OPTIMAL SENSOR CONFIGURATION IN REMOTE IMAGE FORMATION

Behzad Sharif, Student Member, IEEE, and Farzad Kamalabadi, Member, IEEE

EDICS Category: GEO-SENS

Abstract

Determination of optimal sensor configuration is an important issue in many remote imaging modalities such as tomographic and interferometric imaging. In this paper, a statistical optimality criterion is defined and a search is performed over the space of candidate sensor locations to determine the configuration that optimizes the criterion over all candidates. To make the search process computationally feasible, a modified version of a previously proposed suboptimal backward greedy algorithm is used. A statistical framework is developed which allows for inclusion of several widely used image constraints. Computational complexity of the proposed algorithm is discussed and a fast implementation is described. Connections of the method to the deterministic backward greedy algorithm for the subset selection problem are presented and two application examples are described. Four compelling optimality criteria are considered and their performance is investigated through numerical experiments for a tomographic imaging scenario. In all cases, it is verified that the chosen configuration by the proposed algorithm performs better than wisely chosen alternatives.

I. INTRODUCTION

In many image formation scenarios involving multiple sensors, where the relationship between the set of observations and the unknown field can be adequately characterized by a linear observation model, the image reconstruction problem can be formulated by the Fredholm integral equation of the first kind [1]:

\[ Y(r, s) = \iint_{\Omega} a(r, s; r', s') X(r', s') dr' ds' \]

(1)

The authors are with the department of Electrical and Computer Engineering and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (E-mail: {sharif, farzadk}@uiuc.edu).
where a two-dimensional observation geometry is assumed with \( r \) and \( s \) denoting spatial variables, and \( \Omega \subset \mathbb{R}^2 \) is the region of support. Also, \( Y(r, s) \) and \( X(r, s) \) are the measured data and the unknown field respectively. The observation kernel is denoted by \( a(r, s; r', s') \). In practice, the observations are often a discrete sequence of measured data, \( \{y_i\}_{i=1}^m \). Furthermore, for a nonanalytical solution, the unknown field \( X(r, s) \) must be discretized. In what follows, it is assumed that the unknown field can be sufficiently represented by a weighted sum of \( n \) basis functions \( \{\phi_j(r, s)\}_{j=1}^n \) as follows:

\[
X(r, s) = \sum_{j=1}^n x_j \phi_j(r, s) \quad (2)
\]

For instance, \( \{\phi_j(r, s)\}_{j=1}^n \) are often chosen to be the set of unit height boxes corresponding to a two-dimensional array of square pixels. In that case, if a square \( g \times f \) pixel array is used, then \( n=g \cdot f \) and the discretized field is completely described by the set of coefficients \( \{x_j\}_{j=1}^n \), corresponding to the pixel values. Collecting all the observations into a vector \( y \) of length \( m \), and the unknown image coefficients into a vector \( x \) of length \( n \), results in the following observation model in form of a matrix equation:

\[
y = Ax + w \quad (3)
\]

where \( w \) is the additive measurement noise that is assumed to be independent of \( x \), and \( A \in \mathbb{C}^{m \times n} \) where \( m \geq n \) is the observation matrix, comprised of inner products of the basis functions with the corresponding observation kernel:

\[
(A)_{ij} = \int \int_{\Omega} a_i(r', s') \phi_j(r', s') dr' ds', \quad 1 \leq i \leq m, \quad 1 \leq j \leq n \quad (4)
\]

where \( a_i(r', s') = a(r_i, s_i; r', s') \) denotes the kernel function corresponding to the \( i \)-th observation. In case of image reconstruction from projection data, a ray is defined as a line running through the image plane in the absence of diffraction and scattering effects. Let \( y_i \) be the line integral measured by the \( i \)-th ray. Then \((A)_{ij}\) is the weighting factor that represents the contribution of the \( j \)-th cell to the \( i \)-th ray integral.

It is known that the placement of sensors affects the information content of the dataset and has implications on the reconstruction quality. In [2], [3], this relationship is studied by considering the distribution of the singular values of \( A \). In this paper, we approach the problem by developing a statistical framework for finding the optimal sensor configuration in remote imaging. Related works include [4] for optimal sampling in parallel magnetic resonance imaging and [5] for finding the optimal dithering pattern in a rectangular-grid array utilized for image formation from periodic nonuniform samples.

The task is to search over a set of candidate locations for sensors to find the optimal or close-to-optimal subset. Initially, it is assumed that there is a sensor located at each candidate location. It is
assumed that there are $p$ candidate locations and each sensor takes $d$ measurements. Since there are possibly hundreds of candidate locations, the initial observation matrix, denoted by $A^0 \in \mathbb{C}^{m \times n}$ where $m = p \cdot d$, is typically very large. Selecting a subset of candidate locations is equivalent to choosing a subset of rows of $A^0$. An optimality criterion is needed for the choice of rows, and subsequently the resulting combinatorial optimization problem must be tackled. The optimality criterion is designed such that it only depends on the statistics of $x$ and $w$ and the observation kernel. Assuming the statistics of the problem are invariant under different choices of sensor configuration, the cost function will only be a function of the observation kernel. Denoting the set of all candidate locations by $S$ and the observation matrix corresponding to $U \subseteq S$ by $A^{(U)}$, the following optimization problem is reached:

$$S^* = \arg \min_{U \subseteq S, |U|=q} \text{Cost}(A^{(U)})$$

(5)

where $|.|$ denotes the cardinality, $q$ is the desired number of sensors, and $\text{Cost}(A^{(U)})$ is the cost of choosing subset $U$ as the locations for the $q$ sensors. The best subset of rows of $A^0$ can be found by exhaustive search over all possible combinations $\binom{p}{q}$ of rows. But this grows exponentially in $p$ and is impractical even for moderate number of candidate locations. It is possible to exploit the structure of the problem to reduce the size of the search space by means of Branch and Bound-type algorithms [6]. Again, for practical situations, even the restricted search space of Branch and Bound methods becomes too large to handle. In order to avoid the exhaustive search, one has to resort to suboptimal heuristic techniques [7]. In general, in all subset selection problems of this type there is an inherent complexity/performance trade-off [8].

In order to make the search computationally feasible, a modified version of the sequential backward selection (SBS) algorithm [9] is used. The original SBS algorithm eliminates the least important row at each step. The algorithm stops when the desired number of rows remain. Although the approach is suboptimal, it eliminates the exhaustive search required for optimally solving the subset selection problem and has performed consistently well in all the examples tried in the existing literature [6], [9]. Note that our problem is slightly different from the one that SBS aims to solve. Each candidate sensor location contributes several rows to the $A^0$ matrix. Hence, instead of considering each row of $A^0$ independently and deciding whether to eliminate it or not (as done in SBS), a group of rows contributed by each sensor is considered at each step. At every iteration, the cost of removing each group of rows is calculated and the group which incurs the least increase in the cost is eliminated together with its corresponding sensor. The stop condition is when the number of sensors is reduced to the predesigned value. The proposed algorithm is referred to as the clustered SBS (CSBS) algorithm.
The paper is organized as follows. Section II develops a statistical framework which will be used in subsequent sections to formulate the problem. Section III introduces several optimality criteria. In Section IV, we formally introduce the CSBS algorithm and elaborate on its complexity, performance and optimality. In Section V, two important imaging modalities are described as examples that motivate the application of proposed method. Section VI contains the simulation results. Finally, Section VII concludes the paper.

II. Statistical Formulation

In a typical remote image formation application, due to limitations in the observation geometry, the corresponding observation matrix is highly ill-conditioned and the resulting inverse problem formulated in (3) is ill-posed. Tikhonov (quadratic) regularization [10] is one of the more common techniques used to overcome this issue and is equivalent to maximum a posteriori (MAP) estimation assuming Gaussian statistics for both the unknown image and noise [11]. Assuming \( w \sim \mathcal{CN}(0, \Sigma_w) \) and \( x \sim \mathcal{CN}(x_0, \Sigma_x) \), where \( \mathcal{CN}(\mu, \Sigma) \) represents the complex normal distribution with mean \( \mu \) and covariance \( \Sigma \), the MAP estimate is

\[
\hat{x}_{\text{MAP}} = \arg \min_{x \in \mathbb{C}^n} \left[ -\log p(y|x) - \log p(x) \right] = \arg \min_{x \in \mathbb{C}^n} \left[ \|y - Ax\|^2_{\Sigma_w^{-1}} + \|x - x_0\|^2_{\Sigma_x^{-1}} \right] = x_0 + (A^H \Sigma_w^{-1}A + \Sigma_x^{-1})^{-1}A^H \Sigma_w^{-1}(y - Ax_0)
\]

Assuming independent identically distributed (IID) Gaussian noise and taking \( \Sigma_x = \frac{1}{\gamma^2} (L^T L)^{-1} \), we arrive at the well known Tikhonov regularization functional:

\[
\hat{x}_{\text{Tik}} = \arg \min_{x \in \mathbb{C}^n} \left[ \frac{1}{\sigma_w^2} \|y - Ax\|^2 + \gamma^2 \|L(x - x_0)\|^2 \right] = \arg \min_{x \in \mathbb{C}^n} \left[ \|y - Ax\|^2 + \lambda \|L(x - x_0)\|^2 \right] = x_0 + \left( \frac{1}{\sigma_w^2} A^H A + \gamma^2 L^T L \right)^{-1} \frac{1}{\sigma_w^2} A^H (y - Ax_0)
\]

where \( L \) is the positive definite regularization matrix and \( \lambda = (\gamma \sigma_w)^2 \) where \( \sigma_w^2 \) is the variance of the noise samples. A special case is when \( L = I \), which results in \( \lambda \) being inverse of the signal-to-noise ratio. Although we assumed IID noise, more general forms of noise covariance have been applied in remote sensing applications [12], [13].
The statistical framework allows for a closed-form measure of estimation uncertainty through the error covariance matrix \( \Sigma_e = E[ee^T] \), where \( e = x - \hat{x}_{\text{Tikh}} \). For the MAP estimate, the error covariance is the inverse of the corresponding Fisher information matrix and is given by

\[
\Sigma_e = (\mathbf{A}^T \Sigma_w^{-1} \mathbf{A} + \Sigma_x^{-1})^{-1}
\]

where the expected squared error for the \( i \)-th element of \( x \) is the \((i, i)\)-th element of \( \Sigma_e \), denoted by \((\Sigma_e)_{ii}\). Consequently, the error covariance for Tikhonov regularized reconstruction is \( \Sigma_e = (\frac{1}{\sigma_w} \mathbf{A}^T \mathbf{A} + \gamma^2 \mathbf{L}^T \mathbf{L})^{-1} \). It should be noted that with no assumption on the distribution of the image (i.e., non-Gaussian statistics) the estimator in (7) is the linear minimum mean square error (LMMSE) estimator for \( x \) which minimizes \( E[\|x - \hat{x}\|^2] \). Therefore, the results of this paper will in general be valid in the context of LMMSE estimation.

As we will see in the next section, selection of optimality criteria requires considering costs that are functions of \( \Sigma_e \). As shown in (8), this in turn requires estimating \( \Sigma_x \) or equivalently the regularization matrix \( \mathbf{L} \). In what follows, we show how to incorporate four different forms of constraints into the statistical framework established in this section in order to estimate \( \Sigma_x \).

1) **Smoothness Constraint**: The \( \mathbf{L} \) matrix in (7) is typically taken to be a discrete approximation to the gradient operator in order to formulate smoothness of the unknown image. The discrete approximation to the first order horizontal derivative has the following form

\[
\mathbf{L} = \begin{bmatrix}
1 & -1 & 0 & \cdots \\
0 & 1 & -1 & 0 & \cdots \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & \cdots & 1 & -1
\end{bmatrix}
\]

(9)

Using this regularization matrix is equivalent to assuming \( x \) to be a Brownian motion [11]. It is possible to use higher order derivatives or to adapt these matrices to work on vertical neighbors [14]. An alternative class of regularization matrices are discretizations of the two-dimensional Laplacian operator, namely Fried and Hudgin discrete Laplacians [15], [16]. In some applications, multiple regularization matrices are used. One example is when it is desired to weight smoothness in horizontal direction with respect to vertical direction:

\[
\hat{x} = \arg \min_{x \in \mathbb{R}^n} \left[ \|y - \mathbf{A}x\|^2_2 + \lambda_h \|\mathbf{L}_h x\|^2_2 + \lambda_v \|\mathbf{L}_v x\|^2_2 \right]
\]

(10)

This can be written in form of (7) if \( \mathbf{L}^T \mathbf{L} = \lambda_h \mathbf{L}_h^T \mathbf{L}_h + \lambda_v \mathbf{L}_v^T \mathbf{L}_v \) (assuming \( \lambda = 1 \) and \( x_0 = 0 \)). Since the right hand side is symmetric positive definite, the equivalent \( \mathbf{L} \) matrix exists and is equal to the positive definite square root of the right hand side.
2) *Support Constraint:* In some occasions, the unknown image $x \in \mathbb{C}^{m \times n}$ is expected to vanish outside a region $\mathcal{H}$. It is possible to design a matrix $B$ such that:

$$
h(x) = \int_{(r,s) \in \mathcal{H}} |X(r,s)|^2 dr ds \approx \|Bx\|_2^2
$$

(11)

One such construction is a diagonal matrix with $(B)_{ii}$ being one if $x_i$ lies inside $\mathcal{H}$ and zero otherwise. Having $B$, one can formulate the support constraint by adding the term $\lambda_s h(x)$ to penalize the functional in (7). As described in item 1, the added penalty term can be combined with the regularization term to give a new regularization matrix equal to the positive definite square root of $\lambda L^T L + \lambda_s B^T B$ (assuming $x_0 = 0$).

3) *Reference Constraint:* In some applications, it is known that $x \in \mathcal{C}$ where

$$
\mathcal{C} = \{x' \in \mathbb{C}^n : \|x' - x_0\|_2 < \rho\}
$$

(12)

for some reference image $x_0$ and radius $\rho > 0$. This is a convex constraint and is referred to as the reference (prototype) image constraint [17]. If no knowledge of statistics of $x$ exists, using the maximum likelihood framework together with the reference constraint leads to a constrained maximum likelihood (CML) estimation problem [18]. The CML estimation is equivalent to an unconstrained MAP estimation which, assuming Gaussian noise, reduces to Tikhonov regularization:

$$
\hat{x}_{CML} = \arg \min_{x \in \mathcal{C}} [-\log p(y|x)]
= \arg \min_{x \in \mathbb{C}^n} [-\log p(y|x) + \nu(\rho, y)\|x - x_0\|_2^2]
= \arg \min_{x \in \mathbb{C}^n} [\|y - Ax\|_2^2 + \nu(\rho, y)\|x - x_0\|_2^2]
$$

where $\nu(\rho, y) \in \mathbb{R}^+$ is the associated Lagrange multiplier (see Theorem 1, p. 217 of [19]). As can be seen, the CML estimation has the same form as the Tikhonov functional with $L = I_{n \times n}$ and $\lambda = \nu(\rho, y)$. Hence (8) applies here as well.

4) *Energy Constraint:* Energy constraint is a special case of the reference constraint with $x_0 = 0$ and $\rho = \sqrt{E_0}$ where $E_0$ is the energy of the image. Hence, all the results in item 3 apply.

### III. Optimality Criteria

In this section, based on the framework developed in the last section, four compelling choices for the $\text{Cost}(A)$ function in (5) are introduced:

1) *Sum of Squared Errors (SSE):* The expected value of the squared estimation error is given by

$$
\text{Cost}^1(A) = E[\|e\|^2] = \text{tr}(\Sigma_e) = \sum_{i=1}^{n} (\Sigma_e)_{ii}
$$

(13)
It is worth mentioning that if no assumption about the statistics of the signal are made and only
the noise is assumed to be white Gaussian, the SSE criterion will take the form of \( \text{tr}\{(A^H A)^{-1}\} \)
where the inverse operator is usually replaced by the pseudo inverse to avoid instability due to
ill-posedness of the problem [6].

2) **Weighted Sum of Squared Errors (WSSE):** In some applications it is desirable to have small error in
one specific area of the reconstructed image while larger errors could be tolerated in other areas. A
natural approach to designing the cost function for this setting is to weight \( \text{Cost}^1(A) \) as follows:

\[
\text{Cost}^2(A) = \sum_{i=1}^{n} W_i \cdot (\Sigma_e)_{ii}
\]

where \( W_i \in \mathbb{R}^+ \) are the weighting coefficients.

3) **Uniformity of Squared Errors (USE):** In some applications, if the location of the feature of interest
is not known beforehand, it is desirable to place the sensors such that the error is distributed evenly
over all pixels. Intuitively, this will minimize the cost of the worst-case scenario. This goal implies
a different cost function:

\[
\text{Cost}^3(A) = \text{STD}\{(\Sigma_e)_{ii}\}_{i=1}^{n}
\]

where \( \text{STD}\{c_i\}_{i=1}^{n} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (c_i - \frac{1}{n} \sum_{j=1}^{n} c_j)^2} \).

4) **Detection Performance:** In particular applications, the goal is to reliably detect the presence of a
feature in the image rather than accurate reconstruction of the image. One example is detection
of presence of a plasma depletion in tomographic imaging of the ionosphere. In [4], the authors
formulate the problem as a binary hypothesis testing problem as follows. Assume the problem is
to decide whether a feature \( b \) is present in the background \( x \). Both the feature and the background
are unknown and are modeled as uncorrelated Gaussian random vectors distributed as \( \mathcal{CN}(b_0, \Sigma_b) \)
and \( \mathcal{CN}(x_0, \Sigma_x) \) respectively. Also the noise vector \( w \) is a Gaussian distributed as \( \mathcal{CN}(0, \Sigma_w) \)
and is independent of both \( b \) and \( x \). Using the imaging equation (3), the detection problem can be
modeled as the following binary hypothesis testing problem

\[
H_0 : y = Ax + w
\]

\[
H_1 : y = A(x + b) + w
\]

Assuming \( A \in \mathbb{R}^{m \times n} \), the observation vector \( y \) is also a real Gaussian random vector. The
covariance matrix of \( y \) under hypothesis \( H_0 \) is

\[
\Sigma_{y|H_0} = A \Sigma_x A^H + \Sigma_w
\]
and under $H_1$ is
\[ \Sigma_{y|H_1} = \mathbf{A}(\Sigma_x + \Sigma_b)\mathbf{A}^H + \Sigma_w \] (17)

The optimal choice of receiver configuration would be the one that minimizes the probability of error for the above hypothesis testing problem. Unfortunately, the corresponding optimization problem is analytically intractable. One way of tackling this issue is to maximize the divergence between the observation distributions under $H_0$ and $H_1$. In [4], the use of Bhattacharya distance [20] for this purpose is proposed, as given by
\[
Bhat(\mathbf{A}) = \frac{1}{2} \ln \frac{|\Sigma_{y|H_1}| |\Sigma_{y|H_0}|}{|\Sigma|} 
+ \frac{1}{2} \ln \left| \frac{|\Sigma_{y|H_0}|}{\sqrt{|\Sigma_{y|H_1}|}} \right|
\] (18)

where $|.|$ denotes the determinant operator and $\mathbf{m}_i = E[y|H_i]$ and $\Sigma = \frac{1}{2}(\Sigma_{y|H_0} + \Sigma_{y|H_1})$. For the above setting, $\mathbf{m}_0 = \mathbf{A}\mathbf{x}_0$ and $\mathbf{m}_1 = \mathbf{A}(\mathbf{x}_0 + \mathbf{b}_0)$. However, in most of the remote imaging applications the size of matrices are very large and computation of the determinant in (18) is not practical due to limited numerical precision. We propose using the so called J-divergence [20] which does not include any determinant terms and is defined as
\[
J \text{div}(\mathbf{A}) = \frac{1}{2} \text{tr} \left( [\Sigma_{y|H_1} - \Sigma_{y|H_0}] [\Sigma_{y|H_0}^{-1} - \Sigma_{y|H_1}^{-1}] \right) + \frac{1}{2} \text{tr} \left( \left( \Sigma_{y|H_1}^{-1} + \Sigma_{y|H_0}^{-1} \right) (\mathbf{m}_1 - \mathbf{m}_0)(\mathbf{m}_1 - \mathbf{m}_0)^H \right)
\] (19)

Therefore, the cost function corresponding to the detection performance criteria is
\[ \text{Cost}^4(\mathbf{A}) = J \text{div}(\mathbf{A}) \] (20)

IV. THE CSBS ALGORITHM

Denote by $\Gamma$ the set of available sensor locations. Initially, $\Gamma$ contains all of the candidate sensor locations, i.e., $\mathbf{A}(\Gamma) = \mathbf{A}^0$. The CSBS algorithm can be formally written as:
\[
\Gamma \leftarrow \Gamma \setminus \{k^*\} : k^* = \arg \min_{k \in \Gamma} \text{Cost}(\mathbf{A}(\Gamma \setminus \{k\}))
\] (21)
with the stopping criterion being $|\Gamma| = q$.

In Subsection IV-A, first we analyze the computational complexity of the general CSBS algorithm. Next, based on the results in [9], we introduce a fast implementation of the CSBS algorithm with SSE criteria. In Subsection IV-B, We develop upper bounds for the sum of squared errors of the CSBS algorithm. This provides a measure of performance guarantee and insight into the behavior of the proposed algorithm.
Finally in Subsection IV-C, we establish a connection between CSBS and the deterministic backward greedy algorithm and suggest a conjecture on the optimality of the SBS and CSBS algorithms.

A. Computational Complexity and Fast Implementation

In the general CSBS algorithm stated in (21), there can be at most \( p - 1 \) iterations (because \( q > 0 \)) and at \( i \)-th iteration we need to compute the cost function for the existing \( p - i + 1 \) sensors. Therefore, in the worst case we need to compute the cost function \( \sum_{i=2}^{p} i = p(p + 1)/2 - 1 \) times. Consider the SSE criterion. We need to compute the error covariance matrix in (8). Assuming that the constant matrices are computed and stored beforehand, straightforward computation of \( \mathbf{A}^H \Sigma_w^{-1} \mathbf{A} \) is of \( O(mn^2 + nm^2) \) and we need an additional \( O(n^3) \) for inversion of the sum. Having \( m \geq n \), complexity of computing the cost function is \( O(nm^2) \). Therefore, the overall complexity is \( O(nm^2p^2) \). The same result holds for the other two criteria.

Using the Sherman-Morrison matrix inversion formula [21], Reeves et al. in [9] have developed an efficient implementation of the SBS algorithm. Here, we first restate Reeves’ method for SBS and based on that result improve upon the computational complexity of the CSBS algorithm in case of SSE criterion. If we eliminate the \( i \)-th row of \( \mathbf{A} \), denoted by \( \mathbf{a}_i \), the SSE corresponding to the modified matrix is given by [9]

\[
\text{Cost}^1(\mathbf{A}) + \frac{\mathbf{a}_i \Sigma_e^2 \mathbf{a}_i^H}{1 - \mathbf{a}_i \Sigma_e \mathbf{a}_i^H}
\]

where \( \Sigma_e \) is as given in (8). Therefore, SBS can be implemented by choosing the row that minimizes the second term of the right hand side. Based on this fact, a fast implementation of CSBS can be derived as follows:

\[
\mathbf{\Gamma} \leftarrow \mathbf{\Gamma} \setminus \{k^*\} : k^* = \arg \min_{k \in \mathbf{\Gamma}} \sum_{i \in \Pi_k} \frac{\mathbf{a}_i \Sigma_e^2 \mathbf{a}_i^H}{1 - \mathbf{a}_i \Sigma_e \mathbf{a}_i^H}
\]

(22)

where \( \Pi_k \) is the set of indices of rows measured by the \( k \)-th sensor and \( \Sigma_e = (\mathbf{A}(\mathbf{\Gamma})^H \Sigma_w^{-1} \mathbf{A}(\mathbf{\Gamma}) + \Sigma_x^{-1})^{-1} \).

Note that for each iteration of the algorithm in (22), we only need to compute and store \( \Sigma_e \) once which has complexity \( O(nm^2) \) as discussed above. For each iteration, i.e. elimination of one sensor, the summands in (22) can be computed in \( O(n^2p) \). So each iteration is of \( O(nm^2 + n^2p) \) which is of \( O(nm^2) \) since \( m \geq n \) and \( m \geq p \). This results in an overall complexity of \( O(nm^2p) \). Therefore, we have reduced the complexity by a factor of \( p \) which can be significant in practical applications.
B. Upper bounds on performance of CSBS

The CSBS algorithm is greedy and therefore suboptimal. It is known that a greedy observation selection algorithm can give the worst possible combination of rows for some applications [22]. In this section, we provide an upper bound for SSE of the CSBS algorithm (with SSE criterion) that is valid under a certain condition. If the condition is met, the upper bound provides a performance guarantee for the CSBS algorithm. Throughout this section, the \( A \) matrices involved are the initial \( A^0 \) matrix but for notational simplicity we drop the superscript. To proceed we need the following result from [9].

**Lemma 1 (Upper Bound for SBS):** If the initial number of rows is \( d \) and the final number after applying SBS is \( k \), the final SSE is bounded above by

\[
\text{Cost}_1(A) = \text{tr}\{(\tilde{A}^H \tilde{A} + \hat{K})^{-1}\} = \text{tr}\{(\tilde{A}_i^H \tilde{A}_i + P + K)^{-1}\}
\]

where \( \tilde{A}_i \in \mathbb{C}^{d \times n} \) is the initial matrix.

**Proof:** See Theorem 1 in [9] for proof. \( \square \)

Since \( \Sigma_w^{-1} \) is positive definite Hermitian, it has a positive definite square root, namely \( \sqrt{\Sigma_w^{-1}} \). With \( \hat{K} = \Sigma_x^{-1} \) and \( \hat{A} = \sqrt{\Sigma_w^{-1}}A \), Lemma 1 provides a performance guarantee for the SBS algorithm [9].

Considering the CSBS algorithm and defining \( \hat{A} = \sqrt{\Sigma_w^{-1}}A \), the SSE in (8) can be written as:

\[
\Sigma_e = (\hat{A}^H \hat{A} + \Sigma_x^{-1})^{-1}
\]  

(23)

Denote by \( A_i \) the matrix formed by all rows of \( A \) with indices in \( \Pi_i \). In other words, \( A_i \in \mathbb{C}^{d \times n} \) is the collection of all measurements by the \( i \)-th sensor. According to (23), the SSE can be written as:

\[
\text{Cost}_1(A) = \text{tr}\{(\tilde{A}_i^H \tilde{A}_i + \hat{K})^{-1}\} = \text{tr}\{(\hat{A}_i^H \hat{A}_i + P + K)^{-1}\}
\]

where \( \hat{A}_i = \sqrt{\Sigma_w^{-1}}A_i \) and \( P = \sum_{j \neq i} \hat{A}_j^H \hat{A}_j \) and \( K = \Sigma_x^{-1} \).

Now we are in a position to obtain a closed-form expression for SSE after the \( i \)-th sensor is removed. By assigning \( \hat{A} = \hat{A}_i \) and \( \hat{K} = P + K \), Lemma 1 gives the upper bound for SSE if \( d - k \) rows of \( \hat{A}_i \) are removed. But eliminating a sensor means that all the corresponding rows should be removed. Therefore, \( k \) is zero which gives the following SSE for a fixed \( i \):

\[
\frac{d - n + 1 + \text{tr}\{(\hat{A}_i^H \hat{A}_i + K)^{-1}(P + K)\}}{-n + 1 + \text{tr}\{(\hat{A}_i^H \hat{A}_i + K)^{-1}(P + K)\}}
\]

(24)
where we used $\tilde{A}_i^H \tilde{A}_i + P = \sum_{j=1}^p \tilde{A}_j^H \tilde{A}_j = \tilde{A}^H \tilde{A}$. The expression in (24) is in effect an upper bound for $\text{tr}\{(P + K)^{-1}\}$ since no columns of $\tilde{A}_i$ are left after removing sensor $i$. Noting that

$$\text{tr}\{(\tilde{A}^H \tilde{A} + K)^{-1}(\tilde{A}^H \tilde{A} + K - \tilde{A}_i^H \tilde{A}_i)\} = n - \text{tr}\{\tilde{A}_i^H \tilde{A}_i(\tilde{A}^H \tilde{A} + K)^{-1}\}$$

we can write (24) as

$$\frac{d + 1 - \text{tr}\{\tilde{A}_i^H \tilde{A}_i(\tilde{A}^H \tilde{A} + K)^{-1}\} \text{tr}\{(\tilde{A}^H \tilde{A} + K)^{-1}\}}{1 - \text{tr}\{\tilde{A}_i^H \tilde{A}_i(\tilde{A}^H \tilde{A} + K)^{-1}\}}$$

(25)

It should be noted that the bound in Lemma 1 is valid only if the denominator is positive. In fact, considering all $1 \leq i \leq p$, the expression in (25) is a valid upper bound for a SSE if

$$\text{tr}\{\tilde{A}_i^H \tilde{A}_i(\tilde{A}^H \tilde{A} + K)^{-1}\} < 1 \text{ for } 1 \leq i \leq p$$

(C.1)

The following lemma provides an easily verifiable sufficient condition under which (C.1) holds.

**Lemma 2:** With $\Sigma_x = \frac{1}{\gamma}(L^T L)^{-1}$, a sufficient condition for (C.1) to be satisfied is

$$\gamma^2 \geq \max_{1 \leq i \leq p} \left( \text{tr}\{\tilde{A}_i^H \tilde{A}_i(L^T L)^{-1}\} \right)$$

(26)

And assuming IID Gaussian noise (26) is equivalent to

$$\lambda \geq \max_{1 \leq i \leq p} \left( \text{tr}\{A_i^H A_i(L^T L)^{-1}\} \right)$$

(27)

**Proof:** The proof is provided in Appendix I.

Therefore, if the reconstruction is regularized enough then (C.1) holds. Under (C.1), the main result of this subsection that is stated in the following theorem provides an upper bound (independent of $i$) for (25).

**Theorem 1 (Upper Bound for CSBS):** With $A \in \mathbb{C}^{p \times n}$ as the initial matrix, the SSE after removing one sensor using the CSBS algorithm (with SSE criterion) is bounded above by

$$\frac{(d + 1)p - n + \text{tr}\{\tilde{A}_i^H \tilde{A}_i(\tilde{A}^H \tilde{A} + K)^{-1}\}}{p - n + \text{tr}\{(\tilde{A}^H \tilde{A} + K)^{-1}\}} \text{tr}\{(\tilde{A}^H \tilde{A} + K)^{-1}\}$$

$$= \frac{(d + 1)p - n + \text{tr}\{\Sigma^{-1}_x \Sigma_e\}}{p - n + \text{tr}\{\Sigma^{-1}_e\}} \text{tr}\{\Sigma_e\}$$

provided that (C.1) holds.

**Proof:** The proof is by contradiction. Assume that for all $1 \leq i \leq p$ the bound on SSE after removing one sensor (stated in (25)) is larger than the bound in Theorem 1. We will show that this will result in a contradiction. Therefore, there exists at least one $1 \leq i \leq p$ that violates the assumption. Since
CSBS picks the best \( i \) (in SSE sense), it has to violate the assumption which means that it will satisfy the bound in Theorem 1. To proceed, we assume that for all \( 1 \leq i \leq p \) we have that
\[
\frac{d + 1 - \text{tr}\{\tilde{A}^H_i \tilde{A}_i (\tilde{A}^H \tilde{A} + K)^{-1}\}}{1 - \text{tr}\{\tilde{A}^H_i \tilde{A}_i (\tilde{A}^H \tilde{A} + K)^{-1}\}} \text{tr}\{(\tilde{A}^H \tilde{A} + K)^{-1}\}
\]
(28)
is larger than
\[
\frac{(d + 1)p - n + \text{tr}\{\hat{A}^H_i \hat{A}_i (\hat{A}^H + K)^{-1}\}}{p - n + \text{tr}\{\hat{A}^H_i \hat{A}_i (\hat{A}^H + K)^{-1}\}} \text{tr}\{(\hat{A}^H \hat{A} + K)^{-1}\}
\]
(29)
The common term in the expressions above is positive since it is a SSE. The numerator and denominator is (28) are both positive because of (C.1). The numerator in (29) is positive since \( d \cdot p = m \) which was assumed to be larger that \( n \). And the matrix inside the trace is positive definite since it is a multiplication of two square positive definite matrices. As stated and proved in Appendix II (Lemma 3), the denominator of (29) is also positive under (C.1). Therefore, eliminating the common term and rearranging gives:
\[
(p - n + \text{tr}\{GK\})(d + 1 - \text{tr}\{\hat{A}^H_i \hat{A}_i G\}) > (1 - \text{tr}\{\hat{A}^H_i \hat{A}_i G\})(d + 1)p - n + \text{tr}\{GK\}
\]
(30)
where \( G = (\hat{A}^H \hat{A} + K)^{-1} \). Since (30) holds for all \( 1 \leq i \leq p \), its summation should hold as well:
\[
(p - n + \text{tr}\{GK\}) \sum_{i=1}^{p} (d + 1 - \text{tr}\{\hat{A}^H_i \hat{A}_i G\}) > ((d + 1)p - n + \text{tr}\{GK\}) \sum_{i=1}^{p} (1 - \text{tr}\{\hat{A}^H_i \hat{A}_i G\})
\]
(31)
But
\[
\sum_{i=1}^{p} (\text{tr}\{\hat{A}^H_i \hat{A}_i G\}) = \text{tr}\{G \sum_{i=1}^{p} \hat{A}^H_i \hat{A}_i\} = \text{tr}\{G \hat{A}^H \hat{A}\}
\]
(32)
and
\[
\text{tr}\{G(\hat{A}^H \hat{A})\} = \text{tr}\{G(\hat{A}^H \hat{A} + K - K)\} = \text{tr}\{(\hat{A}^H \hat{A} + K)^{-1}(\hat{A}^H \hat{A} + K - K)\} = n - \text{tr}\{GK\}
\]
(33)
Plugging (32) and (33) in (31) gives:
\[
(p - n + \text{tr}\{GK\})((d + 1)p - n + \text{tr}\{GK\}) > ((d + 1)p - n + \text{tr}\{GK\})(p - n + \text{tr}\{GK\})
\]
which is a contradiction. Hence, there exists at least one \( i \) that violates (30). As explained in the beginning of the proof, this establishes the claim.
C. Conjecture on Optimality of SBS and CSBS

The arsenal of greedy algorithms for subset selection can be classified into two categories: forward greedy and backward greedy. In a forward greedy algorithm such as matching pursuit [21] and its variations [23], the idea is to start by finding the row of $A^0$ closest to $y$ and then to proceed by adding, at each step, the column that gives the largest drop in the least squares residual until $q$ columns are selected. In contrast, the deterministic backward greedy (DBG) algorithm [24] starts with all of the rows of $A^0$ as the set of candidates. At each step, the row that minimizes the increment in the least squares residual is removed. Both SBS and CSBS algorithms are also backward greedy. The main difference between SBS or CSBS and DBG is that they do not use any knowledge of the measured data $y$ and instead utilize the error covariance matrix computed in (8). In other words, the DBG algorithm works in the Euclidean Hilbert space whereas the SBS and CSBS algorithms (with the SSE criterion) work in the Hilbert space of random variables. In fact, the SSE criterion in (13) can be written as:

$$\text{Cost}^1(A) = \min \|x - \hat{x}\|_E^2$$

(34)

where $\|z\|_E^2 = \sum_{i=1}^n E[|z_i|^2]$ is the induced norm of the Hilbert space of zero-mean, finite-variance random variables. In [24], it has been shown that the DBG algorithm is guaranteed to select the correct subset of rows of $A^0$ if the noise level is small enough. This suggests that there exists a similar optimality result for both SBS and CSBS. Thus, we propose the following conjecture:

**Conjecture (Optimality of SBS and CSBS algorithms)**: If the statistical assumptions mentioned in Section II hold (if (8) applies) and under benign constraints on the condition number of the $A^0$ matrix, the SBS and CSBS algorithms (with SSE criterion) select the optimal sensor configuration (in SSE sense) provided that the noise variance is small enough.

V. Examples of Imaging Modalities

A. Ionospheric Tomography

In ionospheric tomography, the goal is to reconstruct the ionospheric electron density from a set of projection data which are typically total electron content (TEC) in the case of radio tomography or photometric brightness measurements in the case of optical tomography. In ground-based radio tomography, coherent transmissions from a low earth orbit (LEO) satellite are tracked by an array of ground sensors [25]. The goal of this technique is to reconstruct horizontal and vertical structures within the two-dimensional slice of the ionosphere above the ground sensors. In Figure 1, a scenario with three ground sensors and a LEO satellite at a 735-km orbit is shown. Each line in the figure illustrates one
integral path between the location of the satellite and one of the ground sensors at a fixed time. A grid (in polar coordinates) is defined in the vertical plane above the receiver chain. The components of the $A$ matrix correspond to the distance that each radio link travels through each pixel. The corresponding inverse problem can be formulated into a linear set of equations as in (3).

B. Interferometric Imaging in Radio Astronomy and Radar Remote Sensing

Radio interferometric imaging is the common imaging modality in radio astronomy, the most well-known example of which is the Very Large Array astronomical radio observatory [26]. Radar remote sensing of ionospheric plasma irregularities represents another relevant scenario. Huge antenna arrays such as Jicamarca Radio Observatory near Lima, Peru are used for these purposes and interferometric imaging techniques are used for this soft target detection application [27]. In both applications, the objective is to image the intensity of an unknown two-dimensional sky. In radio astronomy, the antenna are typically fixed and the earth rotation places them at new effective look positions [28]. In radar imaging, the antenna array beam can be steered to form different look positions (see Ch. 7 of [29]). Assume that the discretized sky is made up of $n$ pixels, $x = [x_1 \cdots x_n]^T \in \mathbb{C}^n$. Suppose the two-dimensional antenna array has $p$ elements and we take data at $d$ look positions. Denote by $\phi_j$ and $\theta_j$ the azimuth and zenith angles that the $j$-th pixel form relative to the center of the array. Assuming that $\theta_j$ and $\phi_j$ are small (less than 0.1 radian) so that small-angle approximations to trigonometric functions apply, each antenna sees the signal from the $j$-th pixel with a phase shift of $\exp(j2\pi[r_k(i)\theta_j + s_k(i)\phi_j])$, where $(r_k(i), s_k(i))$ are the
coordinates of the \( k \)-th antenna at \( i \)-th look position. Therefore, the data received at \( k \)-th antenna is:

\[
y_k(i) = \sum_{j=1}^{n} x_j e^{j2\pi[r_k(i) \delta_i + s_k(i) \phi_j]} + w_k(i) \tag{35}
\]

for \( i = 1, \cdots, d \) where \( w_k(i) \) represents the sum of the sensor noise and the atmospheric turbulence noise both assumed to be white, zero-mean complex Gaussian and independent of sky values. Collecting the data for all of the antenna at look position \( i \) into a vector, the imaging equation can be expressed as:

\[
y_i = A_i x + w_i \tag{36}
\]

where \( A_i \in \mathbb{C}^{p \times n} \) is the corresponding observation operator that consists of the complex exponentials in (35). As can be seen, in this modality we have multiple independent measurements and the formulation in Section II needs to be adapted accordingly. The error covariance for the case of \( m \) independent measurements in (36) is given by

\[
\Sigma_e = (\Sigma_{x}^{-1} + \sum_{i=1}^{d} A_i^H \Sigma_{w,i}^{-1} A_i)^{-1} \tag{37}
\]

Besides, a typical astronomical image consists of sparse point sources on a smooth background. The background can be estimated and subtracted from the measurements, leaving an image which is typically sparse, i.e., only a small percentage of pixel values are nonzero. Injecting this form of \textit{a priori} knowledge is known as sparsity regularization [30], [31]. Sparsity regularization is accomplished by adding a function, e.g., \( c(x) \), that is monotonously increasing with the number of nonzero elements in \( x \) (which can also be thought of as the complexity of \( x \)) to the data fidelity term [24]. Recently, it has been shown that using a function of the form \( c(x) = \lambda \|x\|_1 \) performs very well even in the presence of noise [32]. However, in the MAP estimation framework, having \( \ell_1 \)-norm as the regularization functional translates into Laplacian distribution for the unknown image coefficients. Using \( c(x) = \lambda \|x\|_1 \) gives \( p(x) \propto e^{-\lambda \|x\|_1} \) where \( p(x) \) represents the probability distribution function of the random vector \( x \). But this violates the statistical assumptions in Section II and makes the computation of error covariance intractable. Nonetheless, one can derive the LMMSE estimator of \( x \) and use (8) together with any of the optimality criteria in Section III. The LMMSE estimator is

\[
\hat{x}_{LMMSE} = x_0 + \left( \frac{1}{\sigma^2_w} \sum_{i=1}^{d} A_i^H A_i + \gamma^2 I \right)^{-1} \left( \frac{1}{\sigma^2_w} \sum_{i=1}^{d} A_i^H (y_i - A_i x_0) \right)
\]

where \( \gamma = \frac{\lambda}{\sqrt{2}}. \) The corresponding error covariance is \( \Sigma_e = \left( \frac{1}{\sigma^2_w} \sum_{i=1}^{d} A_i^H A_i + \gamma^2 I \right)^{-1}.\)
VI. Results

Simulation results are presented for the application of ground-based ionospheric radio tomography described in Section V-A using the optimality criteria introduced in Section III. In all the experiments, the regularization matrix was chosen to be the Fried discrete Laplacian corresponding to the two dimensional curvature operator.

In the first experiment, the optimality criterion is chosen to be SSE. The set of candidate locations, $S$ in (5), was $31$ equispaced points within normalized longitude range of $-2.5$ to $2.5$. The desired number of sensors, $q$ in (5), is set to $2$. The discretization grid has a span of $-1$ to $1$. Figure 2(a) shows the original profile that is a two-dimensional section of the electron density in ionosphere. The observations corresponding to a typical limited-angle geometry similar to Figure 1 were computed and IID Gaussian noise was added to achieve an SNR (defined as $10 \log_{10} \frac{\text{Var}(x)}{\text{Var}(w)}$) of $20$dB. The regularization parameter $\lambda$ was chosen to yield the best possible Tikhonov reconstruction for a reasonable guess of optimal locations taken to be $\{-1, 1\}$. The reconstruction result is shown in Figure 2(b). The CSBS algorithm returned $\{-0.5, 0.17\}$ as the configuration for the pair of sensors. The SSE for the CSBS choice is $1.26$ whereas
for the alternative configuration it is 1.70. The normalized error defined as $\frac{\|\hat{x} - x_{true}\|_2}{\|x_{true}\|_2}$ for CSBS-chosen and alternative configurations are 9.3% and 15.1% respectively. Besides, as can be seen in Figure 2, the reconstruction result with the CSBS-chosen configuration (Figure 2(c)) is superior to the alternative configuration of $\{-1, 1\}$ (Figure 2(b)). Another way of measuring the performance of the CSBS algorithm is to compare the evolution of its SSE to another ad hoc algorithm as the number of sensors decrease. Eliminating the sensors from right side of the array to its left side is a viable candidate for the ad hoc method and is referred to as the right-to-left algorithm. Figure 3(a) compares the SSE (average of 50 Monte-Carlo runs each with a different noise realization) as a function of number of sensors for the CSBS and right-to-left algorithms. As can be seen, CSBS performs consistently better (in SSE sense) for any desired number of sensors. Figure 3(b) shows the normalized error for the two algorithms (average of 50 Monte-Carlo runs). Overall, it can be deduced that as the number of receivers decrease the superiority
of the CSBS algorithm becomes more noticeable.

In the second experiment, the WSSE optimality criterion is used. The desired number of sensors, SNR, and regularization parameter were set as in the first experiment. Figure 4(a) shows the predesigned weight matrix ($W_i$'s in Eq. (14)). According to the weight profile, the desirable observation geometry should result in smaller error in the region of interest compared to the rest of the profile. The set of candidate locations was 21 points spread evenly in the range $-2.5$ to $2.5$. The CSBS algorithm returned $\{0.25, 2.5\}$ as the sensor configuration. Figure 4(c) shows the error covariance for the CSBS-chosen configuration alongside an alternative configuration of $\{-1, 1\}$ shown in panel (b). As can be seen, the error covariance of the CSBS-chosen locations is smaller in the region of interest. Quantitatively, $\text{Cost}^2(A)$ for the CSBS choice is 0.29 whereas for the alternative configuration it is 0.45. Furthermore, Figure 5 shows the reconstructed profiles for CSBS-chosen configuration alongside the alternative configuration. Comparing Figures 5(b) and 5(c), it can be seen that reconstruction of the region of interest by CSBS-chosen geometry is closer to the original profile shown in Figure 5(a). Particularly, the high to low density transition in the middle
Fig. 5. (a) The original ionospheric profile, the black box shows the boundaries of the region of interest (b) The reconstructed profile with the two sensors at \{-1, 1\} (a reasonable alternative) (c) The reconstructed profile with the two sensors at \{0.25, 2.5\} (CSBS choice for WSSE).

of the region of interest is much better expressed in Panel (c) compared to Panel (b). Figure 6 compares the WSSE (average of 50 Monte-Carlo runs) as a function of number of sensors for the CSBS and right-to-left algorithms. As can be seen, CSBS performs consistently better (in WSSE sense) for any desired number of sensors.

In the third experiment, the optimality criterion is chosen to be USE. The desired number of sensors, SNR and regularization parameter were set as in the first experiment. The set of candidate locations was the same as the second experiment. The CSBS algorithm returned \{0.40, 2.5\} as the configuration for the pair of sensors. The error covariance matrix for the CSBS choice is shown in Figure 7. As can be seen, the error covariance matrix looks smoother than the alternative shown in Figure 4(c). Quantitatively, \(\text{Cost}^{3}(A)\) for the CSBS choice is \(6.3 \times 10^{-5}\) whereas for the alternative configuration it is \(11.8 \times 10^{-5}\).

In the fourth experiment, the detection performance criterion is studied using the J-divergence cost function defined in (19). The SNR and regularization parameter were set as in the first experiment. The set of candidate locations was 17 points spread evenly between \(-2.5\) to 2.5. But this time, in order to get meaningful results, the number of desired receivers was chosen to be three \((q = 3\) in (5)). The feature
(vector $b$) is taken to be a plasma depletion artificially impinged on the background ionosphere as shown in Figure 8(a). It was assumed that the feature can only be present in the right half of the image. This knowledge might be available from physics-based models or other prior measurements. The covariance matrix of the feature, $\Sigma_b$, was numerically estimated assuming that the feature location follows a uniform distribution.

The CSBS algorithm returned $\{-2.5, 0, 2.5\}$ as the configuration for the three sensors. Figures 8(b) and (c) shows the reconstructed profiles for the alternative configuration of $\{-1, 0, 1\}$ and the CSBS choice respectively. As can be seen, although the overall reconstruction is worse in panel (c), but the presence of the feature is more pronounced compared to panel (b). Quantitatively, the J-divergence for CSBS choice is 6369 whereas for the alternative configuration it is 4537. Finally, Figure 9 compares the performance of the CSBS and the right-to-left algorithms and clearly demonstrates the effectiveness of the CSBS algorithm for the detection criterion.

VII. SUMMARY AND CONCLUSIONS

We have introduced a modified version of a previously proposed backward greedy algorithm for the problem of determining the optimal sensor configuration in remote imaging. Several statistical criteria were introduced and the applicability of the proposed technique (CSBS) was verified through numerical simulations. The developed framework allows for inclusion of smoothness and other convex constraints on the image space. It should be noted that the results of this paper will in general be valid in the context of LMMSE estimation. Computational complexity of the proposed algorithm is discussed and a fast implementation is described. Furthermore, upper bounds on the sum of squared error of the proposed algorithm are derived. Connections of CSBS to the deterministic backward greedy algorithm are described and a conjecture on its optimality is given. Considering the promising simulation results, the proposed framework can contribute to optimal design of real world remote image formation applications.

APPENDIX I
PROOF OF LEMMA 2

Proof: Using the so called Woodbury matrix inversion formula [21] we can write:

$$(\Sigma_x^{-1} + \tilde{A}^H \tilde{A})^{-1} = \Sigma_x - \Sigma_x \tilde{A}^H (I + \tilde{A} \Sigma_x \tilde{A}^H)^{-1} \tilde{A} \Sigma_x$$
Fig. 6. Weighted sum of squared error (in log scale) as a function of number of sensors for CSBS (with WSSE criterion) and right-to-left algorithms.

Fig. 7. The error covariance for the reconstructed profile with the two sensors at \{0.40, 2.5\} (CSBS choice for USE).

Multiplying both sides by \( \tilde{A}_i^H \tilde{A}_i \), taking the trace and rearranging the terms inside the trace on the right hand side gives:

\[
\text{tr}\{\tilde{A}_i^H \tilde{A}_i (\tilde{A}_i^H \tilde{A}_i + \Sigma_x^{-1})^{-1}\} = \\
\text{tr}\{\tilde{A}_i^H \tilde{A}_i \Sigma_x\} - \text{tr}\{\tilde{A}_i \Sigma_x \tilde{A}_i^H (I + \tilde{A}_i \Sigma_x \tilde{A}_i^H)^{-1} \tilde{A}_i \Sigma_x \tilde{A}_i^H\}
\] (38)

The second term on the right hand side can be written as \( \text{tr}\{(I + \Psi^H \Phi)^{-1} \Phi^H \Phi\} \) where \( \Psi = \sqrt{\Sigma_x} \tilde{A}_i^H \) and \( \Phi = \tilde{A}_i \Sigma_x \tilde{A}_i^H \). Now, since \( (I + \Psi^H \Psi)^{-1} \) is positive definite and \( \Phi^H \Phi \) is nonnegative definite, the matrix inside the trace is nonnegative definite. But, trace of a nonnegative definite matrix is nonnegative. Hence, the second term on the right hand side of (38) is nonpositive which results in the following inequality:

\[
\text{tr}\{\tilde{A}_i^H \tilde{A}_i (\tilde{A}_i^H \tilde{A}_i + \Sigma_x^{-1})^{-1}\} \leq \text{tr}\{\tilde{A}_i^H \tilde{A}_i \Sigma_x\}
\] (39)
Fig. 8. (a) The ionospheric profile for the fourth experiment (b) The reconstructed profile with with the three sensors at \{-1, 0, 1\} (a reasonable alternative) (c) The reconstructed profile with the sensors at \{-2.5, 0, 2.5\} (CSBS choice for detection criterion).

If the right hand side of (39) is less than or equal to 1 then (C.1) will be satisfied which, after plugging in \(\Sigma_x = (\gamma^2 L^T L)^{-1}\), is equivalent to:

\[
\gamma^2 \geq \max_{1 \leq i \leq p} \left( \text{tr}\{\tilde{A}_i^H \tilde{A}_i (L^T L)^{-1}\} \right)
\]

Assuming IID Gaussian noise, we have \(\tilde{A}_i^H \tilde{A}_i = \frac{1}{\sigma_w^2} A_i^H A_i\) and \(\lambda = (\gamma \sigma_w)^2\). In this case, (40) is equivalent to

\[
\lambda \geq \max_{1 \leq i \leq p} \left( \text{tr}\{A_i^H A_i (L^T L)^{-1}\} \right)
\]

\[
\Box
\]

**APPENDIX II**

**LEMMA 3**

Lemma 3: If (C.1) is satisfied, i.e., for all \(1 \leq i \leq p\) we have

\[
\text{tr}\{\tilde{A}_i^H \tilde{A}_i (\tilde{A}_i^H + K)^{-1}\} < 1
\]
then

\[ p - n + \text{tr}\{\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \mathbf{K}\}^{-1}\mathbf{K}\} > 0 \quad (43) \]

\textbf{Proof: } Summing up both sides of (42) from \( i = 1 \) to \( i = p \) gives:

\[ \sum_{i=1}^{p} \text{tr}\{\tilde{\mathbf{A}}_i^H \tilde{\mathbf{A}}_i(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \mathbf{K})^{-1}\} < \sum_{i=1}^{p} 1 \]

So,

\[ \text{tr}\{(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \mathbf{K})^{-1} \sum_{i=1}^{p} \tilde{\mathbf{A}}_i^H \tilde{\mathbf{A}}_i\} < p \quad (44) \]

Having \( \sum_{i=1}^{p} \tilde{\mathbf{A}}_i^H \tilde{\mathbf{A}}_i = \tilde{\mathbf{A}}^H \tilde{\mathbf{A}} \) and noting that

\[ \text{tr}\{(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \mathbf{K})^{-1}(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}})\} = \text{tr}\{(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \mathbf{K})^{-1}(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \mathbf{K} - \mathbf{K})\} = \text{tr}\{\mathbf{I}_{n \times n}\} - \text{tr}\{\mathbf{K}(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \mathbf{K})^{-1}\} = n - \text{tr}\{\mathbf{K}(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \mathbf{K})^{-1}\} \]

we can rewrite (44) as

\[ n - \text{tr}\{\mathbf{K}(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \mathbf{K})^{-1}\} < p \]

which after rearranging gives (43) and completes the proof. \( \square \)

\textbf{ACKNOWLEDGMENT}

The authors would like to thank Dr. Nitin Aggarwal, Dr. Mathews Jacob, and Prof. Minh N. Do for their helpful remarks.
REFERENCES


