Predicted Sensitivity for establishing well-posedness conditions in stochastic inversion problems

Mélanie Blazère & Nicolas Bousquet
Institut de Mathématique de Toulouse & EDF Lab Chatou, France

We consider a stochastic inversion problem defined by the knowledge of observations $\mathbf{y_n} = (y_i^*)_{i\in\{1,\ldots,n\}}$ living in a $q-$dimensional space, which are assumed to be realizations of a random variable Y^* such that

$$\begin{aligned} Y^* &= Y + \varepsilon, \\ Y &= g(X) \end{aligned}$$

where X is a $d-$dimensional random Gaussian variable $X \sim \mathcal{N}(\mu, \Sigma)$ with unknown $\theta = (\mu, \Sigma)$, ε is a (experimental or/and process) noise with know distribution f_ϵ, and g is some deterministic function from $\mathbb{R}^p$ to $\mathbb{R}^q$ (possibly a black-box computer model) . This inversion problem (ie., estimating θ) can be solved in frequentist [3,1] or Bayesian [4] frameworks (possibly by linearizing g [1]), using missing data algorithms. In both frameworks, inferring on θ requires that several conditions of well-posedness and identifiability are gathered.

The first one is Hadamard's well-posedness condition, which states that the solution $\hat{\theta}$ of the inversion/calibration problem should exist, be unique and be continuously dependent on observations according to a reasonable topology. In the case where g is linear or can be linearized, namely if there exists a linear operator H such that $Y^* = HX + \varepsilon$, this condition is traduced by a low value of the *condition number* of H [2]. The second condition is the identifiability of the input model $X \sim \mathcal{N}(\mu, \Sigma)$. In similar cases of linearity or linearization, this condition states that H must be injective ($\text{rank}(H) = d$) and $d \leq nq$.

However, additionally to Hadamard's condition, and independently of the availability of experimental data $\mathbf{y}^*$, a second condition of well-posedness is, to our knowledge, never evoked in practice, while it seems to be of primary importance in the specific framework of stochastic inversion. This condition arise from (let us say) *predictive sensitivity analysis*. Imagine that the problem is solved and θ is known. Any sensitivity study, for instance based on celebrated Sobol' indices [5], should highlight that the main source of uncertainty, explaining the variations of Y^*, is X and not ε. In practice, this kind of diagnostic is established a posteriori, as a check for an estimated solution $\hat{\theta}$ (or a posterior distribution $\pi(\theta|\mathbf{y_n})$ in a Bayesian context). However, this property is more than desirable and should be converted into a modelling constraint for the estimation of θ. In a Bayesian inversion context, such a constraint would apply on the prior elicitation of (the parameters of) θ, and could help to define better reference measures when no other prior information is available on θ nor X. Such a study requires a formal definition of what "the main source of uncertainty" means.

Several answers to the problem of well-defining a stochastic inversion problem, by formalizing a new condition on θ with respect to the features of g and f_ε, are proposed in this talked. They are successively based on Sobol' indices, entropic indices then comparisons of Fisher information. The case when g is linear or linearizable is considered in this study, since it appears as a minimal framework to establish such a rule (or possibly several rules) of well-posedness. Strongly nonlinear cases often by definition present more contrasted behavior between observations and noise, and it is likely that the ratio between signal and noise be more in favor of the signal.

In this regard, one-dimensional linear then linearizable models are studied. Then a general rule of well-posedness is presented and results are derived for multivariate linearizable models. A theoretical link is done with Sobol'indices. Extensions to nonlinear cases are discussed, as well as stochastic metamodels as Gaussian kriging used in general stochastic inversion [4]. Furthermore, the control of bias arising in linearizable contexts is evoked, as a new source of uncertainty that should be monitored in the same way as the noise ε. Finally, the approach is tested over toy examples and a simplified hydraulical example, and compared with usual stochastic inversion methodologies that do not consider this well-posedness condition a priori.

References:

[1] P. Barbillon, G. Celeux, A. Grimaud, Y. Lefebvre, and E. Rocquigny (de). "Non linear methods for inverse statistical problems". Computational Statistics and Data Analysis 55, 132-142, 2011.

[2] D.A. Belsley, E. Kuh, and R.E. Welsch. "The Condition Number". In: *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* New York: John Wiley & Sons, 1980.

[3] G. Celeux, A. Grimaud, Y. Lefebvre, and E. Rocquigny (de). "Identifying intrinsic variability in multivariate systems through linearised inverse methods". Inverse Problems in Engineering 18, 401-415, 2010.

[4] S. Fu, G. Celeux, N. Bousquet, and M. Couplet. "Bayesian inference for inverse problems occurring in uncertainty analysis". International Journal for Uncertainty Quantification 5, 73-98, 2015.

[5] B. Iooss and P. Lemaître. "A review on global sensitivity analysis methods". In: *Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications*, C. Meloni and G. Dellino (eds.), Springer, 2015