23rd International Congress on Modelling and Simulation, Canberra, ACT, Australia, 1 to 6 December 2019
mssanz.org.au/modsim2019

A functional strategy for nonlinear functionals

R. S. Anderssen [a], **B. Haak** [b] **and M. Hegland** [c]

[a]*Data61 CSIRO, Canberra, Australia*
[b]*Institut de Mathématiques de Bordeaux, Université Bordeaux 1, France*
[c]*Mathematical Sciences Institute, Australian National University, Australia*
Email: Markus.Hegland@anu.edu.au

Abstract: The linear functional strategy introduced by the first author in 1986 provided a shift in the way inverse problems were solved. It is based on the fact that for applications one is interested in specific properties of the solution of an inverse problem. These properties or quantities of interest are usually obtained by applying a functional to the solution of the inverse problem. The linear functional strategy avoided the need to solve the full inverse problem by solving the adjoint problem for the functional instead. The solution to this adjoint problem is a functional which, when applied to the data returns the quantity of interest. In some cases, the adjoint problem can be solved exactly. In any case, the adjoint problem does not need to deal explicitly with data errors.

In this paper we review the original approach. It is noted that any method which is able to produce an approximation to the solution of the adjoint problem which is continuous leads to a linear dependence of the error in the quantity of interest with respect to the data error.

Most of the paper considers the application of advances in computational and applied mathematics in the last 30 years to the functional strategy. We define a general (nonlinear) functional strategy and illustrate how this problem is solved. We define a generalised adjoint problem for nonlinear functionals and inverse problems. This adjoint problem is shown to be linear. Furthermore, we observe that nonlinear functionals which are Lipschitz continuous are stable with respect to data errors. The solution of the adjoint problem constrained to Lipschitz continuous functionals leads to Tikhonov regularisation.

We indicate how to implement the functional strategy for a simple example and provide links to modern functional analysis.

Keywords: *Functional strategy, inverse problems*

1 THE FUNCTIONAL STRATEGY

Many applied inverse problems are of the form:

- *Determine $q = \phi(u)$ (a quantity of interest) where ϕ is a given functional and where u is the solution of a (usually ill-posed) equation $F(u) = f$ for some data f.*

The *functional strategy* solves this problem by introducing a functional θ which computes the quantity of interest q directly from the data by $q = \theta(f)$. From the definition of q and the equation $F(u) = f$ one then gets an equation relating θ to ϕ as

$$\theta(F(u)) = \phi(u). \tag{1}$$

If one can find a Lipschitz continuous functional θ which solves equation (1) then one has solved the applied inverse problem. By the Lipschitz continuity of θ measured data f_δ satisfying $\|f_\delta - f\| \le \delta$ can replace f in the evaluation of θ and one gets $q_\delta = \theta(f_\delta)$ satisfying

$$|q_\delta - q| \le L_\theta\, \delta$$

where L_θ is the Lipschitz constant of θ. In practice one often faces two challenges:

i. θ can not be determined explicitly;
ii. the Lipschitz constant L_θ is large or even infinite.

We will in the following discuss how these challenges can be overcome. While we do not have the space to discuss the numerical solution in detail, we remark that all numerical approximations reduce both the functional ϕ and the the operator F to finite dimensional approximations ϕ_h and F_h, respectively. Rather than discussing the impact of this discretisation we will here assume that the original ϕ and F are finite dimensional, as are u and f.

An alternative to the functional strategy consists of solving the inverse problem $F(u) = f$ for the data f_δ using a nonlinear inverse problem solver. Then q is approximated by evaluation of the functional for the regularised solution u_α, i.e., $q \approx \phi(u_\alpha)$. The functional strategy has practical advantages over this approach while it seems to compare favorably regarding accuracy:

- the determination of θ can be done offline independently of the data using substantial computational resources
- once θ is known, it can be evaluated relatively efficiently and this approach is thus suitable to online delivery, for example in control systems, the operating theatre and on exploration vehicles
- while further investigation of the accuracy of the approach needs to be done, our initial computational simulations suggest that in most cases the functional strategy performs as well as the approach based on regularisation of the inverse problem
- the alternative approach does not take any properties of ϕ into account but this has been addressed by Louis (1996) and Mathé and Pereverzev (2002).

1.1 The linear functional strategy

The functional strategy sketched above was originally introduced in Anderssen (1986) for the case of linear functionals ϕ and linear operators F; see also Anderssen (2004). In the finite dimensional case one then has $u \in \mathbb{R}^d$ and $f \in \mathbb{R}^n$. The quantity of interest is then $q = \phi(u) = b^T u$ for some $b \in \mathbb{R}^d$ and $f = F(u) = Au$ for $A \in \mathbb{R}^{n\times d}$. If $\theta(f) = c^T f$, equation (1) takes the form $c^T Au = b^T u$ which is satisfied for all u if

$$A^T c = b. \tag{2}$$

In practice, the dimension d of the approximation space does not have to be equal to the data size n thus equation (2) typically does not have a unique solution. One may use the least squares solution with minimal norm which is $c = \left(A^T\right)^+ b$. As the norm of c is equal to the Lipschitz constant of the functional θ the minimal norm solution chooses the most stable functional from the ones which minimise the norm of the residual $A^T c - b$. However, even this choice might lead to a norm which is too high. One thus might wish to trade off some error in the equation to get better stability. Assuming that we accept a Lipschitz constant L for the functional one gets the following quadratic optimisation problem with a quadratic constraint for c:

- *Minimise* $\|A^T c - b\|^2$ *under the constraint* $\|c\| \leq L$.

The solution of this problem amounts to *Tikhonov regularisation* and one has the approximation

$$c = (AA^T + \lambda I)^{-1} Ab \tag{3}$$

for some parameter $\lambda > 0$ which is obtained from the constraint and depends on the Lipschitz constant desired.

1.2 A polynomial functional strategy

As in the linear case we consider $u \in \mathbb{R}^d$ and $f \in \mathbb{R}^n$. The functionals ϕ and θ are here polynomials in u and f. For example, one has

$$\phi(u) = \sum_{\alpha \in I} b_\alpha u^\alpha = \underline{b}^T \underline{u}$$

where $\alpha = (\alpha_1, \ldots, \alpha_d)$ is a multi-index and $u^\alpha = u_1^{\alpha_1} \cdots u_d^{\alpha_d}$. The sum is over a finite set $I \subset \mathbb{N}^d$. The vectors $\underline{b} = (b_\alpha)_{\alpha \in I}$ and $\underline{u} = (u^\alpha)_{\alpha \in I}$ are elements of $\mathbb{R}^I$. The same notation is used for the functional θ

$$\theta(f) = \sum_{\beta \in J} c_\beta f^\beta = \underline{c}^T \underline{f}$$

where $J \subset \mathbb{N}^n$ and $\underline{c}$ and $\underline{f}$ are elements of $\mathbb{R}^J$. The operator F is also assumed to be polynomial, i.e., the i-th component $f_i = F_i(u)$ is a linear combination of u^α. It then follows that f^β is a linear combination of powers u^α and thus we can introduce a matrix $\underline{A} \in \mathbb{R}^{I,J}$ such that $f^\beta = \sum_{\alpha \in I} A_{\beta,\alpha} u^\alpha$ or in matrix notation

$$\underline{f} = \underline{A}\, \underline{u}.$$

Like in the linear functional strategy one then has

$$\theta(f) = \underline{c}^T f = \underline{c}^T \underline{A}\, \underline{u}$$

and from this we get

$$\underline{A}^T \underline{c} = \underline{b}. \tag{4}$$

Note the difference between the polynomial and the linear case: the polynomial functionals are linear in the vector $\underline{u}$ of monomials u^α but nonlinear in the vector u of components u_i. The functionals are differentiable which is useful for the determination of the Lipschitz constants. As θ is differentiable, the Lipschitz constant of θ is equal to the maximal norm of the gradient. The squared norm of the gradient of $\theta(u)$ is

$$\|\nabla \theta(f)\|^2 = \sum_{i=1}^n \left(\sum_\beta \beta_i c_\beta f^{\beta - e_i} \right)^2 = \sum_{\beta,\gamma} c_\gamma M_{\gamma,\beta}(f) c_\beta$$

where

$$M_{\beta,\gamma}(f) = \sum_{i=1}^n \gamma_i \beta_i f^{\beta+\gamma-2e_i}$$

and where e_i is the standard basis vector with all components zero except for the i-th component which is one. It then follows that the Lipschitz constant of θ is bounded by

$$L_\theta \leq \sup_{f \in \mathrm{dom}(\theta)} \|M(f)\| \, \|\underline{c}\|$$

where the matrix $M(f) \in \mathbb{R}^{J,J}$ has the matrix elements $M_{\gamma,\beta}(f)$.

A simple regularised solution of equation (4) is then

$$\underline{c} = \left(\underline{A}\,\underline{A}^T + \lambda I\right)^{-1} \underline{A}\,\underline{b}.$$

2 A SIMPLE EXAMPLE

To illustrate our new approach we use a simple finite dimensional example. The data is obtained from discretising the Volterra integral equation $\int_0^x u(t) = f(x)$ to get a linear system of equations for the vector of values of u at equidistant grid points. (We denote this vector also by u and similarly for f.) The quantity of interest is the squared L_2 norm of u.

Here $u, f \in \mathbb{R}^d$ such that $Au = f$ where $A \in \mathbb{R}^{d,d}$. Furthermore let $q = \phi(u) = u^T B u$ for some symmetric matrix B. We approximate q by $\theta(f_\delta)$ with $\theta(f) = f^T C f$ and where f_δ is an observed version of f with a data error such that $\|f_\delta - f\| \leq \delta$. The error of this approximation is then

$$|\theta(f_\delta) - \phi(u)| \leq |\theta(f_\delta) - \theta(f)| + |\theta(f) - \phi(u)|.$$

The error relating to the data error is bounded by

$$|\theta(f_\delta) - \theta(f)| \leq 2\|C\|_F \|f\| \delta + \|C\|_F \delta^2$$

and the approximation error is (as $f = Au$)

$$|\theta(f) - \phi(u)| \leq \|A^T C A - B\|_F \|u\|^2.$$

We use a nonlinear optimisation solver to determine C which miminises the objective

$$\Psi(C) = \|A^T C A - B\|_F^2$$

and satisfies the constraint

$$\|C\|_F \leq L$$

for some appropriate L. L is chosen depending on the data error. Computational tests have shown this method to perform as well as the conventional method based on solving $Au = f$ for observed data f_δ using regularisation.

Consider now the special case obtained from discretising the functional $\phi(u) = \int_0^1 u(s)^2 \, ds$ and the Volterra integral equation $\int_0^t u(s) \, ds$ using the midpoint quadrature rule. Specifically, we will consider the matrix

$$B = hI \in \mathbb{R}^{n,n}$$

where $h = 1/n$ and $A \in \mathbb{R}^{n,n}$ is a lower triangular matrix with nonzero elements $a_{i,j} = h$. The problem is then to minimise the Tikhonov functional

$$\frac{1}{2}\|A^T C A - B\|_F^2 + \alpha \|C\|_F^2$$

that is, computing the matrix C for given A and B.

We have implemented this using the singular value decomposition of A:

$$A = U \Sigma V^T.$$

Introducing $X = U^T C U$ and $Z = V B V^T$ we get a transformed functional for X:

$$\|\Sigma^T X \Sigma - Z\|_F^2 + \alpha \|X\|_F^2$$

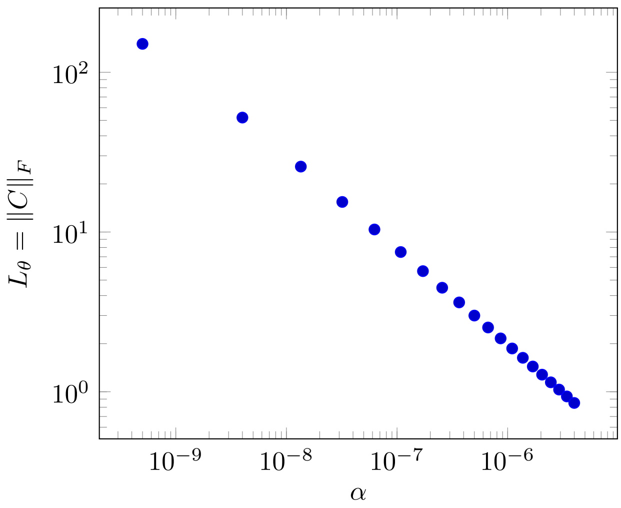

Figure 1. Lipschitz constant as function of α

or with $x = \mathrm{vec}(X)$ and $z = \mathrm{vec}(Z)$ we get a functional for x:

$$\|\Sigma \otimes \Sigma\, x - z\|^2 + \alpha \|x\|^2.$$

The minimiser of this functional is

$$x = \left(\Sigma^2 \otimes \Sigma^2 + \alpha I\right)^{-1} \Sigma \otimes \Sigma\, z.$$

We have solved this for multiple values of α and in figure 1 we display the Lipschitz constant L_θ as a function of the regularisation parameter α. One can see that even with a small regularisation parameter one gets no error amplification (Lipschitz constant one).

3 A FUNCTIONAL STRATEGY FOR GENERAL LIPSCHITZ CONTINUOUS FUNCTIONS – SOME MATHEMATICS

The mathematical theory of the functional strategy relies on functional analysis. In particular, we consider here a Hilbert space H and a dense subspace $V \subset H$. Let $F \in \mathrm{Lip}(H, V)$ be injective and $F(0) = 0$. Note that inverting F is a nonlinear ill-posed problem and in general, the extension of F^{-1} to H is unbounded. Using the natural inclusion one can interpret F as an element of $\mathrm{Lip}(H, H)$ which we will also call F.

Following Benyamini and Lindenstrauss (2000), we introduce the space of Lipschitz continuous functionals

$$H^\sharp = \{\theta \in \mathrm{Lip}(H, \mathbb{R}) : \theta(0) = 0\}.$$

Endowed with the Lipschitz norm

$$\|\theta\|_{H^\sharp} = \sup_{f \neq g} \frac{|\theta(f) - \theta(g)|}{\|f - g\|}$$

the space $H^\sharp$ becomes a Banach space.

We now assume that we are given an F as specified above and a $\phi \in H^\sharp$. The inverse function strategy aims to determine a functional $\theta \in H^\sharp$ such that

$$\theta(F(u)) = \phi(u), \quad \text{for all } u \in H.$$

We now introduce the (linear) pullback operator $F^\sharp : H^\sharp \to H^\sharp$ with

$$(F^\sharp \theta)(u) := \theta(F(u)), \quad u \in H.$$

Notice that $F^\sharp$ is a bounded linear operator as:

$$\|F^\sharp \theta\|_{H^\sharp} \leq \sup_{u \neq v} \frac{\theta(F(u) - F(v))}{\|u - v\|} \leq L_F \|\theta\|_{H^\sharp}.$$

Thus one has the linear ill-posed problem

$$F^\sharp \theta = \phi$$

for the determination of θ given ϕ and $F^\sharp$.

The Hilbert space framework is useful to define a sampling approach to solve this problem. In particular we define a Gaussian measure on H using an appropriate covariance operator and samples $u^{(i)}$. We choose the covariance to model the (smoothness) of our expected functions $u \in H$. We then apply F to get $y^{(i)} = F(u^{(i)})$ for $i = 1, \ldots, m$. Then an approximation to θ is obtained from minimising the objective

$$\Psi(\theta) = \frac{1}{m} \sum_{i=1}^{m} (\theta(y^{(i)}) - \phi(u^{(i)}))^2$$

combined with the constraint on the Lipschitz norm

$$\|\theta\|_{H^\sharp} \leq C.$$

As usual, the least squares sum is the empirical surrogate for the expected squared deviation

$$E\left[(\theta(Y) - \phi(U)^2\right] = \int (\theta(F(u)) - \phi(u))^2 p(u) du.$$

Here U and $Y = F(U)$ are the random variables underlying the samples and $p(u)$ is the normal density function of U. Note that this is a quadratic optimisation problem with a quadratic constraint.

ACKNOWLEDGEMENT

This work was done while the second author was visiting the ANU supported by a French-Australian collaborative grant.

REFERENCES

Anderssen, R. S. (1986). The linear functional strategy for improperly posed problems. In *Inverse problems (Oberwolfach, 1986)*, Volume 77 of *Internat. Schriftenreihe Numer. Math.*, pp. 11–30. Birkhäuser, Basel.

Anderssen, R. S. (2004). Inverse problems: a pragmatist's approach to the recovery of information from indirect measurements. *ANZIAM J. 46*((C)), C588–C622.

Benyamini, Y. and J. Lindenstrauss (2000). *Geometric nonlinear functional analysis. Vol. 1*, Volume 48 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI.

Louis, A. K. (1996). Corrigendum: "Approximate inverse for linear and some nonlinear problems" [Inverse Problems **11** (1995), no. 6, 1211–1223; MR1361769 (96f:65068)]. *Inverse Problems 12*(2), 175–190.

Mathé, P. and S. V. Pereverzev (2002). Direct estimation of linear functionals from indirect noisy observations. *J. Complexity 18*(2), 500–516. Algorithms and complexity for continuous problems/Algorithms, computational complexity, and models of computation for nonlinear and multivariate problems (Dagstuhl/South Hadley, MA, 2000).