1 ODE estimation

The estimation problem starts with an ODE

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}\left(t, \mathbf{x}, \boldsymbol{\beta}\right), \quad \mathbf{x}, \mathbf{f} \in R^m, \quad \boldsymbol{\beta} \in R^p,$$

and a set of observations

$$\mathbf{y}_i = \mathcal{O}\mathbf{x}^*\left(t_i\right) + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \cdots, n, \quad \mathbf{y} \in R^q.$$

Here "*" indicates the "true" solution, and the noise components (observational error) $\boldsymbol{\varepsilon}_i \sim N\left(0, \sigma^2 I_q\right)$ are assumed independent. The aim is to estimate the parameter vector $\boldsymbol{\beta}^*$ and/or the solution $\mathbf{x}^*\left(t\right)$ by minimizing the objective

$$\Phi = \sum_{i=1}^{n} \left\|\mathbf{y}_i - \mathcal{O}\mathbf{x}\left(t_i, \boldsymbol{\beta}\right)\right\|_2^2.$$

subject to the constraint on $\mathbf{x}$ imposed by the differential equation.

Two main approaches are currently used to solve the estimation problem. The embedding method imposes boundary conditions on the ODE and these introduce m extra parameters which must be adjusted as part of the estimation procedure. The resulting boundary value problem must be solved for each evaluation of Φ and this could result in additional overhead. The simultaneous method poses the estimation problem as a constrained optimisation problem where the equality constraints are obtained by discretizing the differential equation. This approach could be more efficient in general as the solution of the ODE is not obtained until the process has converged in contrast to the embedding method. However, this means that questions such as error control in the ODE solution become more obscure. Questions of interest include the equivalence of the two approaches (formally they look rather different), and convergence rates for algorithms. There is particular interest in large sample (large n) properties. The equivalence question, at the level of the necessary conditions, suggests a possible new algorithm which looks like a bit like hybrid scheme.

2 Polyhedral optimization

The homotopy algorithm of Osborne, Presnel and Turlach for minimizing a sum of squares subject to an l_1 constraint on the variables (the so called lasso - a particular case of a polyhedral constraint) has proved spectacularly successful in applications to a range of variable selection problems in which the aim is the selection of a minimum set of variables in linear models compatible with adequate model representation of the observed signal. In many cases it takes just k steps to select k variables. Applications include compressed sensing where economical signal representation has realized bandwidth reductions of up to ten times. There is still work to be done on the original problem for objectives other than a sum of squares and for other forms of polyhedral constraints. A piecewise linear homotopy trajectory requires that the objective be locally at most quadratic. For maximum likelihood estimation in the case of linear models this requires piecewise linear or piecewise quadratic approximation of the log density terms as a function of their arguments. The piecewise nature means that homotopy slope discontinuities occur not only where the variable set changes (a selection step) but also where the local representation changes. There has been some work on the locally quadratic case. A new question is what is required for accuracy control of the local approximation of the objective in a variable selection exercise. Also, it should be possible to make at least some progress for some nonlinear models by using low order (≤ 2) multivariate splines for example.