On Optimal Sensor Array Conguration in Remote Image Formation

Farzad Kamalabadi

University of Illinois at Urbana-Champaign

June 18, 09 – i3S/CNRS
Joint Work with: Behzad Sharif
IEEE Trans. Image Proc., Vol 17, No. 2, Feb 2008

Outline

1. Motivation and Background
2. Image Formation Model
3. Clustered Sequential Backward Selection
4. Statistical Formulation and Optimality Criteria
5. Numerical Experiments
6. Computational Complexity and Fast Implementation
7. Upper Bound on the Performance of CSBS

Background and Problem Statement

- The quality of the reconstructed images in tomographic and interferometric imaging is influenced by the choice of the placement of array elements.
- The problem of optimizing the array configuration has most commonly been approached from a transform-domain analysis by investigating the impact of the sensor placement on the sampling of multi-dimensional Fourier space, or alternatively by considering the distribution of the singular values of the imaging operator.
- In this work, a combinatoric optimization of possible configurations based on minimizing a statistical error measure is proposed. In particular, a statistical framework is developed which allows for inclusion of several widely used image model constraints, and a search is performed over the space of candidate sensor locations to determine the conguration that optimizes the desired statistical optimality criterion over all candidates.

Example Imaging Modality: Tomography

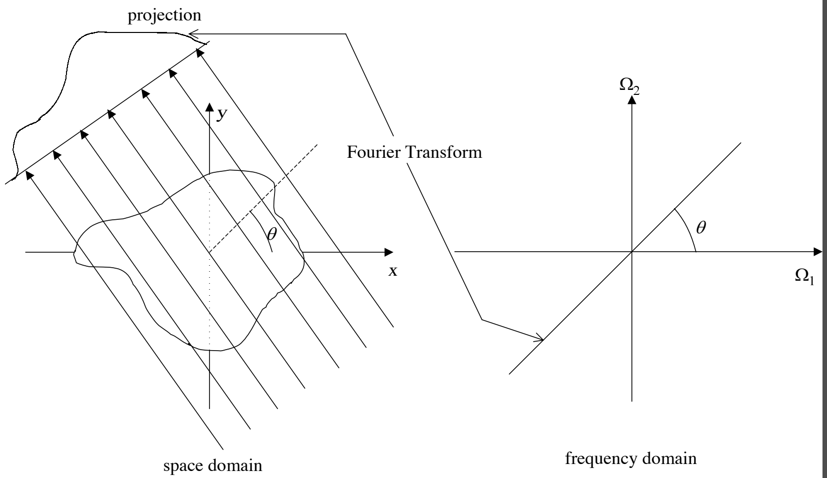


Image formation model

- 2D model: In image formation scenarios involving multiple sensors, the relationship between the set of observations and the unknown field can often be adequately characterized by a linear observation model. The observation/state relationship can thus be formulated by the Fredholm integral equation of the first kind [?]:

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

  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 $X(r,s)$ denote the measured data (observation) and the unknown field (state) respectively. The observation kernel or system point spread (response) function is denoted by $h(r,s;r',s')$.

Discrete Model (1/2)

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 adequately 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.

Discrete Model(2/2)

Collecting all the observations into a vector $\boldsymbol{y}$ of length M, and the unknown image coefficients into a vector $\boldsymbol{x}$ of length N, results in the following observation model in the form of a matrix equation:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{3}$$

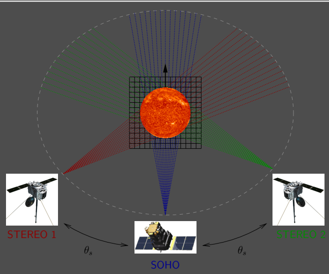
where $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ is the linear operator relating the unknown field to the observations, comprised of inner products of the basis functions with the corresponding observation kernel:

$$(\boldsymbol{A})_{ij} = \iint_{\Omega} h_i(r', s')\phi_j(r', s')dr'ds',\ 1 \leq i \leq M,\ 1 \leq j \leq N \tag{4}$$

where $h_i(r', s') = h(r_i, s_i; r', s')$ denotes the kernel function corresponding to the i-th observation.

3D Image Formation Model Example

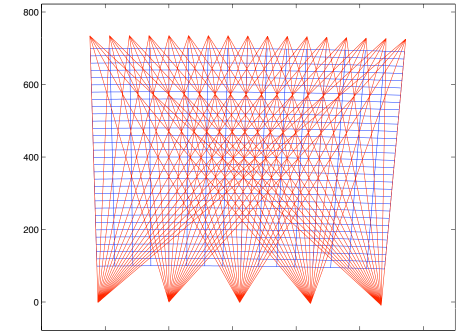
$$Y(r,s,t) = \iiint_\Omega h(r,s,t;r',s',t')X(r',s',t')dr'ds'dt' \qquad (5)$$



(a) $\boldsymbol{y}_1 = \boldsymbol{A}_1\boldsymbol{x} + \boldsymbol{v}_1$ (b) $\boldsymbol{y}_2 = \boldsymbol{A}_2\boldsymbol{x} + \boldsymbol{v}_2$

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_{14} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{A}_{14} \end{pmatrix}\boldsymbol{x} + \begin{pmatrix} \boldsymbol{v}_1 \\ \vdots \\ \boldsymbol{v}_{14} \end{pmatrix} \implies \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{v}$$

Examples of Imaging Modalities: Ionospheric Tomography



Examples of Imaging Modalities: Ionospheric Tomography

- 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.
- The goal of this technique is to reconstruct horizontal and vertical structures within the two-dimensional slice of the ionosphere above the ground sensors.

Examples of Imaging Modalities: Interferometric Imaging (1/2)

Assume that the discretized sky is made up of n pixels, $\mathbf{x} = [x_1 \cdots x_n]^T \in \mathbb{C}^n$. Suppose the two-dimensional antenna array has p elements and we take data at d look positions. Denote by ϕ_j and θ_j the azimuth and zenith angles that the j-th pixel form relative to the center of the array. Assuming that θ_j and ϕ_j are small (less than 0.1 radian) each antenna sees the signal from the j-th 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 k-th antenna at i-th look position. Therefore, the data received at k-th antenna is:

$$y_k(i) = \sum_{j=1}^{n} x_j e^{\mathbf{j}2\pi[r_k(i)\theta_j + s_k(i)\phi_j]} + w_k(i) \tag{6}$$

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.

Examples of Imaging Modalities: Interferometric Imaging (2/2)

Collecting the data for all of the antenna at look position i into a vector, the imaging equation can be expressed as:

$$\mathbf{y}_i = \mathbf{A}_i \mathbf{x} + \mathbf{w}_i \tag{7}$$

where $\mathbf{A}_i \in \mathbb{C}^{p \times n}$ is the corresponding observation operator that consists of the complex exponentials in (6). As can be seen, in this modality we have multiple independent measurements and the formulation needs to be adapted accordingly. The error covariance for the case of m independent measurements in (7) is given by

$$\boldsymbol{\Sigma}_{\mathbf{e}} = (\boldsymbol{\Sigma}_{\mathbf{x}}^{-1} + \sum_{i=1}^{d} \mathbf{A}_i^H \boldsymbol{\Sigma}_{\mathbf{w}_i}^{-1} \mathbf{A}_i)^{-1} \tag{8}$$

Combinatoric Search

Search over a set of candidate locations for sensors to find the optimal or close-to-optimal subset. Assume 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 $\mathbf{A}^0$. 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:

$$\mathcal{S}^* = \arg \min_{\mathcal{U} \subseteq \mathcal{S},\ |\mathcal{U}| = q} \mathrm{Cost}(\mathbf{A}^{(\mathcal{U})}) \tag{9}$$

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.

Optimality Criteria

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.

$$\mathcal{S}^* = \arg \min_{\mathcal{U} \subseteq \mathcal{S},\ |\mathcal{U}|=q} \mathrm{Cost}(\mathbf{A}^{(\mathcal{U})}) \tag{10}$$

The best subset of rows of $\mathbf{A}^0$ can be found by exhaustive search over all $\binom{p}{q}$ possible combinations of rows. But this grows exponentially in p and is impractical even for moderate number of candidate locations.

Statistical Formulation

Maximum *a posteriori* (MAP) estimation assuming Gaussian statistics for both the unknown image and noise. Assuming $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{w})$ and $\mathbf{x} \sim \mathcal{CN}(\mathbf{x}_0, \boldsymbol{\Sigma}_\mathbf{x})$, where $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ represents the complex normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, the MAP estimate is

$$\begin{aligned}
\hat{\mathbf{x}}_{\mathrm{MAP}} &= \arg\min_{\mathbf{x}\in\mathbb{C}^n} \left[-\log p(\mathbf{y}|\mathbf{x}) - \log p(\mathbf{x})\right] \\
&= \arg\min_{\mathbf{x}\in\mathbb{C}^n} \left[\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2_{\boldsymbol{\Sigma}_\mathbf{w}^{-1}} + \|\mathbf{x} - \mathbf{x}_0\|^2_{\boldsymbol{\Sigma}_\mathbf{x}^{-1}}\right] \\
&= \mathbf{x}_0 + (\mathbf{A}^H \boldsymbol{\Sigma}_\mathbf{w}^{-1} \mathbf{A} + \boldsymbol{\Sigma}_\mathbf{x}^{-1})^{-1} \mathbf{A}^H \boldsymbol{\Sigma}_\mathbf{w}^{-1} (\mathbf{y} - \mathbf{A}\mathbf{x}_0)
\end{aligned} \tag{11}$$

Equivalent Variational Form

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

$$\hat{\mathbf{x}}_{\text{Tik}} = \arg\min_{\mathbf{x}\in\mathbb{C}^n} \left[\frac{1}{\sigma_w^2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \gamma^2\|\mathbf{L}(\mathbf{x}-\mathbf{x}_0)\|_2^2\right] \tag{12}$$

$$= \arg\min_{\mathbf{x}\in\mathbb{C}^n} \left[\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{L}(\mathbf{x}-\mathbf{x}_0)\|_2^2\right]$$

$$= \mathbf{x}_0 + \left(\frac{1}{\sigma_w^2}\mathbf{A}^H\mathbf{A} + \gamma^2\mathbf{L}^T\mathbf{L}\right)^{-1}\frac{1}{\sigma_w^2}\mathbf{A}^H(\mathbf{y}-\mathbf{A}\mathbf{x}_0) \tag{13}$$

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.

Measure of Estimation Uncertainty

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 $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$. For the MAP estimate, the error covariance is given by:

$$\mathbf{\Sigma_e} = (\mathbf{A}^H \mathbf{\Sigma_w}^{-1} \mathbf{A} + \mathbf{\Sigma_x}^{-1})^{-1} \tag{14}$$

where the expected squared error for the i-th 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 $\mathbf{\Sigma_e} = (\frac{1}{\sigma_w^2}\mathbf{A}^H\mathbf{A} + \gamma^2 \mathbf{L}^{\mathbf{T}}\mathbf{L})^{-1}$. It should be noted that with no assumption on the distribution of the image (i.e., non-Gaussian statistics) the estimator in (12) 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 work will in general be valid in the context of linear MMSE estimation.

Incorporating Convex Constraints into the Statistical Framework (1/4)

- *Smoothness Constraint*: The $\mathbf{L}$ matrix in (12) is typically taken to be a discrete approximation to the gradient operator in order to formulate smoothness of the unknown image. Using this regularization matrix is equivalent to assuming $\mathbf{x}$ to be a Brownian motion. An alternative class of regularization matrices are discretizations of the two-dimensional Laplacian operator, namely Fried and Hudgin discrete Laplacians.
  In some applications, multiple regularization matrices are used.

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

  This can be written in form of (12) 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.

Incorporating Convex Constraints into the Statistical Framework (2/4)

- *Support Constraint*: In some scenarios, 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 (12) 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 (13), it is desirable to design a matrix $\mathbf{B}$ such that:

$$h(\mathbf{x}) = \int_{(r,s)\notin\mathcal{H}} |X(r,s)|^2 dr ds \approx \|\mathbf{B}\mathbf{x}\|_2^2 \qquad (15)$$

  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. 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 $\mathbf{x}_0 = \mathbf{0}$).

Incorporating Convex Constraints into the Statistical Framework (3/4)

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

$$\mathcal{C} = \{\mathbf{x}' \in \mathbb{C}^n : \|\mathbf{x}' - \mathbf{x}_0\|_2 < \rho\} \tag{16}$$

  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. 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. The constrained ML estimation is equivalent to an unconstrained MAP estimation which, assuming Gaussian noise, reduces to Tikhonov regularization:

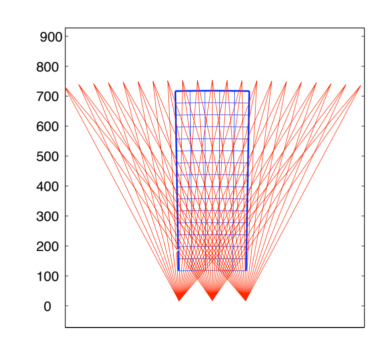
$$\begin{aligned} \hat{\mathbf{x}}_{\mathrm{CML}} &= \arg\min_{x \in \mathcal{C}} \left[-\log p(\mathbf{y}|\mathbf{x})\right] \\ &= \arg\min_{x \in \mathbb{C}^n} \left[-\log p(\mathbf{y}|\mathbf{x}) + \nu(\rho, \mathbf{y})\|\mathbf{x} - \mathbf{x}_0\|_2^2\right] \end{aligned}$$

Incorporating Convex Constraints into the Statistical Framework (4/4)

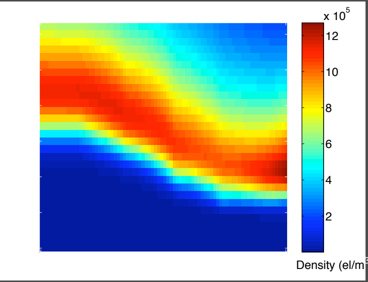
- *Reference Constraint (continued)*: where $\nu(\rho, \mathbf{y}) \in \mathbb{R}^+$ is the associated Lagrange multiplier. As can be seen, the constrained ML estimation has the same form as the Tikhonov functional with $\mathbf{L} = \mathbf{I}_{n\times n}$ and $\lambda = \nu(\rho, \mathbf{y})$. Hence (14) applies here as well.
- *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 this work apply.

Numerical Experiments

A typical scenario for ionospheric radio tomography



(a) Measurement Geometry.



(b) Original Ionosphere.

The set of candidate locations, $\mathcal{S}$ in (10), was 31 equispaced points within normalized longitude range of -2.5 to 2.5. The desired number of sensors, q in (10), is set to 2. The discretization grid has a span of -1 to 1.

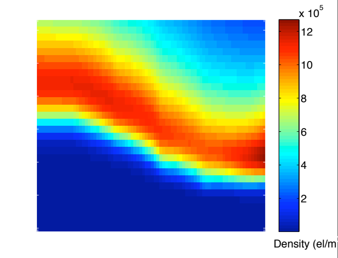
Optimality Criteria (1/5)

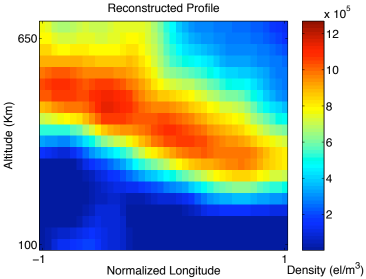
- *Sum of Squared Errors*: The sum of squared errors (SSE) or expected value of the squared estimation error, i.e., $E[\mathbf{e}^H\mathbf{e}]$, is given by

$$\mathrm{Cost}^1(\mathbf{A}) = \mathrm{tr}(\mathbf{\Sigma_e}) = \sum_{i=1}^{n} (\mathbf{\Sigma_e})_{ii} \tag{17}$$

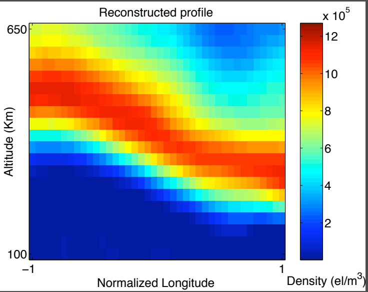
  If no assumption about the statistics of the signal are made and only the noise is assumed to be white Gaussian, the SSE criterion will take the form of $\mathrm{tr}\{(\mathbf{A}^H\mathbf{A})^{-1}\}$ where the inverse operator is usually replaced by the pseudo inverse to avoid instability due to ill-posedness of the problem.

Numerical Experiments



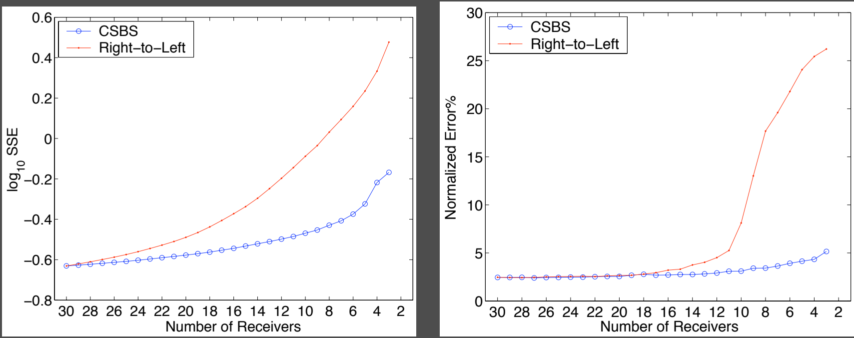


(c) Reconstruction with uniform spacing $\{-1, 1\}$



(d) Reconstruction with CSBS spacing $\{-0.5, 0.17\}$

Numerical Experiments



(e) SSE (f) Normalized error

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.

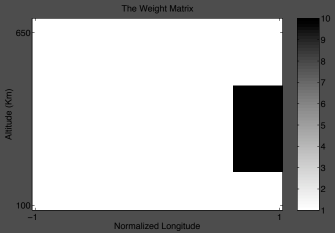
Optimality Criteria (2/5)

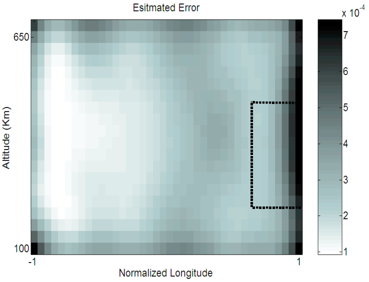
- *Weighted Sum of Squared Errors*: In some applications, it is desirable to have small error in one specific area of the reconstructed image while larger errors could be tolerated in other areas. A reasonable approach to designing the cost function for this setting is to weight $\mathrm{Cost}^1(\mathbf{A})$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

$$\mathrm{Cost}^2(\mathbf{A}) = \sum_{i=1}^{n} (\mathbf{\Sigma_e})_{ii} W_i \tag{18}$$

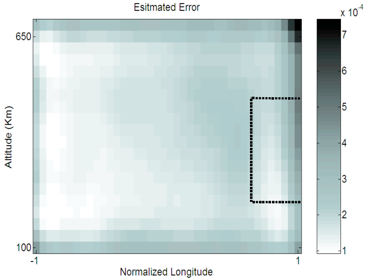
where $W_i \in \mathbb{R}^+$ are the weighting coefficients.

Numerical Experiments



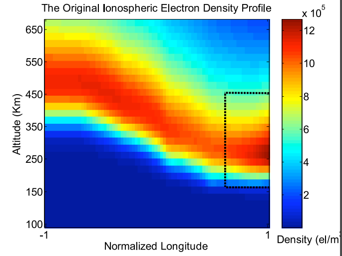


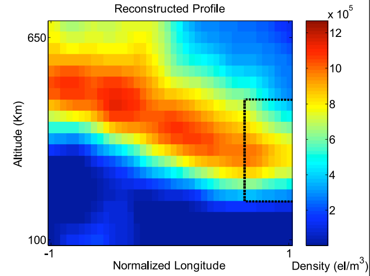
(g) Error Cov for uniform spacing



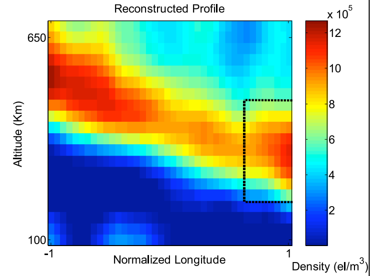
(h) Error Cov for CSBS spacing $\{0.25, 2.5\}$

Numerical Experiments



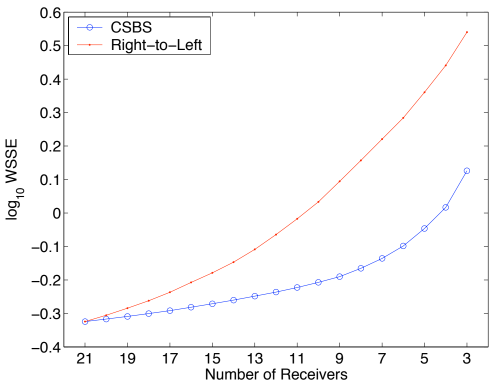


(i) Reconstruction with uniform spacing



(j) Reconstruction with CSBS spacing

Numerical Experiments



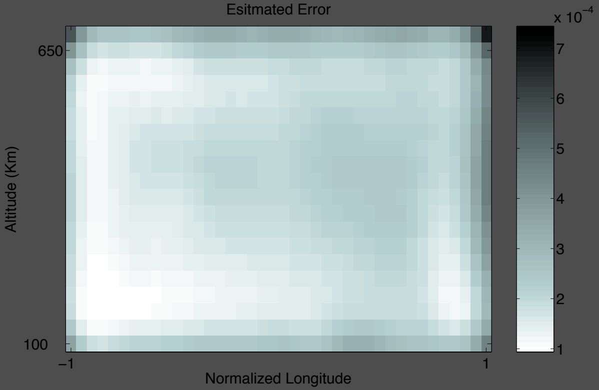
Optimality Criteria (3/5)

- *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:

$$\mathrm{Cost}^3(\mathbf{A}) = \mathrm{STD}\left\{(\mathbf{\Sigma_e})_{ii}\right\}_{i=1}^{n} \tag{19}$$

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

Numerical Experiments



CSBS: $\{0.25, 2.5\}$

Optimality Criteria (4/5)

- *Detection Performance*: 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 $\mathcal{CN}(\mathbf{b}_0, \mathbf{\Sigma}_\mathbf{b})$ and $\mathcal{CN}(\mathbf{x}_0, \mathbf{\Sigma}_\mathbf{x})$ respectively. The noise vector $\mathbf{w}$ is a Gaussian distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_\mathbf{w})$, independent of $\mathbf{b}$ and $\mathbf{x}$. The detection problem model: (binary hypothesis testing)

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

Assuming $\mathbf{A} \in \mathbb{R}^{m \times n}$, the observation vector $\mathbf{y}$ is also a real Gaussian random vector. The covariance matrices of $\mathbf{y}$ under hypotheses H_0 and H_1 are:

$$\mathbf{\Sigma}_{\mathbf{y}|H_0} = \mathbf{A}\mathbf{\Sigma}_\mathbf{x}\mathbf{A}^H + \mathbf{\Sigma}_\mathbf{w} \quad (20)$$

$$\mathbf{\Sigma}_{\mathbf{y}|H_1} = \mathbf{A}(\mathbf{\Sigma}_\mathbf{x} + \mathbf{\Sigma}_\mathbf{b})\mathbf{A}^H + \mathbf{\Sigma}_\mathbf{w} \quad (21)$$

Optimality Criteria (4/5)

- *Detection Performance*: The optimal choice of receiver configuration: minimizes the probability of error for the above hypothesis testing problem. However, the corresponding optimization problem is analytically intractable. Instead, maximize the divergence between the observation distributions under H_0 and H_1 that is the distance between the conditional densities. Such distance measures include the Bhattacharyya distance, the I-divergence, the J-divergence, and the Chernoff distance.

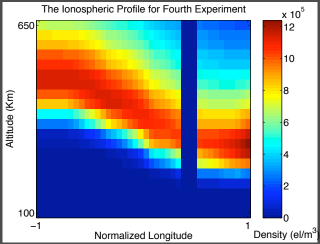
$$Jdiv(\mathbf{A}) = \frac{1}{2}\mathrm{tr}\left(\left[\boldsymbol{\Sigma}_{\mathbf{y}|H_1} - \boldsymbol{\Sigma}_{\mathbf{y}|H_0}\right]\left[{\boldsymbol{\Sigma}_{\mathbf{y}|H_0}}^{-1} - {\boldsymbol{\Sigma}_{\mathbf{y}|H_1}}^{-1}\right]\right) + \tag{22}$$

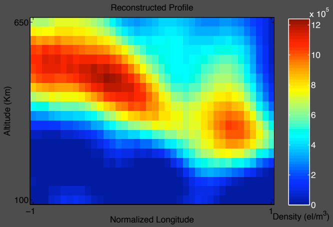
$$\frac{1}{2}\mathrm{tr}\left(\left[{\boldsymbol{\Sigma}_{\mathbf{y}|H_1}}^{-1} + {\boldsymbol{\Sigma}_{\mathbf{y}|H_0}}^{-1}\right](\mathbf{m}_1 - \mathbf{m}_0)(\mathbf{m}_1 - \mathbf{m}_0)^H\right) \tag{23}$$

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 is

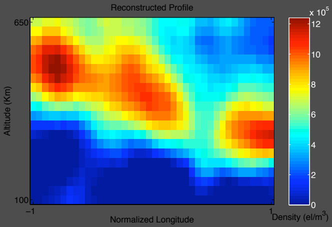
$$\mathrm{Cost}^4(\mathbf{A}) = -Jdiv(\mathbf{A}) \tag{24}$$

Numerical Experiments



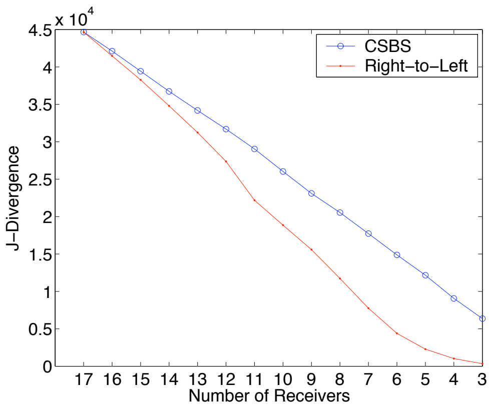


(k) Reconstruction with uniform spacing $\{-1, 0, 1\}$



(l) Reconstruction with CSBS spacing $\{-2.5, 0, 2.5\}$

Numerical Experiments



Optimality Criteria (5/5)

- *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. 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:

$$I(\mathbf{x};\mathbf{y}) = h(\mathbf{x}) + h(\mathbf{y}) - h(\mathbf{x},\mathbf{y}) \tag{25}$$

  where $h(.)$ denotes the differential entropy and $h(\mathbf{x},\mathbf{y})$ is the joint entropy of random vectors $\mathbf{x}$ and $\mathbf{y}$. It is a measure of general statistical dependence between variables. It does not commit to a specific error metric as opposed to the aforementioned criteria and is 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 $\mathbf{y}$.

Optimality Criteria (5/5)

- *Mutual Information (Computable form)*:

Lemma 1:

Under our statistical framework, $I(\mathbf{x};\mathbf{y})$ can be simplified as

$$I(\mathbf{x};\mathbf{y}) = \frac{1}{2}\ln(|\pi e \mathbf{\Sigma_x}|) - \frac{1}{2}\ln(|\pi e \mathbf{\Sigma_e}|) \tag{26}$$

where $|.|$ denotes the determinant operator.

Note that the first term on the right hand side in (26) is independent of observations and thus the set of most informative sensors maximizes the second term. According to (26), in order to maximize the mutual information $I(\mathbf{x};\mathbf{y})$, we need to minimize $\frac{1}{2}\ln(|\pi e \mathbf{\Sigma_e}|)$ or equivalently $|\mathbf{\Sigma_e}|$. Hence,

$$\mathrm{Cost}^5(\mathbf{A}) = |\mathbf{\Sigma_e}| \tag{27}$$

Note the connection between the SSE cost function in (17) and that of mutual information in (27).

Optimality Criteria (5/5)

- *Mutual Information*:

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 $\frac{1}{2}\ln|\pi e \mathbf{\Sigma}|$. Hence:

$$\begin{aligned} I(\mathbf{x};\mathbf{y}) &= h(\mathbf{x}) + h(\mathbf{y}) - h(\mathbf{x},\mathbf{y}) \\ &= \frac{1}{2}\ln\left(\frac{|\pi e \mathbf{\Sigma_x}||\pi e \mathbf{\Sigma_y}|}{|\pi e \mathbf{C}|}\right) \end{aligned} \quad (28)$$

□

Optimality Criteria (5/5)

- *Mutual Information (continued)*:
  Using the identity for determinant of two block by two block matrices it follows that:

$$\begin{aligned}|\mathbf{C}| &= |\boldsymbol{\Sigma}_{\mathbf{y}}||\boldsymbol{\Sigma}_{\mathbf{x}} - \boldsymbol{\Sigma}_{\mathbf{c}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\boldsymbol{\Sigma}_{\mathbf{c}}^{H}| \\ &= |\boldsymbol{\Sigma}_{\mathbf{y}}||\boldsymbol{\Sigma}_{\mathbf{e}}| \end{aligned} \tag{29}$$

  where we used the identity $\boldsymbol{\Sigma}_{\mathbf{e}} = \boldsymbol{\Sigma}_{\mathbf{x}} - \boldsymbol{\Sigma}_{\mathbf{c}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\boldsymbol{\Sigma}_{\mathbf{c}}^{H}$ from MMSE estimation theory. By combining (28) and (29), it follows that:

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

Clustered Sequential Backward Selection Algorithm

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:

$$\mathbf{\Gamma} \leftarrow \mathbf{\Gamma} \backslash \{k^*\} : k^* = \arg\min_{k \in \mathbf{\Gamma}} \mathrm{Cost}(\mathbf{A}^{(\mathbf{\Gamma} \backslash \{k\})}) \tag{30}$$

with the stopping criterion being $|\mathbf{\Gamma}| = q$ where q is the desired number of sensors as mentioned in (10).

Computational Complexity

In the general CSBS algorithm stated in (30), there can be at most $p-1$ iterations (because $q>0$) and at the i-th 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 (14). Assuming that the constant matrices are computed and stored beforehand, straightforward computation of $\mathbf{A}^H \mathbf{\Sigma}_{\mathbf{w}}^{-1} \mathbf{A}$ is of $O(mn^2+nm^2)$ and we need an additional $O(n^3)$ flops for inversion of the sum. Having $m \geq n$, complexity of computing the cost function is $O(nm^2)$. Therefore, the overall complexity is $O(nm^2p^2)$.

Fast Implementation (1/2)

Using the Sherman-Morrison matrix inversion formula, we first state a 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

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

where $\boldsymbol{\Sigma}_{\mathbf{e}}$ is as given in (14). Therefore, SBS can be implemented by choosing the row that minimizes the second term of the right hand side.

Fast Implementation (2/2)

Based on this fact, a fast implementation of CSBS can be derived as follows:

$$\mathbf{\Gamma} \leftarrow \mathbf{\Gamma} \backslash \{k^*\} : k^* = \arg\min_{k \in \mathbf{\Gamma}} \sum_{i \in \Pi_k} \frac{\mathbf{a}_i \mathbf{\Sigma}_{\mathbf{e}}^2 \mathbf{a}_i^H}{1 - \mathbf{a}_i \mathbf{\Sigma}_{\mathbf{e}} \mathbf{a}_i^H} \tag{31}$$

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

Note that for each iteration of the algorithm in (31), we only need to compute and store $\mathbf{\Sigma}_{\mathbf{e}}$ once, which has complexity $O(nm^2)$ as discussed above. For each iteration, i.e., elimination of one sensor, the summands in (31) can be computed in $O(n^2p)$ flops. 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.

Upper Bound on Performance of CSBS

Throughout the remainder, 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 $\mathbf{\Sigma}_{\mathbf{w}}^{-1}$ is positive definite Hermitian, it has a positive definite square root, namely $\sqrt{\mathbf{\Sigma}_{\mathbf{w}}^{-1}}$. Considering the CSBS algorithm and defining $\tilde{\mathbf{A}} = \sqrt{\mathbf{\Sigma}_{\mathbf{w}}^{-1}}\mathbf{A}$, the SSE in (14) can be written as:

$$\mathbf{\Sigma}_{\mathbf{e}} = (\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \mathbf{\Sigma}_{\mathbf{x}}^{-1})^{-1} \tag{32}$$

Let us assume that the set of p candidate sensor locations is indexed as $\Gamma = \{1, 2, \cdots, p\}$. 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 (17) and (32), the SSE can be written as:

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

where $\tilde{\mathbf{A}}_i = \sqrt{\mathbf{\Sigma}_{\mathbf{w}}^{-1}}\mathbf{A}_i$, $\mathbf{G}_i = \sum_{\substack{1 \leq j \leq p \\ j \neq i}} \tilde{\mathbf{A}}_j^H \tilde{\mathbf{A}}_j$, 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, consider the SBS algorithm result.

Upper Bound on Performance of CSBS

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:

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

Upper Bound on Performance of CSBS

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 i-th 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 i-th sensor location:

$$\frac{d - n + 1 + \operatorname{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}(\mathbf{G}_i + \boldsymbol{\Sigma}_{\mathbf{x}}^{-1})\}}{-n + 1 + \operatorname{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}(\mathbf{G}_i + \boldsymbol{\Sigma}_{\mathbf{x}}^{-1})\}} \operatorname{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\} \tag{33}$$

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

$$\operatorname{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}(\mathbf{G}_i + \boldsymbol{\Sigma}_{\mathbf{x}}^{-1})\} = \tag{34}$$

$$\operatorname{tr}\{(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \boldsymbol{\Sigma}_{\mathbf{x}}^{-1})^{-1}(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} - \tilde{\mathbf{A}}_i^H \tilde{\mathbf{A}}_i)\} = \tag{35}$$

$$n - \operatorname{tr}\{\tilde{\mathbf{A}}_i^H \tilde{\mathbf{A}}_i \boldsymbol{\Sigma}_{\mathbf{e}}\} \tag{36}$$

Therefore, we can write (33) as

$$\frac{d + 1 - \operatorname{tr}\{\tilde{\mathbf{A}}_i^H \tilde{\mathbf{A}}_i \boldsymbol{\Sigma}_{\mathbf{e}}\}}{\ldots} \operatorname{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\} \tag{37}$$

Upper Bound on Performance of CSBS

It should be noted that the bound in Lemma 2 is valid only if the denominator is positive. In fact, considering all $1 \le i \le p$, the expression in (37) is a valid upper bound for a SSE only if

$$\mathrm{tr}\{\tilde{\mathbf{A}}_i^H \tilde{\mathbf{A}}_i \mathbf{\Sigma_e}\} < 1 \quad \text{for } 1 \le i \le p \qquad (\mathrm{C}.1)$$

The following lemma provides an easily verifiable sufficient condition under which $(\mathrm{C}.1)$ holds.

Lemma 3:

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

$$\gamma^2 \ge \max_{1 \le i \le p} \left(\mathrm{tr}\{\tilde{\mathbf{A}}_i^H \tilde{\mathbf{A}}_i (\mathbf{L}^T \mathbf{L})^{-1}\} \right) \qquad (38)$$

and assuming IID Gaussian noise, (38) is equivalent to

$$\lambda \ge \max_{1 \le i \le p} \left(\mathrm{tr}\{\mathbf{A}_i^H \mathbf{A}_i (\mathbf{L}^T \mathbf{L})^{-1}\} \right) \qquad (39)$$

Upper Bound on Performance of CSBS

Proof.

Using the so called Woodbury matrix inversion formula we can write $\mathbf{\Sigma_e}$ as:

$$(\mathbf{\Sigma}_{\mathbf{x}}^{-1} + \tilde{\mathbf{A}}^H\tilde{\mathbf{A}})^{-1} = \mathbf{\Sigma}_{\mathbf{x}} - \mathbf{\Sigma}_{\mathbf{x}}\tilde{\mathbf{A}}^H(\mathbf{I} + \tilde{\mathbf{A}}\mathbf{\Sigma}_{\mathbf{x}}\tilde{\mathbf{A}}^H)^{-1}\tilde{\mathbf{A}}\mathbf{\Sigma}_{\mathbf{x}}$$

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:

$$\mathrm{tr}\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}} + \mathbf{\Sigma}_{\mathbf{x}}^{-1})^{-1}\} = \quad (40)$$

$$\mathrm{tr}\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\mathbf{\Sigma}_{\mathbf{x}}\} - \mathrm{tr}\{\tilde{\mathbf{A}}_i\mathbf{\Sigma}_{\mathbf{x}}\tilde{\mathbf{A}}^H(\mathbf{I} + \tilde{\mathbf{A}}\mathbf{\Sigma}_{\mathbf{x}}\tilde{\mathbf{A}}^H)^{-1}\tilde{\mathbf{A}}\mathbf{\Sigma}_{\mathbf{x}}\tilde{\mathbf{A}}_i^H\} \quad (41)$$

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 $\mathbf{\Psi} = \sqrt{\mathbf{\Sigma}_{\mathbf{x}}}\tilde{\mathbf{A}}^H$ and $\mathbf{\Phi} = \tilde{\mathbf{A}}_i\mathbf{\Sigma}_{\mathbf{x}}\tilde{\mathbf{A}}^H$. 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.

□

Upper Bound on Performance of CSBS

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:

$$\mathrm{tr}\{\tilde{\mathbf{A}}_i^H \tilde{\mathbf{A}}_i (\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \boldsymbol{\Sigma}_{\mathbf{x}}^{-1})^{-1}\} \leq \mathrm{tr}\{\tilde{\mathbf{A}} i^H \tilde{\mathbf{A}}_i \boldsymbol{\Sigma}_{\mathbf{x}}\} \tag{42}$$

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

$$\gamma^2 \geq \max_{1 \leq i \leq p} \left(\mathrm{tr}\{\tilde{\mathbf{A}}_i^H \tilde{\mathbf{A}}_i (\mathbf{L}^T \mathbf{L})^{-1}\} \right) \tag{43}$$

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

$$\lambda \geq \max_{1 \leq i \leq p} \left(\mathrm{tr}\{\mathbf{A}_i^H \mathbf{A}_i (\mathbf{L}^T \mathbf{L})^{-1}\} \right) \tag{44}$$

Upper Bound on Performance of CSBS

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 by

$$\frac{(d+1)p - n + \mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\boldsymbol{\Sigma}_{\mathbf{e}}\}}{p - n + \mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\boldsymbol{\Sigma}_{\mathbf{e}}\}} \mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\} \quad (45)$$

provided that $(\mathrm{C}.1)$ holds.

Proof.

The proof is by contradiction. Assume that the bound on SSE after removing one sensor (stated in (37)) for all sensor locations $i \in \Gamma$ is larger than the bound (45) 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 upperbounded by (45). And since CSBS picks the best choice (in SSE sense) of sensor locations, it will choose the one with index i^*.

Upper Bound on Performance of CSBS

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

$$\frac{d+1-\mathrm{tr}\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\}}{1-\mathrm{tr}\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\}}\mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\} > \frac{(d+1)p-n+\mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\}}{p-n+\mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\}}\mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\} \tag{46}$$

In the inequality above, the common term $\mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\}$ is positive since it is a SSE. The numerator and denominator of the left hand side are both positive because of $(\mathrm{C}.1)$. Lemma **??** (stated and proved in Appendix **??**) proves that the denominator of the right hand side is positive under $(\mathrm{C}.1)$ which also implies that the numerator is positive.

Upper Bound on Performance of CSBS

Therefore, rearranging (46) and eliminating the common term, we arrive at:

$$(p - n + \mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\})(d + 1 - \mathrm{tr}\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\}) >$$
$$(1 - \mathrm{tr}\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\})((d+1)p - n + \mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\}) \quad (47)$$

Since (47) holds for all $1 \leq i \leq p$, its summation over i should hold as well:

$$(p - n + \mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\}) \sum_{i=1}^{p}(d + 1 - \mathrm{tr}\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\}) >$$
$$((d+1)p - n + \mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\}) \sum_{i=1}^{p}(1 - \mathrm{tr}\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\}) \quad (48)$$

Upper Bound on Performance of CSBS

But

$$\sum_{i=1}^{p}(\mathrm{tr}\{\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\boldsymbol{\Sigma}_{\mathbf{e}}\}) = \mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\sum_{i=1}^{p}\tilde{\mathbf{A}}_i^H\tilde{\mathbf{A}}_i\} = \mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\tilde{\mathbf{A}}^H\tilde{\mathbf{A}}\} \quad (49)$$

and

$$\mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}})\} = \mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}} + \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{x}}^{-1})\} \quad (50)$$

$$= \mathrm{tr}\{(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}} + \boldsymbol{\Sigma}_{\mathbf{x}}^{-1})^{-1}(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}} + \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{x}}^{-1})\} \quad (51)$$

$$= n - \mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{e}}\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\} \quad (52)$$

Substituting (49) and (52) in (48) gives:

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

which is a contradiction. Hence, there exists at least one sensor location i^* that violates (47). As explained in the beginning of the proof, this establishes the claim.

Conclusion

- We have introduced a statistical framework based on MAP estimation for sensor array imaging that allows us to analyze the impact of the placement of array elements on reconstructed images.
- We have considered several widely used image model constraints and optimality criteria.
- To perform the search over the space of candidate sensor locations to determine the conguration that optimizes the desired statistical optimality criterion over all candidates efficiently, we have introduced a computationally efficient algorithm and analyzed its computational complexity.
- Since the search is suboptimal, we have derived an upper bound on the performance of the suboptimal algorithm
- Through numerical experiments, we have verified the utility of this approach in determining a near-optimal sensor configuration.