Another look at estimating parameters in systems of ordinary differential equations via regularization

Ivan Vujačić*
VU University Amsterdam, Amsterdam, The Netherlands - i.vujacic@vu.nl

Seyed Mehdi Mahmoudi
University of Groningen, Groningen, The Netherlands - seyed.m.mahmoudi@gmail.com

Ernst Wit
University of Groningen, Groningen, The Netherlands - e.c.wit@rug.nl

Abstract

We consider estimation of parameters in systems of ordinary differential equations (ODEs). This problem is important because many processes in various fields of science are modelled by a system of ODEs. Since the system usually contains unknown parameters it is of interest to estimate them. The problem is approached from the viewpoint of M-estimation. In general, for a given parameter the true solution of the system is unavailable, therefore any M-criterion function is necessarily defined via an approximation of the solution. We define an approximation by viewing the system of ODEs as an operator equation and exploiting the connection with the regularization theory. Combining introduced regularized solution with M-criterion function we lay out a general framework for estimating parameters in ODEs which can handle partially observed systems. If M-criterion function is log-likelihood choosing suitable regularized solution yields estimator which is consistent and asymptotically efficient. Connection with the generalized profiling procedure is made.

Keywords: M-estimation; Quasisolution; Generalized Tikhonov regularizer; Asymptotic efficiency.

1. Introduction

Consider the system of ordinary differential equations of the form

$$\begin{cases} \boldsymbol{x}'(t) = \boldsymbol{f}(\boldsymbol{x}(t), t; \boldsymbol{\theta}), \ t \in [0, T], \\ \boldsymbol{x}(0) = \boldsymbol{\xi}, \end{cases} \tag{1}$$

where $\boldsymbol{x}(t)$ takes values in $\mathbb{R}^d$, $\boldsymbol{\xi}$ in $\Xi \subset \mathbb{R}^d$, $\boldsymbol{\theta}$ in $\Theta \subset \mathbb{R}^p$ and $\boldsymbol{f}$ is known function. Given the values of $\boldsymbol{\xi}$ and $\boldsymbol{\theta}$, we denote the solution of (**??**) by $\boldsymbol{x}(t; \boldsymbol{\theta}, \boldsymbol{\xi})$. Let us assume that a process is modelled by ODEs (**??**) with unknown parameters $\boldsymbol{\xi}_0$ and $\boldsymbol{\theta}_0$. For simplicity, assume that we have noisy observations $y_i(t_j)$, $j = 1, \ldots, n$ of the first $1 \leq d_1 \leq d$ states $x_i(t; \boldsymbol{\theta}_0, \boldsymbol{\xi}_0)$, $i = 1, \ldots, d_1$ at time points $t_j \in [0, T]$, $j = 1, \ldots, n$:

$$y_i(t_j) = x_i(t_j; \boldsymbol{\theta}_0, \boldsymbol{\xi}_0) + \varepsilon_i(t_j), \quad i = 1, \ldots, d_1; j = 1, \ldots, n. \tag{2}$$

The problem is to estimate $\boldsymbol{\theta}_0$ from the data $\mathbf{Y}$, where $\mathbf{Y} = (y_i(t_i))_{ij}$ denotes the matrix that contains all the observations. A general way to estimate the unknown parameter is to minimize some known function M_n of the parameter and the data; the obtained estimator is called M-estimator(**?**). In the problem we consider M_n depends on the solution of the ODEs system. In general the solution is not available, therefore its approximation has to be used. Approximation can be deterministic or stochastic; those used in the literature include:

1. Numerical solution given by a numerical ODEs solver (**?**).
2. Classical smoothers like cubic splines, kernel estimators, regression splines, local polynomials, step function estimators (**?????**).
3. Specially designed smoothers that use ODEs model, like model based smoothing (**?**) or reproducing kernel hilbert space based smoother (**?**).

In this paper, we follow the idea described above and estimate the parameter as follows:

$$1.\ \widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathcal{X}_m}{\operatorname{argmin}}\, \mathcal{T}_{\alpha,\gamma}(\boldsymbol{x}), \tag{3}$$

$$2.\ \widehat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}}\, M_n(\boldsymbol{\theta}|\widehat{\boldsymbol{x}}, \mathbf{Y}). \tag{4}$$

Here $\mathcal{T}_{\alpha,\gamma}$ is a functional with parameters $\alpha \geq 0$ and $\gamma \geq 0$, which is optimized over some finite-dimensional subspace $\mathcal{X}_m$ of the solution space. M_n is a criterion function to be optimized, for example the log-likelihood criterion. The main goal of the paper is to show how to define the functional $\mathcal{T}_{\alpha,\gamma}$ by using regularization theory. We will call $\mathcal{T}_{\alpha,\gamma}$ generalized Tikhonov functional and its minimizer generalized Tikhonov regularizer. The proposed framework can handle fully and partially observed systems and is not based on numerical integration of the system. The trade-off is that the numerical integration is substituted with numerical optimization. Some issues related to the proposed methodology that are also present in similar methods are selection of α, γ and the dimension m of the space $\mathcal{X}_m$.
The rest of the paper is organized as follows. A review of the regularization theory is provided in the following section. In Section 3 the described theory is used to define generalized Tikhonov functional for the ODEs system. Section 4 contains theoretical results for the estimator defined by (**??**) and (**??**) with M-criterion function being the log-likelihood. Some comparison with the generalized profiling procedure of (**?**) is also provided. The final section contains discussion.

2. Ill-posed problems, quasisolutions and regularization

Let $F\colon \mathcal{X} \to \mathcal{Y}$ where $\mathcal{X}, \mathcal{Y}$ are linear normed spaces and consider the operator equation

$$F(x) = y, \tag{5}$$

$x \in \mathcal{X}$, $y \in \mathcal{Y}$. In the regularization theory, $\mathcal{X}$ is called *solution space* and $\mathcal{Y}$ *data space* . The problem (**??**) is *well-posed in the sense of Hadamard* on the pair of normed spaces $\mathcal{X}$ and $\mathcal{Y}$ if the solution of (**??**) exists, it is unique and it is continuous with respect to y. The problem(**??**) is *ill-posed* on the pair of normed spaces $\mathcal{X}$ and $\mathcal{Y}$ if at least one of the three well-posedness conditions does not hold.
Equation (**??**) can be solved on the set $S \subset \mathcal{X}$ by finding the minimum of the *objective functional*

$$\mathcal{J}(x) = \|F(x) - y\|^2, \tag{6}$$

on S. This idea dates back to the works of A. M. Legendre and K. Gauss from the beginning of 19th century, who proposed the least squares method for solving systems of linear algebraic equations. (**?**) In this regard, Ivanov introduced a concept of *quasisolution* (or pseudo solution or least squares solution) of equation (**??**) on $S \subset \mathcal{X}$ — quasisolution is any minimizer of (**??**) on S (**?**, Sec 1.2). For numerical optimization of $\mathcal{J}$ problem has to be discretized; computation can be performed only with finite-dimensional spaces. Moreover, the finite dimensional subspace of $\mathcal{X}$ has to be chosen from the family of subspaces $\mathcal{X}_1 \subset \mathcal{X}_2 \subset \ldots$ such that their union is dense in $\mathcal{X}$. This is minimal assumption for the sequence of minimum norm minimizers x_n of (**??**) on $\mathcal{X}_n$ to converge to the minimum norm minimizer x of (**??**) on $\mathcal{X}$. When regularization is achieved by a finite dimensional approximation alone it is called *self-regularization* (**?**) or *regularization by projection.* (**?**)
Tikhonov regularization involves further regularization by minimizing so-called *Tikhonov functional*

$$\mathcal{T}_\alpha(x) = \mathcal{J}(x) + \alpha\Omega(x - x_0), \tag{7}$$

where x_0 is *the trial solution*, $\alpha \geq 0$ is *the regularization parameter* and Ω is *the stabilizing functional* which is usually given by a norm or a semi-norm on $\mathcal{X}$. Stabilizing functional incorporates a priori information on the smoothness of the solution x. Some possible choices are square of the norm $\Omega(x) = \|x\|^2$ and total variation $\Omega(x) = V_0^T(x)$.
A priori information on values of the solution may be available, which allows to single out solutions which satisfy physical requirements. This information can be incorporated by adding additional functional $\mathcal{S}$ in (**??**) (**??**), which measures closeness of the solution to the aforementioned a priori information:

$$\mathcal{T}_{\alpha,\gamma}(x) = \mathcal{J}(x) + \alpha\Omega(x - x_0) + \gamma\mathcal{S}(x). \tag{8}$$

Here $\gamma \geq 0$ is the *penalty parameter*. In noise free case, **?** discussed information given via equality or inequality constraints; the functional $\mathcal{S}$ was termed the penalty functional and the resulting method was called generalized regularization method. In the presence of noise, **?** used the term the similarity functional for $\mathcal{S}$ and named the resulting method multi modal Tikhonov regularization. We will use the terms *similarity functional* and *generalized Tikhonov regularization*. Thus, *generalized Tikhonov regularization* finds an approximation to the solution of (**??**) by minimizing (**??**) over some finite-dimensional subspace of $\mathcal{X}$. We will refer to the functional $\mathcal{T}_{\alpha,\gamma}(x)$ as *generalized Tikhonov functional* and its minimizer as generalized Tikhonov regularizer.

3. Generalized Tikhonov regularization in ODE estimation setting

The problem (**??**) is well-posed. Indeed, the first two conditions of the definition of the well-posed problem in the sense of Hadamard follow from existence and uniqueness theorems for ordinary differential equations, and the third condition follows from the theorem of continuous dependence of solution on initial conditions and parameters (**?**). However, some states may not be observed in which case the initial conditions are not known. In this case the solution is not unique and the problem becomes ill-posed. Even if the initial conditions are known, non-uniqueness can still be introduced through discretization. (**?**) Thus, to deal with this issue regularization can be employed. To this aim for system (**??**) without initial condition we define generalized Tikhonov regularizer by using the theory described in the previous section. For notational simplicity, we suppress the dependence on $\boldsymbol{\theta}$.

3.1 Discretization

The first step is to discretize the problem; here we assume that the solution of the system (**??**) lies in $(C^1[0,T])^d$. For simplicity we assume that each component of $\boldsymbol{x}$ is approximated by an element from the same finite dimensional function space $\mathcal{X}_m$ of dimension m with basis $\{h_1, \ldots, h_m\}$. Any $x_i \in \mathcal{X}_m$, $i = 1 \ldots, d$ can be written as a linear combination of basis functions:

$$x_i(t) = \sum_{k=1}^{m} \beta_{ik} h_k(t) = \boldsymbol{\beta}_i^\top \boldsymbol{h}(t), \tag{9}$$

where $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{im})^\top$ and $\boldsymbol{h}(t) = (h_1(t), \ldots, h_m(t))^\top$. Commonly used basis functions are B-splines; they yield sequence of spaces $\mathcal{X}_m^d$ whose union is dense in $(C^1[0,T])^d$ (see Lemma **??** in Section 4).

3.2 Objective functional

For fixed $\boldsymbol{\theta}$ ODEs system (**??**) without initial condition is equivalent to the operator equation $F(\boldsymbol{x}) = \mathbf{0}$, where $F(\boldsymbol{x}(\cdot)) = \boldsymbol{x}'(\cdot) - \boldsymbol{f}(\boldsymbol{x}(\cdot), \cdot, \boldsymbol{\theta})$. The corresponding objective functional is

$$\mathcal{J}(\boldsymbol{x}) = \|\boldsymbol{x}' - \boldsymbol{f}(\boldsymbol{x}, \cdot, \boldsymbol{\theta})\|_{2,\boldsymbol{w}}^2, \tag{10}$$

where $\boldsymbol{w} = (w_1, \ldots, w_d)$, $w_i > 0$ for $i = 1, \ldots, d$ and $\|\boldsymbol{x}\|_{2,\boldsymbol{w}} = \sqrt{\sum_{i=1}^d w_i \int_0^T x_i^2(t)\mathrm{d}t}$ is the norm on the space of d vector valued functions. Depending on the application other norms in $\mathcal{J}$ can be chosen.

3.3 Stabilizing functional

Stabilizing functional enforces smoothing conditions on the solution $\boldsymbol{x}$. The common choice in the literature is the total curvature on $[0,T]$ measured by the integral $\int_0^T \{x''(t)\}^2 \mathrm{d}t$. In regard to this measure one choice for stabilizing functional is

$$\Omega(\boldsymbol{x}) = \sum_{i=1}^{d} v_i \int_0^T \{x_i''(t)\}^2 \mathrm{d}t \tag{11}$$

where v_i, $i = 1, \ldots, d$ are nonnegative constants. Other choices are possible, like total variation $V_0^T(x)$.

3.4 Similarity functional

For simplicity let us assume that $d = 1$. Let $p(y(t)|x(t;\boldsymbol{\theta}_0, \boldsymbol{\xi}_0))$ be the distribution of the data based on the model (**??**) and g be the true distribution of the data. The observations $y(t_i)$ represent the data for the problem of the estimation of $\boldsymbol{\theta}_0$ but they are a priori information for the problem of finding the solution $\boldsymbol{x}(t;\boldsymbol{\theta}_0, \boldsymbol{\xi}_0)$ of (**??**). The solution is measured with noise therefore the distribution of the data should be close to a priori distribution of the solution — distribution based on the model. One measure of closeness between distributions is the Kullback-Leibler (KL) divergence, which belongs to the class of f-divergences (**?**). Since

g is not known KL divergence can only be estimated from the data by approximating g with its empirical density:

$$KL(g(\cdot); p(\cdot|x(\cdot;\boldsymbol{\theta}_0,\boldsymbol{\xi}_0)) \approx -\frac{1}{n}\sum_{i=1}^{n}\log p(y(t_i)|x(t_i;\boldsymbol{\theta}_0,\boldsymbol{\xi}_0)).$$

Because the solution $x(t_i;\boldsymbol{\theta}_0,\boldsymbol{\xi}_0)$ is unknown by using the approximation $p(y(t_i)|x(t;\boldsymbol{\theta}_0,\boldsymbol{\xi}_0)) \approx p(y(t_i)|\widehat{x}(t;\boldsymbol{\beta}))$ we obtain the similarity functional which for the system with d_1 observed components is

$$\mathcal{S}(\boldsymbol{x}) = -\sum_{i=1}^{d_1}\sum_{j=1}^{n}\log p(y_i(t_j)|\widehat{x}_i(t_j;\boldsymbol{\beta})). \tag{12}$$

Here the scale $1/n$ is omitted because it can be subsumed in the penalty parameter γ which multiplies $\mathcal{S}$ in (**??**). The weakness of this approach is that the employed approximation ignores the statistical uncertainty in the regularizer (**?**).

3.5 Generalized Tikhonov functional

From the previous subsections it follows that the generalized Tikhonov functional for the equation $F(\boldsymbol{x}) = \mathbf{0}$ is

$$\mathcal{T}_{\alpha,\gamma}(\boldsymbol{x}(\boldsymbol{\beta})) = \mathcal{J}(\boldsymbol{x}(\boldsymbol{\beta})) + \alpha\Omega(\boldsymbol{x}(\boldsymbol{\beta}) - \boldsymbol{x}_0) + \gamma\mathcal{S}(\boldsymbol{x}(\boldsymbol{\beta})), \tag{13}$$

where the functionals $\mathcal{J}$, Ω and $\mathcal{S}$ are defined in (**??**), (**??**) and (**??**), respectively. The regularized solution is found by optimizing (**??**) over $\mathcal{X}_m^d$ parametrized by $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_d^\top)^\top$. This can be achieved by optimizing (**??**) with respect to $\boldsymbol{\beta}$ over $\mathbb{R}^{dm}$:

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^{dm}}{\operatorname{argmin}}\, \mathcal{T}_{\alpha,\gamma}(\boldsymbol{x}(\boldsymbol{\beta})),$$

and applying (**??**).

4. Generalized Tikhonov regularizer with log-likelihood criterion and connection with the generalized profiling procedure

Combining generalized Tikhonov regularizer (**??**) with the log-likelihood criterion

$$M_n(\boldsymbol{\theta}) = -\sum_{i=1}^{d_1}\sum_{j=1}^{n}\log p(y_i(t_j)|\widehat{x}_i(t_j;\boldsymbol{\theta})),$$

we obtain an estimator which is asymptotically efficient. The proof of asymptotic properties of the estimator follow like in (**?**), with slight modifications. More specifically, let $\boldsymbol{x}^o$ and $\boldsymbol{x}^u$ denote the observed and unobserved part of $\boldsymbol{x}$, respectively and let $\boldsymbol{\xi}_0^o$ and $\boldsymbol{\xi}_0^u$ be their corresponding initial conditions. The following results, similar to Lemma 1, Theorem 3.2 and Theorem 3.3 of **?**, hold.

Lemma 1. *Under Assumption 2 of* **?**, *there exist a sequence of finite-dimensional subspaces $\mathcal{X}_n$ of $C^1[0,T]$ such that for any compact subset Θ_0 of Θ and any compact subset Ξ_0 of Ξ, it holds*

$$\lim_{n\to\infty} r_n = 0,$$

where

$$\begin{aligned} r_n \;=\; \max\Bigg[& \sup_{(\boldsymbol{\theta},\boldsymbol{\xi})\in\Theta_0\times\Xi_0}\ \inf_{\boldsymbol{w}\in\mathcal{X}_n,\boldsymbol{w}(0)=\boldsymbol{\xi}_0^o}\left\{\|\boldsymbol{x}^o(\boldsymbol{\theta},\boldsymbol{\xi},\cdot)-\boldsymbol{w}\|_\infty \vee \left\|\frac{\mathrm{d}\boldsymbol{x}^o}{\mathrm{d}t}(\boldsymbol{\theta},\boldsymbol{\xi},\cdot)-\frac{\mathrm{d}\boldsymbol{w}}{\mathrm{d}t}\right\|_\infty \vee \left\|\frac{\mathrm{d}^2\boldsymbol{x}^o}{\mathrm{d}t^2}(\boldsymbol{\theta},\boldsymbol{\xi},\cdot)-\frac{\mathrm{d}^2\boldsymbol{w}}{\mathrm{d}t^2}\right\|_\infty\right\}, \\ & \sup_{(\boldsymbol{\theta},\boldsymbol{\xi})\in\Theta_0\times\Xi_0}\ \inf_{\boldsymbol{v}\in\mathcal{X}_n,\boldsymbol{v}(0)=\boldsymbol{\xi}_0^u}\left\{\|\boldsymbol{x}^u(\boldsymbol{\theta},\boldsymbol{\xi},\cdot)-\boldsymbol{v}\|_\infty \vee \left\|\frac{\mathrm{d}\boldsymbol{x}^u}{\mathrm{d}t}(\boldsymbol{\theta},\boldsymbol{\xi},\cdot)-\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}\right\|_\infty \vee \left\|\frac{\mathrm{d}^2\boldsymbol{x}^u}{\mathrm{d}t^2}(\boldsymbol{\theta},\boldsymbol{\xi},\cdot)-\frac{\mathrm{d}^2\boldsymbol{v}}{\mathrm{d}t^2}\right\|_\infty\right\}\Bigg]. \end{aligned}$$

Theorem 1. *Let Assumptions 1-5 from* **?** *hold. If* $r_n \to 0$, $\alpha_n \to 0$ *and* $\gamma_n \to 0$, *as* $n \to \infty$ *then*

$$\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = o_P(1).$$

Theorem 2. *Let Assumptions 1-6 from* **?** *hold. If* $r_n = o(n^{-1})$, $\alpha_n = o(n^{-2})$ *and* $\gamma_n = o(n^{-2})$ *as* $n \to \infty$ *then the estimator* $\widehat{\boldsymbol{\theta}}_n$ *is asymptotically efficient.*

Model based smoother used in generalized profiling (**?**) can be viewed as a generalized Tikhonov regularizer. Indeed, the inner fitting criterion J defined in **?** is

$$J(\boldsymbol{x}) = -\sum_{i=1}^{d_1}\sum_{j=1}^{n} \log p(y_i(t_j)|x_i(\boldsymbol{\theta})) + \sum_{i=1}^{d} \lambda_i \int_0^T \{x_i'(t) - f_i(\boldsymbol{x}(t), t, \boldsymbol{\theta})\}^2 \mathrm{d}t. \tag{14}$$

With representation $\lambda_i = \lambda w_i$, where w_i, $i = 1, \dots d$ are some constants, we obtain

$$J(\boldsymbol{x}) = \lambda \left\{ \frac{1}{\lambda}\mathcal{S}(\boldsymbol{x}) + \mathcal{J}(\boldsymbol{x}) \right\} = \lambda \mathcal{T}_{0,1/\lambda}(\boldsymbol{x}), \tag{15}$$

and consequently the model based smoother is a minimizer of $\mathcal{T}_{0,1/\lambda}$. It can be seen from (**??**) and (**??**) that the roles of the penalty terms and parameters in $\mathcal{T}_{0,1/\lambda}$ and J are reversed; relationship between the parameters is $\gamma = 1/\lambda$. For the solutions of the dynamic systems the fidelity to the ODEs is of the major concern and data term is of secondary importance (**?**). In regard to this, the regularization formulation seems more natural since the objective functional $\mathcal{J}$, which measures the fidelity to the ODEs, is not the penalty in $\mathcal{T}_{\alpha,\gamma}$ but the main term. On the other hand, in J the ODEs fidelity term is the penalty. The consequence of this is that to force the penalty to zero λ must approach ∞, which lead to ill conditioning in the optimization (**?**). This is avoided in the regularization formulation; here γ must approach to zero.

5. Discussion

Generalized Tikhonov regularizer $\mathcal{T}_{\alpha,\gamma}$ is broad because it allows to incorporate various a priori information available. It involves several layers of regularization, the first one being self-regularization (see Section 2). A practical issue is to determine the amount of regularization in $\mathcal{T}_{\alpha,\gamma}$. Table **??** gives a list of different regularizers for the ODEs estimation problem. The first four regularizers in the table all satisfy the conditions of Theorems 1 and 2 when $\alpha, \gamma \to 0$. Thus, they are all asymptotically efficient. The most sensible choices seem to be the model based smoother (**?**) and Ivanov's quasi solution. Indeed, the similarity functional $\mathcal{S}$ is more informative then Ω since $\mathcal{S}$ gives information on the values of the minimizer of $\mathcal{J}$ while Ω narrows down the class of functions to which the minimizer of $\mathcal{J}$ should belong. This class is already restricted through self-regularization and hence further regularization via Ω may not be necessary. This will be examined in simulation studies elsewhere.

Parameters	$\mathcal{T}_{\alpha,\gamma}(\boldsymbol{x})$	$\widehat{\boldsymbol{x}} = \operatorname{argmin}_{\boldsymbol{x}\in\mathcal{X}_m} \mathcal{T}_{\alpha,\gamma}(\boldsymbol{x})$
$\alpha > 0$, $\gamma > 0$	$\mathcal{J}(\boldsymbol{x}) + \alpha\Omega(\boldsymbol{x} - \boldsymbol{x}_0) + \gamma\mathcal{S}(\boldsymbol{x})$,	Generalized Tikhonov's regularizer
$\alpha = 0$, $\gamma = 0$	$\mathcal{J}(\boldsymbol{x})$,	Ivanov's quasi solution
$\alpha > 0$, $\gamma = 0$	$\mathcal{J}(\boldsymbol{x}) + \alpha\Omega(\boldsymbol{x} - \boldsymbol{x}_0)$,	Tikhonov's regularizer
$\alpha = 0$, $\gamma > 0$	$\mathcal{J}(\boldsymbol{x}) + \gamma\mathcal{S}(\boldsymbol{x})$	model based smoother of (**?**)
$\alpha = \infty$, $\gamma = 0$	$\mathcal{J}(\boldsymbol{x}_0)\delta(\boldsymbol{x} - \boldsymbol{x}_0)$	trial solution $\boldsymbol{x}_0$

Table 1: Generalized Tikhonov regulariser and its special cases. The last row should be interpreted as $\mathcal{T}_{\alpha,0}(\boldsymbol{x}) \to \mathcal{J}(\boldsymbol{x}_0)\delta(\boldsymbol{x} - \boldsymbol{x}_0)$ as $\alpha \to +\infty$, where δ is the Dirac's delta function.

Acknowledgements

The first author thanks Bartek Knapik for useful discussions on inverse and ill-posed problems and for providing useful references.