

Inverse problems: A pragmatist's approach to the recovery of information from indirect measurements

R. S. (Bob) Anderssen*

(Received 17 January 2005, revised 2 June 2005)

Abstract

Within the class of inverse problems, it is the subclass of indirect measurement problems that characterize the nature of inverse problems that arise in applications. With very few exceptions, measurements only record some indirect aspect of the phenomenon of interest (for example, X-rays and tomographic images in medical applications; telescope images in astronomy; stereological assessment of biological structure and processes; signatures in geophysical prospecting). Even when the direct information is measured such as weight or length, it is measured as a correlation against a standard and this correlation can be quite indirect, such as the measurement of weight by the extension (compression) of a spring.

*CSIRO Mathematical and Information Sciences, GPO Box 664, Canberra, ACT 2601, AUSTRALIA. <mailto:Bob.Anderssen@csiro.au>

See <http://anziamj.austms.org.au/V46/CTAC2004/Ande> for this article, © Austral. Mathematical Soc. 2005. Published July 9, 2005. ISSN 1446-8735

The recovery of information about a phenomenon from indirect measurements is a piecemeal process. Any class of indirect measurements can only recover certain information about the phenomenon. In order to formulate realistic mathematical models that relate the indirect measurements to the specific information from the phenomenon that is to be recovered, there is a need to invoke simplifying assumptions (for example, radial or axial symmetry). The required information about the phenomenon is often only vaguely contained in the available indirect measurements, and this will be reflected in the nature of the mathematical equations which model the relationship between the indirect measurements and the phenomenon of interest.

All these aspects influence how the recovery of the information can be performed. The choice of methodology is not limited. The challenge is to perform the recovery in a way that correctly reflects the underlying nature of the problem context. It is not a matter of blindly applying some form of quadratic regularization for which algorithms and packages are readily available. Though such tools are useful for initial exploratory analysis, the crucial characterization of the information to be recovered is hidden in the mathematical model that relates the indirect measurements to the phenomenon within the problem context.

When recovering information from indirect measurements, the question that focuses the problem solving comes from the need for decision making to have answers to specific matters. The data available for the associated decision support will be indirect measurements of the phenomenon under investigation. As a consequence of the applications context, the recovery of information of the associated inverse problem will be constrained by practical challenges including:

1. In a given situation, how does one decide on the indirect measurements to be performed?
2. How are some practical people able to solve indirect measurement problems without having to perform an explicit regularization?
3. Is there any advantage in combining different indirect measure-

ments of the same phenomenon?

4. What are the alternatives, when there is only a (very) limited amount of data?
5. How does one proceed when a mathematical model is not available or is too complex to formulate?

We examine, in terms of practical problems, how such challenges can be accommodated, as well as explore the wide range of mathematical matters and considerations that arise when solving inverse problems.

Contents

1	Introduction	C591
2	The phenomenology of indirect measurement problems	C593
3	A useful practical classification for indirect measurement problems	C597
3.1	Mathematical models for which the inverse is known explicitly	C597
3.2	Mathematical models where the inverse is not known explicitly	C600
3.3	Indirect measurement problems with no known mathematical model	C602
4	A brief review of key methods	C604
4.1	The linear functional strategy	C606
4.2	Joint inversion	C608
5	A practical perspective: the rheology of grain hardness	C609
6	On the stabilization of the linear functional strategy	C614
	References	C615

1 Introduction

Since the beginning of time, solving inverse problems has played a fundamental role in the biological evolution of life on planet Earth and in the development of human civilization. It is the essential *modus operandi* that underpins the “survival-of-the-fittest” ansatz as the plethora of possible solutions are subject to trial-and-error testing. Both plants and animals, as well as humans, have implicitly exploited various trial-and-error strategies to sort through the possible solutions to a problem to identify a more appropriate subset. For the animals and plants, the motivation has always been adapting to the changing environmental circumstances in order to guarantee the survival of the species. For the plants, a key goal is the storage of food for the next generation (Simmonds [53]). For the animals, a key goal is finding the next meal (Zimmer [61]). Birds solved the problems of flight implicitly, long before humans resolved some of the issues explicitly (Altshuler, Dudley and Ellington [1], Spedding [55]).

For human civilization, the motivation has not only been survival. Others include the quest for the meaning of life and existence, understanding the nature of the physical universe of which we are a part, the control and exploitation of resources and the accumulation of knowledge involving the sorting through of the various possibilities to find a new understanding, interpretation and/or opportunity. Consequently, from a phenomenological perspective, the need to solve an inverse problem arises the moment the problem under consideration involves multiple solutions which must be assessed with respect to some selection criterion to identify optimal choices.

Only in recent times have the implicit trial-and-error strategies been formalized and, where appropriate, placed on a rigorous footing. Historically, identifying and defining the concept of an inverse problem has been an inverse problem. First came the mathematical P^3 -concept of a *properly posed problem* as any mathematical formulation for which the following three conditions hold:

Existence (Performance) — inner consistency which guarantees a non-trivial “realistic” solution.

Uniqueness (Perfection) — only one solution so that there is no ambiguity about the outcome.

Continuous dependence on the data (Parsimony) — inner stability in the sense that small changes in the output (the measured data) correspond to only small changes in the input (the solution).

The intuition behind the alternative P³-concept terminology of Performance, Perfection and Parsimony is to give an informal heuristic conceptualization of the meaning of a properly posed problem for the non-mathematical reader. The idea of “Performance” relates to the need for the problem’s mathematical representation to support sensible solutions, while “Perfection” requires that there be only one solution. The resulting solution must exhibit “Parsimony” in that the data required to generate an approximation to the solution must have a sensible economically “dry” relationship to the solution.

The formal mathematical definition of an inverse problem, as an improperly posed (or ill-posed) problem, followed naturally by saying that any formulation that failed to satisfy existence, uniqueness and/or continuous dependence was an *inverse problem*. Then followed the gradual realization that any problem involving the recovery of information from measurements is an inverse problem because, in one way or another, it will involve some smoothing (indirect encapsulation) of the essential structure of the phenomenon under investigation. The concept of regularization appeared to resolve the situation as it guaranteed a P³-resolution for inverse problems that could be formulated mathematically. In this way, the reality of actual indirect measurements broadened our understanding to the point where the universality of inverse problems is now quite widely appreciated.

There is a comprehensive literature on inverse problems. Sources include the journal *Inverse Problems*, and the books by Anderssen, de Hoog

and Lukas [13], Gladwell [31], Golberg [33], Engl, Hanke and Neubauer [30], Groetsch [34], Morozov [48], Murio [49] and Vogel [58].

The motivation for this paper is an illustration of this universality in terms of a variety of practical indirect measurement problems, and the wide range of mathematical considerations that arise when solving inverse problems. The remainder of the paper has been organized in the following manner. Section 2 gives some background about the nature of indirect measurement problems. A mathematical framework for their classification is the subject of Section 3. Various methodologies are briefly surveyed in Section 4 including the linear functional strategy and joint inversion. An illustration of the challenges that arise, when solving indirect measurement problems, is the subject Section 5 where the importance of the choice and interpretation of rheological measurements for grain hardness is discussed. A deeper aspect of the linear functional strategy is reviewed in Section 6.

2 The phenomenology of indirect measurement problems

Mathematically, the starting point for the definition of a *properly posed problem* is a mathematical relationship (model) between some specific property u of the phenomenon of interest (the solution; the input) and the indirect measurements s (the data; the output). It is normally formalized mathematically as

$$Au = s, \quad A : F_1 \rightarrow F_2, \quad (1)$$

where A denotes some linear or non-linear mapping from some function space F_1 (for example, Banach or Hilbert) to the same or another function space F_2 .

Definition 1 *The mathematical model $Au = s$ corresponds to a properly posed problem, if*

1. *it has at least one solution (existence),*
2. *it has at most one solution (uniqueness) and*
3. *the data (output) depends continuously on the solution (input) (continuous dependence).*

Definition 2 *The mathematical model $Au = s$ corresponds to an improperly posed problem, if it fails to satisfy at least one of the conditions for it to be a properly posed problem.*

Phenomenologically, one must turn to specific examples to obtain the conceptualization and intuition (compare with Anderssen [6], Engl, Hanke and Neubauer [30], Groetsch [34]). Popular choices include numerical differentiation, the Riemann–Lebesgue Lemma, first kind Fredholm integral equations with smooth kernels and first kind Volterra integral equations such as the Abel integral equation.

In the theoretical and mathematical literature, attention tends to focus on the failure of continuous dependence by assuming that the range $\underline{R}(A)$ of the operator A is not closed. The theory for linear operators is quite substantial and comprehensive (Engl, Hanke and Neubauer [30]), whereas that for non-linear operators is still under development and involves many deep and novel challenges and considerations (Burger and Osher [23]).

A quite simple illustration of the failure of existence is any situation where the model and the output data are inconsistent such as occurs when half the base multiplied by the height of a triangle does not equal the given data about its area. At one level, statistics is concerned with such situations where the inconsistency is due to measurement errors and is not a fundamental failure in the formulation of the model and problem. Measurement error inconsistency is resolved by working with over-determined systems where there are more observations (measurements) than unknowns.

Non-uniqueness arises when the formulated model is under-determined, when there are more unknowns than observations. This is often the situation in the formulation of practical problems, especially when the observations are expensive to make (such as in astronomical observations of distant stars, and in medical dose response data) or correspond to specific functionals defined on some continuous representation of the phenomenon of interest (such as the mass and moment-of-inertia of a sphere with a radially symmetric density distribution). In such situations, practitioners often implicitly orchestrate stabilization by modelling the solution as a parametric model with a very small number of unknowns (for example, less than 5) so that the initial under-determined formulation becomes a highly constrained over-determined system. In this way, the inherent improperly posedness is circumvented without its existence being acknowledged.

An instructive illustration of the fundamental nature and consequences of the failure of continuous dependence is encapsulated in the Riemann–Lebesgue Lemma which states

$$\lim_{t \rightarrow \infty} \int_I u(x) \exp(itx) dx \rightarrow 0,$$

where u denotes an arbitrary Lebesgue integrable function and I is any interval of the real line. It clearly formalizes a number of quite generic and important points:

1. The solution of an improperly posed problem (taking the solution to be the order one $\exp(itx)$ function) does not necessarily depend continuously on the measured data (the value of the integral).
2. The structure of (information in) the solution (now taking the solution to be $u(x)$) of an improperly posed problem can have a very tenuous link to the measured data (the value of the integral).
3. The conclusions [1](#) and [2](#) are independent of the domain (the interval I) over which the relationship between input and output (the integration) holds.

The most popular illustration of the pervasive and practical consequences of the failure of continuous dependence is numerical differentiation. Its importance is that it illustrates the effect of the measurement errors on the numerical solution of improperly posed problems, since, in any practical situation, the available data d_i , $i = 1, 2, \dots, n$, is not continuous but discrete. In some situations, it will be a mixture of discrete functionals (for example, mass and moment-of-inertia) and observations of a continuous process at a discrete set of spatial-temporal locations. When the data values measure an underlying continuous signal $s(t)$ at some discrete set of locations, they will take the form

$$d_i = s(t_i) + \epsilon_i, \quad t_1 < t_2 < \dots < t_n,$$

where the ϵ_i denote the actual measurement errors.

In this way, numerical differentiation formalizes the following two generic points, which supplement and complement 1, 2 and 3 above:

4. Small measurement errors can be the source for unacceptable perturbations in the solution of an inverse problem.
5. Except when limited parametrization can be perceptively applied (for example, fit a quadratic curve to the data and differentiate the resulting quadratic equation), discretization of an improperly posed problem will tend to accentuate the improperly posedness, and will certainly not improve it.

A continuous data $d(t)$ conceptualization with

$$d(t) = s(t) + A \sin \omega t,$$

where the error is modelled as $A \sin \omega t$, pin-points the situation on observing that the derivative takes the form

$$\dot{d}(t) = \dot{s}(t) + A\omega \cos \omega t.$$

It illustrates that, even when the information about $s(t)$ is quite strongly captured in $d(t)$, because the amplitude A of the perturbation is minute, the information about $\dot{s}(t)$ will be very weakly captured in $\dot{d}(t)$, when the value of the frequency ω is considerably greater than the value of A .

Another important aspect of numerical differentiation is that it can be used to conceptualize the role of averaging in the solution of indirect measurement problems (compare with Anderssen and de Hoog [9], Anderssen, de Hoog and Hegland [11] and Anderssen and Hegland [12]).

3 A useful practical classification for indirect measurement problems

A cursory survey of the inverse problems literature quickly indicates that the way that an inverse problem is solved depends heavily on the mathematical properties of the relationship that defines the dependence of the output on the input. It therefore represents a way of classifying inverse problems. It also yields a clearer conceptualization of the diverse nature of inverse problems from a mathematical and problem-solving perspective. Within the higher level classification given below, one could invoke various subclassifications.

3.1 Mathematical models for which the inverse is known explicitly

The classical examples of this class of improperly posed problems are the first kind Volterra integral equations known as the Abel (integral) equations which take various forms including (Anderssen and de Hoog [10])

$$s(y) = \int_y^a (x^2 - y^2)^{-1/2} u(x) dx .$$

For this form, the following two inversion formulas (as well as others) hold

$$u(x) = -\frac{1}{\pi} \int_x^a (y^2 - x^2)^{-1/2} \frac{ds(y)}{dy} dy,$$

and

$$u(x) = -\frac{1}{\pi} \frac{d}{dx} \left\{ \int_x^a (x^2 - y^2)^{-1/2} y s(y) dy \right\}.$$

Interestingly, Abel equations arise naturally in a wide variety of applications for quite independent mathematical reasons. For example, in physics (compare with Groetsch [34], Engl, Hanke and Neubauer [30]), the input $u(x)$ corresponds of a radial symmetric density structure in a cylinder whereas the output $s(y)$ corresponds to its X-ray projection onto a plane (photographic) plate. In stereology (compare with Kendall and Moran [37], Moran [47], Jakesman and Anderssen [36]), for a slightly different version of the above Abel equation, the input $u(x)$ corresponds to the size distribution of spheres distributed randomly in three dimensions while the output $s(y)$ corresponds to the size distribution of the circles observed on random plane section through the three dimensional structure.

Analytically, the solution of the Abel integral equation corresponds to a half differentiation of the output $s(y)$. The importance in having different inversion formulas is that each can play a different role from both theoretical and numerical analysis perspectives. In particular, because the measured data $\{d_i\}$ are often observed on a non-even grid, the second inversion formula is the more appropriate because the indefinite integration can be used to integrate the observational data $\{d_i\}$ onto an even grid so that standard even-grid numerical differentiation algorithms can be applied.

All the different inversion formulas for an Abel equation have the same intrinsic degree of improperly posedness (namely, the half differentiation). In fact, the order of the numerical differentiation (including fractional), to which the inversion of the relationship (1) corresponds, is often exploited to compare the degree of improperly posedness of different indirect measurement

problems. As explained above, as well as being the basis for proving various theoretical results, the various inversion formulas represent the means for adapting the actual inversion for given observational data to best fit the particular situation being modelled and analysed. However, the essential nature of the improperly posedness does not change as it is the mathematical essence of the relationship between the input and output of the process under investigation. Anderssen and de Hoog [8] discuss a situation where it was claimed that one inversion formula was better posed than the others.

Various integral equations have inversion formulas. Some remain undiscovered. Unfortunately, there is no single compendium. Such information is scattered over a range of sources. A useful reference is Schmeidler [52].

An interesting Australian illustration is given by the foliage angle distribution model (Anderson [2, 3], Warren Wilson [60], Smith et al. [54]):

$$\begin{aligned}
 f(\beta) &= \int_0^{\pi/2} k(\alpha, \beta)g(\alpha) d\alpha, \quad 0 \leq \alpha \leq \pi/2, \\
 k(\alpha, \beta) &= \begin{cases} \cos \alpha \sin \beta, & \alpha \leq \beta, \\ \cos \alpha \sin \beta [1 + 2/\pi(\tan \theta - \theta)], & \alpha \geq \beta, \end{cases} \\
 \theta &= \cos^{-1}(\tan \beta / \tan \alpha), \quad 0 \leq \theta \leq \pi/2.
 \end{aligned}$$

It was formulated in Australia by Warren Wilson [60] and colleagues. The existence of the following inversion formula was established by the Australian mathematician John Miller [46]:

$$\begin{aligned}
 g(\alpha) &= \tan \alpha \sec^3 \alpha \int_{\alpha}^{\pi/2} \frac{3 \cos^2 \tau \sin \tau (f(\tau) + f''(\tau))}{(\tan^2 \tau - \tan^2 \alpha)^{1/2}} \\
 &\quad - \frac{\cos^3 \tau (f'(\tau) + f'''(\tau))}{(\tan^2 \tau - \tan^2 \alpha)^{1/2}} d\tau.
 \end{aligned}$$

On the basis of an analogy with the existence of the second inversion formula for the Abel equation (given above), Anderssen and Jackett [14]

derived the alternative and more compact formula

$$g(\alpha) = -\sec \alpha \frac{d}{d\alpha} \left\{ \int_{\alpha}^{\pi/2} \frac{\sin \beta (f(\beta) + f''(\beta))}{(\tan^2 \beta - \tan^2 \alpha)^{1/2}} d\beta \right\},$$

and subsequently developed numerical methods which exploited this structure for the implementation of the linear functional strategy (Anderssen and Jackett [15], and Anderssen, Jackett and Jupp [16]).

I stress that the existence of an inversion formula does not necessarily guarantee the basis for a good numerical algorithm. The existence of alternative analytic results can be exploited in various ways. For example, for some integral equations, even when no inversion formula is known, sets of basis functions for $u(x)$ and $s(y)$, complete in some function space setting, are known. Such information can be utilized to construct pseudo-analytic method algorithms (Anderssen and de Hoog [10]), and apply adjoint operator methods (compare with Golberg [32] and Donoho [29]).

3.2 **Mathematical models where the inverse is not known explicitly**

The classical examples are the first kind integral equations of Fredholm and Volterra (Cochran [25]), where the degree of improperly posedness is controlled by the smoothness of the kernel, which, in turn, is controlled by the rate at which the associated singular values decay to zero (Engl, Hanke and Neubauer [30]). Because a survey of the relevant theory for such integral equations is beyond the scope of the current review, they are best illustrated using representative examples.

A first kind Volterra convolution integral equation from rheology is the Boltzmann causal integral equation model of linear viscoelasticity (Boltz-

mann [24]):

$$\sigma(t) = \int_{-\infty}^t G(t - \tau) \dot{\gamma}(\tau) d\tau = \int_0^{\infty} G(\tau) \dot{\gamma}(t - \tau) d\tau, \quad \dot{\gamma}(t) = \frac{d\gamma(t)}{dt},$$

where $\sigma(t)$ denotes the stress measured at time t in response to a previously applied strain-rate $\dot{\gamma}(\tau)$ with the nature of the viscoelasticity encapsulated in the structure of the relaxation modulus $G(t)$. This equation is of particular interest as the relaxation modulus $G(t)$, in order that this equation encapsulates sensible physics, must have *fading memory* in the sense that changes in the distant past have less effect now than the same changes more recently. The popular choice is to assume that $G(t)$ is a completely monotone function, and thereby involves quite deep mathematical considerations (Anderssen and Loy [18, 19]). A specific choice is the Kohlrausch function $\exp(-t^\beta)$, $0 < \beta \leq 1$, which has sufficient structure to allow some quite interesting analytic results to be derived (Anderssen, Husain and Loy [17], Husain and Anderssen [35]).

A first kind Fredholm integral equation from rheology is the molecular weight distribution mixing rule (Mead [45], Thimm et al. [56]) which Anderssen and Loy [18, 19] have generalized to

$$G(t) = G_N^0 \left(\int_0^{\infty} \exp(-\theta(t)/\tau(m)) w(m) dm \right)^r, \\ \dot{\theta} \sim \text{completely monotone}, \quad r \geq 1,$$

where G_N^0 is the plateau modulus constant, $G(t)$ the above mentioned relaxation modulus, and $w(m)$ the molecular weight distribution of the linear viscoelastic material for which $G(t)$ has been determined. The exponent r is a characterization of the nature of the polymer dynamics that occurs when a deformation is applied to a material, with $r = 1$ corresponding to single reptation (Doi and Edwards [28]), and $r = 2$ to double reptation (des Cloizeaux [27] and Tsenoglou [57]). Whether r can be fractional remains an open question in terms of how the above equation should be interpreted from

a polymer dynamics perspective. The function $\theta(t)$ was introduced by Andersen and Loy [19] to allow theoretically for different types of fading memory in the molecular weight distributions' response to an applied deformation.

3.3 Indirect measurement problems with no known mathematical model

In many modern applications, such as the near infrared (NIR) spectroscopic prediction of the protein content of wheat, though measurements of the input $u(x)$ and the output $s(y)$ are available, the relationship connecting them is either unknown or too complex to formulate as a set of mathematical equations. In such situations, various forms of the calibration-and-prediction (CAP) ansatz (Martens and Naes [40] and Naes et al. [50]) have been successfully utilized to perform the underlying inversion.

The essence of the CAP ansatz can be explained in the following manner:

The problem context. One is given: (a) samples of some material M , for which the values of some property P are required, and (b) a rapid and inexpensive experimental protocol for determining some related property Y of the material which is, in one way or another, an indirect encapsulation of the property P .

The experimental framework for the model formulation. The *modus operandi* is to measure, for each member of a representative set S of samples of the material, the indirect characterizations Y (the NIR spectra) along with the corresponding values Z of the property (the protein content of wheat).

The Goal. Determine a relationship r between the Y 's and the Z 's such that, for the indirect measurement Y^* of a new sample S^* , this relationship becomes the predictor for the corresponding value Z^* of the required property.

The essential CAP methodology The application of the CAP methodology is a two-step process which involves:

The model formulation as a calibration. Model (linearly) the direct measurements Z as a function of the indirect measurements Y to generate the predictor r for subsequent utilization.

The solution process as a prediction. For the new sample S^* , measure Y^* and substitute into the relationship r to determine the corresponding value Z^* .

The essential algebra of the calibration. Clearly, the key step is the calibration. With the indirect measurements

$$Z_j = \theta_j + \epsilon_j, \quad j = 1, 2, \dots, n,$$

modelled as a vector Z , where the θ_j denote the true values of the property P and the ϵ_j the corresponding measurement errors, and with the indirect measurements

$$Y_{jk} = g_j(\lambda_k) + \delta_{jk}, \quad k = 1, 2, \dots, m, \quad m \ll n,$$

modelled as a matrix Y with n rows and m columns, where the g_j denote the exact spectra and the λ_j the wavelengths at which the indirect measurements are recorded, the linear relationship r takes the form

$$Y\beta = Z,$$

where the vector β becomes the parametric characterization of r .

The solution ansatz. Though a wide variety of methods have been proposed (Martens and Naes [40] and Naes et al. [50]), many of the successful algorithms use a matrix Q of regression factors to determine the unknown β ; namely,

$$Y = QP, \quad Z = Q\beta.$$

The prediction. For the spectrum Y^* (vector) of a new sample S^* , the corresponding value Z^* of the property P is given by the predictor as

$$Z^* = \beta^T Y^* .$$

Except for Anderssen, Osborne and Wesley [21], the improperly posed nature of CAP has not been examined in the detail that it merits.

4 A brief review of key methods

Regularization, in the sense of a simultaneous minimization of some measure of the residual error (between the measured and the model predicted output) plus some constraint on the smoothness of the solution, is the most popular methodology from both theoretical and practical perspectives. This type of regularization has the canonical form

$$\min_{u \in \mathcal{H}} \{ \|Au - s\|_*^2 + \alpha \Lambda(u) \} , \quad \alpha > 0 ,$$

where α denotes the regularization parameter, which determines the trade-off between the minimization of the residual error and the smoothness constraint, $\|\cdot\|_*$ an appropriate norm in which to assess the size of the residual error, $\Lambda(u)$ is either a norm or seminorm, and \mathcal{H} defines the admissible set over which the minimization is to be performed. A popular choice for $\Lambda(u)$ is the norm $\|Tu\|_{**}^2$, where T denotes a differential operator, the null space of which controls the properties of the strongly smoothed regularized solutions. Normally, the smoothness norm $\|\cdot\|_{**}$ is chosen to be the same as the residual norm $\|\cdot\|_*$, but this is not mandatory.

Special cases include:

1. Tikhonov, when the operator T is second order differentiation and the two norms are L_2 function space norms (Engl, Hanke and Neubauer [30]);

2. the fitting of cubic splines of given data (non-parametric curve fitting), when $A = I$ is the identity operator, $\|\cdot\|_*$ an appropriate discrete norm, and the smoothness constraint corresponds to that for Tikhonov regularization (Wahba [59]);
3. generalized maximum entropy, in the sense proposed by Landl and Anderssen [39], where the smoothness constraint takes the form

$$\int_{\Omega} Tu \ln(Tu) dx ,$$

where T is a differential operator and Ω the domain on which the solution u is defined.

The term “regularization” is often used more generically to simply mean some form of stabilization. Here it is used in the more restricted sense outlined above. Because of the strong control exerted by the smoothness criterion, regularization will, in general, yield a unique solution. As a direct consequence, the smoothness criterion should be representative of the problem context and not chosen on the basis of some convenient or philosophical expedient. For example, the choice of maximum entropy regularization should be based on the type of considerations outlined in Landl and Anderssen [39] and not on purely philosophical grounds. Koch and Anderssen [38] have given an example where maximum entropy regularization is not the appropriate choice, by showing that the strongly regularized solution will asymptotically approach zero. Indirectly, such results identify the importance, in the study of regularization, of the need to understand the small and large regularization parameter asymptotics of the regularized solution.

From a pragmatic perspective, regularization should be viewed as an exploratory tool, which allows one to investigate the structure of an indirect measurement problem from both theoretical and practical aspects. In a way, the more important methodologies are the ones that focus on the nature of the information required for decision-support, as examined below in Section 4.1. In many applications, it is not the solution of the inverse problem

that is required for decision support, but some property of the solution which often takes the form of a linear functional (Anderssen [4, 5, 6]). The direct evaluation of such functionals represents a form of stabilization known as the linear functional strategy. Furthermore, as illustrated in McNabb and Wake [43] and McNabb et al. [44], the importance of identifying the linear functionals required by decision support extends more widely than simply the stabilized solution of inverse problems.

In other applications, different types of measurements can or have been performed on the same input and the utilization of this joint information again represents a form of stabilization which is discussed in Section 4.2 under the heading of “Joint Inversion”.

4.1 The linear functional strategy

In some ways, the origins of the linear functional strategy can be traced back to the method of adjoints applied to the solution of operator equations (Golberg [32]), such as equation (1) with $F_1 = F_2 = H$, where H is a Hilbert space. The starting point is the fact that, with respect to some complete orthonormal system $\{\psi_j\}$ in the Hilbert space H with inner product and norm (\cdot, \cdot) and $\|\cdot\|$, the input u will take the form

$$u = \sum_{j=1}^{\infty} (u, \psi_j) \psi_j.$$

On assuming that, with respect to the adjoint A^* of A , there exists a complete set of functions ϕ_j such that

$$A^* \phi_j = \psi_j,$$

and on utilising the duality relationship $(u, A^* \phi_j) = (Au, \phi_j)$, which corresponds to a change in the order of integration, the previous equation becomes

$$u = \sum_{j=1}^{\infty} (Au, \phi_j) \psi_j = \sum_{j=1}^{\infty} (s, \phi_j) \psi_j.$$

The advantage and appeal of this result is the fact that, if the $\{\phi_j\}$ are known for a representative set of $\{\psi_j\}$, then the determination of this representation for the solution reduces to the evaluation of the coefficients $\{(s, \phi_j)\}$ on the output data.

The *linear functional strategy* represents a specialization of this result in the sense that only one coefficient (a linear *solution-functional*) (u, ψ) of the input u is required for the decision support. For the first kind integral equation

$$Au = \int_a^b k(x, \tau)u(\tau) d\tau = s(x),$$

if one starts with the data-functional

$$L_\phi(s) = \int_a^b \phi(x)s(x) dx = (\phi, s),$$

replaces the output s with the above integral transformation of the input u , and changes the order of integration, then the data-functional transforms to

$$L_\phi(s) = \int_a^b \left(\int_a^b \phi(x)k(x, \tau)dx \right) u(\tau) d\tau = \int_a^b \theta(\tau)u(\tau) d\tau,$$

with

$$\theta(\tau) = \int_a^b k(x, \tau)\phi(x) dx = A^*\phi.$$

The corresponding solution-functional is then

$$L_\theta(u) = \int_a^b \theta(\tau)u(\tau) d\tau.$$

This result has the following dual interpretation:

The Forward Properly Posed Aspect: Given $\phi(x)$, evaluate the integral defining $\theta(\tau)$. Here, the only regularity is the requirement that $\phi(x)$, which characterizes the data-functional, and $k(x, \tau)$ are such that the integral defining $\theta(\tau)$ exists.

The Inverse Improperly Posed Aspect: Given $\theta(\tau)$, solve $A^*\phi = \theta$ to find $\phi(x)$. Here, the regularity constraint is quite crucial as $A\phi = \psi$ will only be solvable when

$$\psi \in \underline{\mathbf{R}}(A^*).$$

A summary of the advantages of the linear functional strategy is:

1. Though solving $A^*\phi = \theta$ is an improperly posed problem, the output $\theta(\tau)$ is now an analytic function.
2. Both θ and $k(x, \tau)$ are known analytic functions.
3. The evaluation of $L_\phi(s) = (\phi, s)$ reduces to an appropriate smoothing of the data s , which is of a great advantage when s is only available as discrete noisy observations (Mathe and Pereverzev [42]).
4. It avoids computing functionals on the solution for which there are no counterpart as functionals on the data. However, as shown by Davies and Anderssen [26], the linear functional strategy can be generalized to situations where the above regularity condition does not hold through an appropriate mollification strategy. This will be examined in some detail in Section 6.

4.2 Joint inversion

Another way to induce stability is to limit the size of the class of possible solutions. As noted above, regularization achieves this by limiting the class to a single unique solution. An alternative approach, which only reduces the size of the class of non-unique possibilities, is to simultaneously utilize different types of indirect measurements of the same phenomenon of interest. A simple illustration is when both the mass and moment of inertia of a distant planet are used to infer information about its internal density structure.

In the geophysical literature, Jupp and Vozoff (1975) introduced the terminology *joint inversion* when they examined the advantage of including both electric and magnetic induction data in the recovery of electrical conductivity structure.

Such ideas are also applied in medical applications. A representative example is the joint utilization of emission and transmission data in order to improve the accuracy of SPECT imaging (compare with Bailey, Hutton and Walker [22]).

Anderssen, Carter, Osborne and Wesley [20] introduced the concept of leap frog calibration-and-prediction in order to extend joint inversion of CAP situations. In their work, they successfully applied this methodology to the joint inversion of NIR and Raman spectroscopic measurements of wheat to predict various properties such as protein and moisture content as well as the gliadin-glutenin ratio.

5 **A practical perspective: the rheology of grain hardness**

Many practical inverse problems do not fall naturally into the framework of equation (1). The role of calibration-and-prediction (CAP) has been discussed above as the appropriate *modus operandi* when it is not possible to formulate a mathematical model for the relationship between the input and output. A different situation arises when the goal is to determine the best way to characterize some concept in an appropriately explicit manner, in order to construct a suitable framework in which to investigate, analyse and resolve some practical matter, and/or place some practical deliberations on a firm and rigorous footing. In a way, such situations correspond to the *conceptualization* that normally precedes the formulation of a mathematical model. The fact that the formulation of mathematical models is not a

framework free process is often overlooked in the discussion of mathematical modelling. In fact, the construction of such a framework is itself an inverse problem. As a consequence, such situations must be examined in their own right in terms of the available instrumentation that can perform the indirect measurements, the interpretation of the information that is contained in the associated indirect measurements, the nature of the practical problem under investigation, the specific questions to be investigated and answered, and the history of previous mathematical modelling.

Such a situation arises in the recovery of information about wheat grain hardness from rheological measurements. Because grain hardness is genetic, a historic fact which dates back many centuries, it plays the role of a link concept between end-product quality (for example, volume of a bread loaf and soft texture of a biscuit) and plant breeding.

Historically, research into the assessment of wheat blends led to the development of the Single Kernel Characterization System (SKCS) instrument and the associated software (Martin, Rousser and Brabec [41], Osborne and Anderssen [51]). From the measurement of the crushing of a large number of wheat kernels (~ 300), the software determines a Hardness Index (HI) value. A reading of the literature (Martin, Rousser and Brabec [41]) establishes that the underlying mathematical modelling is based on a multivariate analysis prediction formula applied, with the assistance of a Gompertz model

$$s(x) = A^{B^x}, \quad 0 < A, B \leq 1,$$

applied to the histogram of incremental changes (Figure 1) occurring during their crushing. Interestingly, the Gompertz model was formulated in 1825 for the modelling of the population of New York, because, for $0 < A, B < 1$, $s(x)$ is a normalized cumulative distribution, and because the estimation of the parameters A and B can be reduced to the use of logarithms and the fitting of a straight line to the resulting logarithmic data. This background does not represent the reasons for the success of the HI values in predicting grain hardness in a consistent manner. On reflection, it is clear that the reason is

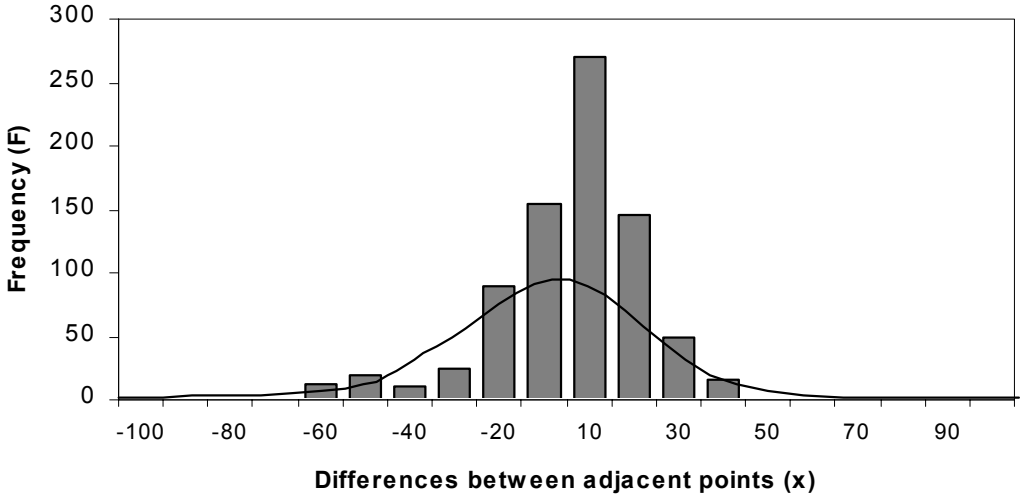
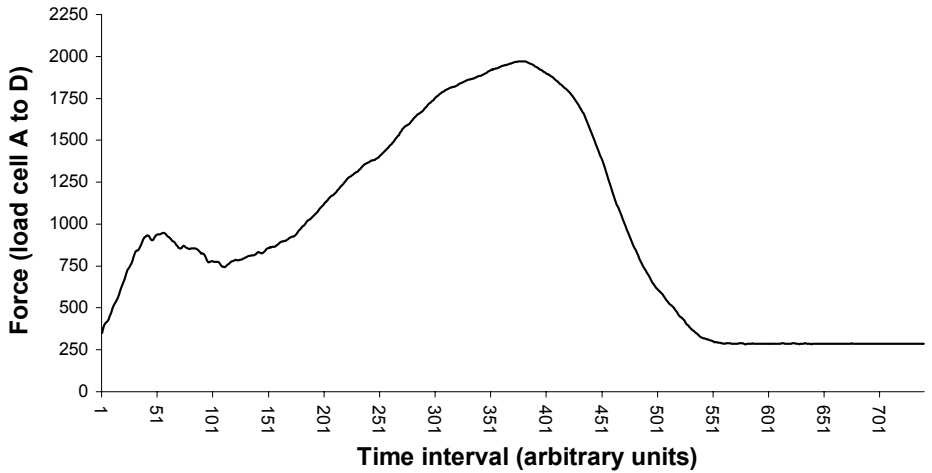


FIGURE 1: Incremental Change Histogram.

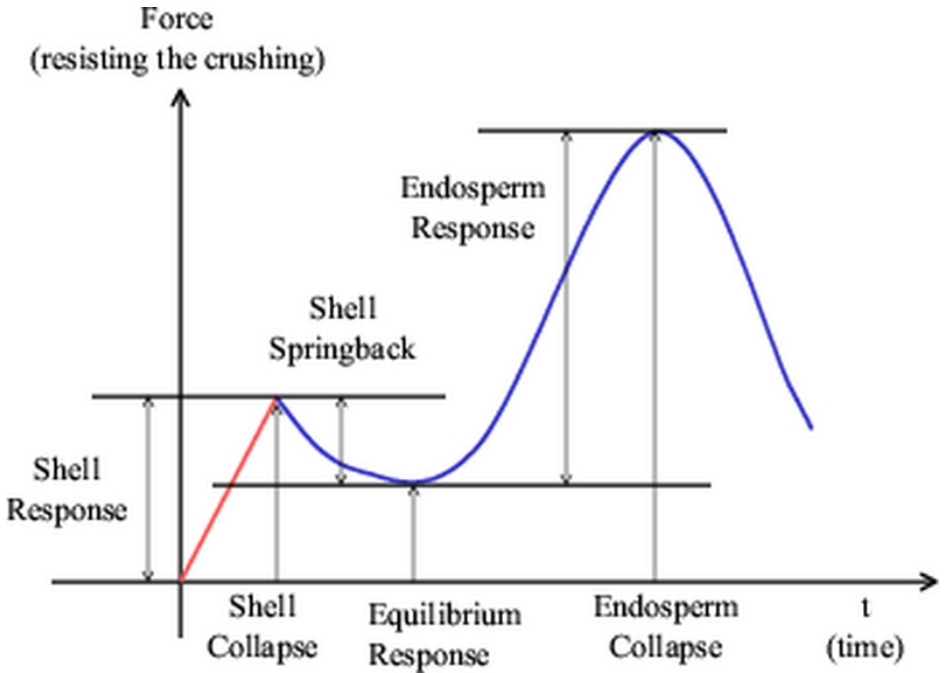
not the fancy analysis applied to the incremental change histogram, but the more basic fact that grain hardness is so definite that even the methodology implemented in the SKCS software sees a difference.

In the algorithm that calculates the HI value, all of the specific rheological information about the crushing of the various layers within the wheat has been lumped into the incremental change analysis and lost. The SKCS development in 1963 was clearly ahead of its time, and the nature of microprocessors that could be coupled to the instrumentation limited the sophistication of the software development. Because the crushing of each kernel was being recorded as part of the determination of the incremental changes, the rheology of the crushing could be recovered as a stress-time (stress-pseudostrain) curve (crush response profile (CRP)). The average CRP of three hundred such curves is shown in Figure 2(a).

It is clear from a comparison of Figures 2(a) and 2(b) that the rheological phases in the structure of the CRPs correlate with the successive crushing of



(a)



(b)

FIGURE 2: (a) Average crush response profile for 300 wheat kernels. (b) The generic structure of the crush response profile of a wheat kernel.

**Elastic
Response
of the Shell**

**Viscoelastic
Response of
the
Compressed
Endosperm**

**Collapse of
the Shell
and Internal
Porosity**

CSIRO Mathematical and Information Sciences

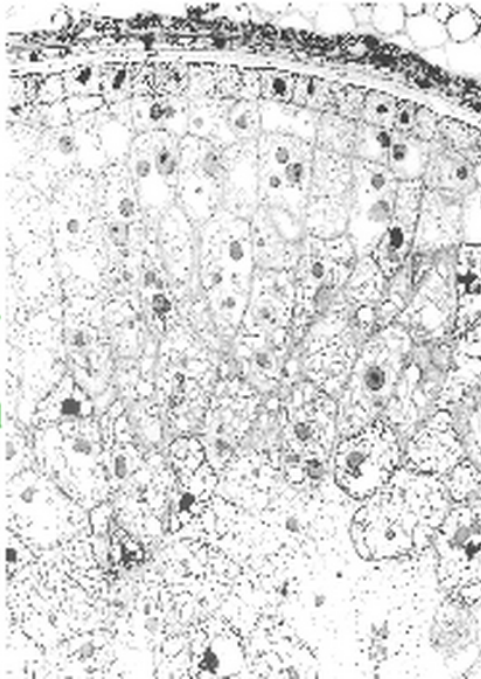


FIGURE 3: Relationship of crush response phases to the internal structure in wheat.

the structure in a wheat kernel. Consequently, the analysis of grain hardness should be performed with respect to the information in the various phases of the crushing (Osborne and Anderssen [51]). Such phases represent quite specific phenotypes which can be given molecular interpretations. This leads naturally to the concept of “molecular rheology” as an appropriate way for relating quantitative concepts, which can be modelled mathematically, to molecular processes.

6 On the stabilization of the linear functional strategy

Without loss of generality, attention is restricted to operators A that map a Hilbert space H into itself ($F_1 = F_2 = H$ in equation (1)) with $\underline{R}(A)$, and hence $\underline{R}(A^*)$, in H .

A very excellent theory has been developed by a number of authors (Donoho [29], Mathe and Pereverzev [42]) to explain the extent and limitations of the stabilization of the linear functional strategy. In general, this theory only applies to solution functionals (u, θ) for $\theta \in \underline{R}(A^*)$. Even when $\theta \notin \underline{R}(A^*)$, a meaning can be given to the corresponding solution functional by exploiting the properties of weak convergence. For a given $\theta \notin \underline{R}(A^*)$, a sequence $\{\theta_n\}$ contained in $\underline{R}(A^*)$ is chosen which converges strongly to θ in H . The fact that strong convergence automatically guarantees weak convergence ensures that

$$\gamma_n = (u, \theta_n)$$

defines a convergent sequence with limit $\gamma = (u, \theta)$. However, from the perspective of the implementation of the linear functional strategy, it is not $\gamma_n = (u, \theta_n)$ that is evaluated, but

$$\gamma_n = (f, \phi_n), \quad A^* \phi_n = \theta_n.$$

The existence of ϕ_n is automatically guaranteed because $\theta_n \in \underline{R}(A^*)$. However, the convergence of the $\{\phi_n\}$ to a limit is quite problematic, because $\theta \notin \underline{R}(A^*)$ implies that the corresponding limit ϕ , if it exists, will be in some larger space and, on occasions, in some quite esoteric space.

An illustration of this fact can be found in the paper by Davies and Anderssen [26] which examines the evaluation of the partial viscosities from oscillatory shear data. As a consequence, as outlined in Anderssen and Davies [7] and Davies and Anderssen [26], special algorithms must be constructed if one wishes to work directly with such ϕ .

In a real application, the output will have been measured only at a finite set of points on a finite interval of the independent variable. It is therefore necessary to understand how this modifies theoretically the solution functional. In part, this can be analysed by looking at the solutions of $A^*\phi = \theta$, where either ϕ or θ is the box-car function $B_{(a,b)}(x)$ which takes the value one on the interval $[a, b]$ and is zero outside. Again, there are forward and inverse aspects to this problem:

The Forward Aspect. The mapping $A^*B_{(a,b)} \rightarrow \theta_B$ determines how the information on the interval of definition of the input is redistributed into the output. For smooth A^* , this mapping is properly posed.

The Inverse Aspect. The mapping $A^*\phi_B \rightarrow B_{(a,b)}$ must be inverted in order to determine how the information on the interval of definition of the output has come from the input. For smooth A^* , this mapping is improperly posed.

An illustration of this fact can be found in the paper by Davies and Anderssen [26] where the sampling localization associated with oscillatory shear data in the recovery of a relaxation spectrum is resolved.

References

- [1] D. L. Altshuler, R. Dudley and C. P. Ellington, Aerodynamic forces of revolving hummingbird wings and wing models, *J. Zoology* 264 (2004), 327–332. <http://dx.doi.org/10.1017/S0952836904005813> C591
- [2] M. C. Anderson, Studies of the woodland climate. I. The photographic computation of light conditions, *J. Ecol.* 52 (1964), 27–41. C599

- [3] M. C. Anderson, Radiation and crop structure, in *Plant Photosynthetic Production, Manual of Methods* (Eds J. Catsky and P. G. Jarvis), Junk, The Hague, 1971, p. 412–466. [C599](#)
- [4] R. S. Anderssen, On the use of linear functionals for Abel-type integral equations in applications, in *Applications and Numerical Solution of Integral Equations* (Eds. R. S. Anderssen, F. R. de Hoog and M. A. Lukas), Sijthoff and Noordhoff, The Netherlands, 1980, pp. 195–221. [C606](#)
- [5] R. S. Anderssen, The linear functional strategy for improperly posed problems, in *Inverse Problems* (Eds. J. R. Cannon and U. Hornung; Proceedings of the Oberwolfach Conference, May, 1986), Birkhäuser Verlag, Basel, 1986, pp. 11–30. [C606](#)
- [6] R. S. Anderssen, The pragmatics of solving industrial (real-world) inverse problems with exemplification based on the molecular weight distribution problem, *Inverse Problems* 15 (1999), R1-R40. [C594](#), [C606](#)
- [7] R. S. Anderssen and A. R. Davies, Simple moving-average formulae for the direct recovery of the relaxation spectrum, *J. Rheol.* 45 (2001), 1–27. <http://dx.doi.org/10.1122/1.1332787> [C614](#)
- [8] R. S. Anderssen and F. R. de Hoog, On the method of Chan and Lu for Abel's integral equation, *J. Phys. A: Math. Gen.* 14 (1981), 3117–3121. [C599](#)
- [9] R. S. Anderssen and F. R. de Hoog, Finite difference methods for the numerical differentiation of non-exact data, *Computing* 33 (1984), 259–267. [C597](#)
- [10] R. S. Anderssen and F. R. de Hoog, Abel integral equations, in *Numerical Solution of Integral Equations* (Ed. M. A. Golberg), Plenum Press, New York, 1990, pp. 373–410. [C597](#), [C600](#)

- [11] R. S. Anderssen, F. R. de Hoog and M. Hegland, A stable finite difference ansatz for higher order differentiation of non-exact data, *Bull. Austral. Math. Soc.* 58 (1998), 223–232. C597
- [12] R. S. Anderssen and M. Hegland, For numerical differentiation, dimensionality can be a blessing! *Math. Comp.* 68 (1999), 1121–1141. C597
- [13] R. S. Anderssen, F. R. de Hoog and M. A. Lukas, *The Application and Numerical Solution of Integral Equations*, Sijthoff & Noordhoff, Alphen aan den Rijn, The Netherlands, 1980. C593
- [14] R. S. Anderssen and J. R. Jackett, Linear functionals of the foliage angle density, *J. Aust. Math. Soc., Series B* 25 (1984), 431–442. C599
- [15] R. S. Anderssen and D. R. Jackett, Use of the functional strategy for assessing the status of plant canopies, Chapter 7 in *Industrial Numerical Analysis* (Editors: S. McKee and C. M. Elliott), Clarendon Press, Oxford, 1986. C600
- [16] R. S. Anderssen, D. R. Jackett and D. L. B. Jupp, Linear functionals of the foliage angle distribution as tools to study the structure of plant canopies, *Aust. J. Bot.* 32 (1984), 147–156. C600
- [17] R. S. Anderssen, S. A. Husain and R. J. Loy, The Kohlrausch function: properties and applications, *ANZIAM J.* 45(E) (2004), C800–C816. <http://anziamj.austms.org.au/V45/CTAC2003/Ande> C601
- [18] R. S. Anderssen and R. J. Loy, Completely monotone fading memory relaxation moduli, *Bull. Austral. Math. Soc.* 65 (2002), 449–460. C601
- [19] R. S. Anderssen and R. J. Loy, Rheological implications of completely monotone fading memory, *J. Rheol.* 46 (2002), 1459–1472. <http://dx.doi.org/10.1122/1.1514203> C601, C602

- [20] R. S. Anderssen, E. Carter, B. J. Osborne and I. J. Wesley, Joint inversion of multi-modal spectroscopic data of wheat flours, *Applied Spectroscopy* (in press) C609
- [21] R. S. Anderssen, B. J. Osborne and I. J. Wesley, The application of localisation to near infrared calibration and prediction through partial least squares regression, *J. Near Infrared Spectrosc.* 11 (2003), 39–48. C604
- [22] D. L. Bailey, B. F. Hutton and B. J. Walker, Improved SPECT using simultaneous emission and transmission tomography, *J. Nuclear Med.* 28 (1987), 844–851. C609
- [23] M. Burger and S. Osher, Convergence rates of convex variational regularization, *Inverse Problems* 20 (2004), 1411–1421. C594
- [24] L. Boltzmann, Zur Theorie der elastischen Nachwirkung, *Ann. Phys. Chem.* 7 (1876), 624–657. C601
- [25] J. A. Cochran, *Applied Mathematics* Wadsworth, Belmont, California, 1982. C600
- [26] A. R. Davies and R. S. Anderssen, Sampling localization in determining the relaxation spectrum, *J. Non-Newtonian Fluid Mech.* 73 (1997), 163–179. C608, C614, C615
- [27] J. des Cloizeaux, Double reptation vs single reptation in polymer melts, *Europhys Lett* 5 (1988), 437–442; 6 (1988), 475. C601
- [28] M. Doi and S. F. Edwards, Dynamics of concentrated polymer systems. Part 1. Brownian motion in the equilibrium state, *J. Chem. Soc., Faraday Trans.* 274 (1978), 1789–1801. C601
- [29] D. Donoho, Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition, *Appl. Comp. Harmonic Anal.* 2 (1995), 101–126. <http://dx.doi.org/10.1006/acha.1995.1008> C600, C614

- [30] H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Boston, 1996. C593, C594, C598, C600, C604
- [31] G. M. L. Gladwell, *Inverse Problems in Vibration*, Martinus Noordhoff (Kluwer), The Netherlands, 1986. C593
- [32] M. Golberg, A method of adjoints for solving some ill-posed inverse problems, *Appl. Math. Comput.* 5 (1979), 123–130. C600, C606
- [33] M. A. Golberg, *Numerical Solution of Integral Equations*, Plenum Press, New York, 1990. C593
- [34] C. W. Groetsch, *Inverse Problems in the Mathematical Sciences*, Vieweg, Wiesbaden, 1993. C593, C594, C598
- [35] S. A. Husain and R. S. Anderssen, Modelling the relaxation modulus of linear viscoelasticity using Kohlrausch functions, *J. Non-Newtonian Fluid Mech.* 125 (2005), 159–170.
<http://dx.doi.org/10.1016/j.jnnfm.2004.11.002> C601
- [36] A. J. Jakeman and R. S. Anderssen, Abel type integral equations in stereology. I. General discussion, *J. Microscopy* 105 (1975), 121–133. C598
- [37] M. G. Kendall and P. A. P. Moran, *Geometric Probability*, Griffins Statistical Monographs, No. 10, C. Griffin, London, 1963. C598
- [38] I. Koch and R. S. Anderssen, A direct surface smoothing procedure for Fourier image reconstruction in radio-physics, *Astron. Astrophys.* 183 (1987), 170–176. C605
- [39] G. Landl and R. S. Anderssen, Non-negative differentially constrained entropy-like regularization, *Inverse Problems* 12 (1996), 35–53. C605
- [40] H. Martens and T. Naes, *Multivariate Calibration*, John Wiley & Sons, Chichester, UK, 1991. C602, C603

- [41] C. R. Martin, R. Rousser and D. L. Barbec, Development of a single-kernel characterization system, *Trans. ASAE* 36 (1993), 1399–1404. **C610**
- [42] P. Mathe and S. V. Pereverzev, Direct estimation of linear functionals from indirect noisy observations, *J. Complexity* 18 (2002), 500–516. <http://dx.doi.org/10.1006/jcom.2001.0614> **C608**, **C614**
- [43] A. McNabb and G. C. Wake, Heat conduction and finite measures for transition times between steady states, *IMA J. Appl. Maths.* 47 (1991), 193–260. **C606**
- [44] A. McNabb, G. C. Wake, R. D. Lambourne and R. S. Anderssen, The theoretical derivation of rules-of-thumb for freezing times, *Inverse Problems* 7 (1991), 633–642. **C606**
- [45] D. W. Mead, Determination of molecular weight distributions of linear flexible polymers from linear viscoelastic material functions, *J. Rheol.* 38 (1994), 1797–1827. <http://dx.doi.org/10.1122/1.550527> **C601**
- [46] J. B. Miller, An integral equation from Phytology, *J. Aust. Math. Soc.* 4 (1963), 397–402. **C599**
- [47] P. A. P. Moran, The probabilistic basis of stereology, *Special Supplement Adv. Appl. Probability* 4 (1972), 69–91. **C598**
- [48] V. A. Morozov, *Regularization Methods for Ill-Posed Problems*, CRC Press, 1993. **C593**
- [49] D. Murio, *The Mollification Method and the Numerical Solution of Ill-Posed Problems*, John Wiley & Sons, 1993. **C593**
- [50] T. Naes, T. Isaksson, T. Fearn and T. Davies, *A User-Friendly Guide to Multivariate Calibration and Classification*, NIR Publications, Chichester, UK, 2002. **C602**, **C603**

- [51] B. G. Osborne and R. S. Anderssen, Single-kernel characterization principles and applications, *Cereal Chem.* 80 (2003), 613–622.
<http://dx.doi.org/10.1098/rstb.2003.1352> C610, C613
- [52] W. Schmeidler, *Integralgleichungen mit Anwendungen in Physik und Technik*, Geest & Portig, Leipzig, 1950. C599
- [53] D. H. Simmonds, *Wheat and Wheat Quality in Australia*, CSIRO Publishing, 1989. C591
- [54] J. A. Smith, R. E. Oliver and J. K. Berry, A comparison of two photographic techniques for estimating foliage angle distribution, *Aust. J. Bot.* 25 (1977), 545–553. C599
- [55] G. R. Spedding, Comparing fluid mechanics models with experimental data, *Phil. Trans. Proc. Roy. Soc. B* 358 (2003), 1567–1576. C591
- [56] W. Thimm, C. Friedrich, M. Marth and J. Honerkamp, An analytic relation between relaxation time spectrum and molecular weight distribution, *J. Rheol.* 43 (1999), 1663–1672.
<http://dx.doi.org/10.1122/1.551066> C601
- [57] C. Tsenoglou, Molecular weight polydispersity effects on the viscoelasticity of entangled polymers, *Macromolecules* 24 (1991), 1762–1767. C601
- [58] C. Vogel, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002 C593
- [59] G. Wahba, *Spline Models for Observational Data*, SIAM, Philadelphia, 1990. C605
- [60] J. Warren Wilson, Inclined point quadrats, with Appendix by J. E. Reeve, *The New Phytologist* 59 (1960), 1–8. C599

- [61] C. Zimmer, By a whisker, harbour seals catch their prey, *Science* 293 (2001), 29–31. <http://dx.doi.org/10.1126/science.293.5527.29b>
C591