Basic Linear Inverse Method Theory - DRAFT NOTES

Peter P. Jones[1]

[1] *Centre for Complexity Science, University of Warwick, Coventry CV4 7AL, UK**

(Dated: 21st June 2012)

BASIC LINEAR INVERSE METHOD THEORY

Eigen-Decompositions

Drawing on the notes in [1, §7.7.10], assume we have a forward problem that is linear in form, $\mathbf{Ax} = \mathbf{b}$, and presume that $\mathbf{A}$ is square and not singular. If $\mathbf{A}$ is square it implies that $\mathbf{x}$ and $\mathbf{b}$ are the same length. Then, in a given basis, $\mathbf{A}$ has an eigen-decomposition (also called a spectral decomposition) $\text{Eig}(\mathbf{A}) = (\{\lambda_n\}, \{\varphi_n\})$ with eigenvalues λ_n and eigenvectors φ_n, such that we can describe the vector $\mathbf{x}$ with:

$$\lambda_n \langle \varphi_n, \mathbf{x} \rangle = b_n \tag{1}$$

$$\langle \varphi_n, \mathbf{x} \rangle = \frac{b_n}{\lambda_n} \tag{2}$$

$$x_n = \frac{b_n}{\lambda_n} = \frac{\langle \varphi_n, \mathbf{b} \rangle}{\lambda_n} \tag{3}$$

$$\Rightarrow \mathbf{x} = \sum_{n=1}^{N} \frac{\alpha_n}{\lambda_n} \varphi_n \qquad \text{where } \alpha_n := \langle \varphi_n, \mathbf{b} \rangle,\ \lambda_n \neq 0 \tag{4}$$

Here $\langle \cdot, \cdot \rangle$ is notation for the inner product, and N may be finite or infinite, depending on the space in question. Unless stated otherwise, from now on we will consider only N finite, to match with numerical discretisation. If we now generalise further by using $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ and $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_N)$ such that:

$$\mathbf{AX} = \mathbf{B}, \qquad \text{with } \mathbf{A}, \mathbf{X}, \mathbf{B} \in \mathbb{R}^{N \times N} \tag{5}$$

Then in a manner similar to (4) we can say:

$$\mathbf{x}_i = \sum_{n=1}^{N} \frac{\xi_n^{(i)}}{\lambda_n} \varphi_n \qquad \text{where } \xi_n^{(i)} := \langle \varphi_n, \mathbf{b}_i \rangle,\ \lambda_n \neq 0,\ \mathbf{b}_i \in \mathbf{B},,\ \mathbf{x}_i \in \mathbf{X},\ i = 1, 2, \ldots, N \tag{6}$$

The above equations (4) and (6) could be considered the core equations of inverse method theory. They assert that if $\mathbf{A}$ has an eigen-decomposition with non-zero eigenvalues, then for any $\mathbf{x}$ bounded in some domain X such that the operation $\mathbf{Ax}$ results in some definite $\mathbf{b}$, then given such a $\mathbf{b}$ it is possible to recover $\mathbf{x}$ directly. In reality, however, matters rarely work so neatly. In practice, the eigen-decomposition of $\mathbf{A}$ can result in eigenvalues that are too close to zero to obtain good solutions by computational methods (which is another way of saying that $\mathbf{A}$ doesn't have a useful inverse).

Moreover, let us for the sake of argument consider $\mathbf{b}$ to reside in a space $B = X$. It is possible for some operator $\mathbf{A}$ that even if $\mathbf{x}$ can vary continuously in the space X that the results $\mathbf{b}$ are nonetheless distributed unevenly in that space (so that there can be 'gaps' between different results for neighbouring input values of $\mathbf{x}$). Thus there is a set of values $\{\mathbf{b}'\} \in B$ that are not in the range of $\mathbf{A}$ given $\mathbf{x}$. If the data one receives from an experiment is $\mathbf{b}'$ then it is no longer clear that an $\mathbf{x} \in X$ could be recovered by direct inverse solution, or whether neighbouring values of $\mathbf{b}'$ would respect any notions of continuous variation of $\mathbf{x}$. Which is another way of saying that even if an answer could be recovered, it might not make a lot of sense within the chosen forward model of the phenomena represented by $\mathbf{A}$. Even if $\mathbf{A}$ is highly successful as a forward model in $\mathbf{Ax} = \mathbf{b}$ the above can still be the case.

There are essentially three tactics that can be deployed in such a situation. The first is to hope that physical constraints mean that the results $\mathbf{b}$ are sufficiently well separated in X that suitable ensemble averaging will remove noise. That is, for $\mathbf{B}_p = (\mathbf{b'}_1, \mathbf{b'}_2, \ldots, \mathbf{b'}_p)$, as $p \to M$ where M is suitably large, then $\langle \mathbf{b'}_i \rangle = \mathbf{b}$. The second tactic is to employ some method that presumes that (if results can be suitably ordered) then $\mathbf{b}_1 \preccurlyeq \mathbf{b'} \prec \mathbf{b}_2$ can be coerced to resolve in a sufficiently neat way such that $\mathbf{x}_1 \preccurlyeq \mathbf{x'} \prec \mathbf{x}_2$ is the case. With the exception of Bayesian methods, most inverse methods for working with real data are variations on this second tactic.

The third tactic is to use Bayesian methods, which take in some $\mathbf{B}_p = (\mathbf{b'}_1, \mathbf{b'}_2, \ldots, \mathbf{b'}_p)$ and then attempt to construct a probability density function (p.d.f.) for $\mathbf{b'}$ as the input to the method. The output is then a distribution on $\mathbf{x'}$ conditioned on the p.d.f. of $\mathbf{b'}$. This is a useful approach when it is unclear whether the true $\mathbf{b}$ are neatly separated in space B, or where the lack of evidence for suitable ordering prevents the use of the second tactic above. When the input data $\mathbf{b'}$ is known to have errors that conform to a Gaussian distribution and there is little or no knowledge about any prior distribution of $\mathbf{x}$ (so a flat prior is used) then it is known that Tikhonov-Regularised Least Squares methods return the maximum *a posteriori* (MAP) value that would be obtained from a Bayesian approach[2]. (A comprehensive discussion of Bayesian treatments is beyond the scope of these notes. Though the interested reader is referred to [3] for a pragmatic introduction, with [2] and [4] for more advanced coverage.)

The second tactic above requires that there be an appropriate constraint on $\mathbf{x'}$ that attempts to satisfy the desire for an ordering of potential solutions. Typically this might be to seek the minimum norm of $\mathbf{x'}$ achievable with some Least-Squares fit.

$$\min_{\mathbf{x'}} \|\mathbf{A}\mathbf{x'} - \mathbf{b'}\|^2 \qquad \text{such that } \min \|\mathbf{x'}\|^2 \tag{7}$$

The Least-Squares approach has the advantage that it doesn't require either that $\mathbf{A}^{-1}$ exist in a useful sense, nor that eigen-decomposition be undertaken. However, some algorithms used in implementations can require that the differential,

$$\frac{\partial}{\partial x'_i} \|\mathbf{A}\mathbf{x'} - \mathbf{b'}\|^2 < \infty, \quad \forall x'_i \in \mathbf{x} \tag{8}$$

(and possibly the second-order differential) exist and be sufficiently bounded in order to operate.

Singular Value Decomposition and the Generalised Inverse

Another way to talk about the way such linear inverse methods work in practice is to discuss the Singular Value Decomposition (SVD) of the matrix $\mathbf{A}$. The approach is similar to the discussion of eigen-decompositions earlier, but it reveals the link between such decompositions and the Least-Squares method more clearly. **The following account draws heavily from [3, Ch. 4]**.

It is known that every matrix has a singular value decomposition (SVD). Given an $m \times n$ matrix $\mathbf{A}$, it can be factored into the decomposition:

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}} \tag{9}$$

Where,

- $\mathbf{U}$ is an $m \times m$ orthonormal matrix, whose columns are basis vectors that span the data space $\mathbb{R}^m$.
- $\mathbf{V}$ is an $n \times n$ orthogonal matrix, whose columns span the model space $\mathbb{R}^n$.
- $\mathbf{S}$ is an $m \times n$ matrix with entries only on the diagonal, which are the non-negative singular values. (Non-negative implies some singular values may be zero.)

(In some numerical software packages $\mathbf{V}$ is also an orthonormal matrix, and the entries in $\mathbf{S}$ are therefore rescaled appropriately.) The singular values, $\{s\} \in \mathbf{S}$ are typically ordered with decreasing value, largest first to smallest last. If p values on the diagonal of $\mathbf{S}$ are non-zero, then $\mathbf{S}$ can be written in block form as,

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{10}$$

In which case $\mathbf{A}$ can be described by,

$$\mathbf{A} = \begin{bmatrix}\mathbf{U}_p, & \mathbf{U}_0\end{bmatrix} \begin{bmatrix}\mathbf{S}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} \begin{bmatrix}\mathbf{V}_p, & \mathbf{V}_0\end{bmatrix}^{\mathrm{T}} \tag{11}$$

Where the notation $\mathbf{U}_p$ represents the first p columns of $\mathbf{U}$, and $\mathbf{U}_0$ represents the remaining columns. Similarly for $\mathbf{V}$. Which implies there is a compact form,

$$\mathbf{A} = \mathbf{U}_p \mathbf{S}_p \mathbf{V}_p^{\mathrm{T}} \tag{12}$$

Therefore, for some $\mathbf{y}$ in the range of $\mathbf{A}$, we have,

$$\begin{aligned}\mathbf{y} &= \mathbf{A}\mathbf{x} \\ &= \mathbf{U}_p \mathbf{S}_p \mathbf{V}_p^{\mathrm{T}} \mathbf{x}\end{aligned} \tag{13}$$

The SVD can be used to compute the generalised inverse of $\mathbf{A}$, also known as the Moore-Penrose Pseudoinverse,

$$\mathbf{A}^{\dagger} = \mathbf{V}_p \mathbf{S}_p^{-1} \mathbf{U}_p^{\mathrm{T}} \tag{14}$$

Hence, given a data vector $\mathbf{b}$ we can obtain the pseudoinverse solution as,

$$\begin{aligned}\mathbf{x}_{\dagger} &= \mathbf{A}^{\dagger}\mathbf{b} \\ &= \mathbf{V}_p \mathbf{S}_p^{-1} \mathbf{U}_p^{\mathrm{T}} \mathbf{b}\end{aligned} \tag{15}$$

For any matrix the pseudoinverse $\mathbf{A}^{\dagger}$ always exists, hence $\mathbf{x}_{\dagger}$ always exists, and $\mathbf{x}_{\dagger}$ is also a (minimum length) least squares solution.

Referring once again to the approach in [3, Ch. 4], the results of using the pseudoinverse can be classified into four cases depending upon the respective forms of the data null space $\mathcal{N}(\mathbf{A}^{\mathrm{T}})$ and the model null space $\mathcal{N}(\mathbf{A})$.

Model Null $\mathcal{N}(\mathbf{A})$	**Data Null $\mathcal{N}(\mathbf{A}^{\mathrm{T}})$**	**$\mathbf{A}^{\dagger}$ Result**
Trivial $= \{0\}$	Trivial $= \{0\}$	$\mathbf{A}^{\dagger} = \mathbf{A}^{-1}$. Result is exact.
Non-trivial	Trivial $= \{0\}$	An exact data fit, but non-unique solution. The solution is also a minimum length Least Squares solution.
Trivial $= \{0\}$	Non-trivial	$\mathbf{x}_{\dagger}$ is the Least Squares solution such that if $\mathbf{A}\mathbf{x}_{\dagger} = \mathbf{b}'$ then $\mathbf{b}'$ is nearest to $\mathbf{b}$. If $\mathbf{b}$ is actually in Range($\mathbf{A}$) then $\mathbf{x}_{\dagger}$ is an exact solution. The solution is unique but cannot fit general data exactly.
Non-trivial	Non-trivial	$\mathbf{x}_{\dagger}$ minimizes both $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$ and $\|\mathbf{x}\|_2$ as a Least Squares solution, but it is not unique and cannot fit general data exactly.

In geometric terms, a non-trivial null space for an operator $\mathbf{T}$ implies that some of the dimensions in a vector $\mathbf{v}$ are effectively projected by the transformation $\mathbf{T}\mathbf{v}$ on to a lower dimensional subspace of the vector space, and the resulting loss of dimensionality (because some dimensions are now connected together) permits cancellation among coefficients of the basis vectors for certain values of $\mathbf{v} \neq \mathbf{0}$. This is one way to interpret the notion of the linear dependence of the columns of a linear operator. As a consequence, it is possible that scalings and rotations or reflections of $\mathbf{v}$ can also become solutions of $\mathbf{T}\mathbf{v} = \mathbf{0}$. Hence it is also possible that, for example, $\mathbf{T}\mathbf{w} = \mathbf{b}$ can exhibit similar non-uniqueness of its solutions $\mathbf{w}$, as if all that has really happened is a shifting of the origin with respect to which the transformation takes place. Hopefully, this explanation makes it easier to see why the second row of the table makes sense.

In order to explain the third row of the table, one needs to know about the normal equations for the Least Squares problem (see [3, p.16 and p.237] and [5]). From these, the Least Squares solution appears as $\mathbf{x}_{LS} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$, provided that the inverse of the matrix $(\mathbf{A}^T\mathbf{A})^{-1}$ exists. In this case, if $\mathbf{b}$ is in the null space of $\mathbf{A}^T$ then there is no corresponding non-trivial $\mathbf{x}$ as a solution, which is another of saying that $\mathbf{b}$ is not in Range($\mathbf{A}$). If this is the case, then the Least Squares solution $\mathbf{x}$ obtained will behave as a best minimum length approximation to the true solution, and the behaviours stated in the third row of the table then follow.

It can still be the case in practice that even when both null spaces are trivial that some of the matrix singular values may be so small as to behave on the scale of numerical truncation errors. It then becomes important to check whether these small singular values could be the source of instability in the inversion. The discrete Picard condition is one way to do this. If we write, with singular values $\{s_i\}_{i=1}^{p}$ in $\mathbf{S}$,

$$\begin{aligned}\mathbf{x}_\dagger &= \mathbf{V}_p\mathbf{S}_p^{-1}\mathbf{U}_p^{\mathrm{T}}\mathbf{b} \\ &= \sum_{i=1}^{p}\frac{\mathbf{U}_{\cdot,i}^{\mathrm{T}}\mathbf{b}}{s_i}\mathbf{V}_{\cdot,i}\end{aligned} \tag{16}$$

Then the discrete Picard condition asserts that if the $\mathbf{U}_{\cdot,i}^{\mathrm{T}}\mathbf{b}$ decay faster than the s_i as i increases toward p then any instability is not because of small singular values. However, if the (generalised) inversion fails to meet the discrete Picard condition, this motivates the use of methods like Truncated SVD and others form of regularisation. The presence of non-trivial null spaces is analogous to an automatic failure of the discrete Picard condition, because the coupling of dimensions by the linear transformation means that some b_i is likely to exceed its respective s_i in order for the resulting vector to cancel terms to become $\mathbf{0}$.

In Truncated SVD, any singular values that are too small and thought to be a source of instability are set to zero, and the inversion tested for some rank $q < p$. In an alternative form, the small singular values are boosted by suitable amounts to values at which the discrete Picard condition holds. Both of these approaches are effectively subsets of Tikhonov-Regularised Least Squares which we will discuss briefly below, but they can have independent value as diagnostic tools.

Least Squares Inversion and Regularisation

For the specific problem of inverting from observed data for physical systems, since we expect to start with given data sets we do not have to worry about the (mathematical) existence of solutions to the forward problem. Rather we are concerned with the conditions for well-posed problems posed by Hadamard. A problem is well-posed if (quoting from [6, Ch. 2]):

1. For all admissible data, a solution exists;
2. For all admissible data, the solution is unique;
3. The solution depends continuously on the data.

It is often the case that inverse problems violate one or more of these conditions, in which case they are ill-posed. Violation of the last condition can mean that arbitrarily small changes to the data result in large changes to any obtained solution. Thus, noise in the data, from whatever source, can cause problems. However, merely the construction of the inverse problem itself can involve differentiations that greatly amplify sensitivity (see [6, §1.1] for an instructive example).

Considering the problem of retrieving the kernel of the SCE, an immediate simplification is to ensure that the forward problem can be structured as a linear operation with respect to the kernel, because inversion of linear problems is typically more straightforward. However, it is not always the case that for a given forward problem $\mathbf{Ax} = \mathbf{b}$ that an inverse operator $\mathbf{A}^{-1}$ can be readily constructed. In addition, as shown earlier, though the pseudoinverse $\mathbf{A}^\dagger$ can be constructed, $\mathbf{A}$ or its transpose might have a non-trivial null space. In this case values for the parameters are usually sought from using a Least Squares approach, which seeks to minimise the squared norm of the residual error.

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2^2 \tag{17}$$

This is essentially finding the minimal point of a convex (concave-up) surface, and the techniques to do so often rely on differentiation. Unfortunately, simply minimising the errors of fit resulting from estimations, $\mathbf{x}_E$, of the solution $\mathbf{x}_S$, does not guarantee that a unique $\mathbf{x}_S$ will be obtained. It is possible for an infinite number of solutions to satisfy a set of linear equations in the absence of adequate constraints on the form of the solution.

In practice, if there is a lack of such constraints, the 'simplest' possible solution is sought, where simple usually means some combination (or possibly a trade-off) of small solution norm and smoothness (i.e. some derivative of potential solutions $\mathbf{x}_E$, perhaps the second-order derivative or higher, might be required to be almost constant). A standard technique for implementing this approach is to use Tikhonov regularisation, which takes the form:

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda_r \|\mathbf{L}\mathbf{x}\|_2^2 \tag{18}$$

Where $\mathbf{L}$ is often a linear difference operator employed to approximate numerically some order of differentiation, and λ_r is the regularisation parameter. If $\mathbf{L}$ is the identity matrix, $\mathbf{I}$, then this is known as zeroth-order Tikhonov regularisation. We also note at this point that even this extra machinery cannot fully guarantee that a unique solution will be obtained. It obtains a minimum length Least Squares solution with some degree of smoothness, while implicitly controlling (through the tuning of the value of λ_r) potential instabilities, but there might still be a number of solutions with the same minimum length.

We can expand equation (18) to obtain the form usually deployed in implementations for Least Squares minimisation of ill-posed linear systems.

$$\min_{\mathbf{x}} \quad (\mathbf{A}\mathbf{x} - \mathbf{b})^T(\mathbf{A}\mathbf{x} - \mathbf{b}) + \lambda_r(\mathbf{L}\mathbf{x})^T(\mathbf{L}\mathbf{x}) \tag{19}$$

$$\min_{\mathbf{x}} \quad \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{A}^T\mathbf{b} - \mathbf{b}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{b} + \lambda_r\mathbf{x}^T\mathbf{L}^T\mathbf{L}\mathbf{x} \tag{20}$$

For the ideal solution, it ought to be the case that $\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{A}^T\mathbf{b} - \mathbf{b}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{b} = 0$. Hence,

$$\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{A}^T\mathbf{b} = \mathbf{b}^T\mathbf{A}\mathbf{x} - \mathbf{b}^T\mathbf{b} \tag{21}$$

In addition, since $\mathbf{A}\mathbf{x} = \mathbf{b}$ by definition, then the RHS of this equation should also be equal to zero, and so should the LHS. This implies that the solution can be found from $\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{A}^T\mathbf{b} = 0$. Switching notation, by the linearity of the inner product we can write $\langle \mathbf{x}^T, \mathbf{A}^T\mathbf{A}\mathbf{x} - \mathbf{A}^T\mathbf{b} \rangle = 0$. The criterion for orthogonality then implies that if $\mathbf{x} \neq \mathbf{0}$, then it must be the case that the non-trivial solution for $\mathbf{x}$ can be obtained from the **normal equation** for the system.

$$\mathbf{A}^T\mathbf{A}\mathbf{x} - \mathbf{A}^T\mathbf{b} = \mathbf{0} \tag{22}$$

In practice, applying Least Squares minimisation to the LHS of (22) should then find a good approximation to the correct solution. If regularisation is required, then putting this back in again gives $(\mathbf{A}^T\mathbf{A} + \lambda_r\mathbf{L}^T\mathbf{L})\mathbf{x} - \mathbf{A}^T\mathbf{b}$. If $\mathbf{L} = \mathbf{I}$ then we have the form,

$$\min_{\mathbf{x}} \quad \|(\mathbf{A}^T\mathbf{A} + \lambda_r\mathbf{I})\mathbf{x} - \mathbf{A}^T\mathbf{b}\|_2^2 \tag{23}$$

If, for the sake of exposition, we consider the matrix $\mathbf{A}$ to be a member of $\mathbb{R}^{2\times 2}$, use its SVD, and then ask what the term $(\mathbf{A}^T\mathbf{A} + \lambda_r\mathbf{I})$ looks like, we get,

$$\begin{pmatrix} s_1^2v_1^2 + s_2^2v_2^2 + \lambda_r & s_1^2v_1v_3 + s_2^2v_2v_4 \\ s_1^2v_1v_3 + s_2^2v_2v_4 & s_1^2v_3^2 + s_2^2v_4^2 + \lambda_r \end{pmatrix} \tag{24}$$

Hence, if it turns out that, say, s_2^2 has a magnitude below the threshold for the noise in observations b_i or below the threshold for numerical representation in some numerical calculation, then λ_r compensates for this. In doing so it shifts the minimisation of the residuals to cope with the case when $\mathbf{b}$ is outside the range of the operator $\mathbf{A}$, thus stabilising the process. An appropriately chosen value of λ_r then ensures that the approximate solutions for $\mathbf{x}$ stay closer to the solutions that match the observations than might otherwise be possible.

In practice, there is no requirement that the regularisation be structured as $\lambda_r \|\mathbf{Lx}\|_2^2$. What is required is that the regularisation impose appropriate constraints on the structure of the solution(s) as the minimisation process proceeds. Modern mathematical software packages can often now support the use of a custom-made function for the regularisation, $\lambda_r f(\mathbf{x})$, when this is appropriate to a specific real inverse problem.

For further details about use of the Regularised Least Squares method, we refer the reader to the excellent pragmatic introduction in [3], the practical and theoretical treatment in [7, Ch. 3], and the thorough theoretical treatment in [6].

* p.p.jones@warwick.ac.uk

[1] *Oxford Users' Guide to Mathematics*, edited by E. Zeidler, with W. Hackbusch, and H. R. Schwarz (Oxford University Press, 2004).
[2] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Applied Mathematical Sciences No. 160 (Springer, 2005).
[3] R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter Estimation and Inverse Problems* (Elsevier Academic Press, 2005).
[4] A. Tarantola, *Inverse Problem Theory* (Society for Industrial and Applied Mathematics, 2005).
[5] D. C. Lay, *Linear Algebra and its Applications*, 4th ed. (Addison-Wesley, 2012).
[6] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Methods*, Mathematics and its Applications, Vol. 375 (Springer, 1996).
[7] A. Doicu, T. Trautmann, and F. Schreier, *Numerical Regularization for Atmospheric Inverse Problems* (Springer, 2010).