Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Things Arising from Electrocardiographic Imaging: Toward a theory of partial inverse problems

Fred Greensite

Abstract— **The task of Electrocardiographic Imaging is an ill-posed inverse problem, requiring regularization. However, it has special features, firstly because it is a "non-stationary" inverse problem, and secondly because the inherent dynamical variety (e.g., epicardial breakthroughs, arrhythmias, ischemic changes) may preclude a fruitful nontivial process model. Importantly, its structure places it in the category of "partial inverse problems" - a theory that arises from this setting. Surprising features of the resulting regularization methodology include the ability to fashion nontrivial regularization matrices in part (and sometimes entirely) from the data. There is evidence that these theoretical results can have significant practical benefits.**

I. INTRODUCTION

It has long been recognized that Electrocardiographic Imaging is an ill-posed inverse problem, and that regularization is required. It is also a "non-stationary" inverse problem [7], i.e., a time series of distinct inverse problems. The latter realization suggests imposition of a Kalman smoother format, particularly as individual regularization of each constituent inverse problem independently leads to undesirable noise in the solution time series. Unfortunately, there is no evident nontrivial process model for this problem (e.g., depolarization on the epicardium proceeds from numerous epicardial breakthroughs "appearing out of nowhere", with markedly varying local wavefront propagation depending on the proximity to a breakthough, local relative curvature of depolarization wavefront and epicardial surface, ischemia, arrhythmias, etc.). Random walk Kalman (the default) simply corresponds to incorporation of a temporal second-order Tikhonov regularization operator [9]. Other schemes that address the time domain, such as Joint Regularization [9], have substantially *ad hoc* features (noted in [3]).

However, a re-evaluation of the problem in light of the structural differences of the unknown compared to that of an "ordinary" inverse problem, leads to a distinct approach, which forms the basis of a more general theory of "partial inverse problems". Application of elementary consequences of this theory appears to lead to significant improvements in Electrocardiographic Imaging solution quality [8], but we will here concentrate on the general theory - and its favored nature in principle, rather than in specific applications.

II. PARTIAL INVERSE PROBLEMS

In order to make the generalization to higher dimensions obvious and simplify derivations, we use tensor indices notation (and the Einstein summation convention).

F. Greensite is with the Department of Radiological Sciences, University of California, Orange, CA USA fred.greensite@uci.edu

A linear ordinary inverse problem, discretized for numerical analysis, and with an additive noise model, takes the form

$$Y^\alpha = F^\alpha_{\ \mu} X^\mu + N^\alpha, \tag{1}$$

where $Y^\alpha, X^\mu, N^\alpha$ are data, source, and noise, respectively, and $F^\alpha_{\ \mu}$ represents the transfer matrix. From a Bayesian standpoint, Y^α, X^μ, and N^α represent random vectors. Assuming these to be zero mean jointly Gaussian, with independent signal and noise, the solution is given by the autocovariance of random m-vector X^μ, since the posterior distribution (and resulting formulae for the minimum-mean-square-error estimate and it's associated covariance) then follow from a (presumed given) noise model and data realization of Y^α. The solution set $\mathcal{V}_m$ of zero mean Gaussian random m-vectors (defined by their $m \times m$ positive definite autocovariance matrices) is a homogeneous space, since the orbit of any member of this set under the action of the general linear group encompasses $\mathcal{V}_m$. Accordingly, cross-covariance matrices between any two members are defined, so that $\mathcal{V}_m$ is a vector space (i.e., given $(X_1)^\mu, (X_2)^\mu \in \mathcal{V}_m$, and scalars a, b, we have $(aX_1)^\mu + (bX_2)^\mu \in \mathcal{V}_m$ since the latter has zero mean, and its autocovariance can be supplied).

The above directly generalizes to the case of a discretized linear inverse problem having two (or more) variables in the unknown (and an additive noise model). In the two variable setting, (1) is replaced by

$$Y^{\alpha\beta} = F^{\alpha\beta}_{\ \ \mu\nu} X^{\mu\nu} + N^{\alpha\beta} = \sum_i (A_i)^\alpha_{\ \mu} (B_i)_\nu^{\ \beta} X^{\mu\nu} + N^{\alpha\beta}, \tag{2}$$

where N represents additive noise (we reserve lower case Roman letter subscripts for distinguishing between different tensors, and lower case Greek letter subscripts and superscripts for tensor component indices). The second equality in (2) follows from the ability to express any fourth rank tensor as a linear combination of tensor products of rank 2 tensors. The unknown is now a random second rank tensor $X^{\mu\nu}$ with indices μ, ν running from 1 to m and 1 to n, respectively. So, for a discretized partial inverse problem in two variables, the unknown is associated with a member of the tensor product of spaces $\mathcal{V}_m \otimes \mathcal{V}_n$. Each member of $\mathcal{V}_m \otimes \mathcal{V}_n$ is a finite linear combination of tensor products of members of $\mathcal{V}_m$ with members of $\mathcal{V}_n$. Thus, we can write $X^{\mu\nu} = \sum_i (s_i)^\mu (t_i)^\nu$, with $(s_i)^\mu \in \mathcal{V}_m$ and $(t_i)^\nu \in \mathcal{V}_n$. We assume that the components of members of $\mathcal{V}_m$ are jointly Gaussian with the components of members of $\mathcal{V}_n$. Since the members of $\mathcal{V}_m \otimes \mathcal{V}_n$ have zero mean, it follows that for $s^\mu \in \mathcal{V}_m$ and $t^\nu \in \mathcal{V}_n$, $\mathcal{E}[s^\mu t^\nu] = 0$, so that s^μ

is independent of t^ν. If we have no constraints regarding the entanglement of the two variables, all contractions of members of $\mathcal{V}_m \otimes \mathcal{V}_n$ with respect to their second index must lead to random vectors that are proportional to each other (due to "insufficient knowledge" [6], manipulations with respect to the second variable cannot influence utilized statistics related to the first variable). That is, there is an $s^\mu \in \mathcal{V}_m$ such that for any random n-covector $R_\nu \in \mathcal{V}_n^*$,

$$R_\nu X^{\mu\nu} = \sum_i \left(R_\nu (t_i)^\nu\right)(s_i)^\mu = a s^\mu,$$

where a is a scalar dependent on R_ν. This is only possible if either $(t_i)^\nu \propto (t_j)^\nu$ for all i, j, or $(s_i)^\mu \propto (s_j)^\mu$ for all i, j. In either case, this implies the existence of $s^\mu \in \mathcal{V}_m$ and $t^\nu \in \mathcal{V}_n$ such that $X^{\mu\nu} = s^\mu t^\nu$. Thus,

$$\begin{aligned}
&\mathcal{E}[X^{\gamma\sigma}X_{\gamma\sigma}]\mathcal{E}[X^{\mu\nu}X_{\xi\eta}] \\
&\quad = \mathcal{E}[s^\gamma t^\sigma s_\gamma t_\sigma]\mathcal{E}[s^\mu t^\nu s_\xi t_\eta] \\
&\quad = (\mathcal{E}[s^\gamma s_\gamma]\mathcal{E}[t^\sigma t_\sigma])(\mathcal{E}[s^\mu s_\xi]\mathcal{E}[t^\nu t_\eta]) \\
&\quad = (\mathcal{E}[s^\gamma s_\gamma]\mathcal{E}[t^\nu t_\eta])(\mathcal{E}[s^\mu s_\xi]\mathcal{E}[t^\sigma t_\sigma]) \\
&\quad = \mathcal{E}[X^{\gamma\nu}X_{\gamma\eta}]\mathcal{E}[X^{\mu\sigma}X_{\xi\sigma}], \qquad (3)
\end{aligned}$$

where the second and fourth equalities above follow from the previously noted independence of members of $\mathcal{V}_m$ with respect to members of $\mathcal{V}_n$. Equation (3) has two important immediate consequences.

1) The autocovariance of $X^{\mu\nu}$ is

$$\mathcal{E}[X^{\mu\nu}X_{\xi\eta}] = \frac{\mathcal{E}[X^{\mu\sigma}X_{\xi\sigma}]\mathcal{E}[X^{\gamma\nu}X_{\gamma\eta}]}{\mathcal{E}[X^{\omega\rho}X_{\omega\rho}]}, \qquad (4)$$

2) Non-random tensors effecting a contraction of an autocovariance simply act as scale factors,

$$\begin{aligned}
&\mathcal{E}[P_\lambda{}^\xi Q^\lambda{}_\mu X^{\mu\nu}X_{\xi\eta}] \\
&\quad = \left(P_\lambda{}^\xi Q^\lambda{}_\mu \frac{\mathcal{E}[X^{\mu\sigma}X_{\xi\sigma}]}{\mathcal{E}[X^{\omega\rho}X_{\omega\rho}]}\right)\mathcal{E}[X^{\gamma\nu}X_{\gamma\eta}]. \qquad (5)
\end{aligned}$$

Note that the parenthetical expression on the right-hand-side above is a scalar. Also note that an analogous equation follows if the contraction is with respect to ν, η rather than μ, ξ.

Equations (4) and (5) apply to regularization of problems such as defined by (2).

So-called non-stationary inverse problems [7] of the form $y_i = Ax_i + e_i$, $i = 1, 2, \ldots, n$, can be written as $Y = AX + N$ where the i-th columns of matrices Y, X, N are y_i, x_i, e_i respectively. Expressing this as $Y^{\alpha\beta} = A^\alpha{}_\mu \delta^\beta_\nu X^{\mu\nu} + N^{\alpha\beta}$, we see that we are dealing with the simplest nontrivial example of (2), i.e., the case in which

$$F^{\alpha\beta}{}_{\mu\nu} = A^\alpha{}_\mu \delta^\beta_\nu. \qquad (6)$$

Setting $P' = Q = A$ in (5) leads to [2]

$$\frac{\mathcal{E}[X^{\gamma\nu}X_{\gamma\beta}]}{\mathcal{E}[X^{\omega\rho}X_{\omega\rho}]} = \frac{\mathcal{E}[Y^{\alpha\nu}Y_{\alpha\beta}] - \mathcal{E}[N^{\alpha\nu}N_{\alpha\beta}]}{\mathcal{E}[Y^{\omega\rho}Y_{\omega\rho}] - \mathcal{E}[N^{\omega\rho}N_{\omega\rho}]}. \qquad (7)$$

In this case, we now we see that the problem of estimating X can be treated by specifying the inverse of the regularization matrix as (4), where $\mathcal{E}[X^{\mu\sigma}X_{\xi\sigma}]$ is chosen (up to a scalar magnitude) to be the autocovariance of x_i (the inverse of the spatial regularization matrix), with the remaining term in (4) supplied by (7). In utilizing (7), $\mathcal{E}[Y^{\alpha\nu}Y_{\alpha\beta}]$ is taken to be the observed (data derived) realization $Y'Y$ (since $Y'Y = \mathcal{E}[Y'Y] + O(\epsilon^2)$, assuming $\mathcal{E}[N'N] = O(\epsilon^2)$). In actuality, a multiple regularization parameter approach is preferable [2].

More generally, (2) and (5) imply that

$$\begin{aligned}
&\mathcal{E}[Y^{\alpha\beta}Y_{\alpha\beta'}] \\
&= \left\{\sum_{i,j}\left((A_i)_\lambda{}^\xi (A_j)^\lambda{}_\mu \frac{\mathcal{E}[X^{\mu\sigma}X_{\xi\sigma}]}{\mathcal{E}[X^{\omega\rho}X_{\omega\rho}]}\right)(B_i)^\eta{}_{\beta'}(B_j)_\nu{}^\beta\right\} \\
&\qquad \cdot\mathcal{E}[X^{\gamma\nu}X_{\gamma\eta}] + \mathcal{E}[N^{\alpha\beta}N_{\alpha\beta'}]. \qquad (8)
\end{aligned}$$

If there is a k such that B_k is well conditioned, then the inverse of the term in braces on the right-hand-side above is stable, and thus $\mathcal{E}[X^{\gamma\nu}X_{\gamma\eta}]$ (i.e., the second factor in the numerator on the right-hand-side of (4)) can be stably estimated (given values for the scalars embodied by the large parenthetical expression inside the summation on the right-hand-side of (8)). If there is also a r such that A_r is well conditioned, then an expression analogous (8) can be used to estimate $\mathcal{E}[X^{\mu\sigma}X_{\xi\sigma}]$, i.e.,

$$\begin{aligned}
&\mathcal{E}[Y^{\alpha\beta}Y_{\alpha'\beta}] \\
&= \left\{\sum_{i,j}\left((B_i)^\lambda{}_\nu (B_j)_\lambda{}^\eta \frac{\mathcal{E}[X^{\gamma\nu}X_{\gamma\eta}]}{\mathcal{E}[X^{\omega\rho}X_{\omega\rho}]}\right)(A_i)^\alpha{}_\mu (A_j)_{\alpha'}{}^\xi\right\} \\
&\qquad \cdot\mathcal{E}[X^{\mu\sigma}X_{\xi\sigma}] + \mathcal{E}[N^{\alpha\beta}N_{\alpha'\beta}]. \qquad (9)
\end{aligned}$$

In this case, the entire regularization matrix can be estimated via a maximum likelihood calculation (an example will be provided in Section III). It is important to note that this stable approach, which we call the "Adaptive Isotropy Method" (AIM), is desirable only so long as the coupled equations are not subject to excessive noise.

If all the A_i and B_i are ill conditioned, one can regularize (8) and (9) to obtain the requisite regularizing functional. However, one has now lost the previous advantage in that a hyperprior must now be imposed.

Note that the simultaneous equations (8) and (9) are coupled nonlinearly via the scalars appearing as the expression inside the large parentheses inside the respective summations on their right-hand-sides. The proper way to deal with this nonlinearity is one of several implementation questions considered in the next section.

III. ALGORITHMIC FEATURES

Though a tensorial decomposition of the operator in (2) might indicate that only "one dimensional" algorithmic methodology is applicable, if any of the component simple tensors of the decomposition have a well conditioned factor, then Section II indicates that a quite different algorithm is desirable - one which has no analogue in the theory of ordinary inverse problems. However, along with this potentially

more accurate approach, come a number of related questions. To introduce these, we will deal with an operator having the following decomposition,

$$F^{\alpha\beta}{}_{\mu\nu} = A^{\alpha}{}_{\mu}\delta^{\beta}_{\nu} + \delta^{\alpha}_{\mu}B_{\nu}{}^{\beta}, \tag{10}$$

where A and B are ill-conditioned (treatment of the simpler case $F^{\alpha\beta}{}_{\mu\nu} = A^{\alpha}{}_{\mu}\delta^{\beta}_{\nu}$ is the subject of [2]). Equation (10) corresponds to problems of the form $Y = AX + XB + N$. We will take the "unknown" X to consist of a couple of closely spaced sharp peaks, on a subtle damped wave-like background. A, B will be taken to be Gaussian-shaped convolution matrices. However, similar results are also obtained with quite different choices of A, B.

The following phenomena will be observed:

1) Consistent with expectations, the new methodology is capable of greater accuracy than zero-order Tikhonov regularization, assuming computation of the regularization tensor via (8) and (9) is not overwhelmed by noise,
2) The Discrete Picard Condition [4] is violated,
3) It is necessary to use a GSVD-based algorithm,
4) Despite the apparent nonlinear coupling of the second kind equations (8), (9) arising in the regularization tensor computation, the problem can be effectively treated linearly in practice.

We will consider these in reverse order.

A. The problem of regularization tensor computation is linear in practice

The relevant equations for computation of the regularization matrix are (8) and (9). The scalars within the summation in each equation apparently couple the simultaneous equations nonlinearly. In (8) (respectively, (9)), these scalars are the Frobenius product of $(A_i)_{\lambda}{}^{\xi}(A_j)^{\lambda}{}_{\mu}$ (respectively, $(B_i)^{\lambda}{}_{\nu}(B_j)_{\lambda}{}^{\eta}$) with a unit trace version of $\mathcal{E}[X^{\mu\sigma}X_{\xi\sigma}]$ (respectively, $\mathcal{E}[X^{\gamma\nu}X_{\gamma\eta}]$). Rather than devising a scheme to directly address this nonlinear problem, one might instead choose the scalars as resulting from Frobenius products with unit trace versions of the identity matrix (instead of $\mathcal{E}[X^{\mu\sigma}X_{\xi\sigma}]$ or $\mathcal{E}[X^{\gamma\nu}X_{\gamma\eta}]$). Thus, in the context of (10), the scalars are supplied by the traces of $A'A$, A, $B'B$, and B. Figure 1 demonstrates that this choice performs quite well compared with the case where the scalars are computed from actual knowledge of the unknown. Other obvious choices substituting for $\mathcal{E}[X^{\mu\sigma}X_{\xi\sigma}]$ or $\mathcal{E}[X^{\gamma\nu}X_{\gamma\eta}]$ (such as $A'A$ or $B'B$, respectively) also perform well for computation of the scalars in question, and also typically lead to improvements over standard Tikhonov regularization. Even a very naive choice (e.g., where $\mathcal{E}[Y^{\mu\sigma}Y_{\xi\sigma}]$ substitutes for $\mathcal{E}[X^{\mu\nu}X_{\mu'\nu}]$) performs better than Tikhonov regularization in this particular example.

For the simpler problem $Y = AX + N$, it is known that the choice of the "correct" term for the analogous scalar does not give the solution of lowest relative error, when compared with a multiple regularization parameter approach [2]. Thus, it is possible one can do even better by some multiple regularization parameter selection scheme in the treatment of (2), (10). This is an open question.

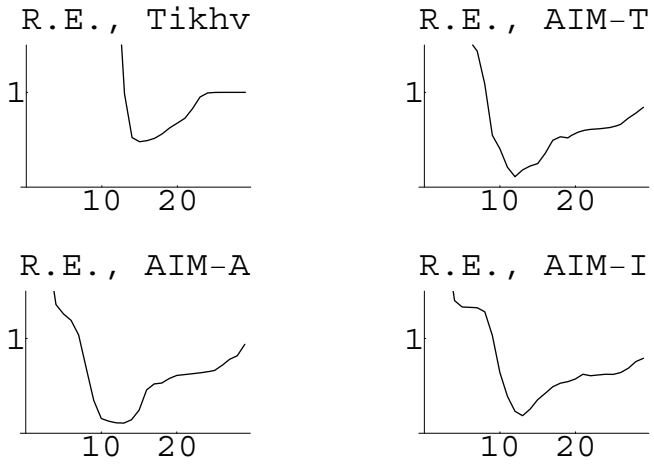

Fig. 1. Above are four graphs of relative error (R.E.) versus regularization parameter variation over 29 orders of magnitude. The unknown scalars arising in (8) and (9) are calculated using choices derived from A, B (AIM-A), or the identity matrices (AIM-I), as noted in the text. These give relative error curves comparable to that obtained with use of the true scalars as derived from knowledge of X (AIM-T). Each of these can perform better than zero-order Tikhonov (Tikhv).

B. The algorithm should employ the generalized singular value decomposition

As noted in [5], a GSVD analysis leads to the conclusion that the L-curve should be monotonic decreasing. However, this is not necessarily observed in practice if the GSVD components are not explicitly incorporated into the regularization algorithm. In fact, for treatment of (2), (10), rather curious solution instabilities can be noted as the regularization parameter is varied, when the algorithm does not explicitly utilize the GSVD components. Greater accuracy is achieved with a GSVD-based algorithm (we use a SVD-based algorithm to implement zero-order Tikhonov regularization as well).

C. The Discrete Picard Condition appears to be violated

Singular value expansion of a function will not be convergent if the Picard Condition is violated. This motivates the Discrete Picard Condition [4]. If the generalized singular values are arranged from smallest to largest (in conformance with usual convention), this condition posits that a particular regularization matrix will be useful only if the terminal portion of the sequence increases more steeply than the terminal portion of the sequence of corresponding Fourier components of the data. Interestingly, this condition is apparently not satisfied for AIM, as is clear from Figure 2.

This indicates the substantial difference between this regularization methodology and prior approaches.

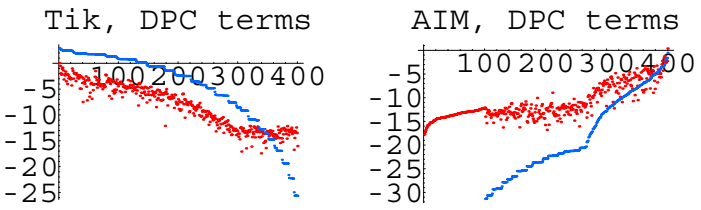

Fig. 2. *Left:* Superimposed graphs of logarithms of the operator singular values versus logarithms of the absolute values of the associated data Fourier coefficients, for the problem defined by (2) and (10) treated by zero-order Tikhonov. *Right:* Superimposed graphs of the logarithms of the generalized singular values versus logarithms of the absolute values of the associated Fourier coefficients, for application of AIM to this problem. With AIM, the Discrete Picard Condition is violated. On both the left and right above, the plots having greater local point variability pertain to the Fourier coefficients.

D. Increased accuracy of AIM versus Tikhonov regularization

Figure 3 shows results when A, B are Gaussian convolution matrices (having different standard deviations). AIM results in higher resolution and accuracy than Tikhonov regularization. This advantage decreases with increases in either noise power or variance of the convolving Gaussians. Some of this diminution in efficacy is presumably due to the effect of the noise on computation of the covariance components obtained from treatment of the simultaneous equations (8), (9).

IV. DISCUSSION

For an ill-posed problem, $F^{\alpha\beta}{}_{\mu\nu}$ in (2) models a compact operator. Thus, there cannot exist a particular k such that A_k and B_k are both well conditioned (unless the resulting term is effectively cancelled by other terms in the sum). However, if there exist k, j with $k \neq j$ such that A_k and B_j are well conditioned, then the entire regularizing functional can be derived from the solution of a pair of what are essentially discretized second kind Fredholm equations. Since the solution estimate is equivalent to identification of the regularizing functional and regularization parameter, the solution follows from treatment of this nominally well posed problem - without the need for prior imposition of what are typically *ad hoc* regularization matrices. Even if only one of the tensor factors on the right-hand-side of (2) is well conditioned (e.g., there is a j such that B_j is a well conditioned matrix), then a "well-posed" portion of the problem can be factored out - which has the effect of supplying "half" of the requisite regularization matrix in a stable fashion. It is easy to supply examples for which substantially improved solution estimates are thereby provided. The most elementary cases of these impact important classes of applied inverse problems, including much of what have been termed "non-stationary" inverse problems [7]. However, the implications are potentially more general.

While standard Tikhonov regularization can be thought of as an a example of a parametric emperical Bayesian method (the regularization parameter is computed from the data),

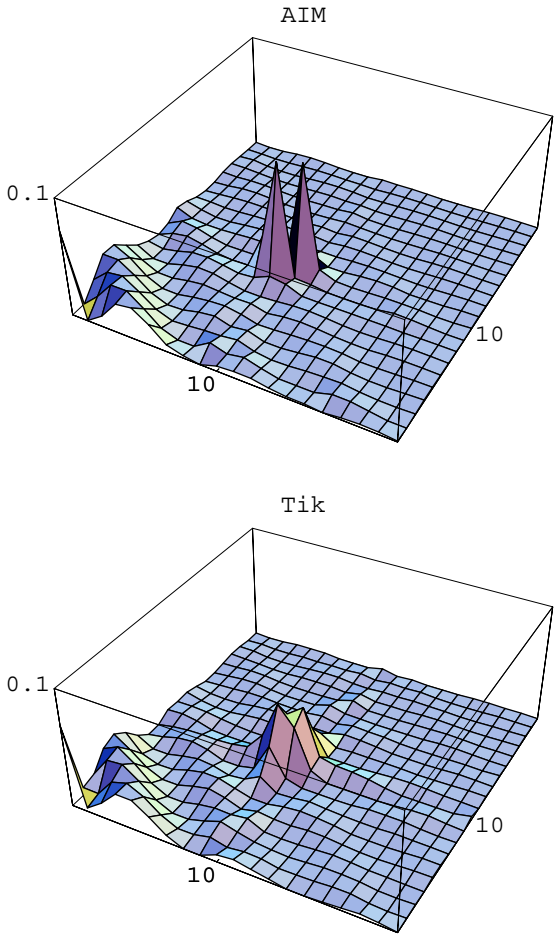

Fig. 3. In this example, the optimal AIM estimate is substantially superior to that obtained with optimal zero-order Tikhonov. This effect is greatest for cases where the unknown has dimension of at least moderate size ($X^{\mu}{}_{\nu}$ represents a 20×20 matrix in this example).

AIM seems to fit more in the category of a nonparametric emperical Bayesian method.

References

[1] R. Bellman, *Introduction to Matrix Analysis*, McGraw-Hill, New York, 1960, p. 230.

[2] F. Greensite, "Inverse problems with $I \otimes A$ Kronecker structure," *SIAM J. Matrix Anal. Appl.*, vol. 27, pp. 218-237, 2005.

[3] F. Greensite, "Partial inverse problems," *Inverse Problems*, vol. 22, pp. 461-479, 2006.

[4] P. C. Hansen, "The discrete Picard condition for discrete ill-posed problems," *BIT*, vol. 30, pp. 658-672, 1990.

[5] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Review*, vol. 34, 1992, pp 561-580.

[6] E. T. Jaynes, *Probability Theory*, Cambridge University Press, Cambridge, 2003, p. 388.

[7] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Springer, New York, 2005, p. 115.

[8] Z. Liu, C. Liu, and B. He, "Noninvasive reconstruction of three dimensional ventricular activation sequence from the inverse solution of distributed curent density," *IEEE Transactions on Medical Imaging*, in press.

[9] Y. Zhang, A. Ghodrati, and D. H. Brooks, "An analytical comparision of three spatio-temporal regularization methods for dynamic linear inverse problems in a common statistical framework,", *Inverse Problems*, vol. 21, pp. 357-382, 2005.