A Sensitivity Matrix Based Methodology for Inverse Problem Formulation

Ariel Cintrón-Arias, H. T. Banks, Alex Capaldi and Alun L. Lloyd

Center for Research in Scientific Computation
and
Center for Quantitative Sciences in Biomedicine
North Carolina State University
Raleigh, NC 27695

April 10, 2009

Abstract

We propose an algorithm to select parameter subset combinations that can be estimated using an ordinary least-squares (OLS) inverse problem formulation with a given data set. First, the algorithm selects the parameter combinations that correspond to sensitivity matrices with full rank. Second, the algorithm involves uncertainty quantification by using the inverse of the Fisher Information Matrix. Nominal values of parameters are used to construct synthetic data sets, and explore the effects of removing certain parameters from those to be estimated using OLS procedures. We quantify these effects via a score for a vector parameter defined using the norm of the vector of standard errors for components of estimates divided by the estimates. In some cases the method leads to reduction of the standard error for a parameter to less than 1% of the estimate.

Key words: Inverse problems, ordinary least squares, sensitivity matrix, Fisher Information matrix, parameter selection, standard errors.

1 Introduction

The question of parameter identifiability/estimation in the context of parameter determination from system observations or output is at least forty years old and received much attention in the zenith years of linear system and control theory in the investigation of observability, controllability and detectability [3, 9, 10, 22, 23, 28, 30, 34, 36]. These early investigations and results were focused primarily on engineering applications, but much interest in other areas (e.g, oceanography, biology) has prompted more recent inquiries for both linear and nonlinear dynamical systems [2, 8, 18, 21, 26, 32, 41, 43, 44, 45]. In some of the earliest results, Bellman and Astrom [9] defined the concept of structural identifiability, and provided a theoretical framework to address, a priori, whether or not it is possible to determine estimates of unknown parameters from experimental data. Specifically they showed that controllablility (in the sense of the Kalman [28] controllability matrix possessing full rank) implies identifiability, thereby establishing one of the earliest linear algebraic

tests for identifiability. In another important early linear algebraic effort, Reid [34] defined the term *sensitivity identifiability* . If $z(\theta)$ denotes the output of a model depending on a parameter vector θ, then Reid explains *sensitivity identifiability* in the following way. Let $\Delta\theta$ denote a local perturbation about a nominal θ_0, i.e., $\Delta\theta = \theta - \theta_0$, which gives rise to local (small) perturbation Δz in the output, i.e., $\Delta z = z(\theta) - z(\theta_0)$. Suppose that $\chi = \frac{\partial z}{\partial \theta}$ denotes the sensitivity matrix, i.e., the Jacobian matrix of the output, being evaluated at θ_0 [4, 6]. Then the first order Taylor approximation (exact for linear dependence on the parameter)

$$\Delta z \approx \chi \Delta\theta. \tag{1}$$

relates the perturbations. A parameter vector is defined as *sensitivity identifiable* if equation (1) can be solved uniquely (in the local sense) for $\Delta\theta$ [18, 34]. In their review, Cobelli and DiStefano [18] explain that a sufficient condition for sensitivity identifiability is the nonsingularity of the matrix $\chi^T\chi$ or equivalently

$$\det(\chi^T\chi) \neq 0.$$

From this one sees immediately that parameter estimation depends inherently on the condition number of the *Fisher Information Matrix* (FIM) $F = \chi^T\chi$. Not surprisingly, subsequent investigations of parameter estimation (in applied mathematics, engineering, and statistics) have focused on the role of the FIM. It is now well known that this matrix and its condition number play a fundamental role in a range of useful ideas such as model comparison [12] (the Akaike Information Criteria, the Takeuchi Information Criteria, etc.), generalized sensitivity functions [6, 7, 40] and experimental design (duration, frequency, quality, etc., of observations required to reliably estimate parameters) as well as computation of standard errors and confidence intervals [4, 5, 7, 19].

Brun, et al., [11] and Burth, et al., [13] proposed analyses that use submatrices of the FIM $\chi^T\chi$. Burth, et al., implement a *reduced-order* estimation by determining which parameter axes lie closest to the ill-conditioned directions of $\chi^T\chi$, and then by fixing the associated parameter values at prior estimates throughout an iterative estimation process. Brun, et al., determined identifiability of parameter combinations using the eigenvalues of submatrices that result from only using some columns of $\chi^T\chi$. Motivated by these efforts and those on the relationship between ill-conditioning of the FIM and quality of parameter estimates investigated in [5, 6, 7], we here use the sensitivity matrix χ to develop a methodology to assist one in parameter estimation or inverse problem formulations.

In particular, in this paper we investigate the problem of finding multiple solutions for unknown parameters from observations with a statistical error structure (a more practical setting than one assuming noise free observations). We address parameter identifiability by exploiting properties of *both* the sensitivity matrix and uncertainty quantifications in the form of standard errors. We propose an algorithm inspired by [11, 13], to select parameter combinations (vectors) in two stages. In the first stage, all possible parameter combinations (i.e., subsets of all parameters) are considered and only those with a full rank sensitivity matrix are selected. In the second stage, a score involving uncertainty quantification (standard errors) is calculated for each parameter vector selected in the first stage. Then parameter subset combinations are examined in view of their score and the condition number of corresponding sensitivity matrices. We believe that some form of this type of *practical identifiability analysis* could be carried out a priori, i.e., before any attempt to solve inverse problems (from experimental observations) is made. We illustrate the ideas and methodology with a seasonal epidemic model.

This manuscript is organized in the following manner. Section 2 introduces the seasonal epidemic model. In Section 3 we explain the statistical model for the observation process; we define the ordinary least squares (OLS) estimator and define precisely the Fisher information matrix. Using a first order Taylor expansion of the model output we compute the OLS estimator in terms of the sensitivity matrix singular values and the error in the observation process. Section 4 contains the proposed subset selection algorithm. In Section 5 some illustrations of the algorithm are discussed, in light of using both synthetic and observational data sets. We conclude with a brief discussion in the last section.

2 Motivating seasonal SEIRS model with demography

We introduce a model, a standard *Susceptible-Exposed-Infective-Recovered-Susceptible* (SEIRS) model, to illustrate the methodology we discuss in this paper. In particular we consider a seasonal model for disease spread and progression in a population. Seasonal patterns of disease incidence are observed in epidemics of influenza [20], meningococcal meningitis [35], measles [1], and rubella [42], to mention a few. Many temporal factors play a role in the formation of cyclical patterns, for instance [25]: (i) survival of the pathogen outside the host, (ii) host behavior and (iii) host immune function.

Cyclical incidence patterns are often modeled with a transmission parameter being a function of time. We denote the time-dependent transmission parameter by $\beta(t)$; it is traditionally defined by [20, 27]

$$\beta(t) = \beta_0 \left[1 + \beta_1 \cos(2\pi(t - t_0))\right], \tag{2}$$

where β_0 is called the baseline level of transmission, β_1 is known as the amplitude of seasonal variation or simply the strength of seasonality, and t_0 denotes the transmission parameter phase shift. We may, for convenience, derive an equivalent formulation. Because

$$\beta_1 \cos(2\pi(t - t_0)) = a_1 \cos(2\pi t) + b_1 \sin(2\pi t),$$

where $a_1 = \beta_1 \cos(2\pi t_0)$ and $b_1 = \beta_1 \sin(2\pi t_0)$, we may re-write equation (2) as

$$\beta(t) = \beta_0 \left(1 + a_1 \cos(2\pi t) + b_1 \sin(2\pi t)\right). \tag{3}$$

The time-dependent transmission parameter $\beta(t)$, as defined in equation (3), is used in the seasonal epidemic model introduced here. Four main epidemiological events are described: latent infection, active infection, recovery, and loss of immunity. It is assumed that individuals becoming infected undergo latency, a period of time during which they are incapable of effectively transmitting the infectious agent, before progressing into active infection. People recover from active infection and develop temporary immunity (they will eventually become susceptible once again). Four epidemiological classes are considered, and at time t the number of: susceptible is denoted by $S(t)$; latent or exposed is denoted by $E(t)$; infectious is denoted by $I(t)$; and recovered or

temporarily immune is denoted by $R(t)$. The nonlinear differential equations [29, 39]

$$\begin{aligned}
\frac{dS}{dt} &= \frac{1}{P}N + \frac{1}{L}R(t) - \beta(t)S(t)\frac{I(t)}{N} - \frac{1}{P}S(t) && (4)\\
\frac{dE}{dt} &= \beta(t)S(t)\frac{I(t)}{N} - \frac{1}{M}E(t) - \frac{1}{P}E(t) && (5)\\
\frac{dI}{dt} &= \frac{1}{M}E(t) - \frac{1}{D}I(t) - \frac{1}{P}I(t) && (6)\\
\frac{dR}{dt} &= \frac{1}{D}I(t) - \frac{1}{L}R(t) - \frac{1}{P}R(t) && (7)\\
N &= S(t) + E(t) + I(t) + R(t) && (8)\\
S(t_0) &= S_0 && (9)\\
E(t_0) &= E_0 && (10)\\
I(t_0) &= I_0 && (11)\\
R(t_0) &= N - S_0 - E_0 - I_0, && (12)
\end{aligned}$$

define the epidemic dynamics known as an SEIRS model. This formulation takes into account demographic processes (the birth rate is N/P and the average life span is P) while assuming the total population size N remains constant.

The mean latency period is denoted by M, while the average length of active infection is denoted by D. It is also assumed immunity lasts an average of L units of time.

In this paper we consider a scenario where the initial conditions of the SEIRS model (S_0, E_0, and I_0) may be unknown, and may need to be estimated, along with all the other model parameters. We apply inverse problem methodologies to determine estimates of the vector parameter

$$\theta = (S_0, E_0, I_0, N, L, D, M, P, \beta_0, a_1, b_1)^T \in \mathbb{R}^p = \mathbb{R}^{11}, \tag{13}$$

according to an ordinary least squares criterion (defined in the next section).

3 Statistical model for the observation process

The observation process is formulated assuming the SEIRS model, together with a particular choice of parameters (the "true" parameter vector denoted as θ_0) describes the epidemic process exactly, but that the n longitudinal observations $\{Y_j\}_{j=1}^n$ are affected by random deviations (such as measurement errors) from this underlying process. More precisely, if $z(t_j;\theta_0)$ denotes the number of new cases of active infection (also referred to as the model output) between the observation time points t_{j-1} and t_j, which is defined as

$$z(t_j;\theta_0) = \int_{t_{j-1}}^{t_j} \frac{1}{M}E(t;\theta_0)dt, \tag{14}$$

then the statistical model for the observation process is

$$Y_j = z(t_j;\theta_0) + \mathcal{E}_j \qquad \text{for } j = 1, \ldots, n. \tag{15}$$

The errors $\mathcal{E}_j$ are assumed to be random variables satisfying the following assumptions:

(i) the errors $\mathcal{E}_j$ have mean zero: $E[\mathcal{E}_j] = 0$;

(ii) the errors $\mathcal{E}_j$ have finite common variance: $\text{var}(\mathcal{E}_j) = \sigma_0^2 < \infty$;

(iii) the errors $\mathcal{E}_j$ are independent (i.e., $\text{cov}(\mathcal{E}_j, \mathcal{E}_i) = 0$ whenever $j \neq i$) and identically distributed.

Under these assumptions, we have that the mean of the observation equals the model output: $E[Y_j] = z(t_j; \theta_0)$ and the variance in the observations is constant in time: $\text{var}(Y_j) = \sigma_0^2$.

3.1 Ordinary least squares (OLS)

We consider an ordinary least squares (OLS) formulation of a generic parameter estimation or inverse problem for a vector parameter (θ) dependent system

$$\frac{dx}{dt}(t) = g(t, x(t;\theta);\theta), \tag{16}$$

$$x(t_0) = x_0 \tag{17}$$

with observation (or model output) process

$$z(t_j) = \mathcal{F}(x(\cdot), \theta), \quad j = 1, \ldots, n. \tag{18}$$

In this context we consider a given vector of observations $Y = (Y_1, \ldots, Y_n)^T$, where each Y_j is defined by equation (15), and the model output vector $z(\theta) = (z(t_1;\theta), \ldots, z(t_n;\theta))^T$ for a given θ. The estimator $\theta_{OLS} = \theta_{OLS}^n$ is a random variable that minimizes the Euclidian norm (in $\mathbb{R}^n$) square of $Y - z(\theta)$, i.e., θ_{OLS}^n minimizes

$$J(\theta|Y) \equiv |Y - z(\theta)|^2 = [Y - z(\theta)]^T [Y - z(\theta)] = \sum_{j=1}^{n} [Y_j - z(t_j;\theta)]^2, \tag{19}$$

which implies θ_{OLS} solves the gradient equation

$$\nabla_\theta(|Y - z(\theta)|^2) = 0. \tag{20}$$

Asymptotic theory can be used to describe the distribution of the estimator θ_{OLS} [4, 19, 37]. Provided that a number of regularity conditions as well as sampling conditions are met (see [37] for details), it can be shown that, asymptotically (i.e., as $n \to \infty$), θ_{OLS} is approximately distributed according to a multivariate normal distribution, i.e.,

$$\theta_{OLS}^n \sim \mathcal{N}_p(\theta_0, \Sigma_0^n), \tag{21}$$

where $\Sigma_0^n = \sigma_0^2[n\Omega_0]^{-1} \in \mathbb{R}^{p\times p}$ and

$$\Omega_0 = \lim_{n\to\infty} \frac{1}{n}\chi^n(\theta_0)^T\chi^n(\theta_0). \tag{22}$$

We remark that the theory requires that this limit exists and that the matrix Ω_0 be non-singular. The matrix Σ_0^n is the $p \times p$ *covariance matrix* $\text{cov}((\theta_{OLS}^n)_i, (\theta_{OLS}^n)_j)$, and the $n \times p$ matrix $\chi(\theta_0) \equiv$

$\chi^n(\theta_0)$ is called the *sensitivity matrix* of the system, and its jth row is equal to $\nabla_\theta z(t_j;\theta_0)$. More precisely,

$$\chi^n_{ji}(\theta_0) = \left.\frac{\partial z(t_j;\theta)}{\partial \theta_i}\right|_{\theta=\theta_0} \qquad 1 \le j \le n,\ 1 \le i \le p. \tag{23}$$

For the motivating SEIRS model, the partial derivatives of the state variable vector $x = (S, E, I, R)^T$ with respect to θ can be readily calculated. If $g = (g_1, g_2, g_3, g_4)^T$ denotes the vector function whose entries are given by the expression on the right sides of equations (4)–(7), then we can write the seasonal SEIRS model in the general vector form (16). The sensitivities $\partial x/\partial\theta$ are calculated, for a given $\theta = \hat{\theta}_{OLS}$ (defined below), by solving (see [4, 17] and the references therein) equation (16) and then

$$\frac{d}{dt}\frac{\partial x}{\partial \theta} \quad = \quad \frac{\partial g}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial g}{\partial \theta}, \tag{24}$$

from $t = t_0$ to $t = t_n$. In equation (24) the matrix $\partial g/\partial x$ is 4×4, while the matrices $\partial x/\partial\theta$ and $\partial g/\partial\theta$ are $4 \times p$.

The solution of equation (20) obtained using a realization $y = (y_1, \ldots, y_n)^T$ of the observation process $Y = (Y_1, \ldots, Y_n)^T$ and denoted as the estimate $\hat{\theta}_{OLS} = \hat{\theta}^n_{OLS}$, provides a realization of the estimator θ_{OLS}. The estimate $\hat{\theta}_{OLS}$ is used in the calculation of the sampling distribution for the parameters. The error variance σ_0^2 is approximated by $\hat{\sigma}^2_{OLS}$, which is calculated as

$$\hat{\sigma}^2_{OLS} = \frac{1}{n-p}\left|y - z(\hat{\theta}_{OLS})\right|^2. \tag{25}$$

The covariance matrix Σ^n_0 is approximated by $\hat{\Sigma}^n_{OLS}$, which is computed by

$$\hat{\Sigma}^n_{OLS} = \hat{\sigma}^2_{OLS}\left[\chi(\hat{\theta}^n_{OLS})^T\chi(\hat{\theta}^n_{OLS})\right]^{-1}. \tag{26}$$

The approximation [19, 37] of the sampling distribution of the estimator is

$$\theta_{OLS} = \theta^n_{OLS} \sim \mathcal{N}_p(\theta_0, \Sigma^n_0) \approx \mathcal{N}_p(\hat{\theta}^n_{OLS}, \hat{\Sigma}^n_{OLS}). \tag{27}$$

The standard errors for $\hat{\theta}^n_{OLS}$ can be approximated by taking the square roots of the diagonal elements of the covariance matrix $\hat{\Sigma}^n_{OLS}$. The standard errors are used to quantify uncertainty in the estimation and are given by

$$SE_k(\hat{\theta}^n_{OLS}) = \sqrt{(\hat{\Sigma}^n_{OLS})_{kk}}, \quad k = 1, \ldots, p. \tag{28}$$

3.2 Fisher information matrix

The matrix

$$F = F^n = \chi^n(\theta_0)^T\chi^n(\theta_0), \tag{29}$$

is known as the Fisher information matrix [6, 18]. Below, we use a linearization argument (similar to that employed in the asymptotic distribution theory for OLS – see Chapter 12 of [37]) to give a heuristic derivation of an approximate expression for the estimator θ_{OLS} in terms of F. This

derivation illustrates the role played by the Fisher information matrix in the estimation of unknown parameters and uncertainty propagation.

We observe that the gradient of $J(\theta|Y)$ as defined in (19) is given by

$$\nabla_\theta J(\theta|Y) = -2\chi^n(\theta)^T \left[Y - z(\theta)\right], \tag{30}$$

because by equation (23), we know that $\nabla_\theta z(t_j;\theta) = \chi^n(\theta_0)^T$. Moreover, the Hessian of $J(\theta|Y)$ is

$$\nabla_\theta^2 J(\theta|Y) = 2\chi^n(\theta)^T\chi^n(\theta) - G(\theta), \tag{31}$$

where

$$G(\theta) = 2\sum_{j=1}^{n} \left[Y_j - z(t_j;\theta)\right] \nabla_\theta^2 z(t_j;\theta).$$

For the next calculations we tacitly assume that $\chi^n(\theta_0)^T\chi^n(\theta_0)$ is nonsingular and $G(\theta_0) = 0$. We consider a linearization of $\nabla_\theta J(\theta|Y)$ around $\theta = \theta_0$, which is given by

$$\mathcal{L}(\theta) \quad = \quad -2\chi^n(\theta_0)^T \left[Y - z(\theta_0)\right] + 2\chi^n(\theta_0)^T\chi^n(\theta_0)(\theta - \theta_0). \tag{32}$$

The solution to $\mathcal{L}(\theta) = 0$ is, to first order, the minimizer θ^n_{OLS}, and we thus have (see equation (2.15) of [37])

$$\theta^n_{OLS} \approx \theta_0 + \left[\chi^n(\theta_0)^T\chi^n(\theta_0)\right]^{-1} \chi^n(\theta_0)^T\mathcal{E}, \tag{33}$$

where $\mathcal{E} = (\mathcal{E}_1, \ldots, \mathcal{E}_n)^T$, with $\mathcal{E}_j = Y_j - z(t_j;\theta_0)$ for $j = 1, \ldots, n$. The propagation of uncertainty from the observation process to the estimator is induced by $\mathcal{E}$ in equation (33).

It is clear from equation (33) that if $F^n = \chi^n(\theta_0)^T\chi^n(\theta_0)$ is nearly singular then θ_{OLS} may be very sensitive to the observation error $\mathcal{E}$. Moreover, equation (26) suggests that near-singularity (or ill-conditioning [24]) of F^n may also affect the approximation of the covariance matrix $\hat{\Sigma}^n_{OLS}$, and consequently the calculation of standard errors and confidence intervals for estimated parameters.

For some time it has been well understood (see [6, 7, 18, 40, 45] and the references therein) that the information content of measurements can be quantified by the Fisher information matrix. Thus, efficient experiments can be designed using the Fisher information matrix F^n. As noted in [6], the three most popular design strategies are: D-optimal design, c-optimal design, and E-optimal design. These strategies involve the determinant, the inverse, and maximum and minimum eigenvalues of F^n. Our approach in this paper relies on properties of the sensitivity matrix $\chi = \chi^n$ rather than F^n as well as asymptotic standard errors (which do depend on F^n) for parameters. In the next section we address rank deficiency and the condition number of the sensitivity matrix χ^n.

3.3 Singular value decomposition of the sensitivity matrix

To motivate the role singular value decomposition plays in uncertainty assessment, we consider another linearization that relates the estimator θ_{OLS} to the singular values of the rectangular sensitivity matrix χ. (Hereafter we shall suppress the superscripts denoting dependence on n when no confusion can occur.)

Suppose the model output $z(\theta)$ is well approximated by its linear Taylor expansion around θ_0, i.e.,

$$z(\theta) \approx z(\theta_0) + \chi(\theta_0)(\theta - \theta_0). \tag{34}$$

This first order Taylor expansion can be used to reduce $Y - z(\theta)$ to an affine transformation of θ, by using equations (34) and (15):

$$Y - z(\theta) = -\chi(\theta_0)(\theta - \theta_0) + \mathcal{E}, \tag{35}$$

where $\chi(\theta_0) \in \mathbb{R}^{n\times p}$, $\theta - \theta_0 \in \mathbb{R}^p$, $\mathcal{E}$ is an $\mathbb{R}^n$-valued random variable, and $n > p$.

The singular value decomposition (SVD) of the sensitivity matrix $\chi(\theta_0)$ is denoted as

$$\chi(\theta_0) = U \begin{bmatrix} \Lambda \\ \mathbf{0} \end{bmatrix} V^T, \tag{36}$$

where U is an $n \times n$ orthogonal matrix, i.e., $U^T U = UU^T = I_n$, with U_1 containing the first p columns of U and U_2 containing the last $n - p$ columns, $U = [U_1\ U_2]$; Λ is a $p \times p$ diagonal matrix defined as $\Lambda = \mathrm{diag}(s_1, \ldots, s_p)$, with $s_1 \geq s_2 \geq \cdots \geq s_p \geq 0$; $\mathbf{0}$ denotes an $(n - p) \times p$ matrix of zeros; and V denotes an orthogonal $p \times p$ matrix, i.e., $V^T V = VV^T = I_p$ (more details about SVD can be found in [24] and references therein).

The Euclidean norm is invariant under orthogonal transformations. In other words, for any vector $w \in \mathbb{R}^n$ we have that $|w|^2 = w^T w = w^T I w = w^T U U^T w = |U^T w|^2$. According to [24, 33] this invariance of the Euclidean norm implies

$$\begin{aligned} |-\chi(\theta_0)(\theta - \theta_0) + \mathcal{E}|^2 &= \left|U^T\left(-\chi(\theta_0)(\theta - \theta_0) + \mathcal{E}\right)\right|^2 && (37)\\ &= \left|-\begin{bmatrix} \Lambda \\ \mathbf{0} \end{bmatrix} V^T(\theta - \theta_0) + \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} \mathcal{E}\right|^2 && (38)\\ &= \left|-\Lambda V^T(\theta - \theta_0) + U_1^T \mathcal{E}\right|^2 + \left|U_2^T \mathcal{E}\right|^2. && (39) \end{aligned}$$

The estimator θ_{OLS} minimizes $|Y - z(\theta)|^2$ and according to equations (35) and (39) can be calculated by solving $\left|-\Lambda V^T(\theta - \theta_0) + U_1^T \mathcal{E}\right|^2 = 0$, for θ and thus obtaining

$$\theta_{OLS} = \theta_0 + V\Lambda^{-1} U_1^T \mathcal{E} = \theta_0 + \sum_{i=1}^{p} \frac{1}{s_i} v_i u_i^T \mathcal{E}, \tag{40}$$

where $v_i \in \mathbb{R}^p$ and $u_i \in \mathbb{R}^n$ denote the ith columns of V and U, respectively (the matrix V has column partitioning $V = [v_1, \ldots, v_p] \in \mathbb{R}^{p\times p}$, while $U = [u_1, \ldots, u_n] \in \mathbb{R}^{n\times n}$).

There is a similarity between equations (33) and (40). Again, the randomness of the observation process is additively propagated into the estimator. In equation (40) we see that if $s_i \to 0$, then the estimator θ_{OLS} is particularly sensitive to $\mathcal{E}$.

At this point we need a couple of definitions. The range of a matrix $C \in \mathbb{R}^{n\times p}$ with column partitioning $C = [c_1, \ldots, c_p]$ is defined as the subspace spanned by its columns, i.e.,

$$\mathcal{R}(C) = \left\{ \sum_{j=1}^{p} q_j c_j \in \mathbb{R}^n : q_j \in \mathbb{R} \right\}. \tag{41}$$

The rank of a matrix $C \in \mathbb{R}^{n\times p}$ is equal to the dimension of $\mathcal{R}(C)$:

$$\mathrm{rank}(C) = \dim(\mathcal{R}(C)). \tag{42}$$

If $\text{rank}(C) < \min\{n, p\} = p$ (because we are assuming there are more observations than parameters, i.e., $n > p$) the matrix $C \in \mathbb{R}^{n\times p}$ is said to be rank deficient. On the other hand, if $\text{rank}(C) = p$ we say the matrix $C \in \mathbb{R}^{n\times p}$ has full (column) rank [24].

For a full rank sensitivity matrix $\chi(\theta_0) \in \mathbb{R}^{n\times p}$ (assuming $\text{rank}(\chi(\theta_0)) = p$ and $s_1 \geq s_2 \geq \cdots \geq s_p > 0$) its *condition number* κ is defined as the ratio of the largest to smallest singular value [24]:

$$\kappa(\chi(\theta_0)) = \frac{s_1}{s_p}. \tag{43}$$

We note that if the matrix $\chi(\theta_0)$ has full rank and a large condition number (a feature known as ill-conditioning [24]), then the Fisher information matrix $F = \chi(\theta_0)^T\chi(\theta_0)$ inherits a large condition number. Equation (36) implies the SVD of $\chi(\theta_0)^T\chi(\theta_0)$ is

$$\chi(\theta_0)^T\chi(\theta_0) = V\Lambda^2V^T, \tag{44}$$

and therefore

$$\kappa(\chi(\theta_0)^T\chi(\theta_0)) = \frac{s_1^2}{s_p^2} = \left[\frac{s_1}{s_p}\right]^2 = \kappa(\chi(\theta_0))^2. \tag{45}$$

As discussed in [24], if the columns of $\chi(\theta_0)$ are nearly dependent then $\kappa(\chi(\theta_0))$ is large. In other words, if $\kappa(\chi(\theta_0))$ is not large (the matrix $\chi(\theta_0)$ is well-conditioned) then the columns of the sensitivity matrix are not nearly dependent, suggesting one could use the condition number of $\chi(\theta_0)$ as a criterium to select parameter combinations.

In the next section we propose an algorithm for parameter selection which is based on the rank and condition number of the sensitivity matrix rather than the Fisher information matrix.

4 Subset selection algorithm

The identifiability analyses developed by Brun, et al., [11], and Burth, et al., [13], motivate the subset selection algorithm introduced in this section. Both of these approaches use submatrices of the Fisher information matrix in their selection procedures. Burth, et al., implemented a reduced-order estimation by determining which parameter axes lie closest to the ill-conditioned directions of the Fisher information matrix, and then by fixing the associated parameter values at priori estimates throughout an iterative estimation process. The subset selection keeps the well-conditioned parameters (those that can be estimated with little uncertainty from given measurements) active in the optimization, subject to having the corresponding Fisher information submatrix with a small condition number. Brun, et al., determine identifiability of parameter combinations using the eigenvalues of submatrices that result from excluding columns out of the Fisher information matrix. They quantify the near dependence of columns in the sensitivity submatrix using the smallest eigenvalue of the Fisher information submatrix.

We propose an algorithm that searches all possible parameter combinations and selects some of them, based on two main criteria: the full rank of the sensitivity matrix, and uncertainty quantification as embodied in asymptotic standard errors.

Our approach is numerical and we illustrate its use with the SEIRS model introduced earlier. To carry out the algorithm we require prior knowledge of nominal variance and nominal parameter

values. We assume the observation error variance is $\sigma_0^2 = 500$, and assume the following nominal parameter values for the SEIRS model:

$$S_0 = 2.78 \times 10^5 (\text{people}),\ E_0 = 1.08 \times 10^{-1} (\text{people}),\ I_0 = 1.89 \times 10^{-1} (\text{people}),$$
$$N = 1.00 \times 10^6 (\text{people}),\ L = 5.00 (\text{years}),\ D = 9.59 \times 10^{-3} (\text{years}),\ M = 5.48 \times 10^{-3} (\text{years}),$$
$$P = 75.00 (\text{years}),\ \beta_0 = 375.00 (\text{years}^{-1}),\ a_1 = 2.00 \times 10^{-2},\ b_1 = -2.00 \times 10^{-2}.$$

Henceforth, we use the terms "parameter combination" and "parameter vector" interchangeably. Parameter vectors $\theta \in \mathbb{R}^p$ will be considered for different fixed values of p. When $p = 11$ the parameter combination

$$\theta = (S_0, E_0, I_0, N, L, D, M, P, \beta_0, a_1, b_1) \in \mathbb{R}^{11}, \tag{46}$$

with the nominal parameter values given above, produces a rank deficient sensitivity matrix $\chi(\theta)$ for the SEIRS model. For $p = 3$ the only parameter combination considered here is that of the transmission parameters, i.e.,

$$\theta = (\beta_0, a_1, b_1) \in \mathbb{R}^3. \tag{47}$$

Other parameter vectors for fixed values of $p = 4, \ldots, 10$ are considered in the following way. For each fixed $j = 1, \ldots, 7$, and therefore fixed $p = 3 + j$, we explore parameter vectors of the form

$$\theta = (\lambda_1, \lambda_2, \ldots, \lambda_j, \beta_0, a_1, b_1) \in \mathbb{R}^p, \tag{48}$$

where for $k = 1, \ldots, j$,

$$\lambda_k \in \{S_0, E_0, I_0, N, L, D, M, P\} = \mathcal{I},$$

such that no entries of θ in equation (48) are repeated.

The set

$$\mathcal{S}_p = \{\theta = (\lambda_1, \lambda_2, \ldots, \lambda_j, \beta_0, a_1, b_1) \in \mathbb{R}^p = \mathbb{R}^{3+j} |\ \lambda_k \in \mathcal{I},\ \lambda_k \neq \lambda_m \forall\ k, m = 1, \ldots, j\} \tag{49}$$

collects the parameter vectors explored by a combinatorial search.

We define the set

$$\Theta_p = \{\theta |\ \theta \in \mathcal{S}_p \subset \mathbb{R}^p,\ \text{rank}(\chi(\theta)) = p\}, \tag{50}$$

where $\chi(\theta)$ denotes the $n \times p$ sensitivity matrix, and its rank is defined by equation (42). By construction, the elements of Θ_p are parameter vectors that give sensitivity matrices with independent columns.

An important step in the selection procedure involves the calculation of standard errors (uncertainty quantification) using the asymptotic theory described in Section 3.1. For every $\theta \in \Theta_p$, we define a vector of *coefficients of variation* $\nu(\theta) \in \mathbb{R}^p$ such that for each $i = 1, \ldots, p$,

$$\nu_i(\theta) = \frac{\sqrt{(\Sigma(\theta))_{ii}}}{\theta_i},$$

and

$$\Sigma(\theta) = \sigma_0^2 \left[\chi(\theta)^T \chi(\theta)\right]^{-1} \in \mathbb{R}^{p \times p}.$$

In other words, the components of the vector $\nu(\theta)$ are the ratios of each standard error for a parameter to the corresponding nominal parameter value. These ratios are dimensionless numbers

that allow comparison even when parameters have substantially different units and scales (e.g., N is on the order of 10^6, while a_1 is on the order of 10^{-2}). Next, define

$$\alpha(\theta) = |\nu(\theta)|\,.$$

We call $\alpha(\theta)$ the *parameter selection score*, and remark that $\alpha(\theta)$ near zero indicates lower uncertainty possibilities in the estimation while large values of $\alpha(\theta)$ suggest that one could expect to find wide uncertainty in at least some of the estimates.

In the optimization literature the term "feasible" usually denotes a vector satisfying inequality or equality constraints. Here we use this term in the context of identifiability: a feasible parameter vector denotes a combination that can be estimated from data with reasonable to little uncertainty. More precisely, we say a given $\theta \in \Theta_p$ is a *feasible parameter vector* if both $\alpha(\theta)$ and $\kappa(\chi(\theta))$ are relatively small.

We summarize the steps of the algorithm as follows:

1. **Combinatorial search.** For a fixed $j = 1, \ldots, 7$, and hence fixed $p = 3 + j$, calculate the set

$$\mathcal{S}_p = \{\theta = (\lambda_1, \lambda_2, \ldots, \lambda_j, \beta_0, a_1, b_1) \in \mathbb{R}^p |\ \lambda_k \in \mathcal{I},\ \lambda_k \neq \lambda_m \forall\ k, m = 1, \ldots, j\}.$$

   The set $\mathcal{S}_p$ collects all the parameter vectors obtained from a combinatorial search.

2. **Full rank test**. Calculate the set of viable parameters Θ_p as

$$\Theta_p = \{\theta |\ \theta \in \mathcal{S}_p \subset \mathbb{R}^p,\ \mathrm{rank}(\chi(\theta)) = p\}.$$

3. **Standard error test.** For every $\theta \in \Theta_p$ calculate a vector of coefficients of variation $\nu(\theta) \in \mathbb{R}^p$ by

$$\nu_i(\theta) = \frac{\sqrt{(\Sigma(\theta))_{ii}}}{\theta_i},$$

   for $i = 1, \ldots, p$, and $\Sigma(\theta) = \sigma_0^2 \left[\chi(\theta)^T \chi(\theta)\right]^{-1} \in \mathbb{R}^{p \times p}$. Calculate the parameter selection score as $\alpha(\theta) = |\nu(\theta)|\,.$

To illustrate the algorithm we consider several values of p, while using the MATLAB (The Mathworks, Inc.) routine `rank` (this routine computes the number of singular values that are greater than "machine tolerance").

Results for $p = 5$ (using the nominal parameter values) are displayed in Figure 1 (on logarithmic scales), where $\alpha(\theta)$ is depicted as a function of $\kappa(\chi(\theta))$ for all $\theta \in \Theta_5$. The pairs in the lower-left corner of Figure 1 correspond to feasible parameter vectors, because $\alpha(\theta)$ and $\kappa(\chi(\theta))$ are here relatively small.

The subset selection algorithm was applied for $p = 4, \ldots, 10$, while using the nominal variance and parameter values. We find that there is not a single parameter combination with $p = 10$ that has a full rank sensitivity matrix. For $p = 9$, only three parameter vectors pass the full rank test, and none of which can be considered feasible. We summarize the feasible parameter vectors in Table 1 for $p = 4, \ldots, 8$, where each feasible $\theta \in \Theta_p$ is displayed along with $\kappa(\chi(\theta))$ and $\alpha(\theta)$. The cutoffs used to select the parameter combinations in Table 1 were somewhat arbitrary but relative to the smallest values computed for the two criteria (condition number and selection score) in each example.

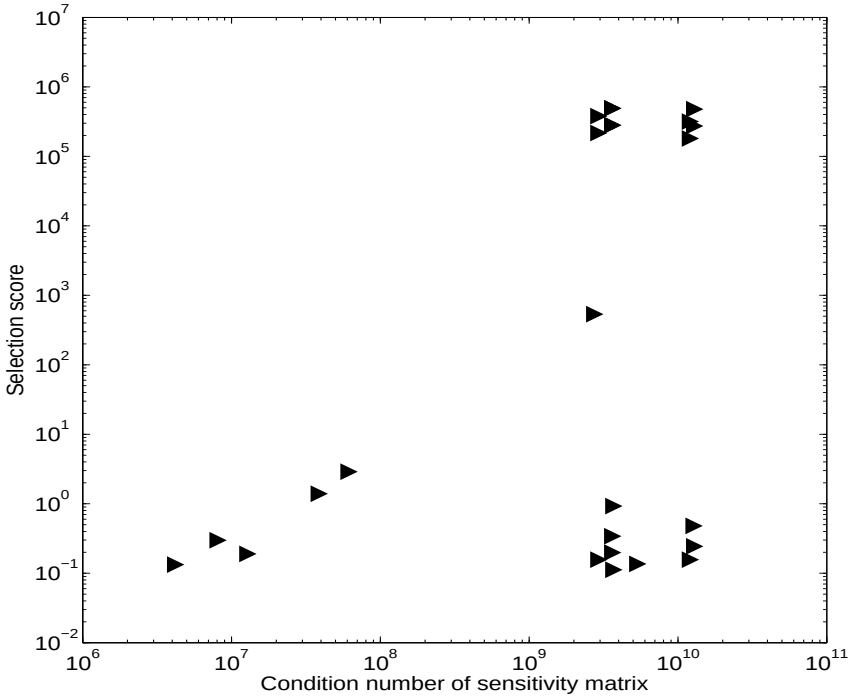

Figure 1: Parameter selection score $\alpha(\theta)$ versus the condition number $\kappa(\chi(\theta))$ of the $n \times p$ sensitivity matrix, for all parameter vectors $\theta \in \Theta_p$ with $p = 5$. Logarithmic scales are used on both axes.

5 Applications of the subset selection algorithm to synthetic and observed data sets

The subset selection algorithm is illustrated first by solving inverse problems from synthetic observations. To construct a synthetic data set we suppose a nominal parameter vector and a nominal error variance are equal to θ_0 (true parameter vector) and σ_0^2 (true variance), respectively. Random noise is then added to the model output as follows:

$$Y_j = z(t_j; \theta_0) + \sigma_0 V_j, \tag{51}$$

where V_j is a standard normal random variable, i.e., $V_j \sim \mathcal{N}(0, 1)$. A realization y_j of the observation process Y_j, is calculated by drawing independent samples v_j from the standard normal distribution so that

$$y_j = z(t_j; \theta_0) + \sigma_0 v_j \quad \text{for } j = 1, \ldots, n.$$

The OLS inverse problems were solved by implementing a subspace trust region method (based on an interior-reflective Newton method [33]). We used the MATLAB (The Mathworks, Inc.) routine `lsqnonlin`. For the purposes of this demonstration we initialized every optimization routine at the nominal parameter vector θ_0.

Table 1: Feasible parameter vectors obtained while applying the subset selection algorithm for $p = 4, \ldots, 8$, using nominal values as listed earlier in the text. For each selected parameter vector $\theta \in \Theta_p$ the condition number of the sensitivity matrix $\kappa(\chi(\theta))$, and the selection score $\alpha(\theta)$ are displayed.

Parameter vector θ	Condition number $\kappa(\chi(\theta))$	Selection score $\alpha(\theta)$
(L, β_0, a_1, b_1)	2.047×10^{5}	5.019×10^{-2}
(M, β_0, a_1, b_1)	1.420×10^{5}	6.386×10^{-2}
(P, β_0, a_1, b_1)	3.176×10^{5}	7.044×10^{-2}
$(L, D, \beta_0, a_1, b_1)$	4.034×10^{6}	1.332×10^{-1}
$(D, M, \beta_0, a_1, b_1)$	1.233×10^{7}	1.897×10^{-1}
$(D, P, \beta_0, a_1, b_1)$	7.781×10^{6}	2.987×10^{-1}
$(N, L, D, \beta_0, a_1, b_1)$	1.829×10^{10}	1.670×10^{-1}
$(S_0, N, D, \beta_0, a_1, b_1)$	1.454×10^{10}	2.026×10^{-1}
$(S_0, L, D, \beta_0, a_1, b_1)$	1.828×10^{10}	2.375×10^{-1}
$(S_0, D, M, \beta_0, a_1, b_1)$	2.152×10^{10}	3.301×10^{-1}
$(S_0, D, P, \beta_0, a_1, b_1)$	1.828×10^{10}	4.832×10^{-1}
$(N, D, M, \beta_0, a_1, b_1)$	2.166×10^{10}	5.739×10^{-1}
$(N, D, P, \beta_0, a_1, b_1)$	1.829×10^{10}	9.658×10^{-1}
$(N, L, D, M, \beta_0, a_1, b_1)$	2.166×10^{10}	5.960×10^{0}
$(S_0, L, D, M, \beta_0, a_1, b_1)$	2.167×10^{10}	5.970×10^{0}
$(N, D, M, P, \beta_0, a_1, b_1)$	2.166×10^{10}	1.153×10^{1}
$(S_0, D, M, P, \beta_0, a_1, b_1)$	2.167×10^{10}	1.159×10^{1}
$(S_0, N, L, D, M, \beta_0, a_1, b_1)$	6.333×10^{12}	5.044×10^{1}
$(S_0, N, D, M, P, \beta_0, a_1, b_1)$	6.561×10^{12}	2.950×10^{2}

The nominal error variance and nominal parameter values are those given in the previous section. The parameter vectors estimated from synthetic data are those appearing on top of each subtable in Table 2, for each value of p, where parameter combinations are sorted in ascending order of their selection score (from top to bottom). In other words, all the parameter vectors estimated from synthetic observations have reasonable condition numbers and relatively small selection scores. Five inverse problems (for $p = 8, 7, 6, 5, 4$) were solved from the same realization of the observation process, to estimate the parameter vectors

$$\theta = (S_0, N, L, D, M, \beta_0, a_1, b_1),$$
$$\theta = (N, L, D, M, \beta_0, a_1, b_1),$$
$$\theta = (N, L, D, \beta_0, a_1, b_1),$$
$$\theta = (L, D, \beta_0, a_1, b_1),$$
$$\theta = (L, \beta_0, a_1, b_1).$$

Results of these numerical experiments are summarized in Table 2.

We analyze the results using the coefficient of variation: standard error (SE) divided by estimate (Est). For instance in Table 2, when $\theta = (S_0, N, L, D, M, \beta_0, a_1, b_1)$ it is seen for D that the standard

Table 2: Results of solving five inverse problems from a single synthetic data set generated as described in the text using nominal values listed earlier. For each parameter combination we display the estimate (Est.), the standard error (S.E.) and the coefficient of variation (standard error divided by the estimate, C.V. = S.E./Est.). For notational convenience we use here the notation e to denote exponentiation to the base 10; i.e., $2.8e^5$ denotes 2.8×10^5, etc.

Parameter vector $\theta = (S_0, N, L, D, M, \beta_0, a_1, b_1)$								
	S_0	N	L	D	M	β_0	a_1	b_1
Est.	$2.8e^5$	$1.0e^6$	$5.0e^0$	$9.6e^{-3}$	$5.5e^{-3}$	$3.7e^2$	$2.0e^{-2}$	-$2.0e^{-2}$
S.E.	$1.5e^6$	$5.0e^6$	$4.5e^1$	$3.1e^{-3}$	$6.2e^{-2}$	$3.4e^3$	$7.7e^{-2}$	$8.4e^{-2}$
C.V.	$5.5e^0$	$5.0e^0$	$9.1e^0$	$3.2e^{-1}$	$1.1e^1$	$9.0e^0$	$3.8e^0$	-$4.2e^0$
Parameter vector $\theta = (N, L, D, M, \beta_0, a_1, b_1)$								
Est.		$1.0e^6$	$5.0e^0$	$9.6e^{-3}$	$5.5e^{-3}$	$3.7e^2$	$2.0e^{-2}$	-$2.0e^{-2}$
S.E.		$2.7e^4$	$2.7e^0$	$2.5e^{-3}$	$2.2e^{-2}$	$5.9e^2$	$3.1e^{-2}$	$2.5e^{-2}$
C.V.		$2.7e^{-2}$	$5.4e^{-1}$	$2.6e^{-1}$	$4.1e^0$	$1.6e^0$	$1.6e^0$	-$1.3e^0$
Parameter vector $\theta = (N, L, D, \beta_0, a_1, b_1)$								
Est.		$1.0e^6$	$5.0e^0$	$9.6e^{-3}$		$3.8e^2$	$2.0e^{-2}$	-$2.0e^{-2}$
S.E.		$2.7e^4$	$1.7e^{-1}$	$5.8e^{-4}$		$1.5e^1$	$1.3e^{-3}$	$1.2e^{-3}$
C.V.		$2.7e^{-2}$	$3.4e^{-2}$	$6.1e^{-2}$		$3.9e^{-2}$	$6.3e^{-2}$	-$6.1e^{-2}$
Parameter vector $\theta = (L, D, \beta_0, a_1, b_1)$								
Est.			$5.0e^0$	$9.6e^{-3}$		$3.8e^2$	$2.0e^{-2}$	-$2.0e^{-2}$
S.E.			$7.4e^{-2}$	$5.8e^{-4}$		$9.8e^0$	$1.2e^{-3}$	$1.2e^{-3}$
C.V.			$1.5e^{-2}$	$6.1e^{-2}$		$2.6e^{-2}$	$6.2e^{-2}$	-$6.0e^{-2}$
Parameter vector $\theta = (L, \beta_0, a_1, b_1)$								
Est.			$5.0e^0$			$3.8e^2$	$2.0e^{-2}$	-$2.0e^{-2}$
S.E.			$1.4e^{-2}$			$2.6e^0$	$2.0e^{-4}$	$7.9e^{-4}$
C.V.			$2.7e^{-3}$			$6.8e^{-3}$	$9.9e^{-3}$	-$4.0e^{-2}$

error is nearly one third of the estimate, suggesting lower uncertainty. For the other parameters S_0, N, L, M, β_0, a_1, and b_1 the standard error can be nearly four times (and up to eleven times) the estimate (for b_1 its SE is $|4 \times \text{Est}|$, because $b_1 < 0$). This feature denotes substantial uncertainty. Figure 2(a) displays the residual plot (see [4] for a discussion of the effective use of residual plots) for this parameter combination: $y_j - z(t_j; \hat{\theta}_{OLS})$ versus time t_j, where $j = 1, \ldots, n$. The temporal pattern in the residuals together with large standard errors suggest that estimation of this parameter combination from observations (with a statistical error structure) would be meaningless.

The residual plots for all the other parameter combinations in Table 2 do not have temporal patterns. For the sake of illustration we display in Figure 2(b) the residuals versus time for $\theta = (L, D, \beta_0, a_1, b_1)$.

Improvements in uncertainty quantification are observed with the removal of some key parameters. We think it is not just reducing the number p of parameters, but rather which parameters are to be estimated that really counts. The near dependence in the columns of the sensitivity matrix χ reflects correlations between parameter estimates which make a parameter combination unsuit-

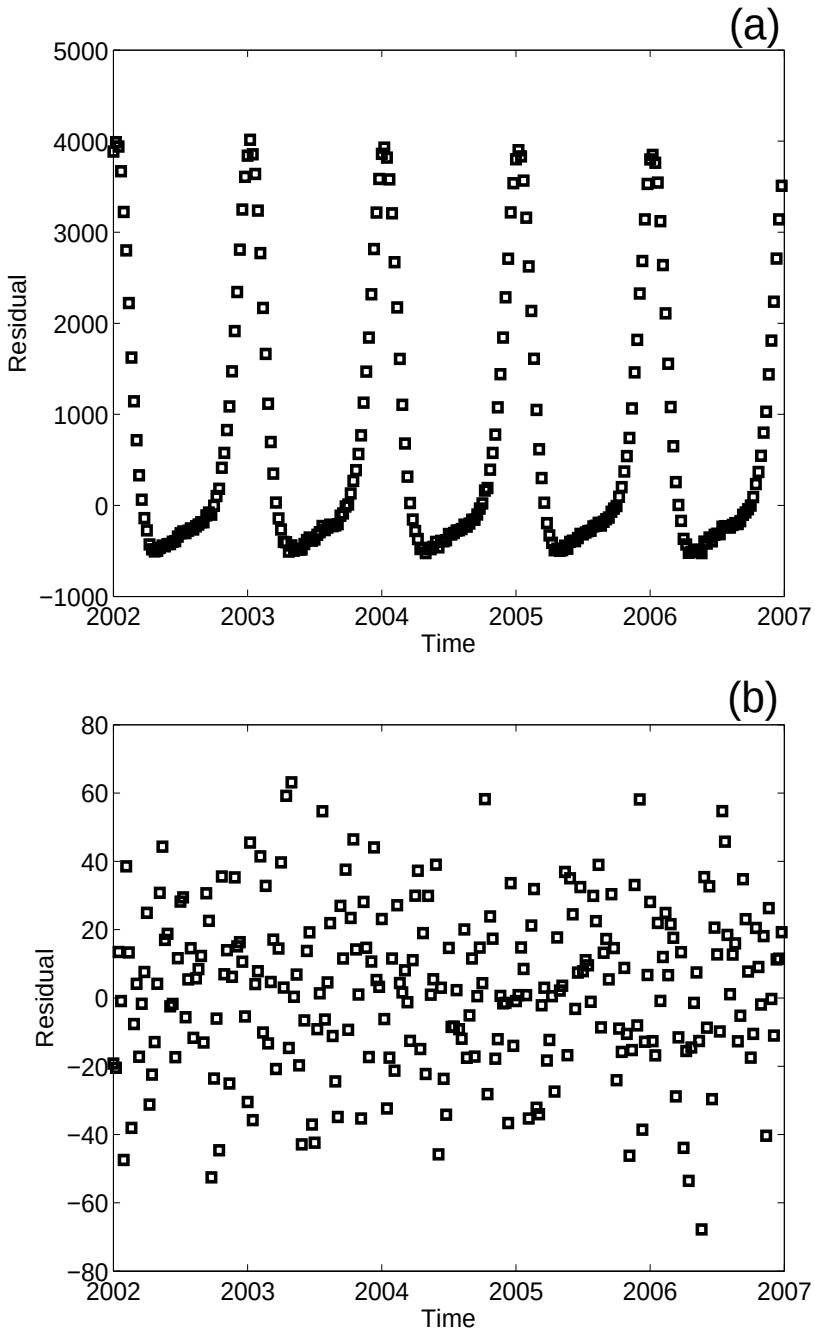

Figure 2: Residual plots: $y_j - z(t_j; \hat{\theta}_{OLS})$, versus time, t_j, for $j = 1, \ldots, n$. Graph (a) displays residuals obtained for $\theta = (S_0, N, L, D, M, \beta_0, a_1, b_1)$, while Graph (b) depicts residuals for $\theta = (L, D, \beta_0, a_1, b_1)$.

able for estimation. For instance, consider the removal of S_0 from the estimation, and compare

$\theta = (S_0, N, L, D, M, \beta_0, a_1, b_1)$ with $\theta = (N, L, D, M, \beta_0, a_1, b_1)$ in Table 2. The standard error for N is seen to drop from 500% to approximately 3% of the estimate. Another substantial improvement when dropping S_0 is obtained for L, for which its standard error reduces from being nine times the estimate to one half of its value. Lower uncertainty improvements are also obtained for the parameters M, β_0, a_1, and b_1.

The next numerical experiment considered here is the removal of S_0 and M. We compare the results for $\theta = (S_0, N, L, D, M, \beta_0, a_1, b_1)$ with those for $\theta = (N, L, D, \beta_0, a_1, b_1)$, in Table 2. There are uncertainty improvements for all parameters. The least (but still substantial) improvement is for D, where its standard error drops from being nearly 30% to being just 6% of the estimate. For the parameters N, L, β_0, a_1, and b_1 an improvement of two orders of magnitude is seen. Improvements in uncertainty are more pronounced after removing S_0, N, and M: for this we compare $\theta = (S_0, N, L, D, M, \beta_0, a_1, b_1)$ and $\theta = (L, D, \beta_0, a_1, b_1)$ in Table 2.

Undoubtedly, the best case scenario of uncertainty quantification we obtained is that of estimating $\theta = (L, \beta_0, a_1, b_1)$ from the same synthetic data set. In Table 2, it is seen that the standard errors reduce to less than 1% of the estimates for L, β_0, and a_1, and to 4% from nearly 400% of the estimate for b_1.

As a final note in this section, we present results obtained from solving the OLS problem while using observations of an influenza-like-illness in France [38]. Some of the parameters were fixed to values suggested in [15, 16, 20]:

$$S_0 = 1.56 \times 10^7 (\text{people}),\ E_0 = 6.44(\text{people}),\ I_0 = 12.88(\text{people}),\ N = 6.40 \times 10^7 (\text{people}),$$
$$L = 6.00(\text{years}),\ D = 1.10 \times 10^{-2}(\text{years}),\ M = 5.50 \times 10^{-3}(\text{years}),\ P = 80.87(\text{years}).$$

The inverse problem was solved with $\theta = (\beta_0, a_1, b_1)$. Simple inspection of the standard errors in Table 3 does not seem to immediately suggest there is a poor fit (not displayed here). Roughly speaking, the standard error is: 13% of the estimate for β_0; 30% of the estimate for a_1; 45% of the estimate for b_1. These calculations give an indication of wide uncertainty, but they are not as extreme as the results for $\theta = (S_0, N, L, D, M, \beta_0, a_1, b_1)$ in Table 2. One can easily be misled by invalid uncertainty quantification in the absence of residual analysis. Residual plots (not displayed here) in this case have systematic patterns, suggesting that either the statistical model (equation (15)) may be incorrect, or more likely, the SEIRS model fails to adequately describe the underlying process.

Table 3: Estimates from influenza-like-illness observations, where $\theta = (\beta_0, a_1, b_1)$. The coefficient of variation is defined as the standard error divided by the estimate.

Parameter	Estimate	Standard error	Unit	Coefficient of variation
β_0	3.100×10^{2}	4.055×10^{1}	years^{-1}	1.308×10^{-1}
a_1	1.539×10^{-2}	4.588×10^{-3}	1	2.981×10^{-1}
b_1	-2.406×10^{-2}	1.090×10^{-2}	1	-4.530×10^{-1}

6 Discussion

We have discussed a computational methodology for inverse problem formulation in the context of parameter identifiability. Using an OLS scheme based on a constant variance statistical model for the observation process and a seasonal SEIRS epidemics model for illustration, we have proposed a prior-analysis algorithm that we believe might profitably precede efforts on parameter estimation from data. The algorithm can be used if reasonable ranges for the sought after parameters are either known a priori, or can be assumed by the user much in the same way one must assume reasonable ranges in inverse problem formulations and initiation of algorithms for the resulting estimation procedures.

The subset selection [31] algorithm given in Section 4 is based on two main criteria for a fixed number of parameters: (i) full rank of the sensitivity matrix; and (ii) calculation of standard errors. We proposed to first select according to the sensitivity matrix rank, because those parameter combinations for which χ has full rank will have a non-singular Fisher information matrix $\chi^T\chi$, and its inverse is used in the calculation of the standard errors (see equation (26)).

The near dependence of the sensitivity matrix columns can be a fingerprint of parameter correlations–a pertinent feature for subset selection [31]. Capaldi, et al., [14] determine identifiability of parameters in a simple SIR model, and show how correlation between parameter estimates can impede the estimation of other parameters and parameter combinations, such as the basic reproductive number. Moreover, Brun, et al., [11] explain that if the columns of χ are nearly dependent, then changes in the model output due to small changes in a single parameter can be compensated by appropriate changes in other parameters.

We have presented illustrations of the how the removal of nearly dependent columns of the sensitivity matrix can provide substantial improvements in uncertainty quantification. This feature involves more than just reducing the number p of parameters, it relates to excluding certain key parameters. For instance, if we assume a linear Taylor expansion of the model output, the estimator $\theta_{OLS} \in \mathbb{R}^p$ is given by equation (40), where the sensitivity matrix $\chi(\theta_0)$ has singular values $s_1 \geq \cdots \geq s_{p-1} \geq s_p > 0$. If $s_p \approx 0$ and $s_{p-1} > 1$, then submatrices with singular values $s_2 \geq \cdots \geq s_p > 0$, and $s_1 \geq \cdots \geq s_{p-1}$, have different conditioning when quantifying the sensitivity of *reduced order* estimations that only involve $p-1$ parameters. The condition number of the former submatrix is s_2/s_p, which is large if $s_p \approx 0$, while for the latter submatrix the condition number satisfies $1 \leq s_1/s_{p-1} < s_1$, because $s_{p-1} > 1$.

In our numerical experiments, we calculate sensitivity matrices $\chi(\theta)$ evaluated at different realizations of the estimator $\theta = \hat{\theta}_{OLS}$. When $\theta = (S_0, N, L, D, M, \beta_0, a_1, b_1)$ the singular values of the sensitivity matrix range from 4.7×10^6 to 4.6×10^{-6} while for $\theta = (L, \beta_0, a_1, b_1)$ the singular values of $\chi(\hat{\theta}_{OLS})$ range from 1.9×10^6 to 9.3×10^0.

The smallest singular value changes from 4.6×10^{-6} to 9.3×10^0 while the largest remain on the order of 10^6. This improvement in conditioning is reflected in the the standard error for L, β_0, and a_1, which reduces to less than 1% of the estimate, from nearly 900% and 380% (see Table 2).

Although in this paper we only discuss OLS, the selection algorithm can be easily applied when using a generalized least squares scheme [4]. We also carried out numerical experiments (for brevity not discussed here) involving use of synthetic nonconstant variance data sets in GLS formulations, and obtained results absolutely consistent with those of the OLS formulation presented here (Section 5).

Acknowledgements

A.C.-A. is grateful to Dr. Sava Dediu for helpful discussions about sensitivity identifiability that led to the initial version of this manuscript. This research was supported in part by Grant Number R01AI071915-07 from the National Institute of Allergy and Infectious Diseases and in part by the Air Force Office of Scientific Research under grant number FA9550-09-1-0226. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIAID, the NIH or the AFOSR.

References

[1] R. Anderson and B.T. Grenfell, Oscillatory fluctuations in the incidence of infectious disease and the impact of vaccination: time series analysis, *J. Hyg. Camb.*, **93** (1984), 587–608.

[2] D.T. Anh, M.P. Bonnet, G. Vachaud, C.V. Minh, N. Prieur, L.V. Duc and L.L. Anh, Biochemical modeling of the Nhue River (Hanoi, Vitenam): practical identifiability analysis and parameter estimation, *Ecol. Model.*, **193** (2006), 182–204.

[3] K.J. Astrom and P. Eykhoff, System identification–A survey, *Automatica*, **7** (1971), 123–162.

[4] H.T. Banks, M. Davidian, J.R. Samuels and K.L. Sutton, An inverse problem statistical methodology summary, Center for Research in Scientific Computation Technical Report CRSC-TR08-1, NCSU, January, 2008; in *Mathematical and Statistical Estimation Approaches in Epidemiology*, (eds. G. Chowell, et. al.), Springer, New York, 2009, pp. 249–302.

[5] H.T. Banks, S. Dediu and S.E. Ernstberger, Sensitivity functions and their uses in inverse problems, *CRSC Tech Report*, CRSC-TR07-12, NCSU, July, 2007; *J. Inverse and Ill-posed Problems*, **15** (2007), 683–708.

[6] H.T. Banks, S. Dediu, S.L. Ernstberger, F. Kappel, A new approach to optimal design problems, Center for Research in Scientific Computation Technical Report CRSC-TR08-12, NCSU, September, 2008.

[7] H. T. Banks, S. L. Ernstberger and S. L.Grove, Standard errors and confidence intervals in inverse problems: sensitivity and associated pitfalls, *J. Inverse and Ill-posed Problems*, **15** (2007), 1–18.

[8] H.T. Banks and J.R. Samuels, Jr., Detection of cardiac occlusions using viscoelastic wave propagation, CRSC-TR08-23, December, 2008; *Advances in Applied Mathematics and Mechanics,* **1** (2009), 1–28.

[9] R. Bellman and K.M. Astrom, On structural identifiability, *Math. Biosci.*, **7** (1970), 329–339.

[10] R. Bellman and R. Kalaba, *Quasilinearization and Nonlinear Boundry Value Problems*, American Elsevier, New York, 1965.

[11] R.Brun, M. Kuhni, H. Siegrist, W. Gujer and P. Reichert, Practical identifiability of ASM2d parameters — systematic selection and tuning of parameter subsets, *Water Res.*, **36** (2002), 4113–4127.

[12] K.P. Burnham and D.R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, Springer-Verlag, New York, 2002.

[13] M. Burth, G.C. Verghese and M. Vélez-Reyes, Subset selection for improved parameter estimation in on-line identification of a synchronous generator, *IEEE T. Power Syst.*, **14** (1999), 218–225.

[14] A. Capaldi, S. Behrend, B. Berman, J. Smith, J. Wright and A.L. Lloyd, Parameter estimation and uncertainty quantification for an epidemic model, in preparation.

[15] Central Intelligence Agency World Factbook (2008), https://www.cia.gov/library/publications/the-world-factbook/index.html, cited 12 Nov 2008.

[16] G. Chowell, M.A. Miller and C. Viboud, Seasonal influenza in the United States, France, and Australia: transmission and prospects for control, *Epidemiol. Infect.*, **136** (2008), 852–864.

[17] A. Cintrón-Arias, C. Castillo-Chávez, L.M.A. Bettencourt, A.L. Lloyd, and H.T. Banks, The estimation of the effective reproductive number from disease outbreak data, Center for Research in Scientic Computation Technical Report CRSC-TR08-08, NCSU, April, 2008; *Math. Biosci. Engr.*, **6** (2009), 261–283.

[18] C. Cobelli and J. J. DiStefano III, Parameter and structural identifiability concepts and ambiguities: a critical review and analysis, *Am. J. Physiol.* **239** (1980), R7–R24.

[19] M. Davidian and D.M. Giltinan, *Nonlinear Models for Repeated Measurement Data*, Chapman & Hall, Boca Raton, 1995.

[20] J. Dushoff, J.B. Plotkin, S.A. Levin and D.J. Earn, Dynamical resonance can account for seasonality of influenza epidemics, *P. Natl. Acad. Sci. USA*, **101** (2004), 16915–16916.

[21] N.D. Evans, L.J. White, M.J. Chapman, K.R. Godfrey and M.J. Chappell, The structural identifiability of the susceptible infected recovered model with seasonal forcing, *Math. Biosci.*, **194** (2005), 175–197.

[22] P. Eykhoff, *System Identification: Parameter and State Estimation*, Wiley & Sons, New York, 1974.

[23] K. Glover and J.C. Willems, Parametrizations of linear dynamical systems: Canonical forms and identifiability, *IEEE Trans. Automat. Contr.*, **AC-19** (1974), 640–645.

[24] G.H. Golub and C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1996.

[25] N.C. Grassly and C. Fraser, Seasonal infectious disease epidemiology, *P. Roy. Soc. B-Biol. Sci.*, **273** (2006), 2541–2550.

[26] A. Holmberg, On the practical identifiability of microbial growth models incorporating Michaelis-Menten type nonlinearities, *Math. Biosci.*, **62** (1982), 23–43.

[27] M. Kalivianakis, S.L.J. Mous and J. Grasman, Reconstruction of the seasonally varying contact rate for measles, *Math. Biosci.*, **124** (1994), 225–234.

[28] R.E Kalman, Mathematical description of linear dynamical systems, *SIAM J. Control*, **1** (1963), 152–192.

[29] Y.A. Kuznetsov and C. Piccardi, Bifurcation analysis of epidemic SEIR and SIR epidemic models, *J. Math. Biol.*, **32** (1994), 109–121.

[30] A.K. Mehra and D.G. Lainiotis, *System Identification*, Academic Press, New York, 1976.

[31] A.J. Miller, *Subset Selection in Regression*, Chapman & Hall, New York, 1990.

[32] I.M. Navon, Practical and theoretical aspects of adjoint parameter estimation and idenrtifiability in meteorology and oceanography, *Dyn. Atmospheres and Oceans*, **27** (1997), 55–79.

[33] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer-Verlag, New York, 1999.

[34] J.G. Reid, Structural identifiability in linear time-invariant systems, *IEEE Trans. Automat. Control*, **22** (1977), 242–246.

[35] F. Riedo, B. Plikaytis and C. Broome, Epidemiology and prevention of meningococcal disease, *Pediatr. Infect. Dis. J.*, **14** (1995), 643–657.

[36] A.P. Sage and J.L. Melsa, *System Identification*, Academic Press, New York, 1971.

[37] G.A.F. Seber and C.J. Wild, *Nonlinear Regression*, John Wiley & Sons, Chichester, 2003.

[38] Sentinelles Influenza-like Illness (2008), http://websenti.b3e.jussieu.fr/sentiweb/?page=maladies&mal=3, cited 12 Nov 2008.

[39] I.B. Schwartz and H.L. Smith, Infinite subharmonic bifurcation in an SEIR epidemic model, *J. Math. Biol.*, **18** (1983), 233–253.

[40] K. Thomaseth and C. Cobelli, Generalized sensitivity functions in physiological system identification, *Annals of Biomedical Engineering*, **27** (1999), 607–616.

[41] L.J. White, N.D. Evans, T.J.G.M. Lam, Y.H. Schukken, G.F. Medley, K.R. Godfrey and M.J. Chappell, The structural identifiability and parameter estimation of a multispecies model for the transmission of mastitis in diary cows, *Math. Biosci.*, **174** (2001), 77–90.

[42] J. Witte, A. Karchmer, M. Case, K.L. Hermann, E. Abrutyn, I. Kassanof, et al.:, Epidemiology of rubella, *Am. J. Dis. Child.*, **118** (1969), 107–111.

[43] H. Wu, H. Zhu, H. Miao and A.S. Perelson, Parameter identifiability and estimation of HIV/AIDS dynamics models, *Bull. Math. Biol.*, **70** (2008), 785–799.

[44] X. Xia and C.M. Moog, Identifiability of nonlinear systems with application to HIV/AIDS models, *IEEE T. Automat. Contr.*, **48** (2003), 330–336.

[45] H. Yue, M. Brown, F. He, J. Jia and D.B. Kell, Sensitivity analysis and robust experimental design of a signal transduction pathway system, *Int. J. Chem. Kinet.*, **40** (2008), 730–741.