\xi, \Sigma_{k}^{-1}\right) \ln \left(p\left(\xi, B_{k}^{-1}\right) / p\left(\xi, \Sigma_{k}^{-1}\right)\right) d \xi \\
& \leq \sum_{k=1}^{K} \tilde{N}_{k} \int_{\xi} p\left(\xi, \Sigma_{k}^{-1}\right)\left(p\left(\xi, B_{k}^{-1}\right) / p\left(\xi, \Sigma_{k}^{-1}\right)-1\right) d \xi=0 \tag{3.14}
\end{align*}
$$

If and only if $p\left(\xi, \Sigma_{k}\right)=p\left(\xi, B_{k}\right)$ with $k=1, \ldots, K$, equality in (3.14) holds true. Hence, $\Sigma_{k}=B_{k}(k=1, \ldots, K)$ solves problem (3.8).

With these preparations, we can easily construct EM algorithm following the line of reasoning shown in Chapter 9 of [8]. For brevity, the details are omitted and we provide the EM algorithm in Algorithm 1 .

Remark 3.3. In Algorithm 1, if the parameters satisfy $N_{d}<N_{s}$, we can usually obtain nonsingular matrixes $\left\{\Sigma_{k}\right\}_{k=1}^{K}$. However, it is time consuming to generate a lot of learning examples $N_{s}$ and, in the mean time, the number of measuring points $N_{d}$ is large for some real world applications [35, 36]. Hence, we may meet the situation $N_{d}>N_{s}$ which makes $\left\{\Sigma_{k}\right\}_{k=1}^{K}$ to be a series of singular matrixes. The determinants of $\Sigma_{k}$ with $k=1, \ldots, K$ will be zero, which is a troublesome problem for calculating the normalization constant appeared in Gaussian probability density function. In order to solve this problem, we adopt a simple strategy that is replace the estimation of $\Sigma_{k}$ in Step 3 by the following formula

$$
\begin{equation*}
\Sigma_{k}^{\text {new }}=\frac{1}{\tilde{N}_{k}} \sum_{n=1}^{N_{s}} \gamma_{n k}\left(e_{n}-\zeta_{k}^{\text {new }}\right)\left(e_{n}-\zeta_{k}^{\text {new }}\right)^{H}+\delta I, \tag{3.15}
\end{equation*}
$$

where $\delta$ is a small positive number named as the regularization parameter.

## Algorithm 1 Complex EM algorithm <br> Step 1: For a series of samples $e_{n} \in \mathbb{C}^{N_{d}}\left(n=1, \ldots, N_{s}\right)$, initialize the means $\zeta_{k}$, covariances $\Sigma_{k}$ and mixing coefficients $\pi_{k}$, and evaluate the initial value of the $\ln$ likelihood. <br> Step 2 (E step): Evaluate the responsibilities using the current parameter values <br> $$
\gamma_{n k}=\frac{\pi_{k} \mathcal{N}_{c}\left(e_{n} \mid \zeta_{k}, \Sigma_{k}\right)}{\sum_{j=1}^{K} \pi_{j} \mathcal{N}_{c}\left(e_{n} \mid \zeta_{j}, \Sigma_{j}\right)}
$$

Step 3 (M step): Re-estimate the parameters using the current responsibilities

$$
\begin{gathered}
\tilde{N}_{k}=\sum_{n=1}^{N_{s}} \gamma_{n k}, \quad \pi_{k}^{\text {new }}=\frac{\tilde{N}_{k}}{N_{s}}, \quad \zeta_{k}^{\text {new }}=\frac{1}{\tilde{N}_{k}} \sum_{n=1}^{N_{s}} \gamma_{n k} e_{n} \\
\Sigma_{k}^{\text {new }}=\frac{1}{\tilde{N}_{k}} \sum_{n=1}^{N_{s}} \gamma_{n k}\left(e_{n}-\zeta_{k}^{\text {new }}\right)\left(e_{n}-\zeta_{k}^{\text {new }}\right)^{H}
\end{gathered}
$$

Step 4: Evaluate the ln likelihood

$$
\sum_{n=1}^{N_{s}} \ln \left\{\sum_{k=1}^{K} \pi_{k} \mathcal{N}_{c}\left(e_{n} \mid \zeta_{k}, \Sigma_{k}\right)\right\}
$$

and check for convergence of either the parameters or the $\ln$ likelihood. If the convergence criterion is not satisfied return to Step 2.
3.2. Adjoint state approach with error compensation. By Algorithm 1, we obtain the estimated mixing coefficients, mean values and covariance matrixes. From the statements shown in Section 2, it is obvious that we need to solve optimization problems as follows

$$
\begin{equation*}
\min _{m \in L^{\infty}(\Omega)}\{-\Phi(m ; d)+\mathcal{R}(m)\} \tag{3.16}
\end{equation*}
$$

where

$$
\begin{aligned}
&-\Phi(m ; d)=-\ln \left\{\sum_{k=1}^{K} \pi_{k} \frac{1}{\pi^{N_{d}} \operatorname{det}\left(\Sigma_{k}+\nu I\right)} \exp \left(-\frac{1}{2}\left\|d-\mathcal{F}_{a}(m)-\zeta_{k}\right\|_{\Sigma_{k}+\nu I}^{2}\right)\right\} \\
& \mathcal{R}(m)=\frac{1}{2}\left\|A^{s / 2} m\right\|_{L^{2}(\Omega)}^{2} \quad \text { or } \quad \mathcal{R}(m)=\lambda\|m\|_{\mathrm{TV}}+\frac{1}{2}\left\|A^{s / 2} m\right\|_{L^{2}(\Omega)}^{2}
\end{aligned}
$$

Different form of functional $\mathcal{R}$ comes from different assumptions of the prior probability measures: Gaussian probability measure or TV-Gaussian probability measure. For the multi-frequency approach of inverse medium scattering problem, the forward operator in each optimization problem is related to $\kappa$. So we rewrite $\mathcal{F}_{a}(m)$ and $\Phi(m ; d)$ as $\mathcal{F}_{a}(m, \kappa)$ and $\Phi(m, \kappa ; d)$, which emphasize the dependence of $\kappa$. We have a series of wavenumbers $0<\kappa_{1}<\kappa_{2}<\cdots \kappa_{N_{w}}<\infty$, and we actually need to solve a series optimization problems

$$
\begin{equation*}
\min _{m \in L^{\infty}(\Omega)}\left\{-\Phi\left(m, \kappa_{i} ; d\right)+\mathcal{R}(m)\right\} \tag{3.17}
\end{equation*}
$$

with $i$ from 1 to $N_{w}$ and the solution of the previous optimization problem is the initial data for the later optimization problem.

Denote $F(m)=-\Phi\left(m, \kappa_{i} ; d\right)$. To minimize the cost functional by a gradient method, it is required to compute Fréchet derivative of functionals $F$ and $\mathcal{R}$. For functional $\mathcal{R}$ used in $(3.16)$, we can obtain the Fréchet derivatives as follows

$$
\begin{equation*}
\mathcal{R}^{\prime}(m)=A^{s} m, \quad \text { or } \quad \mathcal{R}^{\prime}(m)=A^{s} m+2 \lambda \nabla \cdot\left(\frac{\nabla m}{\sqrt{|\nabla m|^{2}+\delta}}\right) \tag{3.18}
\end{equation*}
$$

where we used the following modified version of $\mathcal{R}$

$$
\begin{equation*}
\mathcal{R}(m)=\lambda \int_{\Omega} \sqrt{|\nabla m|^{2}+\delta}+\frac{1}{2}\left\|A^{s / 2} m\right\|_{L^{2}(\Omega)}^{2} \tag{3.19}
\end{equation*}
$$

for the TV-Gaussian prior case and $\delta$ is a small smoothing parameter avoiding zero denominator in (3.18).

Next, we consider the functional $F$ with $\mathcal{F}_{a}$ is the forward operator related to problem 2.18. A simple calculation yields the derivative of $F$ at $q$;

$$
\begin{equation*}
F^{\prime}(m) \delta m=\operatorname{Re}\left(\mathcal{M}(\delta u), \sum_{k=1}^{K} \gamma_{k}\left(\Sigma_{k}+\nu I\right)^{-1}\left(d-\mathcal{F}_{a}\left(m, \kappa_{i}\right)-\zeta_{k}\right)\right) \tag{3.20}
\end{equation*}
$$

where $\delta u$ satisfy

$$
\left\{\begin{array}{l}
\nabla \cdot(s \nabla \delta u)+s_{1} s_{2} \kappa_{i}^{2}(1+m) \delta u=-\kappa^{2} \delta m\left(u^{\mathrm{inc}}+s_{1} s_{2} u_{a}^{s}\right) \quad \text { in } D  \tag{3.21}\\
\delta u=0 \text { on } \partial D
\end{array}\right.
$$

and

$$
\gamma_{k}=\frac{\pi_{k} \mathcal{N}_{c}\left(d-\mathcal{F}_{a}(m) \mid \zeta_{k}, \Sigma_{k}+\nu I\right)}{\sum_{j=1}^{K} \pi_{j} \mathcal{N}_{c}\left(d-\mathcal{F}_{a}(m) \mid \zeta_{j}, \Sigma_{j}+\nu I\right)}
$$

To compute the Fréchet derivative, we introduce the adjoint system:

$$
\left\{\begin{array}{l}
\nabla \cdot(\bar{s} \nabla v)+\bar{s}_{1} \bar{s}_{2} \kappa_{i}^{2}(1+m) v=-\kappa_{i}^{2} \sum_{j=1}^{N_{d}} \delta\left(x-x_{j}\right) \rho_{j} \quad \text { in } D  \tag{3.22}\\
v=0 \quad \text { on } D
\end{array}\right.
$$

where $\rho_{j}\left(j=1, \ldots, N_{d}\right)$ denote the $j$ th component of $\sum_{k=1}^{K} \gamma_{k}\left(\Sigma_{k}+\nu I\right)^{-1}(d-$ $\left.\mathcal{F}_{a}\left(m, \kappa_{i}\right)-\zeta_{k}\right) \in \mathbb{C}^{N_{d}}$. Multiplying equation 3.21 with the complex conjugate of $v$ on both sides and integrating over $D$ yields

$$
\int_{D} \nabla \cdot(s \nabla \delta u) \bar{v}+s_{1} s_{2} \kappa_{i}^{2}(1+m) \delta u \bar{v}=-\int_{D} \kappa^{2} \delta m\left(u^{\mathrm{inc}}+s_{1} s_{2} u_{a}^{s}\right) \bar{v} .
$$

By integration by parts formula, we obtain

$$
\int_{D} \delta u\left(\nabla \cdot(s \nabla \bar{v})+s_{1} s_{2} \kappa_{i}^{2}(1+m) \bar{v}\right)=-\kappa_{i}^{2} \int_{D} \delta m\left(u^{\mathrm{inc}}+s_{1} s_{2} u_{a}^{s}\right) \bar{v}
$$

Taking complex conjugate of equation $\sqrt[3.22]{ }$ and plugging into the above equation yields

$$
-\kappa_{i}^{2} \int_{D} \delta u \sum_{j=1}^{N_{d}} \delta\left(x-x_{j}\right) \bar{\rho}_{j}=-\kappa_{i}^{2} \int_{D} \delta m\left(u^{\mathrm{inc}}+s_{1} s_{2} u_{a}^{s}\right) \bar{v}
$$

which implies

$$
\begin{equation*}
\left(\mathcal{M}(\delta u), \sum_{k=1}^{K} \gamma_{k} \Sigma_{k}^{-1}\left(d-\mathcal{F}_{a}\left(m, \kappa_{i}\right)-\zeta_{k}\right)\right)=\int_{D} \delta m\left(u^{\mathrm{inc}}+s_{1} s_{2} u_{a}^{s}\right) \bar{v} \tag{3.23}
\end{equation*}
$$

Considering both (3.20) and (3.23), we find that

$$
F^{\prime}(m) \delta m=\operatorname{Re} \int_{D} \delta m\left(u^{\mathrm{inc}}+s_{1} s_{2} u_{a}^{s}\right) \bar{v}
$$

which gives the Fréchet derivative as follow

$$
\begin{equation*}
F^{\prime}(m)=\operatorname{Re}\left(\left(\bar{u}^{\mathrm{inc}}+\bar{s}_{1} \bar{s}_{2} \bar{u}_{a}^{s}\right) v\right) \tag{3.24}
\end{equation*}
$$

With these preparations, it is enough to construct mixture Gaussian recursive linearization method (MGRLM) which is shown in Algorithm 2. Notice that for the recursive linearization method (RLM) shown in [4, only one iteration of the gradient descent method for each fixed wavenumber can provide an acceptable recovered function. So we only take one iteration for each fixed wavenumber.

```
Algorithm 2 mixture Gaussian recursive linearization method (MGRLM)
Input: Initialize parameters: \(\sigma_{0}, d_{1}, d_{2}, p,\left\{\zeta_{k}\right\}_{k=1}^{K},\left\{\Sigma_{k}\right\}_{k=1}^{K},\left\{\pi_{k}\right\}_{k=1}^{K}, q\),
    wavenumbers \(\left(\kappa_{1}, \ldots, \kappa_{N_{w}}\right)\) and incident angles \(\left(\mathbf{d}_{1}, \ldots, \mathbf{d}_{N_{m}}\right)\).
    Iteration: for \(i=1,2, \ldots, N_{w}\)
                    for \(j=1,2, \ldots, N_{m}\)
                            solve one forward problem \(\sqrt{2.18}\) with \(\kappa=\kappa_{i}\) and \(\mathbf{d}=\mathbf{d}_{j}\);
                    solve one adjoint problem \(\sqrt{3.22}\) with \(\kappa=\kappa_{i}\) and \(\mathbf{d}=\mathbf{d}_{j}\);
                    compute the Fréchet derivative by formulas (3.18) and (3.24);
                        update the scatterer function;
                    end for
        end for
Output: Final estimation of \(q\).
```


## 4. Numerical examples

In this section, we provide two numerical examples in two dimensions to illustrate the effectiveness of the proposed method. In the following, we assume that $\Omega=$ $\left\{(x, y) \in \mathbb{R}^{2}:-1 \leq x \leq 1\right.$ and $\left.-1 \leq y \leq 1\right\}$ with $\Omega \subset D$ where $D$ is the PML domain with $d_{1}=\bar{d}_{2}=0.15, p=2.5$ and $\sigma_{0}=1.5$. For the forward solver, finite element method (FEM) with first-order elements has been employed. In order to avoid the inverse crime, the scattering data are generated by using adaptive finite element mesh method, whilst uniform triangular mesh has been employed for the learning process and inversion. For the following two examples, we choose $N_{m}=$ 20 and $\mathbf{d}_{j}\left(j=1, \ldots, N_{m}\right)$ are equally distributed around $\partial D$. Equally spaced wavenumbers are used, starting from the lowest wavenumber $\kappa_{\text {min }}=\pi$ and ending at the highest wavenumber $\kappa_{\max }=10 \pi$. Denote by $\Delta \kappa=\left(\kappa_{\max }-\kappa_{\min }\right) / 9=\pi$ the step size of the wavenumber; then the ten equally spaced wavenumbers are $\kappa_{j}=j \Delta \kappa, j=1, \ldots, 10$. We set 400 receivers that equally spaced along the boundary of $\Omega$ as shown in Figure 1 . For the initial guess of the unknown function $m$, there are numerous strategies, i.e., methods based on the Born approximation
[4. 9]. Since the main point here is not on the initial guess, we just set the initial $m$ to be a function always equal to zero for simplicity.

In order to show the stability of the proposed method, some relative random noise is added to the data, i.e.,

$$
\begin{equation*}
\left.u^{s}\right|_{\partial \Omega}:=\left.\left(1+\tilde{\sigma} \operatorname{rand}_{n}\right) u^{s}\right|_{\partial_{D}} \tag{4.1}
\end{equation*}
$$

Here, $\operatorname{rand}_{n}$ gives standard normal distributed random numbers and $\tilde{\sigma}$ is a noise level parameter taken to be 0.02 in our numerical experiments. Define the relative error by

$$
\begin{equation*}
\text { Relative Error }=\frac{\|m-\tilde{m}\|_{L^{2}(\Omega)}}{\|m\|_{L^{2}(\Omega)}} \tag{4.2}
\end{equation*}
$$

where $\tilde{m}$ is the reconstructed scatterer and $m$ is the true scatterer.
For the reader's convenience, we would like to summarize the notations and variables employed in this section.

- $\Omega$ : The square domain $[-1,1]^{2}$ with $\operatorname{supp}(q) \subset[-1,1]^{2}$;
- $D$ : The PML domain containing $[-1,1]^{2}$ with parameters $d_{1}=d_{2}=0.15$, $p=2.5$ and $\sigma=1.5$;
- $\mathbf{d}_{j}\left(j=1,2, \cdots, N_{m}\right)$ : Incident directions equally distributed around $\partial D$ with $N_{m}=20$;
- $\kappa_{j}=j \Delta \kappa(j=1, \ldots, 10)$ : Equally spaced wavenumbers with $\Delta \kappa=\pi$;
- $l_{w}$ : Wavelength related to wavenumber through $l_{w}=2 \pi / \kappa$;
- $m$ : The function of the true scatterer;
- $m_{e}$ : The random function employed to generate learning examples;
- $U\left[b_{1}, b_{2}\right]$ : A uniform distribution with minimum value $b_{1}$ and maximum value $b_{2}$;
- $K$ : The component numbers used in mixture Gaussian probability distribution.
Before giving the two examples, we should provide some additional explanations. The main point of this paper is to provide a new method incorporated learning process into IMSP. For using the proposed method to some specific application areas, a lot of work should be done for collecting historical data or constructing learning examples by some prior knowledge. The proposed method is not restricted to the case of the following two simple examples. Once enough learning examples are available for some specific applications, the proposed method could also be worked for more complicated situations.

During the iterative process, errors are accumulated from a lot of aspects [22], e.g., the FEM approximation, the evaluation of the Fréchet derivative and the evaluation of a large number of point sources appeared in the adjoint equation. These reasons make it difficult to decide the minimum element numbers for FEM should be used for a specific problem. Noticing that, FEM (using second-order elements) with 20000 and 28800 equal triangular elements have been used for Example 1 and Example 2 in [1], respectively. For simplicity, we use similar scatterer models as in [1] and only FEM with first-order elements, which is enough to show the effectiveness of the proposed method. Actually, the proposed method is not relevant to the specific method (e.g., finite element, finite difference) used for solving the forward problem, and just relevant to a rough approximation and an accurate approximation for the forward problem. Based on these considerations, FEM (using first-order elements) with 16204 equal triangular elements will be used to obtain
$\mathcal{S}_{a}(m) u^{\text {inc }}$ and FEM (using first-order elements) with 183198 elements will be used to obtain accurate solutions which we recognized as $\mathcal{S}(m) u^{\text {inc }}$ for both of our two examples.

Example 1: For the first example, let

$$
m_{1}(x, y)=0.3(1-x)^{2} e^{-x^{2}-(y+1)^{2}}-\left(0.2 x-x^{3}-y^{5}\right) e^{-x^{2}-y^{2}}-0.03 e^{-(x+1)^{2}-y^{2}}
$$

and reconstruct a scatterer defined by

$$
m(x, y)=m_{1}(3 x, 3 y)
$$

inside the unit square $\left\{(x, y) \in \mathbb{R}^{2}:-1<x<1\right.$ and $\left.-1<y<1\right\}$.


Figure 2. True scatterer and five typical learning examples

Now, we assume that some prior knowledge of this function $m$ have been known. According to the prior knowledge, we generate 200 learning examples according to the following function

$$
\begin{equation*}
m_{e}(x, y):=\sum_{k=1}^{3}\left(1-x^{2}\right)^{a_{k}^{1}}\left(1-y^{2}\right)^{a_{k}^{2}} a_{k}^{3} \exp \left(-a_{k}^{4}\left(x-a_{k}^{5}\right)^{2}-a_{k}^{6}\left(y-a_{k}^{7}\right)^{2}\right) \tag{4.3}
\end{equation*}
$$

where

$$
\begin{aligned}
& a_{k}^{1}, a_{k}^{2} \sim U[1,3], \quad a_{k}^{3} \sim U[-1,1] \\
& a_{k}^{4}, a_{k}^{6} \sim U[8,10], \quad a_{k}^{5}, a_{k}^{7} \sim U[-0.8,0.8]
\end{aligned}
$$

In order to provide an intuitional sense, we show the true scatterer and several learning examples in Figure 2

Remark 4.1. Although learning examples reflect some key features of the true scatterer, we should notice that the true scatterer cannot be generated from function (4.3). Hence, the true probability density of $m$ is not assumed to be known.

It is instructive to show the residuals of the measurements when rough and accurate FEM approximations are used respectively. In Figure 3, we show the residuals when wavenumber equals to $2 \pi, 5 \pi$ and $7 \pi$, respectively. Differences
for rough and accurate FEM approximations are obviously increasing when the wavenumber becomes large. Hence, compensating these errors should be useful to provide a high resolution reconstruction.


Figure 3. Real part of measuring residuals for rough and accurate FEM approximation with different wavenumbers. Top: residual when $\kappa=2 \pi$; Middle: residual when $\kappa=5 \pi$; Bottom: residual when $\kappa=7 \pi$.

Learning algorithm with $K=4$ proposed in Subsection 3.1 has been used to learn the statistical properties of differences $e_{n}^{i}:=\mathcal{F}\left(m_{n}, \kappa_{i}\right)-\mathcal{F}_{a}\left(m_{n}, \kappa_{i}\right)$ with $\kappa_{i}=i \cdot \pi(i=1, \ldots 10)$ and $m_{n}(n=1, \ldots 200)$ stand for the learning examples. Concerning the regularizing term, we take $A=0.01 \Delta, s=1.5$ and $\lambda=0$, which can be computed by Fourier transform.

Remark 4.2. Here, we take $\lambda=0$ which means that only a Tikhonov type regularization has been used for our numerical examples. For employing an elastic net type regularization (corresponding to the aforementioned TV-Gaussian probability measure), we need to choose $\lambda>0$, which may keep the sharp boundaries of the scatterer and, in the mean time, avoid the staircasing effect 41]. To gain full power of this type of regularization, choosing a proper $\lambda$ is crucial. Actually, how to choose an effective regularization term is a sophisticated problem and has been discussed in a lot of papers, e.g., [25, 28, 45]. However, it is not the main point of our paper. Hence, we will not discuss regularization terms in detail and only employing a simple Tikhonov type regularization which, based on our understanding, is enough to illustrate the effectiveness of our method.

Relative errors of RLM with rough forward solver, RLM with accurate forward solver and MGRLM $(K=4)$ with rough forward solver are shown in Figure 4 , which illustrate the effectiveness of the proposed method. In addition, we also


Figure 4. Relative errors with different parameters: green dashed line are relative errors obtained by using the RLM with 16204 elements; cyan dashed line with circles are relative errors obtained by using the RLM with 183198 elements; cyan dash-dotted line are relative errors obtained by using MGRLM $(K=1)$ with 16204 elements; blue solid line are relative errors obtained by using the MGRLM $(K=4)$ with 16204 elements.
show the relative errors of the recovered functions for MGRLM with $K=1$ and rough forward solver in Figure4. From Figure 4, we can see the following important facts:

- When the wavenumber approaches $8 \pi$, RLM with rough forward solver cannot provide a better result and the errors are accumulated to make the recovered scatterer deviates from the true scatterer.
- Compared with MGRLM with $K=1$, MGRLM with $K=4$ is stable and can provide a better reconstruction. Obviously, MGRLM with $K=1$ may increase the accumulation of errors when the wavenumber is larger than $4 \pi$, which may be explained by Figure 3. When the wavenumber becomes large, the measurement errors are increased and cannot be compensated efficiently by using Gaussian assumption.
- Compared with discontinuous scatterer, we usually can obtain a high resolution reconstruction easier for smooth scatterer, which explains the resolution enhancement is limited for this example. However, the computational costs for Algorithm 2 increase little compared with the classical RLM. Hence, it is paid off to use MGRLM once the learning process has been done for some specific application areas.

Table 1. Relative errors for RLM and MGRLM with different mesh size and wavelength.

| Relative errors | $l_{w}=0.67$ | $l_{w}=0.33$ | $l_{w}=0.29$ | $l_{w}=0.25$ |
| :---: | :---: | :---: | :---: | :---: |
| RLM, Mesh size $=0.0361$ | $51.86 \%$ | $9.78 \%$ | $7.10 \%$ | $7.80 \%$ |
| RLM, Mesh size $=0.0107$ | $42.37 \%$ | $4.45 \%$ | $1.99 \%$ | $0.87 \%$ |
| MGRLM, Mesh size $=0.0361$ | $1.43 \%$ | $1.04 \%$ | $1.92 \%$ | $2.92 \%$ |

Usually, the mesh size and wavelength are used to compare the performance of different methods. Since the mesh we employed is nearly a uniform triangular mesh, the mesh size can be estimated from the size of the domain and the total element number. Notice that the domain employed here is a square with side length of 2.3 . For the rough mesh with 16204 elements and fine mesh with 183198 elements, the mesh size is approximately 0.0361 and 0.0107 , respectively. Employing the formula $l_{w}=2 \pi / \kappa$ with $l_{w}$ represents the wavelength, we can calculate the corresponding wavelength. In order to give a more clear illustration, we show the relative errors and corresponding mesh size and wavelength in Table 1. From Table 1, we can also see the three facts mentioned above. As the wavelength decreases, MGRLM can compensate the numerical errors and make use of some prior information to enhance the inversion. When the wavelength becomes too small, the forward solver with mesh size 0.0361 departs from the accurate forward solver too much. Then both of the RLM and MGRLM become unstable.

At last, we show results obtained by RLM and MGRLM with different parameters in Figure 5. The true scatterer, the results obtained by RLM with rough forward solver $(\kappa=8 \pi)$ and RLM with accurate forward solver $(\kappa=10 \pi)$ are shown on the top left, in the top middle and on the top right, respectively. On the bottom left, in the bottom middle and on the bottom right, there are the results obtained for $\kappa=4 \pi$ by MGRLM with rough forward solver, RLM with rough forward solver and RLM with accurate forward solver, respectively. From these results, we can visually see that the result obtained by MGRLM is comparable with the result obtained by RLM using accurate forward solver.

Example 2: For the second example, let us firstly define the following two squares

$$
\begin{aligned}
& \Omega_{1}:=\left\{(x, y) \in \mathbb{R}^{2}:-0.1 \leq x \leq 0.1 \text { and }-0.1 \leq y \leq 0.1\right\} \\
& \Omega_{2}:=\left\{(x, y) \in \mathbb{R}^{2}:-0.3 \leq x \leq 0.3 \text { and }-0.3 \leq y \leq 0.3\right\}
\end{aligned}
$$

Then the function of the scatterer can be defined as follow

$$
m(x, y):= \begin{cases}0.7 & \text { for }(x, y) \in \Omega_{2} \backslash \Omega_{1}  \tag{4.4}\\ -0.1 & \text { for }(x, y) \in \Omega_{1} \\ 0 & \text { for }(x, y) \in[-1,1]^{2} \backslash \Omega_{2}\end{cases}
$$

As in Example 1, we need to construct some learning examples. Here, we assume that the support of the scatterer is a square in $[-1,1]^{2}$ with unknown position and size. The function values of the scatterer is also unknown. Specifically speaking,


Figure 5. Recovered functions with different parameters. (a) true function; (b) minimum relative error estimate for the RLM with 16204 elements and the wavenumber computed to $8 \pi$; (c) minimum relative error estimate for the RLM with 183198 elements and the wavenumber computed to $10 \pi$; (d) minimum relative error estimate for the MGRLM with 16204 elements and the wavenumber computed to $4 \pi$; (e) recovered function for the RLM with 16204 elements and the wavenumber computed to $4 \pi$; (f) recovered function for the RLM with 183198 elements and the wavenumber computed to $4 \pi$.
the learning examples are generated according to the following function

$$
m_{e}(x, y):= \begin{cases}m_{v} & \text { for }(x, y) \in\left[-L_{1}+x_{1}, x_{1}+L_{1}\right] \times\left[-L_{2}+x_{2}, x_{2}+L_{2}\right] \\ 0 & \text { other areas in }[-1,1]^{2}\end{cases}
$$

where $x_{1} \sim U[-0.5,0.5], x_{2} \sim U[-0.5,0.5], L_{1} \sim U\left[0,0.9-\left|x_{1}\right|\right], L_{2} \sim U[0,0.9-$ $\left.\left|x_{2}\right|\right]$ and $m_{v} \sim U_{d}[-1,0,1] \times 0.5+U[-0.3,0.3]$ with $U_{d}[-1,0,1]$ represents a uniform distribution that only take three values $-1,0,1$. As in Example 1, we generate 200 learning examples. To give the reader an intuitive idea, we show the true scatterer and five typical learning examples in Figure 6 .

Remark 4.3. As in Example 1, learning examples reflect some key features of the true scatterer, but the true scatterer contains a small square which is not appeared in any of the 200 learning examples. Hence, the true probability density of $m$ is not assumed to be known.

As in Example 1, we show the residuals when wavenumber equals to $2 \pi, 5 \pi$ and $7 \pi$ respectively in Figure 7. Differences for rough and accurate FEM approximations are obviously increasing when the wavenumber becomes large. Relative errors of RLM with rough forward solver, RLM with accurate forward solver, MGRLM ( $K=$ 4) with rough forward solver and MGRLM $(K=1)$ with rough forward solver are shown in Figure 8, which illustrate the effectiveness of the proposed method. From Figure 8, we can see the following important facts:


Figure 6. True scatterer and five typical learning examples
Residual of measurements with rough and accurate FEM approximation (Real part)

Figure 7. Real part of measuring residuals for coarse and accurate FEM approximation with different wavenumber. Top: residual when $\kappa=2 \pi$; Middle: residual when $\kappa=5 \pi$; Bottom: residual when $\kappa=7 \pi$.

- When wavenumber approaches $5 \pi$, MGRLM with $K=1$ cannot provide a stable recovery, which indicates that the Gaussian assumption is not suitable for high wavenumber case.
- Similar as in Example 1, when the wavenumber approximates $9 \pi$, RLM with rough forward solver cannot provide a better result and the errors


Figure 8. Relative errors with different parameters: green dashed line are relative errors obtained by using the RLM with 16204 elements; cyan dashed line with circles are relative errors obtained by using the RLM with 183198 elements; cyan dash-dotted line are relative errors obtained by using MGRLM $(K=1)$ with 16204 elements; blue solid line are relative errors obtained by using the MGRLM $(K=4)$ with 16204 elements.

TABLE 2. Relative errors for RLM and MGRLM with different mesh size and wavelength.

| Relative errors | $l_{w}=0.67$ | $l_{w}=0.33$ | $l_{w}=0.25$ | $l_{w}=0.22$ |
| :---: | :---: | :---: | :---: | :---: |
| RLM, Mesh size $=0.0361$ | $75.82 \%$ | $44.40 \%$ | $36.49 \%$ | $37.22 \%$ |
| RLM, Mesh size $=0.0107$ | $66.27 \%$ | $28.53 \%$ | $16.89 \%$ | $12.76 \%$ |
| MGRLM, Mesh size $=0.0361$ | $23.93 \%$ | $13.02 \%$ | $12.06 \%$ | $12.67 \%$ |

are accumulated to make the recovered function deviates from the true scatterer.

- When the wavenumber is smaller than $9 \pi$, MGRLM with rough forward solver provides recovered functions even better than the results obtained by RLM with accurate forward solver. Our understanding of this is that the learning process not only provides a compensation for the numerical errors but also incorporate some prior information since the learning examples reflect some key features of the true scatterer. The prior information incorporated into the mixture Gaussian distribution makes the MGRLM converges even faster than RLM with accurate forward solver.
Here, as in Example 1, we show the relative errors and corresponding mesh size and wavelength in Table 2. From Table 2, we can also see the three facts mentioned above. Similar to Example 1, compared with the classical RLM, MGRLM can provide better estimations. Both of the RLM and MGRLM become unstable when the forward solver with mesh size 0.0361 departs from the accurate forward solver too much.


Figure 9. Recovered functions with different parameters. (a) true function; (b) minimum relative error estimate for the RLM with 16204 elements and the wavenumber computed to $8 \pi$ ); (c) minimum relative error estimate for the RLM with 183198 elements and the wavenumber computed to $10 \pi$; (d) minimum relative error estimate for the MGRLM with 16204 elements and the wavenumber computed to $8 \pi$; (e) recovered function for the RLM with 16204 elements and the wavenumber computed to $8 \pi$; (f) recovered function for the RLM with 183198 elements and the wavenumber computed to $8 \pi$.

Finally, we provide the image of true scatterer on the top left of Figure 9. On the top middle, the best result obtained by RLM with rough forward solver is given. From this image, we can see that it is failed to recover the small square embedded in the large square. The best result obtained by RLM with accurate froward solver is shown on the top right. It is much better than the function obtained by the algorithm with rough forward solver. At the bottom of Figure 9, we show the best result obtained by MGRLM with rough forward solver on the left and show the results obtained by RLM (compute to the same wavenumber as MGRLM) with rough and accurate forward solver in the middle and on the righthand side respectively. The recovered function obtained by MGRLM is not as well as the recovered function obtained by RLM with more than eleven times of elements and higher wavenumber. However, beyond our expectation, it is already capture the small square embedded in the large square, which is not incorporated in our 200 learning examples. If we think the large square to be some human organ and the small square to be a small cancer tissue, the recovered scatterer obtained by MGRLM is good enough to detect the cancer tissue.

In summary, the proposed MGRLM converges much faster than the classical RLM and it can provide a much better result at the same discrete level compared with RLM.

## 5. Conclusions

In this paper, we assume that the modeling errors brought by rough discretization to be a mixture Gaussian random variable. Based on this assumption, we derive a new nonlinear optimization problem by employing the infinite-dimensional Bayes' inverse method and the theories of MAP estimates. Specifically speaking, for the inverse medium scattering problem, well-posedness in the statistical sense has been proved and the related MAP estimate has been obtained. In order to acquire estimates of parameters in the mixture Gaussian distribution, we generalize the EM algorithm with real variables to the case with complex variables, which incorporate the learning process into the classical inverse medium scattering problem. Finally, the adjoint problem has been deduced and the RLM has been generalized to MGRLM based on the previous illustrations. Two numerical examples are given, which demonstrate the effectiveness of the proposed method.

This work is just a beginning, and there are a lot of problems need to be solved. For example, we did not give a principle for choosing parameter $K$ appeared in the mixture Gaussian distribution. In addition, in order to learn the model errors more accurately, we can attempt to design new algorithms to adjust the parameters in the mixture Gaussian distribution efficiently during the inverse iterative procedure.

## 6. Appendix

Here we gather the proofs of various conclusions stated in this paper. Including these proofs in the main text would break the flow of main ideas.

## Proof of Theorem 2.1

Proof. In order to prove this theorem, we need to verify three conditions stated in Assumptions 1 and Theorem 16 in Section 4.1 of 20]. Since

$$
\sum_{k=1}^{K} \pi_{k} \frac{1}{\pi^{N_{d}} \operatorname{det}\left(\Sigma_{k}+\nu I\right)} \exp \left(-\left\|d-\mathcal{F}_{a}(m)-\zeta_{k}\right\|_{\Sigma_{k}+\nu I}^{2}\right) \leq 1
$$

we know that

$$
\begin{equation*}
\Phi(m ; d) \leq 0 \tag{6.1}
\end{equation*}
$$

In the following, we denote

$$
f_{k}(d, m):=\left(d-\mathcal{F}_{a}(m)-\zeta_{k}\right)^{H}\left(\Sigma_{k}+\nu I\right)^{-1}\left(d-\mathcal{F}_{a}(m)-\zeta_{k}\right)
$$

Then, we have

$$
\begin{aligned}
\nabla_{d} f_{k}(d, m) & =\left(d-\mathcal{F}_{a}(m)-\zeta_{k}\right)^{H}\left(\Sigma_{k}+\nu I\right)^{-1}+\overline{\left(d-\mathcal{F}_{a}(m)-\zeta_{k}\right)^{H}\left(\Sigma_{k}+\nu I\right)^{-1}} \\
& =2 \operatorname{Re}\left(\left(d-\mathcal{F}_{a}(m)-\zeta_{k}\right)^{H}\left(\Sigma_{k}+\nu I\right)^{-1}\right)
\end{aligned}
$$

Through some simple calculations, we find that

$$
\begin{equation*}
\nabla_{d} \Phi(m ; d)=-\sum_{k=1}^{K} 2 g_{k} \operatorname{Re}\left(\left(d-\mathcal{F}_{a}(m)-\zeta_{k}\right)^{H}\left(\Sigma_{k}+\nu I\right)^{-1}\right) \tag{6.2}
\end{equation*}
$$

where

$$
\begin{equation*}
g_{k}:=\frac{\pi_{k} \mathcal{N}_{c}\left(d-\mathcal{F}_{a}(m) \mid \zeta_{k}, \Sigma_{k}+\nu I\right)}{\sum_{j=1}^{K} \pi_{j} \mathcal{N}_{c}\left(d-\mathcal{F}_{a}(m) \mid \zeta_{j}, \Sigma_{j}+\nu I\right)} . \tag{6.3}
\end{equation*}
$$

From the expression 6.2 and (1) of Assumption 1, we can deduce that

$$
\begin{equation*}
\left\|\nabla_{d} \Phi(m ; d)\right\|_{2} \leq C\left(1+\|d\|_{2}+\exp \left(\epsilon\|m\|_{X}^{2}\right)\right) \tag{6.4}
\end{equation*}
$$

where the constant $C$ depends on $K,\left\{\Sigma_{k}\right\}_{k=1}^{K}$ and $\left\{\zeta_{k}\right\}_{k=1}^{K}$. Considering (6.4), we obtain

$$
\begin{equation*}
\left|\Phi\left(m ; d_{1}\right)-\Phi\left(m ; d_{2}\right)\right| \leq C\left(1+r+\exp \left(\epsilon\|m\|_{X}^{2}\right)\right)\left\|d_{1}-d_{2}\right\|_{2} \tag{6.5}
\end{equation*}
$$

By our assumptions, the following relation obviously holds

$$
\begin{equation*}
C^{2}\left(1+r+\exp \left(\epsilon\|m\|_{X}^{2}\right)\right)^{2} \in L_{\mu_{0}}^{1}(X ; \mathbb{R}) \tag{6.6}
\end{equation*}
$$

where $L_{\mu_{0}}^{1}(X ; \mathbb{R})$ is the space of functions $f: X \rightarrow \mathbb{R}$ with norm

$$
\|f\|_{L_{\mu_{0}}^{1}(X ; \mathbb{R})}=\int_{X}|f(x)| \mu_{0}(d x) .
$$

At this stage, estimates (6.1), 6.5 and 6.6 verify Assumptions 1 and conditions of Theorem 16 in Section 4.1 of 20]. Employing theories constructed in [20], we complete the proof.

## Proof of Theorem 2.3

Proof. In order to prove this theorem, let us verify the following two conditions concerned with $\Phi(m ; d)$,
(1) For every $r>0$ there exists $M=M(r)>0$ such that, for all $m \in X$ with $\|m\|_{X}<r$ we have $\Phi(m ; d) \geq-M$.
(2) For every $r>0$ there exists $N=N(r)>0$ such that, for all $m_{1}, m_{2} \in X$ with $\left\|m_{1}\right\|_{X},\left\|m_{2}\right\|_{X}<r$ we have $\left|\Phi\left(m_{1} ; d\right)-\Phi\left(m_{2} ; d\right)\right| \leq N\left\|m_{1}-m_{2}\right\|_{X}$.
For the first condition, by employing Jensen's inequality, we have

$$
\begin{aligned}
\Phi(m ; d) & =\ln \left(\sum_{k=1}^{K} \pi_{k} \mathcal{N}_{c}\left(d-\mathcal{F}_{a}(m) \mid \zeta_{k}, \Sigma_{k}+\nu I\right)\right) \\
& \geq \sum_{k=1}^{K} \pi_{k} \ln \left(\frac{1}{\pi^{N_{d}}\left|\Sigma_{k}+\nu I\right|} \exp \left(-\left\|d-\mathcal{F}_{a}(m)-\zeta_{k}\right\|_{\Sigma_{k}+\nu I}^{2}\right)\right) \\
& \geq \sum_{k=1}^{K} \pi_{k}\left(-\left\|d-\mathcal{F}_{a}(m)-\zeta_{k}\right\|_{\Sigma_{k}+\nu I}^{2}-N_{d} \ln (\pi)-\ln \left(\left|\Sigma_{k}+\nu I\right|\right)\right) \\
& \geq-C\left(1+\|d\|_{2}^{2}+\exp \left(\epsilon r^{2}\right)\right)
\end{aligned}
$$

where $C$ is a positive constant depending on $K,\left\{\pi_{k}\right\}_{k=1}^{K},\left\{\Sigma_{k}\right\}_{k=1}^{K},\left\{\zeta_{k}\right\}_{k=1}^{K}$ and $N_{d}$. Now, the first condition holds true by choosing $M=C\left(1+\|d\|_{2}^{2}+\exp \left(\epsilon r^{2}\right)\right)$.

In order to verify the second condition, we denote

$$
f_{k}(d, m):=\left(d-\mathcal{F}_{a}(m)-\zeta_{k}\right)^{H}\left(\Sigma_{k}+\nu I\right)^{-1}\left(d-\mathcal{F}_{a}(m)-\zeta_{k}\right)
$$

then focus on the derivative of $f_{k}$ with respect to $m$. Through some calculations, we find that

$$
\begin{equation*}
\nabla_{m} f_{k}(d, m)=-2 \operatorname{Re}\left(\left(d-\mathcal{F}_{a}(m)-\zeta_{k}\right)^{H}\left(\Sigma_{k}+\nu I\right)^{-1} \mathcal{F}_{a}^{\prime}(m)\right) \tag{6.7}
\end{equation*}
$$

Hence, we have

$$
\begin{equation*}
\nabla_{m} \Phi(m ; d)=-\sum_{k=1}^{K} 2 g_{k} \operatorname{Re}\left(\left(d-\mathcal{F}_{a}(m)-\zeta_{k}\right)^{H}\left(\Sigma_{k}+\nu I\right)^{-1} \mathcal{F}_{a}^{\prime}(m)\right) \tag{6.8}
\end{equation*}
$$

where $g_{k}$ defined as in (6.3). Using Assumption 1 and formula (6.8), we find that

$$
\begin{equation*}
\left|\Phi\left(m_{1} ; d\right)-\Phi\left(m_{2} ; d\right)\right| \leq C L\left(1+\|d\|_{2}+\exp \left(\epsilon r^{2}\right)\right)\left\|m_{1}-m_{2}\right\|_{X} \tag{6.9}
\end{equation*}
$$

Let $N=C L\left(1+\|d\|_{2}+\exp \left(\epsilon r^{2}\right)\right)$, obviously the second condition holds true. Combining these two conditions with 6.1), we can complete the proof by using Theorem 19 in 20.

## Acknowledgments

The authors would like to thank the anonymous referees for their comments and suggestions, which helped to improve the paper significantly. This work was partially supported by the NSFC under grant Nos. 11871392, 11501439, 11771347, 91730306, 41390454 and partially supported by the Major projects of the NSFC under grant Nos. 41390450 and 41390454 , and partially supported by the postdoctoral science foundation project of China under grant no. 2017T100733. The first author would like to thank Prof. Deyu Meng and Dr. Qian Zhao's helpful discussions.

## References

[1] G. Bao, S. N. Chow, P. Li, and H. Zhou. Numerical solution of an inverse medium scattering problem with a stochastic source. Inverse Probl., 26(7):074014, 2010.
[2] G. Bao and P. Li. Inverse medium scattering for the Helmholtz equation at fixed frequency. Inverse Probl., 21(5):1621-1641, 2005.
[3] G. Bao and P. Li. Inverse medium scattering problems for electromagnetic waves. SIAM J. Appl. Math., 65(6):2049-2066, 2005.
[4] G. Bao, P. Li, J. Lin, and F. Triki. Inverse scattering problems with multi-frequencies. Inverse Probl., 31(9):093001, 2015.
[5] G. Bao, J. Lin, and F. Trikic. A multi-frequency inverse source problem. J. Differ. Equ., 249(12):3443-3465, 2010.
[6] G. Bao, S. Lu, W. Rundell, and B. Xu. A recursive algorithm for multi-frequency acoustic inverse source problem. SIAM J. Numer. Anal., 53(3):1608C1628, 2015.
[7] G. Bao and F. Triki. Error estimates for the recursive linearization of inverse medium problems. J. Comput. Math., 28(6):725-744, 2010.
[8] Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, New York, 2006.
[9] N. Bleistein, J. K. Cohen, and J. W. Stockwell, Jr. Mathematics of Multidimensional Seismic Imaging, Migration, and Inversion. Springer-Verlag, New York, 2001.
[10] M. Burger and F. Lucka. Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators. Inverse Probl., 30(11):114004, 2014.
[11] F. Cakoni and D. Colton. Qualitative Methods in Inverse Scattering Theory. Springer-Verlag, Berlin, 2006.
[12] D. Calvetti, M. Dunlop, E. Somersalo, and A. M. Stuart. Iterative updating of model error for Bayesian inversion. Inverse Probl., 34(2):025008, 2018.
[13] Y. Chen. Inverse scattering via Heisenberg's uncertainty principle. Inverse Probl., 13(2):253282, 1997.
[14] Y. Chen and V. Rokhlin. On the riccati equations for the scattering matrices in two dimensions. Inverse Probl., 13(1):1-13, 1997.
[15] D. Colton, J. Coyle, and P. Monk. Recent developments in inverse acoustic scattering theory. SIAM Rev., 42(3):369-414, 2000.
[16] D. Colton and R. Kress. Inverse Acoustic and Electromagnetic Scattering Theory, volume 93 of Applied Mathematical Sciences. Springer, New York, 2013.
[17] S. L. Cotter, M. Dashti, J. C. Robinson, and A. M. Stuart. Bayesian inverse problems for functions and applications to fluid mechanics. Inverse Probl., 25(11):115008, 2009.
[18] M. Dashti, S. Harris, and A. M. Stuart. Besov priors for Bayesian inverse problems. Inverse Probl. Imag., 6(2):183-200, 2012.
[19] M. Dashti, K. J. H. Law, A. M. Stuart, and J. Voss. MAP estimators and their consistency in Bayesian nonparametric inverse problems. Inverse Probl., 29(9):095017, 2013.
[20] M. Dashti and A. M. Stuart. The Bayesian approach to inverse problems. Handbook of Uncertainty Quantification, pages 311-428, 2017.
[21] M. M. Dunlop and A. M. Stuart. MAP estimators for piecewise continuous inversion. Inverse Probl., 32(10):105003, 2016.
[22] A. Fichtner. Full Seismic Waveform Modelling and Inversion. Springer, Berlin Heidelberg, 2011.
[23] N. R. Goodman. Statistical analysis based on a certain multivariate complex Gaussian distribution (An introduction). Ann. Stat., 34(1):152-177, 1963.
[24] M. A. Iglesias, K. Lin, and A. M. Stuart. Well-posed Bayesian geometric inverse problems arising in subsurface flow. Inverse Probl., 30(11):114001, 2014.
[25] K. Ito, B. Jin, and J. Zou. A two-stage method for inverse medium scattering. J. Comput. Phys., 237:211-223, 2013.
[26] J. Jia, J. Peng, and J. Gao. Bayesian approach to inverse problems for functions with a variable-index Besov prior. Inverse Probl., 32(8):085006, 2016.
[27] J. Jia, S. Yue, J. Peng, and J. Gao. Infinite-dimensional Bayesian approach for inverse scattering problems of a fractional Helmholtz equation. J. Funct. Anal., 275(9):2299-2332, 2018.
[28] B. Jin, D. A. Lorenz, and S. Schiffler. Elastic-net regularization: error estimates and active set methods. Inverse Probl., 25(11):115022, 2009.
[29] J. Kaipio and E. Somersalo. Statistical and Computational Inverse Problems. Springer-Verlag, New York, 2006.
[30] J. Koponen, T. Huttunen, T. Tarvainen, and J. P. Kaipio. Bayesian approximation error approach in full-wave ultrasound tomography. IEEE T. Ultrason. Ferr., 61(10):1627-1637, 2014.
[31] S. Lasanen. Non-Gaussian statistical inverse problems. Part I: Posterior distributions. Inverse Probl. Imag., 6:215, 2012.
[32] D. Meng, Q. Zhao, and Z. Xu. Improve robustness of sparse PCA by L1-norm maximization. Patterm Recogn., 45(1):487-497, 2012.
[33] L. Métivier, R. Brossier, Q. Mériǵot, E. Oudet, and J. Virieux. An optimal transport approach for seismic tomography: application to 3 D full waveform inversion. Inverse Probl., 32(11):115008, 2016.
[34] F. Natterer and F. Wubbeling. A propagation-backpropagation method for ultrasound tomography. Inverse Probl., 11(6):1225, 1995.
[35] R-E. Plessix. Three-dimensional frequency-domain full-waveform inversion with an iterative solver. Geophysics, 74(6):WCC149-WCC157, 2009.
[36] R. G. Pratt. Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model. Geophysics, 64(3):888-901, 1999.
[37] F. Qu, J. Yang, and B. Zhang. An approximation factorization method for inverse medium scattering with unknown buried objects. Inverse Probl., 33(3):035007, 2017.
[38] T. Regińska and K. Regiński. Approximate solution of a Cauchy problem for the Helmholtz equation. Inverse Probl., 22(3):975, 2006.
[39] A. M. Stuart. Inverse problems: a Bayesian perspective. Acta Numer., 19:451-559, 2010.
[40] Z. Wang and S. F. Wu. Helmholtz equation-least-squares method for reconstructing acoustic pressure fields. J. Acoust. Soc. Am., 102(5):3090, 1997.
[41] Z. Yao, Z. Hu, and J. Li. A TV-Gaussian prior for infinite-dimensional Bayesian inverse problems and its numerical implementations. Inverse Probl., 32(7):075006, 2016.
[42] H. Yong, D. Meng, W. Zuo, and L. Zhang. Robust online matrix factorization for dynamic background subtraction. IEEE T. Pattern Anal., PP(99):1-1, 2017.
[43] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and Y. Yan. L1-norm low-rank matrix factorization by variational Bayesian method. IEEE T. Neur. Net. Lear., 26(4):825-839, 2015.
[44] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and L. Zhang. Robust principal component analysis with complex noise. In Proceedings of the 31st International Conference on Machine Learning (ICML), pages 55-63, 2014.
[45] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. J. R. Statist. Soc. B, 67:301-320, 2005.

School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China
E-mail address: jjx323@xjtu.edu.cn
School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, 710049, ChiNA

E-mail address: bangyuwu@xjtu.edu.cn
School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China
E-mail address: jgpeng@xjtu.edu.cn
School of Electronic and Information Engineering, Xi’an Jiaotong University, Xi'an 710049, China

E-mail address: jhgao@xjtu.edu.cn


[^0]:    2010 Mathematics Subject Classification. 74G75, 74J20, 47A52.
    Key words and phrases. Inverse medium scattering, Model error learning, Adjoint state approach, Complex mixture Gaussian distribution.

