Parameter Identification by Iterative Constrained Regularization

Journal of Physics: Conference Series **657** (2015) 012002 doi:10.1088/1742-6596/657/1/012002

Parameter Identification by Iterative Constrained Regularization

Fabiana Zama

Department of Mathematics, P.zza Porta S. Donato 5, Bologna, ITALY

E-mail: fabiana.zama@unibo.it

Abstract. Parameter identification from noisy data is an ill-posed inverse problem and data noise leads to poor solutions. Regularization methods are necessary to obtain stable solutions. In this paper we introduce the regularization by means of an iteratively weighted constraint and define an algorithm to compute the weights and solve the constrained problems using as prior information the given measurements. Although this approach is general, in the present work we prove the convergence in the case of least squares data fit with ℓ_2 regularization term. The data reported in the numerical experiments prove the efficiency and good quality of the results.

1. Introduction

Parameter identification is a very important topic in applied sciences and engineering. The modeling of biological or chemical systems requires often to estimate the parameters of differential models from noisy measurements (see for example [1],[2]). Parameter identification is an optimization problem where the objective function, usually non linear, represents the fit to the observed data. The solution process requires to repeatedly solve the differential problem, called *forward problem* or *state equation*. The majority of these identification problems is ill-posed, i.e. the parameters do not depend continuously on the data, which cannot be measured exactly but are affected by noise. Therefore regularization methods have to be used in order to obtain stable solutions. Aim of this work is to present an algorithm, in a general framework, to compute a regularized solution of the parameter identification problem by adding a smoothing constraint with automatic computation of the regularization parameter. The idea of solving identification problems by a sequence of constrained optimization problems is widely present in the literature (see for example [3, 4]). Although the scheme is general, in this paper we analyze the details of choosing the ℓ_2 norm for both the objective function and the penalty term, obtaining therefore the Tikhonov regularization. The application of this regularization method to parameter identification problems has been extensively investigated in the literature [5, 6]. The main difficulty of the Tikhonov regularization is the correct computation of the regularization parameter. In our approach we define an iterative procedure to adjust the weight of the smoothing constraint, without assuming any prior knowledge of the noise level. We define the Constrained Least Squares Identification algorithm (`CLSId`), by extending the principle already used in image deblurring and denoising problems [7, 8]. The regularization method for generic objective and penalty functions is discussed in section 2, while in section 3 we report the algorithmic features for the ℓ_2 norm. The iterative procedure to compute the regularization

weight is analysed is section 4. In the numerical experiments, reported in section 5, we show the results of two parameter identification test problems.

2. The Parameter Identification Algorithm

We consider the discrete finite dimensional problem of estimating the parameter $\mathbf{q} \in \mathbb{R}^K$ of a differential model (state equation) $c(\mathbf{u}, \mathbf{q}) = 0$, usually an ordinary or partial differential equation. We assume that the measurements are obtained by mapping the solution of the state equation $\mathbf{u}(\mathbf{q})$ (state variable) at some measurement points ($\mathbf{y} = F(\mathbf{u}(\mathbf{q}))$) and adding the noise, represented by a discrete normalized random distribution with zero mean and assigned variance δ:

$$\mathbf{y}^\delta = \mathbf{y} + \delta\boldsymbol{\eta} \tag{1}$$

where $\mathbf{y}, \boldsymbol{\eta} \in \mathbb{R}^N$ and $\|\boldsymbol{\eta}\| = 1$. The parameter identification is obtained by minimizing a cost function $J(\mathbf{q})$ representing the data fit of the model measurements $F(\mathbf{u}(\mathbf{q}))$ to the noisy data $\mathbf{y}^\delta$. The majority of such identification problems is ill-posed hence regularization methods have to be used in order to obtain stable solutions. In the present work we introduce a smoothing constraint $\mathcal{R}(\mathbf{q})$ and define the regularized solution $\mathbf{q}_{reg}$ as the solution of the constrained optimization problem:

$$min_{\mathbf{q}} J(\mathbf{q}), \quad \text{s.t.} \quad \mathcal{R}(\mathbf{q}) \leq \gamma \tag{2}$$

where $\gamma > 0$ defines the prescribed smoothness level required by the solution. Our aim is to define a general scheme to compute a suitable value of the regularization constraint γ and to efficiently compute the regularized solution $\mathbf{q}_{reg}$. Assuming $\mathcal{R}(\mathbf{q})$ and $J(\mathbf{q})$ are continuously differentiable and convex, problem (2) can be formulated in the equivalent Lagrangian dual form as:

$$\max_{\lambda} \Phi(\lambda), \quad \text{where} \quad \Phi(\lambda) \equiv \min_{\mathbf{q}} \mathcal{L}(\mathbf{q}, \lambda) \tag{3}$$

where $\mathcal{L}(\mathbf{q}, \lambda) \equiv J(\mathbf{q}) + \lambda\,(\mathcal{R}(\mathbf{q}) - \gamma)$.

After imposing the first order conditions $\nabla_\lambda \mathcal{L}(\mathbf{q}, \lambda) = 0$ we can define the solution $\hat{\lambda}$ of the dual problem (3) as:

$$\text{find} \quad \hat{\lambda} \quad \text{s.t.} \quad \mathcal{R}(\hat{\mathbf{q}}) - \gamma = 0, \tag{4}$$

where $\hat{\mathbf{q}}$ is the solution of the unconstrained problem:

$$\min_{\mathbf{q}} J(\mathbf{q}) + \lambda(\mathcal{R}(\mathbf{q}) - \gamma) \tag{5}$$

with $\lambda = \hat{\lambda}$. Assuming that $\mathcal{R}(\mathbf{q}(\lambda))$ is a continuously differentiable function such that

(i) $\mathcal{R}(\mathbf{q}(\lambda))$ is strictly decreasing in the interval $[\lambda_-, \lambda_+]$ $(0 \leq \lambda_- < \hat{\lambda} < \lambda_+)$.
(ii) $G(\lambda_-) > 0, \quad G(\lambda_+) < 0, \quad \gamma > 0.$

where $\mathbf{q}(\lambda) = argmin_{\mathbf{q}} \mathcal{L}(\mathbf{q}, \lambda)$ and $G(\lambda) \equiv \mathcal{R}(\mathbf{q}(\lambda)) - \gamma$, using proposition 1 in [7] we can define an hybrid bisection-secant method to compute the sequence $\{\lambda_k\}$, $k = 0, 1, \ldots$, that converges to the root $\hat{\lambda}$ of $G(\lambda)$, $\forall \lambda_0 > 0$ such that $\mathcal{R}(\mathbf{q}(\lambda_0)) - \gamma < 0$. Hence for each assigned value $\gamma > 0$, assuming that λ_0 satisfies $\mathcal{R}(\mathbf{q}(\lambda_0)) < \gamma$ and that $\mathcal{R}$ is strictly decreasing, we can compute the sequence $\{\lambda_k, \mathbf{q}_k\}$ converging to the solution of the dual problem $\hat{\lambda}$ and to the solution of the constrained problem (2) $\mathbf{q}_{reg} \equiv \hat{\mathbf{q}}$. The values $\{\lambda_k, \mathbf{q}_k\}$ are computed by means of (4), (5) as follows:

$$\lambda_k = \lambda_{k-1} + \texttt{bisection} - \texttt{secant step} \tag{6}$$

$$\mathbf{q}(\lambda_k) = argmin_{\mathbf{q}} \{J(\mathbf{q}) + \lambda_k(\mathcal{R}(\mathbf{q}) - \gamma)\} \tag{7}$$

Since the unconstrained minimization problem (5) (with $\lambda = \lambda_k$) is solved at each step of the iterations (7), then it is of great importance to use accurate and efficient methods.

3. Least Squares Regularization

The quality of the problem solution depends strongly on the specific choice of the objective function $J(\mathbf{q})$ as well as the penalty term $\mathcal{R}(\mathbf{q})$. Assuming to have Gaussian white noise on the data, the fit function J can be suitably defined by the least squares distance. In this case, in (5), we have:

$$\mathbf{q}(\lambda_k) = argmin_{\mathbf{q}} \left\{ \|F(\mathbf{q}) - \mathbf{y}^{\delta}\|_2^2 + \lambda_k(\mathcal{R}(\mathbf{q}) - \gamma) \right\} \tag{8}$$

A fast and simple way to solve this non linear problem is by the Gauss Newton method. Defining the Jacobian matrix $J_F(\mathbf{q}) = \partial F_i / \partial \mathbf{q}_j$, we compute the solution of (8) as $\mathbf{q}(\lambda_k) = \mathbf{p}^{(m)}$ with $\mathbf{p}^{(0)} = \mathbf{q}(\lambda_{k-1})$ and

$$\mathbf{p}^{(m)} = argmin_{\mathbf{q}} \left\{ \|J_F(\mathbf{p}^{(m-1)})(\mathbf{q} - \mathbf{p}^{(m-1)}) - \mathbf{r}^{(m-1)}\| + \lambda_k \left(\mathcal{R}(\mathbf{q}) - \gamma\right) \right\}, \quad m = 1, 2, \dots \tag{9}$$

where $\mathbf{r}^{(m)} = \mathbf{y}^{\delta} - F(\mathbf{p}^{(m)})$ is the residual vector. The iteration (9) is repeated until $\|\mathbf{p}^{(m)} - \mathbf{p}^{(m-1)}\|_{\infty} < tol \|\mathbf{p}^{(m)}\|_{\infty}$. The solution $\mathbf{p}^{(m)}$ in (9) is computed by applying the Armijo damping rule, to guarantee the decrease of the regularized objective function. The parameter *tol* controls the convergence: using a too small value causes an increases of the algorithm complexity and may not represent a real improvement in the quality of the results. The computation of the residual and Jacobian matrix requires to solve the differential state equation with several values of the parameter. In the present work we solve the state equation by means of a finite difference method using central second order finite differences. The Jacobian matrix is computed by means of forward finite difference approximation. By choosing the regularization function as:

$$\mathcal{R}(\mathbf{q}) = \|L_2 \mathbf{q}\|_2^2 \tag{10}$$

where L_2 is the second order difference matrix, we can prove that $\|L_2 \mathbf{w}(\lambda)\|_2$ is strictly decreasing for each solution $\mathbf{w}(\lambda)$ of the minimization problem (9) with $\lambda = \lambda_k, k = 0, 1, \dots$.

Proposition 1. *Let*

$$\pi(\lambda) = \|L_2 \mathbf{w}(\lambda)\|_2, \quad \alpha = 0, 1, 2$$

where $\mathbf{w}(\lambda)$ *satisfies* (9) *for* $k = 0, 1, \dots$ *then* $\pi(\lambda)$ *is strictly convex on* $[0, \infty)$ *and decays monotonically to zero as* λ *increases from zero.*

Proof. Applying the first order conditions to (9) we can define the solution $\mathbf{w}(\lambda)$ as:

$$\left(J_F(\mathbf{p}^{(m)})^t J_F(\mathbf{p}^{(m)}) + \lambda L_2^t L_2 \right) \mathbf{w}(\lambda) = -J_F(\mathbf{p}^{(m)})^t \mathbf{z}^{(k)} \tag{11}$$

where $\mathbf{z}^{(k)} = \mathbf{r}^{(k)} - J_F(\mathbf{p}^{(m)})\mathbf{p}^{(m)}$. Introducing the Generalized Singular value Decomposition (GSVD) [9, 10] of the matrix

$$B = \begin{bmatrix} J_F(\mathbf{p}^{(m)}) \\ L_2 \end{bmatrix}, \qquad \begin{matrix} J_F((\mathbf{p}^{(m)}) \in \mathbb{R}^{N \times K} \\ L_2 \in \mathbb{R}^{N_L \times K} \end{matrix}$$

in the assumption that $\text{rank}(B) = k \leq K$ and $N \geq K$:

$$J_F = U D_J X^{-1}, \quad D_j = diag(\theta_1, \dots, \theta_K) \geq 0$$

$$L_2 = V D_L X^{-1}, \quad D_L = diag(\beta_1, \dots, \beta_t) \geq 0$$

where $U \in \mathbb{R}^{N \times N}$ and $V \in \mathbb{R}^{N_L \times N_L}$ are orthogonal matrices and $X \in \mathbb{R}^{K \times K}$ is non singular and: $0 \leq \theta_1 \leq \theta_2 \leq \dots \leq \theta_K \leq 1$, $1 \geq \beta_1 \geq \dots \geq \beta_t \geq 0$, $\beta_{t+1} = \dots = \beta_K = 0$ $t = min(N_L, K)$

Using the GSVD in (11) we can write the solution $\mathbf{w}(\lambda)$ as :

$$\mathbf{w}(\lambda) = X^{-1}\left(D_J^t D_J + \lambda D_L^t D_L\right)^{-1} D_J^t U \mathbf{z}^{(k)}$$

and substituting it in $\pi(\lambda)$ we have:

$$\pi(\lambda) = \sqrt{(D_J X \mathbf{w}(\lambda))^t D_J X \mathbf{w}(\lambda)} = \left(\sum_{i=1}^{t} \frac{\mu_i^2}{(\mu_i^2+\lambda)^2}(U_i^t \mathbf{z}^{(k)})^2\right)^{1/2}$$

where $\mu_i = \theta_i/\beta_i$. Now we can use Lemma 2.1 in [11] defining $a_i = \mu_i U_i^t \mathbf{z}^{(k)}$, $b_i = \mu_i^2$ and

$$\pi(\lambda) = \chi(\lambda) \equiv \left(\sum_{i=1}^{t}\left(\frac{a_i}{b_i+\lambda}\right)^2\right)^{1/2}$$

and prove that $\pi(\lambda)$ is strictly convex on $[0,\infty)$ and monotonically decreasing to zero as λ increases from zero. □

This guarantees that assumption (i) at page 2 is fulfilled and the sequence $(\lambda_k, \mathbf{q}_k)$ converges to the solution of the dual problem. Hence we can define the Constrained Least Squares Identification algorithm `CLSId` as follows:

Algorithm `CLSId`(γ)

$k = 0$; $\lambda_0 = 1$,
`repeat`
  $\mathbf{q}(\lambda_k) = argmin_{\mathbf{q}} \{J(\mathbf{q}) + \lambda_k(\mathcal{R}(\mathbf{q}) - \gamma)\}$ (`Gauss Newton + Armijo`)
  $G_k = \mathcal{R}(\mathbf{q}(\lambda_k)) - \gamma$
  $\lambda_{k+1} = \lambda_k$+`hybrid bisection-secant step`
  $k = k + 1$
`until` $|G_k - \gamma| < \tau_r G_0 + \tau_a$ or $|\lambda_k - \lambda_{k-1}| < \tau_a$ or $k > maxit$

where τ_a, τ_r are relative and absolute tolerance parameters.

4. Iterative estimate of the smoothing parameter γ

In order to obtain a good quality solution, a suitable value of the smoothing parameter γ needs to be computed. If γ is too small, the smoothing constraint forces the solution of (2) away from the noisy data, producing a smooth solution $\mathbf{q}$ with a large $J(\mathbf{q})$. On the other hand, if γ is too large, the regularity constraint becomes looser, allowing a better fit to the noisy data (small $J(\mathbf{q})$ value). Since the optimal value γ_{opt} lies in between these two extrema, our idea is to define a starting value $\gamma_0 < \gamma_{opt}$ and increase it on the basis of the behavior of the data fit function, with the aim to get closer to γ_{opt}. Observations performed on different problems show that $J(\mathbf{q})$ decreases steeply when $\gamma \ll \gamma_{opt}$ and tend to become almost flat when γ gets closer to γ_{opt}. Therefore we define the sequence $\gamma_0, \leq \gamma_1 \leq \cdots$ as:

$$\gamma_{\ell+1} = (1+\theta)\gamma_\ell, \quad \text{if } J(\mathbf{q}_\ell) - J(\mathbf{q}_{\ell-1}) > \tau(\gamma_\ell - \gamma_{\ell-1}),\ \gamma_{\ell+1} = \gamma_\ell \text{ otherwise} \tag{12}$$

with θ representing the relative change in $\gamma_{\ell+1}$ and τ the decrease rate of the data fit function at $\mathbf{q}_\ell$, defined as the solution of (2) with $\gamma = \gamma_\ell$. The starting value γ_0 is defined by computing a smooth approximate solution. The least squares fit function allows us to apply the Nonlinear Conjugate gradient method (NCg) to the problem $\min_{\mathbf{q}} J(\mathbf{q})$. The semi-convergence of the NCg iterations is well known and also the strong smoothness of the first NCg iterates. Using this property we compute $\mathbf{q}_{cg}$ by early stopping NCg (2-3 iterations) and define $\gamma_0 = \|L_2 \mathbf{q}_{cg}\|_2^2$. Despite the number of parameter is increased, the tuning of $\tau = 10^{-p}$, $p = 1, 2$ and $\theta \in [1/4, 1/2]$ is easier compared to γ that can change of several orders of magnitudes, depending on the problem.

5. Numerical Experiments

In this section we report some numerical tests where the state equation is a second order linear boundary value problem. The experiments are carried out on Intel Core i5 1.60GHz×4 processor (5.8 GiB RAM), using Matlab 2010a. Test problems:

T1 Identification of the reaction parameter $q(x)$ in the stationary model:

$$u(x)'' - q(x)u(x) = f, \quad u(0) = u(1) = 0, \quad x \in [a, b] \equiv [0, 1] \tag{13}$$

where the function f is defined by setting $u(x) = sin(\pi x)$ and the parameter $q(x) = 10(x^4 - sin(\pi x))$. The starting value $\mathbf{q}_0$ is obtained to guarantee that at least the boundary conditions are fulfilled, by sampling the function: $q_0(x) = 10(2x^3 - (1 + \pi)x^2 + \pi x)$ [12].

T2 Identification of the diffusivity function $c(x)$ in the differential model:

$$(c(x)u(x)')' = f(x), \quad u(0) = u(1) = 0, \quad x \in [a, b] \equiv [0, 1] \tag{14}$$

where the function f is defined by setting $u(x) = \sin(\pi x(1 - x))$ and the parameter $q(x) = 1 + e^{-(10(x-0.3))^2)} + e^{-(10(x-0.7))^2)}$. The starting value $\mathbf{q}_0 = 1$ is used.

The discrete reference solution of the state equation, is computed by the finite difference method, sampling the spatial domain with a uniform grid of M points $x_i = a + (i - 1)(b - a)/(M - 1), \quad i = 1, \ldots, M$. The noiseless measurements $\mathbf{y}$ are obtained by interpolating the reference solution in the measurement points defined by a uniformly spaced grid of N points: $a = \xi_1 < \xi_2 < \ldots \xi_{N-1} < \xi_N = b$. Gaussian white noise, as in (1), is added with different levels $\delta \in [10^{-4}, 10^{-1}]$. The quality of the results is evaluated by the Parameter Relative Error (PRE) and the Residual Norm ($ResN$): PRE $= \|\mathbf{q}_{reg} - \mathbf{q}_{true}\|_2 / \|\mathbf{q}_{true}\|_2$, where $\mathbf{q}_{reg}$ is obtained applying the regularization algorithm `CLSId` and $\mathbf{q}_{true}$ is the discretization of the parameter $q(x)$ at the measurement points $\xi_1, \ldots, \xi_N$. The quality of the reconstructions depends on the discretization values N and M. For values of M and N fixed, the results depend on the level of noise present in the data and on the value of the smoothness level γ. To analyze the features of the regularization algorithm `CLSId`(γ) we focus on the case $M = 1001$, $N = 101$ and with tolerance parameters $\tau_a = \tau_r = 1.e - 4$. The reference value γ_{cg} of the regularization constraint is obtained by stopping the nonlinear Conjugate Gradient iterations as soon as the ratio between the actual residual norm and the initial residual norm is less than 99%. The adapted parameter γ_{adapt} is computed by (12) with $\tau = 0.1$ and $\theta = 0.3$ if $|J(\mathbf{q}_{cg}) - J(0.8\mathbf{q}_{cg})| > 1.e - 5$ and $\theta = 1.3$ otherwise. Table 1 reports γ_{adapt} and the relative error obtained by `CLSId`(γ_{adapt}) ($PRE(\gamma_{adapt})$) for different noise levels. These values are compared to γ_{opt} and $PRE(\gamma_{opt})$ where the optimal parameter γ_{opt} is computed by applying `CLSId`(γ) to a large set of parameters γ and taking the value γ corresponding to the minimum PRE. The computational performance is reported by the number of steps ℓ and the total number of Gauss Newton iterations $GNit$. Figures 1(a) and 2(a) show the reconstructed and true parameter obtained in the case $\delta = 0.1$ and $\delta = 5.e - 3$ respectively. In figure 1(b) the parameter relative errors obtained by choosing the smoothing levels γ_{adapt}, γ_{cg}, $\gamma_0 = \|L\mathbf{q}_0\|_2^2$ are compared to the error obtained by the optimal parameter γ_{opt}. We observe that both $\gamma_c g$ and γ_0 give large reconstruction errors while γ_{adapt} allows us to obtain nearly optimal results. Finally the typical behavior of the parameter error obtained by the iterative procedure (12) is shown in figure 2(b).

6. Conclusions

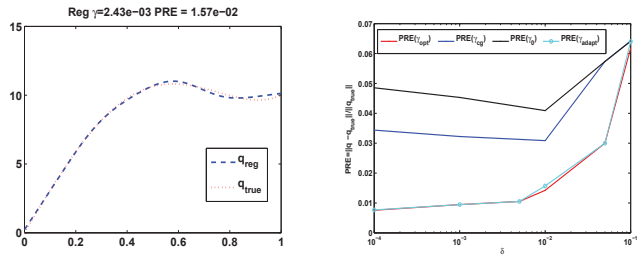
Future work includes the application of different quasi-Newton methods and the extension to non linear time dependent differential models typical of biochemical reactions.

Acknowledgement

The author would like to thank the reviewers for their comments and suggestions.

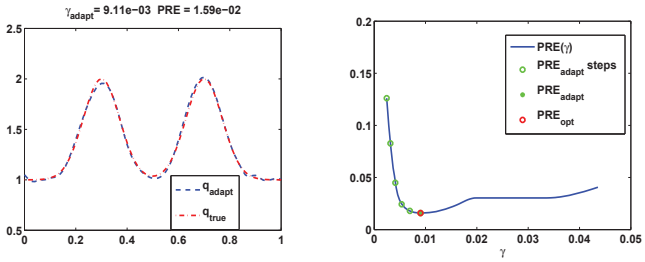
Test	δ	$PRE(\gamma_{adapt})$	γ_{adapt}	$\ell(GNit)$	γ_{opt}	$PRE(\gamma_{opt})$
T1	1.e-3	9.469e-3	2.423e-3	1 (25)	2.436e-3	9.341e-3
	5.e-3	1.051e-2	2.2424e-3	1 (18)	2.232e-3	1.011e-2
	1.e-2	1.571e-2	2.429e-3	1 (17)	2.916e-3	1.357e-2
T2	1e-03	7.8122e-3	9.0088e-3	6 (63)	9.0284e-3	7.8002e-3
	5e-03	1.5877e-2	9.0306e-3	6 (58)	9.0656e-3	1.5863e-2
	1e-02	1.1775e-2	2.9921e-2	7 (68)	6.9674e-03	2.8706e-02

Table 1. `CLSId` results.



(a) Case $\delta = 1.e-2$ (b) $PRE(\gamma)$ vs δ

Figure 1. Test Problem (13)



(a) Case $\delta = 5.e-3$ (b) $PRE(\gamma)$ vs γ

Figure 2. Test Problem (14)

References

[1] Becker R, Braack M and Vexler B 2004 Compust. Theory and Modelling **8**
[2] Zama F, Ciavarelli R, Frascari D and Pinelli D 2013 Numerical parameters estimation in models of pollutant transport with chemical reaction *System Modeling and Optimization* (*IFIP Advances in Information and Communication Technology* vol 391) ed Dietmar H and Fredi T (Springer Berlin Heidelberg) pp 547–556 ISBN 978-3-642-36061-9
[3] Ito K and Kunisch K 1990 *SIAM Journal on Control and Optimization* **28** 113–136
[4] Kunisch K and Sachs E W 1992 *SIAM Journal on Numerical Analysis* **29** 1793–1820
[5] B K, A N and O S 2008 *Iterative Regularization Methods for Nonlinear Problems* (de Gruyter Berlin, New York)
[6] B K, a K and B V 2014 *Inverse Problems* **30** 045002
[7] Piccolomini E L and Zama F 2011 *Applied Mathematics and Computation* **217**
[8] Chen K, Piccolomini E and Zama F 2013 *Numerical Algorithms* 1–20
[9] Paige C and Saunders M 1981 *SIAM Journal on Numerical Analysis* **18** 398–405
[10] Åke Björck 1996 *Numerical Methods for Least Squares Problems* (Society for Industrial and Applied Mathematics)
[11] Cartis C, Gould N and Toint P 2009 *BIT*
[12] Jin B and Maass P 2012 *Inverse Problems* **28** 123001