Optimal Sensor Array Configuration in Remote Image Formation

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

***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. Furthermore, upper bounds on the sum of the squared error of the proposed algorithm are derived. Connections of the method to the deterministic backward greedy algorithm for the subset selection problem are presented, and two application examples are described. Five 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 configuration designed by the proposed algorithm performs better than wisely chosen alternatives.**

***Index Terms*—Image formation, interferometric imaging, remote imaging, remote sensing, sensor configuration, sequential backward selection, subset selection, tomographic imaging.**

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]

$$Y(r,s) = \iint_{\Omega} a(r,s;r',s')X(r',s')dr'ds' \tag{1}$$

where a 2-D 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:

$$X(r,s) = \sum_{j=1}^{n} x_j \phi_j(r,s). \tag{2}$$

For instance, $\{\phi_j(r,s)\}_{j=1}^n$ are often chosen to be the set of unit height boxes corresponding to a 2-D 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 $\mathbf{y}$ of length m, and the unknown image coefficients into a vector $\mathbf{x}$ of length n, results in the following observation model in form of a matrix equation:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \tag{3}$$

where $\mathbf{w}$ is the additive measurement noise that is assumed to be independent of $\mathbf{x}$, and $\mathbf{A} \in \mathbb{C}^{m \times n}$ where $m \geq n$ is the linear operator relating the unknown field to the observations, comprised of inner products of the basis functions with the corresponding observation kernel

$$(\mathbf{A})_{ij} = \iint_{\Omega} a_i(r',s')\phi_j(r',s')dr'ds' \quad 1 \leq i \leq m, \quad 1 \leq j \leq n \tag{4}$$

where $a_i(r',s') = a(r_i,s_i;r',s')$ denotes the kernel function corresponding to the ith observation. In the case of image reconstruction from projection data, a ray is defined as a line through the image plane in the absence of diffraction and scattering effects. Let y_i be the line integral measured by the ith ray. Then $(\mathbf{A})_{ij}$ is the weighting factor that represents the contribution of the jth cell to the ith 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] and [3], this relationship is studied by considering the distribution of the singular values of $\mathbf{A}$. 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

Manuscript received August 25, 2005; revised May 24, 2007. This work was supported in part by the National Science Foundation under Grant ATM 01-35073 to the University of Illinois. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jacques Blanc-Talon.

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@uiuc.edu; farzadk@uiuc.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2007.914225

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 $\mathbf{A}^0 \in \mathbb{C}^{m\times n}$ where $m = p \cdot d$, is typically very large. Selecting a subset of candidate locations is equivalent to choosing a subset of rows of $\mathbf{A}^0$. 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 $\mathbf{x}$ and $\mathbf{w}$ 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 $\mathcal{S}$ and the observation matrix corresponding to $\mathcal{U} \subseteq \mathcal{S}$ by $\mathbf{A}^{(\mathcal{U})}$, the following optimization problem is reached [6]:

$$\mathcal{S}^* = \arg \min_{\mathcal{U}\subseteq \mathcal{S},\ |\mathcal{U}|=q} \mathrm{Cost}(\mathbf{A}^{(\mathcal{U})}) \tag{5}$$

where $|.|$ denotes the cardinality, q is the desired number of sensors, and $\mathrm{Cost}(\mathbf{A}^{(\mathcal{U})})$ is the cost of choosing subset $\mathcal{U}$ as the locations for the q sensors. The best subset of rows of $\mathbf{A}^0$ can be found by exhaustive search over all $\binom{p}{q}$ possible combinations of rows. However, 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 [7]. 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 [8]. In general, in all subset selection problems of this type, there is an inherent complexity/performance trade-off [9].

In order to make the search computationally feasible, a modified version of the sequential backward selection (SBS) algorithm [10] 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 [7], [10]. Note that our problem is different from the one that SBS aims to solve. Each candidate sensor location contributes several rows to the $\mathbf{A}^0$ matrix. Hence, instead of considering each row of $\mathbf{A}^0$ 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 [11] 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 [12]. Assuming $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma_w})$ and $\mathbf{x} \sim \mathcal{CN}(\mathbf{x}_0, \mathbf{\Sigma_x})$, where $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{\Sigma})$ represents the complex normal distribution with mean μ and covariance $\mathbf{\Sigma}$, the MAP estimate is

$$\begin{aligned}\hat{\mathbf{x}}_{\mathrm{MAP}} &= \arg \min_{\mathbf{x}\in\mathbb{C}^n} \left[-\log p(\mathbf{y}|\mathbf{x}) - \log p(\mathbf{x})\right] \\ &= \arg \min_{\mathbf{x}\in\mathbb{C}^n} \left[||\mathbf{y}-\mathbf{A}\mathbf{x}||^2_{\mathbf{\Sigma_w}^{-1}} + ||\mathbf{x}-\mathbf{x}_0||^2_{\mathbf{\Sigma_x}^{-1}}\right] \\ &= \mathbf{x}_0 + \left(\mathbf{A}^H\mathbf{\Sigma_w^{-1}}\mathbf{A} + \mathbf{\Sigma_x^{-1}}\right)^{-1} \\ &\quad \times \mathbf{A}^H\mathbf{\Sigma_w^{-1}}(\mathbf{y}-\mathbf{A}\mathbf{x}_0)\end{aligned} \tag{6}$$

where $\mathbf{A}^H$ denotes the Hermitian of matrix $\mathbf{A}$. Assuming independent identically distributed (IID) Gaussian noise and taking $\mathbf{\Sigma_x} = (1/\gamma^2)(\mathbf{L}^T\mathbf{L})^{-1}$, we arrive at the well-known Tikhonov regularization functional

$$\hat{\mathbf{x}}_{\mathrm{Tik}} = \arg \min_{\mathbf{x}\in\mathbb{C}^n} \left[\frac{1}{\sigma_w^2}||\mathbf{y}-\mathbf{A}\mathbf{x}||_2^2 + \gamma^2||\mathbf{L}(\mathbf{x}-\mathbf{x}_0)||_2^2\right] \tag{7}$$

$$\begin{aligned}&= \arg \min_{\mathbf{x}\in\mathbb{C}^n} \left[||\mathbf{y}-\mathbf{A}\mathbf{x}||_2^2 + \lambda||\mathbf{L}(\mathbf{x}-\mathbf{x}_0)||_2^2\right] \\ &= \mathbf{x}_0 + \left(\frac{1}{\sigma_w^2}\mathbf{A}^H\mathbf{A} + \gamma^2\mathbf{L}^T\mathbf{L}\right)^{-1} \frac{1}{\sigma_w^2}\mathbf{A}^H(\mathbf{y}-\mathbf{A}\mathbf{x}_0)\end{aligned} \tag{8}$$

where $\mathbf{L}$ is the positive definite regularization matrix and $\lambda = (\gamma\sigma_w)^2$ where σ_w^2 is the variance of the noise samples. A special case is when $\mathbf{L} = \mathbf{I}$, which results in λ 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 [13], [14].

The statistical framework allows for a closed-form measure of estimation uncertainty through the error covariance matrix $\mathbf{\Sigma_e} = E[\mathbf{e}\mathbf{e}^H]$, where $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}_{\mathrm{Tik}}$. For the MAP estimate, the error covariance is given by [15]

$$\mathbf{\Sigma_e} = \left(\mathbf{A}^H\mathbf{\Sigma_w^{-1}}\mathbf{A} + \mathbf{\Sigma_x^{-1}}\right)^{-1} \tag{9}$$

where the expected squared error for the ith element of $\mathbf{x}$ is the (i,i)th element of $\mathbf{\Sigma_e}$, denoted by $(\mathbf{\Sigma_e})_{ii}$. Consequently, the error covariance for Tikhonov regularized reconstruction is $\mathbf{\Sigma_e} = ((1/\sigma_w^2)\mathbf{A}^H\mathbf{A} + \gamma^2\mathbf{L^T}\mathbf{L})^{-1}$. 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 (MMSE) estimator for $\mathbf{x}$, which minimizes

$E\left[\|\mathbf{x}-\hat{\mathbf{x}}\|_2^2\right]$. Therefore, the results of this paper will, in general, be valid in the context of linear MMSE estimation.

As we will see in Section III, selection of optimality criteria requires considering costs that are functions of $\mathbf{\Sigma_e}$. As shown in (9), this, in turn, requires estimating $\mathbf{\Sigma_x}$. Many deterministic constraints on images that arise naturally are convex constraints [16]. Examples include nonnegativity, smoothness, support, reference, or energy constraints. In what follows, we show how to incorporate four different forms of convex constraints into the statistical framework established in this section by choosing the appropriate form of $\mathbf{\Sigma_x}$.

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:

$$\mathbf{L}=\begin{bmatrix}1 & -1 & 0 & \cdots & \\ 0 & 1 & -1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & \cdots & 1 & -1 \\ 0 & \cdots & \cdots & 0 & 1\end{bmatrix} \tag{10}$$

where the last row was added to ensure invertibility. Using this regularization matrix is equivalent to assuming $\mathbf{x}$ to be a Brownian motion [12]. It is possible to use higher order derivatives or to adapt these matrices to work on vertical neighbors [17]. An alternative class of regularization matrices are discretizations of the 2-D Laplacian operator, namely Fried and Hudgin discrete Laplacians [18], [19]. 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

$$\hat{\mathbf{x}}=\arg\min_{\mathbf{x}\in\mathbb{R}^n}\left[\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2+\lambda_h\|\mathbf{L_h}\mathbf{x}\|_2^2+\lambda_v\|\mathbf{L_v}\mathbf{x}\|_2^2\right].$$

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 $\mathbf{x}_0=\mathbf{0}$). 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 field $X(r,s)$ is expected to vanish outside a region $\mathcal{H}$. This support constraint can be formulated in the above framework by adding a term $\lambda_s h(\mathbf{x})$ to the functional in (7) penalizing the energy outside of $\mathcal{H}$. The parameter $\lambda_s\in\mathbb{R}^+$ controls the amount of penalization. In order to obtain a closed-form solution as in (8), it is desirable to design a matrix $\mathbf{B}$ such that

$$h(\mathbf{x})=\int_{(r,s)\notin\mathcal{H}}|X(r,s)|^2drds\approx\|\mathbf{B}\mathbf{x}\|_2^2. \tag{11}$$

One such construction is a diagonal matrix with $(\mathbf{B})_{ii}=0$ if $\mathbf{x}_i\in\mathcal{H}$ and $(\mathbf{B})_{ii}=1$ otherwise. 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 $\lambda\mathbf{L}^T\mathbf{L}+\lambda_s\mathbf{B}^T\mathbf{B}$ (assuming $\mathbf{x}_0=\mathbf{0}$).

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

$$\mathcal{C}=\{\mathbf{x}'\in\mathbb{C}^n:\|\mathbf{x}'-\mathbf{x}_0\|_2<\rho\} \tag{12}$$

for some reference image $\mathbf{x}_0$ and radius $\rho>0$. This is a convex constraint and is referred to as the reference (prototype) image constraint [20]. If no knowledge of statistics of $\mathbf{x}$ exists, using the maximum likelihood framework together with the reference constraint leads to a constrained maximum likelihood (ML) estimation problem [21]. The constrained ML estimation is equivalent to an unconstrained MAP estimation which, assuming Gaussian noise, reduces to Tikhonov regularization [22]

$$\begin{aligned}\hat{\mathbf{x}}_{\mathrm{CML}}&=\arg\min_{x\in\mathcal{C}}\left[-\log p(\mathbf{y}|\mathbf{x})\right]\\&=\arg\min_{x\in\mathbb{C}^n}\left[-\log p(\mathbf{y}|\mathbf{x})+\nu(\rho,\mathbf{y})\|\mathbf{x}-\mathbf{x}_0\|_2^2\right]\\&=\arg\min_{x\in\mathbb{C}^n}\left[\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2+\nu(\rho,\mathbf{y})\|\mathbf{x}-\mathbf{x}_0\|_2^2\right]\end{aligned}$$

where $\nu(\rho,\mathbf{y})\in\mathbb{R}^+$ is the associated Lagrange multiplier (see [23, Theorem 1, p. 217]). As can be seen, the constrained ML estimation has the same form as the Tikhonov functional with $\mathbf{L}=\mathbf{I}_{n\times n}$ and $\lambda=\nu(\rho,\mathbf{y})$. Hence, (9) applies here, as well.

4) *Energy Constraint*: Energy constraint is a special case of the reference constraint with $\mathbf{x}_0=\mathbf{0}$ 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 Section II, five compelling choices for the $\mathrm{Cost}(\mathbf{A})$ function in (5) are introduced.

1) *Sum of Squared Errors*: The sum of squared errors (SSE) or expected value of the squared estimation error, i.e., $E[\mathbf{e}^H\mathbf{e}]$, is given by

$$\mathrm{Cost}^1(\mathbf{A})=\mathrm{tr}(\mathbf{\Sigma_e})=\sum_{i=1}^{n}(\mathbf{\Sigma_e})_{ii}. \tag{13}$$

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 $\mathrm{tr}\{(\mathbf{A}^H\mathbf{A})^{-1}\}$ where the inverse operator is usually replaced by the pseudoinverse to avoid instability due to ill posedness of the problem [7].

2) *Weighted Sum of Squared Errors*: 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 reasonable approach to designing the cost function for this setting is to weight $\mathrm{Cost}^1(\mathbf{A})$ as follows:

$$\mathrm{Cost}^2(\mathbf{A})=\sum_{i=1}^{n}(\mathbf{\Sigma_e})_{ii}W_i \tag{14}$$

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

3) *Uniformity of Squared Errors*: If the location of the feature of interest is not known beforehand, it may be 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

$$\text{Cost}^3(\mathbf{A}) = \text{STD}\{(\mathbf{\Sigma_e})_{ii}\}_{i=1}^n \tag{15}$$

where

$\text{STD}\{c_i\}_{i=1}^n = \sqrt{(1/n)\sum_{i=1}^n \left(c_i - (1/n)\sum_{j=1}^n c_j\right)^2}$.

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 the 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 $\mathbf{b}$ is present in the background $\mathbf{x}$. Both the feature and the background are unknown and are modeled as uncorrelated Gaussian random vectors distributed as $\mathcal{CN}(\mathbf{b}_0, \mathbf{\Sigma_b})$ and $\mathcal{CN}(\mathbf{x}_0, \mathbf{\Sigma_x})$, respectively. Also, the noise vector $\mathbf{w}$ is a Gaussian distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{\Sigma_w})$ and is independent of both $\mathbf{b}$ and $\mathbf{x}$. Using the imaging equation (3), the detection problem can be modeled as the following binary hypothesis testing problem

$$H_0 : \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$$
$$H_1 : \mathbf{y} = \mathbf{A}(\mathbf{x} + \mathbf{b}) + \mathbf{w}.$$

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

$$\mathbf{\Sigma}_{\mathbf{y}|H_0} = \mathbf{A}\mathbf{\Sigma_x}\mathbf{A}^H + \mathbf{\Sigma_w} \tag{16}$$

and under H_1 is

$$\mathbf{\Sigma}_{\mathbf{y}|H_1} = \mathbf{A}(\mathbf{\Sigma_x} + \mathbf{\Sigma_b})\mathbf{A}^H + \mathbf{\Sigma_w}. \tag{17}$$

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 that is the distance between the conditional densities [4]. Such distance measures include the Bhattacharyya distance, the I-divergence, the J-divergence, and the Chernoff distance [24]. We choose the J-divergence, which is defined as

$$\begin{aligned} Jdiv(\mathbf{A}) = {} & \frac{1}{2}\text{tr}\left(\left[\mathbf{\Sigma}_{\mathbf{y}|H_1} - \mathbf{\Sigma}_{\mathbf{y}|H_0}\right]\left[\mathbf{\Sigma}_{\mathbf{y}|H_0}{}^{-1} - \mathbf{\Sigma}_{\mathbf{y}|H_1}{}^{-1}\right]\right) \\ & + \frac{1}{2}\text{tr}\left(\left[\mathbf{\Sigma}_{\mathbf{y}|H_1}{}^{-1} + \mathbf{\Sigma}_{\mathbf{y}|H_0}{}^{-1}\right]\right. \\ & \left.\times(\mathbf{m}_1 - \mathbf{m}_0)(\mathbf{m}_1 - \mathbf{m}_0)^H\right) \end{aligned} \tag{18}$$

where $\mathbf{m}_i = E[\mathbf{y}|H_i]$ for $i = 0, 1$. For the above setting, $\mathbf{m}_0 = \mathbf{A}\mathbf{x}_0$ and $\mathbf{m}_1 = \mathbf{A}(\mathbf{x}_0 + \mathbf{b}_0)$. Therefore, the cost function corresponding to the detection performance criterion is

$$\text{Cost}^4(\mathbf{A}) = -Jdiv(\mathbf{A}). \tag{19}$$

5) *Mutual Information*: We may wish to design the imaging system such that it collects only those observations that are most informative of the unknown field [4]. In the sensor configuration problem at hand, one can find the location of those sensors that collect the most informative observations. This idea is facilitated by utilizing the information theoretic quantity of mutual information $I(\mathbf{x};\mathbf{y})$ between the unknown field and observations, defined as [25]

$$I(\mathbf{x};\mathbf{y}) = h(\mathbf{x}) + h(\mathbf{y}) - h(\mathbf{x},\mathbf{y}) \tag{20}$$

where $h(.)$ denotes the differential entropy and $h(\mathbf{x},\mathbf{y})$ is the joint entropy of random vectors $\mathbf{x}$ and $\mathbf{y}$. Mutual information measures general statistical dependence between variables. It has been applied in image registration particularly for medical imaging applications [26] as well as feature selection in machine learning [27]. This criterion does not commit to a specific error metric as opposed to the aforementioned criteria and is, in fact, related to a class of bounds on the reconstruction error.

The goal is to find the set of sensor locations that maximize the mutual information between the unknown image $\mathbf{x}$ and the set of observations $\mathbf{y}$. The following lemma provides a computable form of $I(\mathbf{x};\mathbf{y})$ for our imaging problem.

Lemma 1: Under the statistical framework of Section II, $I(\mathbf{x};\mathbf{y})$ can be simplified as

$$I(\mathbf{x};\mathbf{y}) = \frac{1}{2}\ln(|\pi e\mathbf{\Sigma_x}|) - \frac{1}{2}\ln(|\pi e\mathbf{\Sigma_e}|) \tag{21}$$

where $|.|$ denotes the determinant operator when the argument inside is a matrix.

Proof: Proof is provided in Appendix A. ■

Note that the first term on the right-hand side in (21) is independent of observations, and thus, the set of most informative sensors maximizes the second term. According to (21), in order to maximize the mutual information $I(\mathbf{x};\mathbf{y})$, we need to minimize $(1/2)\ln(|\pi e\mathbf{\Sigma_e}|)$, or equivalently $|\mathbf{\Sigma_e}|$. Hence

$$\text{Cost}^5(\mathbf{A}) = |\mathbf{\Sigma_e}|. \tag{22}$$

Note the connection between the SSE cost function in (13) and that of mutual information in (22). In the former, we minimize the sum of the eigenvalues of the error covariance matrix $\mathbf{\Sigma_e}$, whereas, in the latter, the product of eigenvalues of $\mathbf{\Sigma_e}$ is minimized.

IV. CSBS Algorithm

Let us denote by $\mathbf{\Gamma}$ the set of available sensor locations. Initially, $\mathbf{\Gamma}$ contains all of the p candidate sensor locations, i.e., $\mathbf{A}^{(\mathbf{\Gamma})} = \mathbf{A}^0$. The CSBS algorithm can be formally written as [28]

$$\mathbf{\Gamma} \leftarrow \mathbf{\Gamma}\backslash\{k^*\} : k^* = \arg\min_{k\in\mathbf{\Gamma}} \text{Cost}(\mathbf{A}^{(\mathbf{\Gamma}\backslash\{k\})}) \tag{23}$$

with the stopping criterion being $|\mathbf{\Gamma}| = q$ where q is the desired number of sensors as mentioned in (5).

In Section IV-A, first we analyze the computational complexity of the general CSBS algorithm. Next, based on the results in [10], we introduce a fast implementation of the CSBS algorithm with SSE criterion. In Section 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 Section 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 (23), there can be, at most, $p-1$ iterations (because $q>0$), and at the ith iteration, we need to compute the cost function for 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 (9). Assuming that the constant matrices are computed and stored beforehand, straightforward computation of $\mathbf{A}^H\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{A}$ is of $O(mn^2+nm^2)$ and we need an additional $O(n^3)$ floating point operations for inversion of the sum. Having $m \geq n$, complexity of computing the cost function is $O(nm^2)$. Therefore, the overall complexity is $O(nm^2p^2)$.

Using the Sherman–Morrison matrix inversion formula [29], Reeves *et al.* have developed an efficient implementation of the SBS algorithm [10]. Here, we first restate Reeves' method for SBS and based on that result improve upon the computational complexity of the CSBS algorithm in the case of SSE criterion. If we eliminate the ith row of $\mathbf{A}$, denoted by $\mathbf{a}_i$, the SSE corresponding to the modified matrix is given by [10]

$$\mathrm{Cost}^1(\mathbf{A}) + \frac{\mathbf{a}_i\boldsymbol{\Sigma}_{\mathbf{e}}^2\mathbf{a}_i^H}{1-\mathbf{a}_i\boldsymbol{\Sigma}_{\mathbf{e}}\mathbf{a}_i^H}$$

where $\boldsymbol{\Sigma}_{\mathbf{e}}$ is as given in (9). 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:

$$\boldsymbol{\Gamma} \leftarrow \boldsymbol{\Gamma}\backslash\{k^*\} : k^* = \arg\min_{k\in\boldsymbol{\Gamma}} \sum_{i\in\Pi_k} \frac{\mathbf{a}_i\boldsymbol{\Sigma}_{\mathbf{e}}^2\mathbf{a}_i^H}{1-\mathbf{a}_i\boldsymbol{\Sigma}_{\mathbf{e}}\mathbf{a}_i^H} \tag{24}$$

where Π_k is the set of indices of rows measured by the kth sensor and $\boldsymbol{\Sigma}_{\mathbf{e}} = (\mathbf{A}^{(\boldsymbol{\Gamma})^H}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{A}^{(\boldsymbol{\Gamma})}+\boldsymbol{\Sigma}_{\mathbf{x}}^{-1})^{-1}$.

Note that, for each iteration of the algorithm in (24), we only need to compute and store $\boldsymbol{\Sigma}_{\mathbf{e}}$ once, which has complexity $O(nm^2)$ as discussed above. For each iteration, i.e., elimination of one sensor, the summands in (24) can be computed with $O(n^2p)$ operations. 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, comparing to $O(nm^2p^2)$ above, 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 in the sense that it is not guaranteed to find the combination that minimizes the optimality criterion. Furthermore, it is known that a greedy observation selection algorithm can give the worst possible combination of rows for some applications [30]. Therefore, a legitimate question to ask is whether a certain level of performance can be guaranteed for CSBS. In this section, we provide an upper bound for SSE of the CSBS algorithm (with SSE criterion) that is valid under a certain condition. Throughout this section, the $\mathbf{A}$ matrices involved are the initial $\mathbf{A}^0 \in \mathbb{C}^{p\cdot d\times n}$ matrix and for notational simplicity we drop the superscript. Since $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}$ is positive definite Hermitian, it has a positive definite square root, namely $\sqrt{\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}}$. Considering the CSBS algorithm and defining $\tilde{\mathbf{A}} = \sqrt{\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}}\mathbf{A}$, the SSE in (9) can be written as

$$\boldsymbol{\Sigma}_{\mathbf{e}} = \left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}+\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right)^{-1}. \tag{25}$$

Let us assume that the set of p candidate sensor locations is indexed as $\Gamma = \{1,2,\ldots,p\}$. Each of the sensor locations $1 \leq i \leq p$ contributes a set of rows to $\mathbf{A}$ indexed by Π_i. Denote by $\mathbf{A}_i \in \mathbb{C}^{d\times n}$ the matrix formed by all rows of $\mathbf{A}$ with indices in Π_i. According to (13) and (25), the SSE can be written as

$$\mathrm{Cost}^1(\mathbf{A}) = \mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\} = \mathrm{tr}\left\{\left(\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i+\mathbf{G}_i+\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right)^{-1}\right\}$$

where $\tilde{\mathbf{A}}_i = \sqrt{\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}}\mathbf{A}_i$, $\mathbf{G}_i = \sum_{\substack{1\leq j\leq p\\ j\neq i}} \tilde{\mathbf{A}}_j^H\tilde{\mathbf{A}}_j$, and we used the identity $\tilde{\mathbf{A}}^H\tilde{\mathbf{A}} = \sum_{j=1}^{p}\tilde{\mathbf{A}}_j^H\tilde{\mathbf{A}}_j$. To proceed, we need the following result on performance of the SBS algorithm from [10].

Lemma 2 (Upper Bound for SBS): Assume $\acute{\mathbf{A}} \in \mathbb{C}^{d\times n}$ is the initial matrix of all d candidate rows. Determine the row that when deleted from $\acute{\mathbf{A}}$ has the best value of the SSE criterion $\mathrm{tr}\{(\acute{\mathbf{A}}^H\acute{\mathbf{A}}+\acute{\mathbf{K}})^{-1}\}$ where $\acute{\mathbf{K}}$ is a positive definite Hermitian fixed matrix. Once the worst row has been eliminated, follow the same procedure with the remaining $d-1$ candidate rows. Continue this process until k rows remain. The final SSE is bounded above by

$$\frac{d-n+1+\mathrm{tr}\{(\acute{\mathbf{A}}^H\acute{\mathbf{A}}+\acute{\mathbf{K}})^{-1}\acute{\mathbf{K}}\}}{k-n+1+\mathrm{tr}\{(\acute{\mathbf{A}}^H\acute{\mathbf{A}}+\acute{\mathbf{K}})^{-1}\acute{\mathbf{K}}\}}\mathrm{tr}\{(\acute{\mathbf{A}}^H\acute{\mathbf{A}}+\acute{\mathbf{K}})^{-1}\}.$$

Proof: See [10, Theorem 1] for proof. ∎

By assigning $\acute{\mathbf{A}} = \tilde{\mathbf{A}}_i$ and $\acute{\mathbf{K}} = \mathbf{G}_i+\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}$, Lemma 2 gives an upper bound on the SSE when all but k rows of $\tilde{\mathbf{A}}_i$ are removed. Note that eliminating the ith sensor location means that all the rows of $\tilde{\mathbf{A}}_i$ should be removed. Therefore, substituting $k=0$ and the aforementioned $\acute{\mathbf{A}}$ and $\acute{\mathbf{K}}$ in Lemma 2, we obtain the following bound on the SSE after removing the ith sensor location

$$\frac{d-n+1+\mathrm{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{e}}\left(\mathbf{G}_i+\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right)\right\}}{-n+1+\mathrm{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{e}}\left(\mathbf{G}_i+\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right)\right\}}\mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\} \tag{26}$$

where we used $\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i+\mathbf{G}_i = \sum_{j=1}^{p}\tilde{\mathbf{A}}_j^H\tilde{\mathbf{A}}_j = \tilde{\mathbf{A}}^H\tilde{\mathbf{A}}$. Note that

$$\begin{aligned}&\mathrm{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{e}}\left(\mathbf{G}_i+\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right)\right\}\\ &\quad= \mathrm{tr}\left\{\left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}+\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right)^{-1}\left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}+\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}-\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\right)\right\}\\ &\quad= n-\mathrm{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\right\}.\end{aligned}$$

Therefore, we can write (26) as

$$\frac{d+1-\operatorname{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\right\}}{1-\operatorname{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\right\}}\operatorname{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\} \tag{27}$$

which, for a given i, is an upper bound on the SSE after removal of the ith sensor location. Next, we will use this bound to derive the desired performance bound (independent of i) for CSBS (summarized as Theorem 1 in the following).

It should be noted that the bound in Lemma 2 is valid only if the denominator is positive. In fact, considering all $1 \leq i \leq p$, the expression in (27) is a valid upper bound for a SSE only if

$$\operatorname{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\right\} < 1 \text{ for } 1 \leq i \ \leq p \quad \text{(C.1)}.$$

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

Lemma 3: With $\boldsymbol{\Sigma}_{\mathbf{x}} = (1/\gamma^2)(\mathbf{L}^{\mathbf{T}}\mathbf{L})^{-1}$, a sufficient condition for (C.1) to be satisfied is

$$\gamma^2 \geq \max_{1\leq i\leq p}\left(\operatorname{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i(\mathbf{L}^T\mathbf{L})^{-1}\right\}\right) \tag{28}$$

and assuming IID Gaussian noise, (28) is equivalent to

$$\lambda \geq \max_{1\leq i\leq p}\left(\operatorname{tr}\left\{\mathbf{A}_i^H\mathbf{A}_i(\mathbf{L}^T\mathbf{L})^{-1}\right\}\right). \tag{29}$$

Proof: The proof is provided in Appendix B. ■

From Lemma 3, it follows that if the image reconstruction (i.e., the MAP estimation in Section II) is regularized enough then (C.1) holds. Under (C.1), the main result of this subsection, stated in the following theorem, provides an upper bound for (27).

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

$$\frac{(d+1)p-n+\operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\boldsymbol{\Sigma}_{\mathbf{e}}\right\}}{p-n+\operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\boldsymbol{\Sigma}_{\mathbf{e}}\right\}}\operatorname{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\} \tag{30}$$

provided that (C.1) holds.

Proof: The proof is by contradiction. Assume that the bound on SSE after removing one sensor [stated in (27)] for all sensor locations $i \in \Gamma$ is larger than the bound (30) in Theorem 1. We will show that this will result in a contradiction which implies that there exists at least one sensor location $i^* \in \Gamma$ that violates the assumption, i.e., its SSE is upper-bounded by (30), and since CSBS picks the best choice (in SSE sense) of sensor locations, it will choose the one with index i^*. This means that the CSBS choice will satisfy the bound in Theorem 1 and, hence, will complete the proof.

To proceed, we assume that for all $1 \leq i \leq p$ we have

$$\frac{d+1-\operatorname{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\right\}}{1-\operatorname{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\right\}}\operatorname{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\} > \frac{(d+1)p-n+\operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right\}}{p-n+\operatorname{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\}}\operatorname{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\}. \tag{31}$$

In the inequality above, the common term $\operatorname{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\}$ is positive since it is a SSE. The numerator and denominator of the left-hand side are both positive because of (C.1). Lemma 4 (stated and proved in Appendix C) proves that the denominator of the right-hand side is positive under (C.1) which also implies that the numerator is positive. Therefore, rearranging (31) and eliminating the common term, we arrive at

$$\left(p-n+\operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right\}\right)\left(d+1-\operatorname{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\right\}\right) > \left(1-\operatorname{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\right\}\right)\left((d+1)p-n+\operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right\}\right). \tag{32}$$

Since (32) holds for all $1 \leq i \leq p$, its summation over i should hold, as well

$$\left(p-n+\operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right\}\right)\sum_{i=1}^{p}\left(d+1-\operatorname{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\right\}\right) > \left((d+1)p-n+\operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right\}\right) \times \sum_{i=1}^{p}\left(1-\operatorname{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\right\}\right) \tag{33}$$

but

$$\sum_{i=1}^{p}\left(\operatorname{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\right\}\right) = \operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{e}}\sum_{i=1}^{p}\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\right\} = \operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{e}}\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}\right\} \tag{34}$$

and

$$\begin{aligned}\operatorname{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}})\} &= \operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{e}}\left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}+\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}-\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right)\right\}\\ &= \operatorname{tr}\left\{\left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}+\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right)^{-1}\left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}+\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}-\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right)\right\}\\ &= n-\operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right\}.\end{aligned} \tag{35}$$

Substituting (34) and (35) in (33) gives

$$\left(p-n+\operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right\}\right)\left((d+1)p-n+\operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right\}\right) > \left((d+1)p-n+\operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right\}\right)\left(p-n+\operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right\}\right)$$

which is a contradiction. Hence, there exists at least one sensor location i^* that violates (32). As explained in the beginning of the proof, this establishes the claim. ■

C. Connections to Deterministic Subset Selection

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 [29] and its variations [31], the idea is to start by finding the row of $\mathbf{A}^0$ closest to $\mathbf{y}$ 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 [32] starts with all of the rows of $\mathbf{A}^0$ 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 $\mathbf{y}$ and

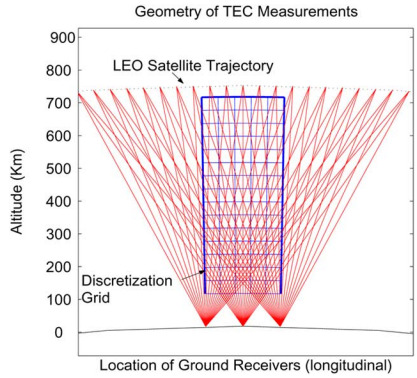

Fig. 1. Typical scenario for ionospheric radio tomography.

instead utilize the error covariance matrix computed in (9). 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

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

where $\|\mathbf{z}\|_E^2 = \sum_{i=1}^n E\left[|z_i|^2\right]$ is the induced norm of the Hilbert space of zero-mean, finite-variance random variables. In [32], it has been shown that the DBG algorithm is guaranteed to select the correct subset of rows of $\mathbf{A}^0$ if the noise level is small enough. However, no such result exists for SBS or CSBS.

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 satellite are tracked by an array of ground sensors [33]. The goal of this technique is to reconstruct horizontal and vertical structures within the 2-D slice of the ionosphere above the ground sensors. In Fig. 1, a scenario with three ground sensors and a low earth orbit 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 $\mathbf{A}$ 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 [34]. 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 [35]. In both applications, the objective is to image the intensity of an unknown 2-D *sky*. In radio astronomy, the antenna are typically fixed and the earth rotation places them at new effective look positions [36]. In radar imaging, the antenna array beam can be steered to form different look positions (see [37, Ch. 7]). Assume that the discretized sky is made up of n pixels, $\mathbf{x} = [x_1 \dots x_n]^T \in \mathbb{C}^n$. Suppose the 2-D antenna array has p elements and we take data at d look positions. Denote by ϕ_j and θ_j the azimuth and zenith angles that the jth pixel form relative to the center of the array. Assuming that θ_j and ϕ_j are small (less than 0.1 radian) so that small-angle approximations to trigonometric functions apply, each antenna sees the signal from the jth pixel with a phase shift of $\exp(\mathbf{j}2\pi[r_k(i)\theta_j + s_k(i)\phi_j])$, where $(r_k(i), s_k(i))$ are the coordinates of the kth antenna at ith look position. Therefore, the data received at kth antenna is

$$y_k(i) = \sum_{j=1}^{n} x_j e^{\mathbf{j}2\pi[r_k(i)\theta_j + s_k(i)\phi_j]} + w_k(i) \qquad (36)$$

for $i = 1, \dots, d$ 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

$$\mathbf{y}_i = \mathbf{A}_i \mathbf{x} + \mathbf{w}_i \qquad (37)$$

where $\mathbf{A}_i \in \mathbb{C}^{p \times n}$ is the corresponding observation operator that consists of the complex exponentials in (36). 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 (37) is given by

$$\mathbf{\Sigma}_\mathbf{e} = \left(\mathbf{\Sigma}_\mathbf{x}^{-1} + \sum_{i=1}^{d} \mathbf{A}_i^H \mathbf{\Sigma}_{\mathbf{w}_i}^{-1} \mathbf{A}_i \right)^{-1}. \qquad (38)$$

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 *a priori* knowledge is known as "sparsity regularization" [38], [39]. Sparsity regularization is accomplished by adding a function, e.g., $c(\mathbf{x})$, that is monotonously increasing with the number of nonzero elements in $\mathbf{x}$ (which can also be thought of as the complexity of $\mathbf{x}$) to the data fidelity term [32]. Recently, it has been shown that using a function of the form $c(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$ performs well even in the presence of noise [40]. However, in the MAP estimation framework, having ℓ_1-norm as the regularization functional translates into Laplacian distribution for the unknown image coefficients. Using $c(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$ gives $p(\mathbf{x}) \propto e^{-\lambda \|\mathbf{x}\|_1}$ where $p(\mathbf{x})$ represents the probability distribution function of the random vector $\mathbf{x}$. However, this violates the statistical assumptions in Section II

and makes the computation of error covariance somewhat intractable. Nonetheless, one can derive the linear MMSE estimator of $\mathbf{x}$ and use (9) together with any of the optimality criteria in Section III. The linear MMSE estimator is

$$\hat{\mathbf{x}}_{\mathrm{LMMSE}} = \mathbf{x}_0 + \left(\frac{1}{\sigma_w^2}\sum_{i=1}^{d}\mathbf{A}_i^H\mathbf{A}_i + \gamma^2\mathbf{I}\right)^{-1}\left(\frac{1}{\sigma_w^2}\sum_{i=1}^{d}\mathbf{A}_i^H(\mathbf{y}_i - \mathbf{A}_i\mathbf{x}_0)\right)$$

where $\gamma = (\lambda/\sqrt{2})$. The corresponding error covariance is $\mathbf{\Sigma_e} = \left((1/\sigma_w^2)\sum_{i=1}^{d}\mathbf{A}_i^H\mathbf{A}_i + \gamma^2\mathbf{I}\right)^{-1}$.

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 2-D curvature operator.

In the first experiment, the optimality criterion was chosen to be SSE. The set of candidate locations, $\mathcal{S}$ 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. Fig. 2(a) shows the original profile that is a 2-D section of the electron density in ionosphere. The observations corresponding to a typical limited-angle geometry similar to Fig. 1 were computed and IID Gaussian noise was added to achieve an SNR (defined as $10\log_{10}(\mathrm{Var}(\mathbf{x})/\mathrm{Var}(\mathbf{w}))$) of 20 dB. The regularization parameter λ 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 Fig. 2(b). The CSBS algorithm returned $\{-0.5, 0.17\}$ as the configuration for the pair of sensors. The SSE for the CSBS design is 1.26, whereas for the alternative configuration it is 1.70. The normalized error defined as $(\|\hat{\mathbf{x}} - \mathbf{x}_{true}\|_2/\|\mathbf{x}_{true}\|_2)$ for CSBS-designed and alternative configurations are 9.3% and 15.1%, respectively. Besides, as can be seen in Fig. 2, the reconstruction result with the CSBS-designed configuration [Fig. 2(c)] is superior to the alternative configuration of $\{-1, 1\}$ [Fig. 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. Fig. 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. Fig. 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 weighted SSE optimality criterion was used. The desired number of sensors, SNR, and regularization parameters were set as in the first experiment. Fig. 4(a) shows the predesigned weight matrix [W_i's in (14)].

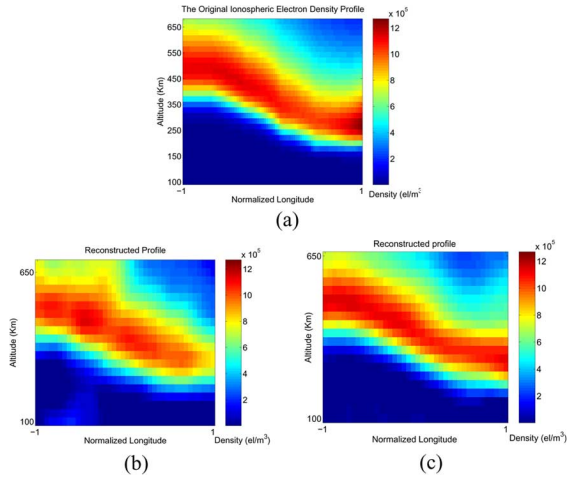

Fig. 2. (a) Original ionospheric profile; (b) reconstructed profile with the two sensors at $\{-1, 1\}$ (a reasonable alternative); (c) reconstructed profile with the two sensors at $\{-0.5, 0.17\}$ (CSBS design for SSE).

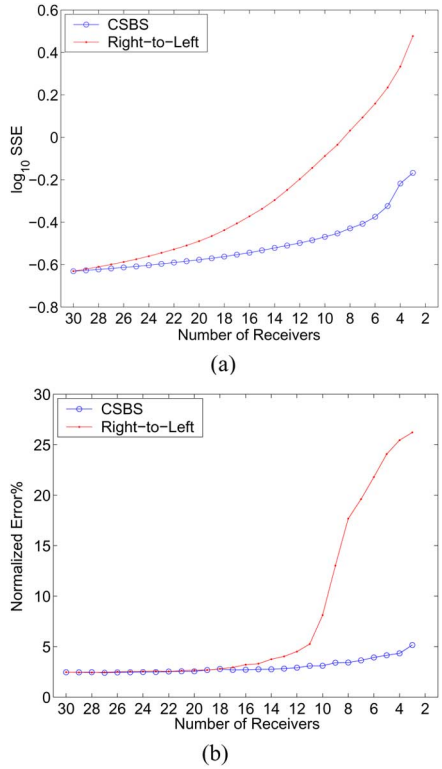

Fig. 3. (a) Sum of squared error (in log scale) as a function of number of sensors for CSBS (with SSE criterion) and right-to-left algorithms; (b) normalized error (average of 50 Monte-Carlo runs each with a different noise realization) as a function of number of sensors for CSBS (with SSE criterion) and right-to-left algorithms.

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. Fig. 4(c) shows the error covariance for the CSBS-designed configuration alongside an alternative configuration of $\{-1, 1\}$ shown in Fig. 4(b). As can be seen, the error covariance of the CSBS-designed locations is smaller in

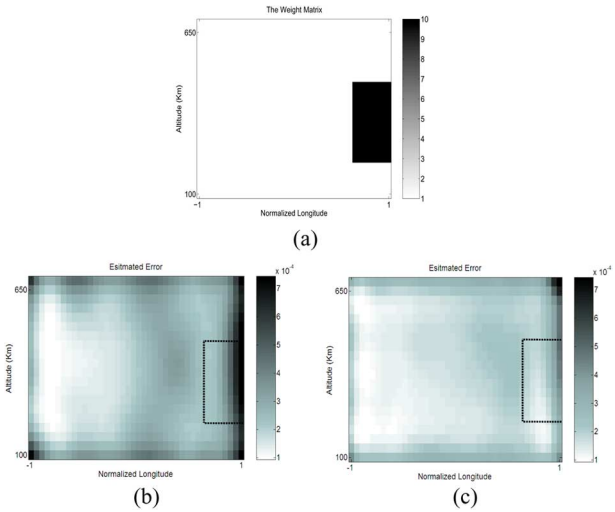

Fig. 4. (a) Predesigned weight matrix [W_i's in (14)]; (b) error covariance for the reconstructed profile with the two sensors at $\{-1, 1\}$ (a reasonable alternative); (c) error covariance for the reconstructed profile with the two sensors at $\{0.25, 2.5\}$ (CSBS design for weighted SSE criterion).

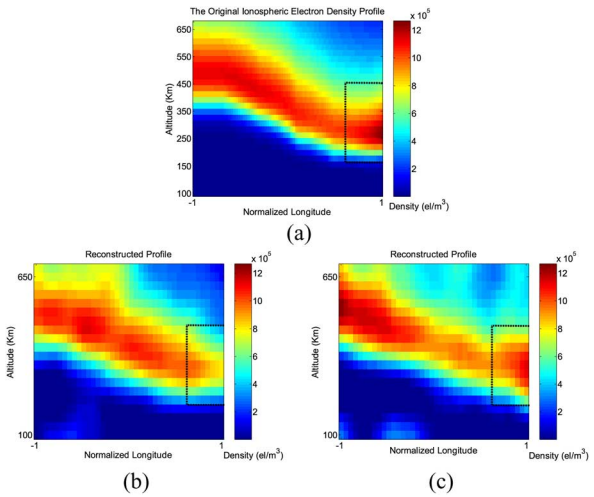

Fig. 5. (a) Original ionospheric profile, the black box shows the boundaries of the region of interest; (b) reconstructed profile with the two sensors at $\{-1, 1\}$ (a reasonable alternative); (c) reconstructed profile with the two sensors at $\{0.25, 2.5\}$ (CSBS design for weighted SSE).

the region of interest. Quantitatively, $\mathrm{Cost}^2(\mathbf{A})$ for the CSBS design is 0.29 whereas for the alternative configuration it is 0.45. Furthermore, Fig. 4 shows the reconstructed profiles for both the CSBS-designed and the alternative configurations. Comparing Fig. 5(b) and (c), it can be seen that reconstruction of the region of interest by the CSBS-designed configuration is closer to the original profile shown in Fig. 5(a). Particularly, the high-to-low-density transition in the middle of the region of interest is much better expressed in Fig. 5(c) compared to Fig. 5(b). Fig. 6 compares the weighted SSE (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 weighted SSE sense) for any desired number of sensors.

In the third experiment, the optimality criterion was chosen to be uniformity of squared errors. 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 design is shown in Fig. 7. As can be seen, the error covariance matrix looks smoother than the alternative shown in Fig. 4(c). Quantitatively, $\mathrm{Cost}^3(\mathbf{A})$ for the CSBS design is 6.3×10^{-5}, whereas for the alternative configuration, it is 11.8×10^{-5}.

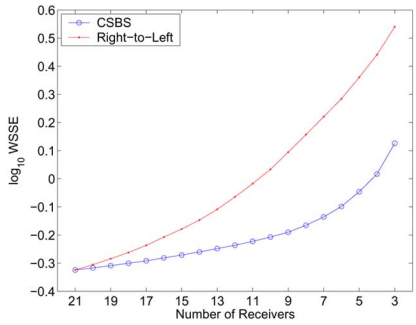

Fig. 6. Weighted sum of squared error (in log scale) as a function of number of sensors for CSBS (with weighted SSE criterion) and right-to-left algorithms.

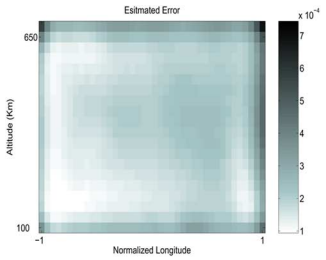

Fig. 7. Error covariance for the reconstructed profile with the two sensors at $\{0.40, 2.5\}$ (CSBS design for uniformity of error criterion).

In the fourth experiment, the detection performance criterion was studied using the J-divergence cost function defined in (18). 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. However, 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 $\mathbf{b}$) is taken to be a plasma depletion artificially impinged on the background ionosphere as shown in Fig. 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, $\mathbf{\Sigma_b}$, 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. Fig. 8(b) and (c) shows the reconstructed profiles for the alternative configuration of $\{-1, 0, 1\}$ and the CSBS-designed configuration, respectively. Comparing the two reconstructions, it can be seen that, although the overall quality of the reconstructed profile in Fig. 8(c) is worse, the presence of the feature (plasma depletion) is more pronounced compared to Fig. 8(b). Quantitatively, the J-divergence for CSBS design is 6369 whereas for the alternative configuration it is 4537. Finally, Fig. 9 compares the performance of the CSBS and the right-to-left algorithms and

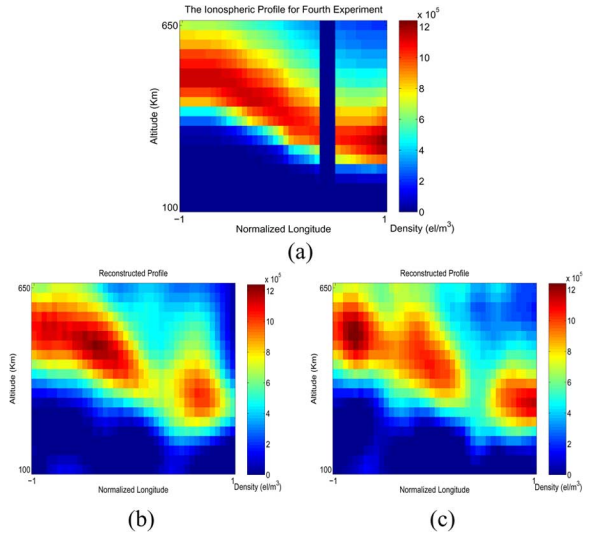

Fig. 8. (a) Ionospheric profile for the fourth experiment; (b) reconstructed profile with the three sensors at $\{-1, 0, 1\}$ (a reasonable alternative); (c) reconstructed profile with the sensors at $\{-2.5, 0, 2.5\}$ (CSBS design for detection criterion).

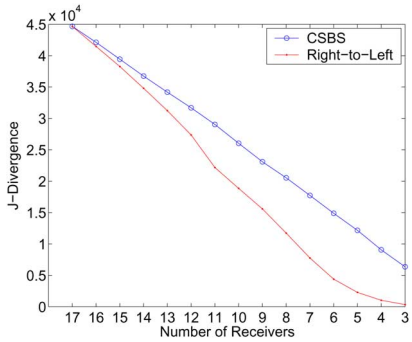

Fig. 9. J-divergence as a function of number of sensors for CSBS (with detection criterion) and right-to-left algorithms.

clearly demonstrates the effectiveness of the CSBS algorithm for the detection criterion.

Finally, the mutual information criterion as defined in (22) was studied. 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 $\{-0.5, 0\}$ as the configuration for the pair of sensors. The reconstructed result was quite similar to that of the SSE criterion in the first experiment shown in Fig. 2(c) and is not shown here for brevity purposes. Quantitatively, the normalized error for the CSBS-designed configuration with mutual information criterion is 9.95% which is almost as good as the SSE result in the first experiment.

VII. Summary and Conclusion

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 linear MMSE 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. Considering the promising simulation results, the proposed framework can contribute to optimal design of real world remote image formation applications.

Appendix A

Proof of Lemma 1

Proof: Define $\mathbf{\Sigma_y}$ to be the auto-covariance matrix of $\mathbf{y}$ and $\mathbf{\Sigma_c}$ to be the cross-covariance of $\mathbf{x}$ and $\mathbf{y}$. The differential entropy of a complex Gaussian random vector with covariance matrix $\mathbf{\Sigma}$ in nats is given by $(1/2)\ln|\pi e\mathbf{\Sigma}|$ where $|.|$ denotes the determinant operator [25]. Hence

$$\begin{aligned} I(\mathbf{x};\mathbf{y}) &= h(\mathbf{x}) + h(\mathbf{y}) - h(\mathbf{x},\mathbf{y}) \\ &= \frac{1}{2}\ln\left(\frac{|\pi e\mathbf{\Sigma_x}||\pi e\mathbf{\Sigma_y}|}{|\pi e\mathbf{C}|}\right) \end{aligned} \tag{39}$$

where

$$\mathbf{C} = \begin{bmatrix} \mathbf{\Sigma_x} & \mathbf{\Sigma_c} \\ \mathbf{\Sigma_c}^H & \mathbf{\Sigma_y} \end{bmatrix}.$$

Using the identity for determinant of two block by two block matrices [29], it follows that

$$\begin{aligned} |\mathbf{C}| &= |\mathbf{\Sigma_y}|\left|\mathbf{\Sigma_x} - \mathbf{\Sigma_c}\mathbf{\Sigma_y}^{-1}\mathbf{\Sigma_c}^H\right| \\ &= |\mathbf{\Sigma_y}||\mathbf{\Sigma_e}| \end{aligned} \tag{40}$$

where we used the identity $\mathbf{\Sigma_e} = \mathbf{\Sigma_x} - \mathbf{\Sigma_c}\mathbf{\Sigma_y}^{-1}\mathbf{\Sigma_c}^H$ from MMSE estimation theory [15, Chapter 11]. By combining (39) and (40), it follows that

$$I(\mathbf{x};\mathbf{y}) = \frac{1}{2}\ln\left(\frac{|\pi e\mathbf{\Sigma_x}|}{|\pi e\mathbf{\Sigma_e}|}\right) = \frac{1}{2}\ln(|\pi e\mathbf{\Sigma_x}|) - \frac{1}{2}\ln(|\pi e\mathbf{\Sigma_e}|)$$

■

Appendix B

Proof of Lemma 3

Proof: Using the so-called Woodbury matrix inversion formula [29], we can write $\mathbf{\Sigma_e}$ as

$$\left(\mathbf{\Sigma_x}^{-1} + \tilde{\mathbf{A}}^H\tilde{\mathbf{A}}\right)^{-1} = \mathbf{\Sigma_x} - \mathbf{\Sigma_x}\tilde{\mathbf{A}}^H\left(\mathbf{I} + \tilde{\mathbf{A}}\mathbf{\Sigma_x}\tilde{\mathbf{A}}^H\right)^{-1}\tilde{\mathbf{A}}\mathbf{\Sigma_x}.$$

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

$$\begin{aligned} &\mathrm{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}} + \mathbf{\Sigma_x}^{-1}\right)^{-1}\right\} \\ &\quad = \mathrm{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\mathbf{\Sigma_x}\right\} \\ &\qquad - \mathrm{tr}\left\{\tilde{\mathbf{A}}_i\mathbf{\Sigma_x}\tilde{\mathbf{A}}^H\left(\mathbf{I} + \tilde{\mathbf{A}}\mathbf{\Sigma_x}\tilde{\mathbf{A}}^H\right)^{-1}\tilde{\mathbf{A}}\mathbf{\Sigma_x}\tilde{\mathbf{A}}_i^H\right\}. \end{aligned} \tag{41}$$

The second term on the right-hand side can be written as $\mathrm{tr}\{(\mathbf{I}+\mathbf{\Psi}^H\mathbf{\Psi})^{-1}\mathbf{\Phi}^H\mathbf{\Phi}\}$, where $\mathbf{\Psi} = \sqrt{\mathbf{\Sigma_x}}\tilde{\mathbf{A}}^H$ and $\mathbf{\Phi} = \tilde{\mathbf{A}}_i\mathbf{\Sigma_x}\tilde{\mathbf{A}}^H$. Now, since $(\mathbf{I}+\mathbf{\Psi}^H\mathbf{\Psi})^{-1}$ is positive definite and $\mathbf{\Phi}^H\mathbf{\Phi}$ is nonnegative definite, the matrix inside the trace is nonnegative definite. Note that trace of a nonnegative definite matrix is nonnegative. Hence, the second term on the right-hand side of (41) is nonpositive which results in the following inequality:

$$\mathrm{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}+\mathbf{\Sigma_x^{-1}}\right)^{-1}\right\} \le \mathrm{tr}\left\{\tilde{\mathbf{A}}i^H\tilde{\mathbf{A}}_i\mathbf{\Sigma_x}\right\}. \quad (42)$$

If the right-hand side of (42) is less than or equal to 1, then (C.1) will be satisfied which, after substituting $\mathbf{\Sigma_x} = (\gamma^2\mathbf{L}^T\mathbf{L})^{-1}$, is equivalent to

$$\gamma^2 \ge \max_{1\le i\le p}\left(\mathrm{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i(\mathbf{L}^T\mathbf{L})^{-1}\right\}\right). \quad (43)$$

Assuming IID Gaussian noise, we have $\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i = (1/\sigma_w^2)\mathbf{A}_i^H\mathbf{A}_i$ and $\lambda = (\gamma\sigma_w)^2$. In this case, (43) is equivalent to

$$\lambda \ge \max_{1\le i\le p}\left(\mathrm{tr}\left\{\mathbf{A}_i^H\mathbf{A}_i(\mathbf{L}^T\mathbf{L})^{-1}\right\}\right). \quad (44)$$

■

APPENDIX C

LEMMA 4 WITH PROOF

Lemma 4: If (C.1) is satisfied, i.e., for all $1 \le i \le p$, we have

$$\mathrm{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}+\mathbf{\Sigma_x^{-1}}\right)^{-1}\right\} < 1 \quad (45)$$

then

$$p - n + \mathrm{tr}\left\{\left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}+\mathbf{\Sigma_x^{-1}}\right)^{-1}\mathbf{\Sigma_x^{-1}}\right\} > 0. \quad (46)$$

Proof: Summing up both sides of (45) from $i = 1$ to $i = p$ gives

$$\sum_{i=1}^{p}\mathrm{tr}\left\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}+\mathbf{\Sigma_x^{-1}}\right)^{-1}\right\} < \sum_{i=1}^{p}1.$$

So

$$\mathrm{tr}\left\{\left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}+\mathbf{\Sigma_x^{-1}}\right)^{-1}\sum_{i=1}^{p}\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\right\} < p. \quad (47)$$

Since $\sum_{i=1}^{p}\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i = \tilde{\mathbf{A}}^H\tilde{\mathbf{A}}$, the left-hand side becomes

$$\begin{aligned}
&\mathrm{tr}\left\{\left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}+\mathbf{\Sigma_x^{-1}}\right)^{-1}(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}})\right\}\\
&\quad= \mathrm{tr}\left\{\left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}+\mathbf{\Sigma_x^{-1}}\right)^{-1}\left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}+\mathbf{\Sigma_x^{-1}}-\mathbf{\Sigma_x^{-1}}\right)\right\}\\
&\quad= \mathrm{tr}\{\mathbf{I}_{n\times n}\} - \mathrm{tr}\left\{\mathbf{\Sigma_x^{-1}}\left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}+\mathbf{\Sigma_x^{-1}}\right)^{-1}\right\}\\
&\quad= n - \mathrm{tr}\left\{\mathbf{\Sigma_x^{-1}}\left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}+\mathbf{\Sigma_x^{-1}}\right)^{-1}\right\}.
\end{aligned}$$

Hence, we can write (47) as

$$n - \mathrm{tr}\left\{\mathbf{\Sigma_x^{-1}}\left(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}+\mathbf{\Sigma_x^{-1}}\right)^{-1}\right\} < p$$

which, after rearranging, gives (46) and completes the proof. ■

ACKNOWLEDGMENT

The authors would like to thank Dr. N. Aggarwal, Dr. M. Jacob, and Prof. M. N. Do for their helpful remarks. They would also like to thank the reviewers for their useful remarks.

REFERENCES

[1] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. New York: IEEE, 1988.

[2] H. Na and H. Lee, "Analysis of fundamental resolution limit of ionospheric tomography," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Mar. 1992, pp. 97–100.

[3] B. Hall and H. Na, "Measures of information content for computerized ionospheric tomography," in *Proc. IEEE Int. Conf. Image Processing*, Oct. 1995, pp. 433–436.

[4] N. Aggarwal and Y. Bresler, "Optimal sampling in parallel magnetic resonance imaging," in *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, Sep. 2003, pp. 737–740.

[5] Y. Gao and S. J. Reeves, "Optimal sampling in array-based image formation," in *Proc. IEEE Int. Conf. Image Processing*, Vancouver, BC, Canada, Sep. 2000, pp. 733–736.

[6] B. Sharif, "Remote image formation from limited data: A hybrid technique, fundamental limitations, and optimal design," M.S. thesis, Dept. Elect. Comput. Eng., Univ. Illinois Urbana-Champaign, Urbana, IL, 2004.

[7] S. J. Reeves and L. P. Heck, "Selection of observations in signal reconstruction," *IEEE Trans. Signal Process.*, vol. 43, no. 3, pp. 788–791, Mar. 1995.

[8] A. J. Miller, *Subset Selection in Regression*. London, U.K.: Chapman & Hall, 1990.

[9] S. F. Cotter, "Subset selection algorithms with applications," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. California, San Diego, CA, 2001.

[10] S. J. Reeves and Z. Zhe, "Sequential algorithms for observation selection," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 123–132, Jan. 1999.

[11] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Sov. Math. (Providence)*, vol. 4, pp. 1035–1038, 1963.

[12] W. C. Karl, "Regularization in image restoration and reconstruction," in *Handbook of Image and Video Processing*, A. Bovik, Ed. San Diego, CA: Academic, 2000, ch. 3.6, pp. 141–160.

[13] T. Nygrén, M. Markkanen, M. Lehtinen, E. D. Tereshchenko, and B. Z. Khudukon, "Stochastic inversion in ionospheric radiotomography," *Radio Sci.*, vol. 32, pp. 2359–2372, 1997.

[14] T. Nygrén, M. Markkanen, M. Lehtinen, and K. Kaila, "Application of stochastic inversion in auroral tomography," *Ann. Geophys.*, vol. 14, pp. 1124–1133, 1996.

[15] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993, vol. 1.

[16] P. L. Combettes, "The foundations of set theoretic estimation," *Proc. IEEE*, vol. 81, no. 2, pp. 182–208, Feb. 1993.

[17] G. Fehmers, L. Kamp, and F. Sluijter, "A model-independent algorithm for ionospheric tomography 1. theory and tests," *Radio Sci.*, vol. 33, pp. 149–149, 1998.

[18] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed. Philadelphia, PA: SIAM, 2003.

[19] C. Vogel and Q. Yang, "Multigrid algorithm for least squares wavefront reconstruction," *Appl. Opt.*, vol. 45, pp. 705–715, 2006.

[20] M. I. Sezan and H. J. Trussell, "Prototype image constraints for set-theoretic image restoration," *IEEE Trans. Signal Process.*, vol. 39, no. 10, pp. 2275–2285, Oct. 1991.

[21] P. Ishwar and P. Moulin, "On the equivalence of set-theoretic and maxent MAP estimation," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 698–713, Mar. 2003.
[22] F. Kamalabadi and B. Sharif, "Robust regularized tomographic imaging with convex projections," in *Proc. Int. Conf. Image Processing*, Genova, Italy, 2005, vol. 2, pp. 205–208.
[23] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.
[24] C.-T. Yu and P. K. Varshney, "Sampling design for Gaussian detection problems," *IEEE Trans. Signal Process.*, vol. 45, no. 9, pp. 2328–2337, Sep. 1997.
[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[26] J. Pluim, J. Maintz, and M. Viergever, "Image registration by maximization of combined mutual information and gradient information," *IEEE Trans. Med. Imag.*, vol. 19, no. 8, pp. 809–814, Aug. 2000.
[27] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Trans. Neural Netw.*, vol. 5, no. 4, pp. 537–550, Jul. 1994.
[28] B. Sharif and F. Kamalabadi, "Optimal sensor configuration in remote image formation," in *Proc. IEEE Int. Conf. Image Processing*, Genova, Italy, 2005, vol. 2, pp. 197–200.
[29] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1989.
[30] T. M. Cover and J. M. V. Campenhout, "On the possible orderings in the measurement selection problem," *IEEE Trans. Syst., Man., Cybern.*, vol. SMC-7, no. 5, pp. 657–661, Sep. 1977.
[31] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
[32] C. Couvreur and Y. Bresler, "On the optimality of the backward greedy algorithm for the subset selection problem," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 3, pp. 797–808, 1999.
[33] J. R. Austen, S. J. Franke, and C. H. Lui, "Ionospheric imaging using computerized tomography," *Radio Sci.*, vol. 23, pp. 299–307, 1988.
[34] A. R. Thompson, J. M. Moran, and G. W. Swenson, Jr, *Interferometry and Synthesis in Radio Astronomy*. New York: Wiley, 2001.
[35] E. Kudeki and F. Surucu, "Radar interferometric imaging of field-aligned plasma irregularities in equatorial electrojet," *Geophys. Res. Lett.*, vol. 18, no. 1, 1991.
[36] E. B. Fomalont, "Earth-rotation aperture synthesis," *Proc. IEEE*, vol. 61, no. 9, pp. 1211–1218, Sep. 1973.
[37] R. E. Blahut, *Theory of Remote Image Formation*. New York: Cambridge Univ. Press, 2004.
[38] G. Harikumar and Y. Bresler, "A new algorithm for computing sparse solutions to linear inverse problems," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, May 1996, pp. 1331–1334.
[39] B. D. Rao, "Signal processing with sparseness constraint," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Seattle, WA, May 1998, pp. 1861–1864.
[40] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

Behzad Sharif (S'99) received the B.Sc. degree (with high distinction) in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2002, and the M.S. degree in electrical and computer engineering (ECE) from the University of Illinois at Urbana-Champaign (UIUC), Urbana, in 2004, where he is currently pursuing the Ph.D. degree.

From August 2002 to May 2005, he was a Research Assistant in the Remote Sensing group at the Coordinated Science Laboratory (CSL), UIUC. Since June 2005, he has been a Research Assistant in the Image Formation and Processing group, CSL. His research interests include theory of image formation and computational imaging, and mutlidimensional and statistical signal processing and their applications to inverse problems in biomedical and remote imaging.

Mr. Sharif received a Silver Medal in the National Iranian Informatics Olympiad in 1998 and the C. R. Allen International Student Award in 2007 from UIUC, in recognition of outstanding research potential and academic record. He was among the top three finalists of the 2007 Lemelson-Illinois student prize for groundbreaking inventiveness. He is a recipient of the Computational Science and Engineering graduate fellowship and the Beckman Institute multidisciplinary graduate fellowship at the University of Illinois.

Farzad Kamalabadi (M'02) received the B.S. degree in computer systems engineering from the University of Massachusetts, Amherst, in 1991, and the M.S. and Ph.D. degrees in electrical engineering from Boston University, Boston, MA, in 1994 and 2000, respectively.

In 2000, he joined the University of Illinois at Urbana-Champaign, Urbana, where he is currently an Associate Professor in the Department of Electrical and Computer Engineering, and a Research Associate Professor at the Coordinated Science Laboratory. In Fall 2002, he was a Visiting Fellow with SRI International, and in Summer 2003, he was a National Aeronautics and Space Administration (NASA) Faculty Fellow with the Jet Propulsion Laboratory, California Institute of Technology. His research interests include multidimensional and statistical signal processing, inverse problems in image formation, sensor array processing, and tomography, and the development of optical and radio sensing and imaging techniques and their applications to space physics, astronomy, and medicine.

Dr. Kamalabadi received a National Science Foundation (NSF) Fellowship in 1994, a Best Student Paper Award in 1997, a NASA fellowship in 1998, a NSF Faculty Early Career Development (CAREER) Award in 2002, and a Xerox Award for Outstanding Faculty Research in 2006. He is currently a member of the Users and Scientific Advisory Committee of the Arecibo Observatory and has served in various organizational capacities, including Session Organizer and Chair for the International Union of Radio Science special sessions on ionospheric imaging in 2003 and 2004.