A General Method for the Solution of Inverse Problems in Transport Phenomena

Marco Vocciante[a], Andrea P. Reverberi[b], Vincenzo G. Dovì*[b]

[a]Dipartimento di Ingegneria Civile, Chimica e Ambientale, Università di Genova, Via Montallegro 1, I-16145 Genova
[b]Dipartimento di Chimica e Chimica Industriale, Università di Genova, Via Dodecaneso 31, I-16146 Genova
dovi@istic.unige.it

The typical inverse problems in transport phenomena are given by partial differential equations with unknown boundary conditions, which are to be estimated from measurements corresponding to solutions of the PDEs or of their gradients. The resulting problem is an ill-posed problem, which can't be solved unless it is adequately regularized, because arbitrarily small errors on the data can give rise to very large deviations in the reconstruction of the unknown boundary conditions. In other words the estimated solution does not depend continuously on the data. The method proposed in this paper is the generalization of methods already applied to a number of problems (diffusion, heat transfer, percolation). The unknown boundary condition is replaced by a piecewise constant (or linear) functions with unknown coefficients. This approximation makes it possible to solve the resulting equation analytically and to estimate the coefficients by comparing theoretical and experimental values by means of linear least squares methods.

The ill-posedness of this class of problems (resulting in singular matrices of the least squares problems) can be tackled using the well-known method proposed by Tikhonov. The value of the parameter contained in Tikhonov's method is estimated from the statistical noise on data by means of the Morozov's discrepancy principle. If additional information on the structure of the solution is available, specialised algorithms can be developed on a case-by-case basis.

The presence of additional unknown non-linear parameters in the partial differential equations (such as diffusion coefficients, conductivities or adsorption coefficients of plant roots) leads to a particular model known as separable least squares. This procedure makes use of the conditional optimality principle for carrying out the overall identification problem in terms of the non-linear parameters only.

1. Introduction

Conservation laws in continuous media take the general form

$$\frac{\partial Q}{\partial t} + \nabla \cdot \left(\boldsymbol{\Phi} + Q\boldsymbol{v} \right) = q \tag{1}$$

where Q is the quantity being conserved, $\boldsymbol{\Phi}$ the flux of Q in the absence of fluid transport, $\boldsymbol{v}$ the bulk velocity and q a source or sink of Q.

Conservation laws are generally coupled with constitutive equations, which provide the necessary analytical expressions for $\boldsymbol{\Phi}$ and q (Bagatin et al., 2014).

As far as mass and energy conservation is concerned, the well-known Fick's (Fourier's) law is generally used for $\boldsymbol{\Phi}$, i.e. $\boldsymbol{\Phi} = -k\nabla Q$. Furthermore, the velocity vector is often determined from separately solved fluid dynamics problems. The conservation of momentum uses Navier-Stokes equations (with laminar and turbulent coefficients) for $\boldsymbol{\Phi}$ with $Q = \rho\boldsymbol{v}$.

For the sake of illustration simplicity only mass and energy conservation are considered in this article. This makes it possible to rewrite Eq. (1) as:

$$\frac{\partial Q}{\partial t} = \mathscr{L}\left(Q \mid t, \boldsymbol{x} / \boldsymbol{\theta}\right) \tag{2}$$

where $\boldsymbol{\theta}$ is the set of unknown parameters contained either in $\boldsymbol{\varPhi}$ or q and $\mathscr{L}$ is a differential operator of the second order.

Equation (2) requires the definition of one initial condition and two boundary conditions:

$$\mathscr{I}(Q) = Q(0, \boldsymbol{x}) = f_0(\boldsymbol{x})$$
$$\mathscr{B}_1\left(Q \mid t, \boldsymbol{x}^{(1)} / \boldsymbol{\theta}\right) = 0 \tag{3}$$
$$\mathscr{B}_2\left(Q \mid t, \boldsymbol{x}^{(2)} / \boldsymbol{\theta}\right) = 0$$

In inverse problem the exact functional form of one of the two boundary conditions has to be estimated from experimentally measured data. It will be supposed (without any loss of generality) that the unknown boundary condition is the one related to the boundary $\boldsymbol{x}^{(2)}$, i.e.

$$Q(0, \boldsymbol{x}) = f_0(\boldsymbol{x})$$
$$\mathscr{B}_1\left(Q \mid t, \boldsymbol{x}^{(1)} / \boldsymbol{\theta}\right) = 0 \tag{4}$$
$$-k\left.\frac{dQ}{dx}\right|_{x=x^{(2)}} + h Q\big|_{x=x^{(2)}} = f(t)$$

Inverse problems generally arise from the necessity of estimating either the very unknown boundary conditions (Solisio et al., 2012) or single physical-chemical parameters in experiments in which boundary conditions are not under the control of the experimenter (Reverberi et al. 2013). Occasionally, both boundary conditions and non-linear parameters are estimated from the data (Palazzi et al., 2014). In all these cases the geometry of the problem is known and is generally selected so as to take advantage of symmetry factors (such as thin disks, cylindrical rods, spherical shells). Thus, the spatial coordinates $\boldsymbol{x}$ reduce to the scalar variable x in the case of small rectangular slabs, to the variable r in spherical shells and long circular conduits. Furthermore the experimental conditions over small intervals of the dependent variables (concentrations or temperature) make it possible to assume a constant coefficient for the Fick's (Fourier's) coefficient, so that $\mathscr{L}$ can be assumed to be a linear differential operator of the second order.

The linearity of $\mathscr{L}$ makes it possible to apply Duhamel's theorem for the solution of Eq. (1) subject to conditions (4). In its simplest formulation Duhamel's principle can be stated as follows (Carslaw and Jaeger, 1959): suppose $Q^*\left(x, t / \boldsymbol{\theta} / \lambda\right)$ is the solution of problem (1)-(4) when $f(t)$ is replaced by $f(\lambda)$, i.e. a time independent parameter. The complete solution $Q\left(t, x / \boldsymbol{\theta} / \lambda\right)$ is then given by:

$$Q(x, t / \boldsymbol{\theta}) = f_0(x) + \int_0^t Q^*(x, t - \lambda / \boldsymbol{\theta} / \lambda) d\lambda \tag{5}$$

The general procedure for the estimation of the unknown boundary condition, which is shown in the next section, is based on a suitable parameterisation of $f(t)$ and on the use of Eq. (5).

2. The general procedure for the estimation of the unknown boundary condition

The assumptions made (linearity, one dimensionality, and time-independence of the unknown boundary condition) make it possible in many cases to evaluate an analytical solution for $Q^*\left(x, t / \boldsymbol{\theta} / \lambda\right)$. A suitable procedure for the evaluation of the integral contained in Eq. (5) is to parametrise the function $f(\lambda)$ as a piecewise polynomial function whose coefficients satisfy the usual smoothness conditions at the boundary of each interval into which the overall range is divided. If M equal intervals of width h=T/M are assumed, $f(\lambda)$ can be rewritten as:

$$f(\lambda) = \sum_{i=1}^{M} c_i f_i(\lambda) \tag{6}$$

where $f_i(\lambda)$ are the polynomials and c_i are adjustable coefficients to be determined by regression. If the $f_i(\lambda)$ are j-th order polynomials, (j-1) continuity/smoothness conditions provide the additional conditions

$$\frac{d^q}{d\lambda^q} f_i(\lambda_{i+1}) = \frac{d^q}{d\lambda^q} f_{i+1}(\lambda_i) \quad \text{at each interval boundary.} \tag{7}$$

The simplest (and generally adequate) assumptions are to consider either a piecewise constant function (in which case no continuity condition is required) or a piecewise linear function. In the latter case introducing the continuity condition provides the general relation:

$$f(\lambda) = f(0) + h \sum_{i=1}^{m-1} c_i + c_m(\lambda - \lambda_{m-1}), \qquad \lambda_{m-1} \le \lambda \le \lambda_m \tag{8}$$

where c_m is the angular coefficient of the linear approximation in the range $\lambda_{m-1} \le \lambda \le \lambda_m$. Since the functional form of $f(t)$ is reconstructed from comparison with measured data, a physical dependence of f on Q is also allowed, i.e. $f(t) = \psi(Q(t), t)$. As shown in previous articles (Dovì et al., 1994) the error introduced by this approximation can be made arbitrarily small by increasing the number of intervals. Thus, not only do traditional numerical schemes lack a general strategy to deal with unknown boundary conditions and to regularize the solution, but they also aren't substantially more accurate.

The overall procedure is exemplified for three important special cases in the next section.

3. Important special cases

a) One dimensional time-dependent heat/mass diffusion between two planes at a distance L

$$\frac{\partial T}{\partial t} = K \frac{\partial T^2}{\partial x^2}$$

$$T(x,0) = T_0$$

$$\left(\frac{\partial T}{\partial x} \right)_{x=L/2} = 0$$

$$T(t, L/2) = f(T, t)$$

The analytical solution corresponding to a piecewise constant approximation of $f(T,t)$ is given by (Dovì et al. 1994)

$$T(L/2, \ t_{r-1} \le t \le t_r) = \sum_{i=1}^{r-1} c_i \left\{ (t_i - t_{i-1}) + \alpha \left[\sum_{n=1}^{\infty} e^{-\beta(2n+1)^2(t-t_{i-1})} - e^{-\beta(2n+1)^2(t-t_i)} / (2n+1)^4 \right] \right\} +$$

$$c_r \left\{ (t - t_{r-1}) + \alpha \left[\sum_{n=1}^{\infty} e^{-\beta(2n+1)^2(t-t_{r-1})} / (2n+1)^4 \right] - \frac{\pi^4}{96} \right\} \qquad \alpha = \frac{32L^2}{K\pi^4}, \quad \beta = \frac{K\pi^2}{4L^2}$$

b) One dimensional time-dependent heat/mass diffusion in cylinders of infinite length or (equivalently) steady state heat/mass diffusion in cylinders of finite length. The velocity distribution $u(r)$ can follow either a laminar or a turbulent distribution law.

$$\rho C_p u(r) \frac{\partial T}{\partial x} = K \frac{1}{r} \frac{\partial}{\partial r} \left[r \frac{\partial T}{\partial r} \right]$$

$$T(0, r) = 1$$

$$\left(\frac{\partial T}{\partial r} \right)_{r=0} = 0$$

$$-K \left(\frac{\partial T}{\partial r} \right)_{r=R} = H[T(x, R)]$$

The analytical solution corresponding to a piecewise linear approximation of $f(t)$ is given by $T(r,x) = \omega(r,x) + \sum_i c_i \xi_i(r,x)$, where the analytical expression of the functions $\omega(r,x)$ and $\xi_i(r,x)$ for both the laminar and turbulent case can be found in (Reverberi et al, 2013).

c) One dimensional percolation through an unsaturated homogeneous soil under a time-dependent surface flux. For practical reasons (Yuan and Lu, 2005), the problem is modelled in terms of capillary pressure Ψ

$$\gamma(\Psi)\frac{\partial \Psi}{\partial t} = -\frac{\partial}{\partial t}\left[K(\Psi)\left(\frac{\partial \Psi}{\partial z}+1\right)\right]$$

$$\Psi(z,0) = \Psi_0$$

$$\left[K(\Psi)\left(\frac{\partial \Psi}{\partial z}+1\right)\right] = -q(t)$$

$$\Psi(0,t) = f(\Psi,t)$$

It can be shown that the analytical solution corresponding to a piecewise constant approximation of $f(\Psi,t)$ is given by $\Psi(z,t_{n-1} < t < t_n) = \frac{1}{\alpha}Log\left[\frac{\alpha}{K_s}\right]\Phi_n(z,t)$,

$$\Phi_n(z,t) = \frac{c_n}{\xi(z)} + \frac{A_1(z)}{\xi(z)} + g_{n-1}(z,t) + (c_1 - c_n)h_n(z)$$

where K_s is the hydraulic conductivity at saturation, α the soil pore-size distribution parameter and $\xi(z), g_n(z), A_1(z)$ and $h_n(z)$ are known function of z an t.

4. Ill-posedness and regularisation techniques

As shown in the previous section experimental measurements are generally predicted by the sum of M terms dependent on the unknown nonlinear parameters, each multiplied by the coefficient c_m that defines the unknown boundary condition. If the nonlinear parameters are known, the determination of the coefficients reduces to ordinary linear least squares, i.e. they can be determined by the minimization of

$$\left\|y' - G(x,t)c\right\|^2 = \sum_k\left[y_k - \sum_i c_i G_i(x_k,t_k)\right]^2 = min \tag{9}$$

where $\sum_i c_i G_i(x_k,t_k)$ is the value of the measured value y_k predicted by the solution (5) when Eq. (6) is used for $f(\lambda)$. The least squares solution of (9) is given by

$$c = \left(G^TG\right)^{-1}G^Ty' \tag{10}$$

provided the symmetric matrix $\left(G^TG\right)$ is full rank. However, due to the ill-posed character of inverse problems of PDEs, (Kabanikhin, 2008), full rankness cannot be assumed for $\left(G^TG\right)$. Thus the determination of c requires the use of adequate regularization techniques for preventing arbitrarily small errors on the data from giving rise to very large deviations in the reconstruction of the vector c. This can be accomplished by replacing $\left(G^TG\right)^{-1}G^Ty'$ by G^+y', where G^+ is the pseudo-inverse of G.

A precise formulation of G^+ can be established using the well-known method proposed by Tikhonov and Arsenine (1976). The value of the parameter contained in Tikhonov's method is estimated from the statistical noise on data by means of the Morozov's discrepancy principle (Morozov, 1984), as shown in section 4.1.
If additional information on the structure of the solution is available, specialised algorithms can be developed on a case-by-case basis. An example is considered in section 4.2.

4.1 Tikhonov-Morozov method

The matrix $G^+(H)$ is approximated by $\left(G^T G + \Psi^T \Psi\right)^{-1} G^T$ where the matrix Ψ is selected so that $\left(G^T G + \Psi^T \Psi\right)$ is invertible and the vector c depends continuously on the data.

This is equivalent to replacing the original minimisation problem $\left\| y' - G(x,t)c \right\|^2 = min$ with the problem

$$\left\| y' - G(x,t)c \right\|^2 + \left\| \Psi c \right\|^2 = min \tag{11}$$

A frequent choice for Ψ is $\Psi = \alpha I$, where the parameter α is selected so as to satisfy some statistical properties. In Morozov's discrepancy method α is iteratively increased until $\left\| y' - G(x,t)c \right\|^2$ is equal to the variance of the data multiplied by the number of data points. Thus, the solution provided, in addition to limiting the norm of c through the introduction of the matrix Ψ, is consistent with the statistical properties of the data. Additionally, some smoothing can be introduced into the solution by introducing tridiagonal and pentadiagonal matrices (Twomey, 1963), although this heuristical method does not rest on a sound statistical basis.

$$\Psi = \alpha I + \varepsilon \begin{pmatrix} -1 & 1 & 0 & \cdots & & & 0 \\ 1 & -2 & 1 & 0 & \cdots & & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ & & & \ddots & & & \\ 0 & \cdots & 0 & 1 & -2 & 1 & 0 \\ 0 & & \cdots & 0 & 1 & -2 & 1 \\ 0 & \cdots & & & 0 & -1 & 1 \end{pmatrix} + \varepsilon \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & & & \\ -2 & 5 & -4 & 1 & 0 & 0 & & \\ 1 & -4 & 6 & -4 & 1 & 0 & & \\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & \\ & & & \ddots & & & & \\ & & & & 0 & 1 & -4 & 5 & -2 \\ & & & & & 0 & 1 & -2 & 1 \end{pmatrix}$$

The tridiagonal matrix tends to reduce the values of the slope, whereas the pentadiagonal matrix tends to hamper oscillations. The value of ε is adjusted until the solution is sufficiently smooth.

4.2 Unimodal regression

Sometimes additional constraints dependent on the particular conditions of the experimentation can be added to the estimation task. A frequent condition is the unimodal distribution of the vector c. In other words, there exists an index i, such that $c_{k-1} \le c_k$ for $k < i$ $c_k \ge c_{k+1}$ for $k \ge i$ (or analogously $c_{k-1} \ge c_k$ for $k < i$ $c_k \le c_{k+1}$ for $k \ge i$). Clearly if $i=1$ or $i=M$ this condition implies a monotonous distribution of c.

The minimization of Eq. (9) is subject now to a set of linear constraints, which change the optimization task into a QP (quadratic programming) problem. If, as is often the case, the value of i is not known, the QP problem has to be repeated for all i ($1 \le i \le M$) until the value of Eq. (9) is consistent with Morozov's principle.

5. Estimation of non-linear parameters

If the non-linear parameters θ are unknown the minimisation problem (9) changes to

$$\left\| y' - G(x,t|\theta)c \right\|^2 = \sum_k \left[y_k - \sum_i c_i G_i \left(x_k, t_k | \theta \right) \right]^2 = min \tag{12}$$

While in principle the problem can be solved by using general non-linear minimization algorithms with respect to both θ and c, the high dimensionality of the resulting task, the difficulty of determining suitable starting values for c and the possibility of multiple local minima reduce the reliability of the solutions determined. However, these difficulties can be overcome if the principle of relative optimality is applied, which in our case

provides optimal values of the vector c for every value of the non-linear parameter θ, i.e. using separable least squares (Golub and Pereyra, 1973). For every value of θ the vector c is computed as $c = G^+(\theta)y'$ and the computation of the non-linear parameter can be carried out by minimising the objective function

$$\left\| y' - G(\theta)G^+(\theta)y' \right\|^2 = \left\| \left\{ I - G(\theta)G^+(\theta) \right\} y' \right\|^2 = \left\| P_G^\perp(\theta)y' \right\|^2 \tag{13}$$

where $P_G^\perp = I - G(\theta)G^+(\theta)$ is the projector on the orthogonal complement of the column space of G with respect to the non-linear parameters only. The typical Gauss-Newton step is given by

$$\theta_{n+1} = \theta_n - \tau_n \left[D\left(P_G^\perp(\theta_n)y' \right) \right]^+ P_G^\perp(\theta_n)y' \tag{14}$$

where the operator D is the Fréchet derivative of matrices and τ_n is an interpolation-extrapolation factor used to locate the minimum of the objective function in the direction $\left[D\left(P_G^\perp(\theta_n)y' \right) \right]^+ P_G^\perp(\theta_n)y'$.

Furthermore $\sigma^2 \left[D\left(P_G^\perp(\theta_n)y' \right) \right]^+$ (where σ^2 is the variance of the data) provides an estimate of the variance of the parameter θ, once convergence at $\theta *$ has been attained.

6. Conclusions

The general method presented is capable of solving a large number of estimation problems related to transport phenomena when the boundary conditions are only partially known.

This situation is frequently encountered in real life problems and is generally dealt with using approximate models or ad-hoc assumptions. Using more realistic models leads to the estimation of more reliable boundary profiles in critical situations (De Rademaeker et al., 2014) and provides more accurate estimations of physical-chemical parameters which can greatly add up to the quality of design of plants and processes.

Acknowledgements

References

Bagatin R., Klemeš J.J., Reverberi A.P., Huisingh D., 2014, Conservation and improvements in water resource management: a global challenge, J Cleaner Prod, 77, 1-9.

Bro, R., Sidiropoulos, N.D., 1998, Least Squares Algorithms under Unimodality and Non-negativity Constraints, J. Chemometrics, 12, 223-247.

Carslaw H.S., Jaeger, J.C., 1959, Conduction of heat in solids, 2nd Editions, Clarendon Press, Oxford, UK.

De Rademaeker E., Suter G., Pasman H.J., Fabiano B., 2014, A review of the past, present and future of the European loss prevention and safety promotion in the process industries, Proc. Saf Environ, 92, 280-291.

Dovì, V.G., Reverberi, A.P., Meshalkin, V.P., 1994, A general procedure for the estimation of diffusivities in solids, Canadian Journal of Chemical Engineering, 72, 1042-1046.

Golub G.H. and Pereyra, V., 1973, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate, SIAM J. Numerical Analysis, 10, 413–432.

Kabanikhin, S. I., 2008, Definitions and examples of inverse and ill-posed problems, J. Inv. Ill-Posed Problems, 16, 317–357.

Morozov, V.A., 1984, Methods for Solving Incorrectly Posed Problems, Springer, New York, US.

Palazzi E., Currò F., Reverberi A., Fabiano B., 2014, Development of a theoretical framework for the evaluation of risk connected to accidental oxygen releases, Process Saf Environ, 92, 357-367.

Reverberi, A.P., Fabiano, B., Dovì, V.G. 2013, Use of inverse modelling techniques for the estimation of heat transfer coefficients to fluids in cylindrical conduits, Int. Comm. Heat and Mass Transfer, 42, 1-94.

Solisio, C., Reverberi, A.P., Del Borghi, A., Dovì, V.G., 2012, Inverse Estimation of Temperature Profiles in Landfills Using Heat Recovery Fluids Measurements, Journal of Applied Mathematics, ID 747410.

Tikhonov, A., Arsenine, V., 1977, Solution of Ill-posed Problems, Winston & Sons, Washington, USA.

Twomey, S., 1963, On the numerical solution of Fredholm integral equations of the first kind by inversion of the linear system produced by quadrature, J. Assoc. Comput. Mach., 10, 97-101.

Yuan F. and Lu Z., 2005, Analytical Solutions for Vertical Flow in Unsaturated, Rooted Soils with Variable Surface Fluxes, Vadose Zone J., 4, 1210-1218.