Subset Selection for Improved Parameter Estimation in On-Line Identification of a Synchronous Generator

MICHAEL BURTH† GEORGE C. VERGHESE ‡ MIGUEL VÉLEZ-REYES‡‡

†Technical University Berlin ‡Massachusetts Institute of Technology
‡‡University of Puerto Rico

Abstract — **This paper examines subset selection for nonlinear least squares parameter estimation, and applies the methodology to a test system previously studied in the power system literature, involving the on-line identification of a synchronous generator model with many parameters. Subset selection partitions the parameters into well-conditioned and ill-conditioned subsets. We show for the test system that fixing the ill-conditioned parameters to prior estimates (even if these prior estimates are substantially in error), and estimating only the remaining parameters, significantly improves the performance of the estimation algorithm and greatly enhances the quality of the estimated parameters. It is shown that attempts to estimate all of the model parameters, as done in the original work with this test system, can yield extremely unreliable results.**

Key Words — **Least Squares, Estimation, Identification, Conditioning, Subset Selection, Synchronous Generator.**

I. INTRODUCTION

Parameter estimation is a key step in fitting a model to measurements, and is therefore central to the task of system identification. The parameter estimates are chosen so as to minimize a measure of the errors between: (i) the model's predictions of what values the available measurements will take, and (ii) the actual values taken by the measurements. Methods of solving this minimization problem are dependent on the structure of the model and on the error criterion. For models whose predictions are linear in the parameters, and with an error measure that is the sum of squared prediction errors, efficient and stable *linear least squares* estimation techniques are available (see, e.g. [1, 2, 3]). For models that are *nonlinear* in the parameters, least squares estimation involves iterative methods, of which the Gauss-Newton *iterated linearized* least squares method is among the most used [1, 4, 5, 6].

Several aspects of the model and measurements affect the performance of a parameter estimation algorithm and determine the quality of the estimates that are produced. What happens very often in the power system setting is that a component model is built up from detailed analysis of the underlying physical phenomena, and therefore involves a relatively large number of physically interesting and interpretable parameters; this is certainly the case with the synchronous generator models found in the literature, [7, 8]. On the other hand, the measurements available from on-line experiments in an interconnected power system — which is the setting that is ultimately of interest — are typically *not* rich enough to adequately reflect the individual effects of all the parameters in the various components of the system. This mismatch between the (high) detail of the models and the (low) richness in the measurements leads to very sensitive or ill-conditioned parameter estimation problems.

The purpose of this paper is to point out the manifestations of such ill-conditioning in the context of parameter estimation for a synchronous generator model, and to suggest a strategy for overcoming the ill-conditioning. Our results apply more broadly than to synchronous generator identification, but this particular identification problem is important enough to have been studied fairly extensively in the power systems literature — see for example [9, 10, 11, 12, 13, 14, 15, 16] and references therein — and therefore provides a fruitful context for our study. In particular, the test system that we use is taken from [16]. Some mention of conditioning in the generator identification problem appears in [11, 13, 14], but the theme does not seem to have been developed in any detail prior to now.

Our strategy for overcoming ill-conditioning is based on the *subset selection* approach proposed in [17, 18] for nonlinear least squares parameter estimation, extending subset selection for linear least squares estimation as described in [2]. Subset selection partitions the model parameters into *well-conditioned* parameters, which are likely to be estimated reliably from the given measurements, and *ill-conditioned* parameters, whose estimates are likely to be unreliable, and whose presence makes the parameter estimation problem very sensitive. Given this partitioning, we proposed in [17, 18] to fix the ill-conditioned parameters at prior estimates, in effect abandoning any attempt to estimate them from the available measurements, and to then solve a reduced-order and well-conditioned parameter estimation problem to determine the remaining parameters. This strategy is successful if the bias introduced by fixing the ill-conditioned parameters to prior estimates is more than made up for by the improvement in estimation of the remaining parameters. The application of this strategy to the estimation of induction machine speed and parameters in [17, 18] led to major performance improvements over full-order estimation.

We demonstrate in this paper, for the synchronous generator

PE-465-PWRS-0-04-1998 A paper recommended and approved by the IEEE Power System Dynamic Performance Committee of the IEEE Power Engineering Society for publication in the IEEE Transactions on Power Systems. Manuscript submitted August 4, 1997; made available for printing April 24, 1998.

model used in the identification experiments of [16], and for qualitatively similar "measurements" (synthesized from the model in [16], since the original data was not available to us), that the proposed strategy leads to a reduced-order estimation procedure and associated parameter estimates that are much better behaved than if all the parameters are estimated together. Section II of the paper briefly reviews the nonlinear least squares problem, highlighting the role of the Jacobian (or gradient or first derivative matrix) of the error vector with respect to the parameter vector in finding the least squares estimate by the Gauss-Newton method; the Hessian (or second derivative matrix) of the error criterion with respect to the parameter vector is also defined. The section then gives the main idea behind the subset selection procedure, applied to the Jacobian or Hessian, and finally specifies the algorithm in more detail. Section III describes the test system, explains how the "measurements" for our identification experiments were synthesized, and then presents the results of our various estimation experiments on the system. Some conclusions are stated in Section IV.

II. Least Squares Estimation and Subset Selection

A. The Nonlinear Least Squares Problem

Least squares fitting of a model to experimental data is a common procedure in engineering. In this method, the parameters of a model are determined such that they minimize the sum of squares of the components of the N-component error (or "residual") vector

$$\mathbf{r}(\theta) = \widehat{\mathbf{y}}(\theta) - \mathbf{y} \,, \tag{1}$$

where θ denotes the n-vector of model parameters, $\widehat{\mathbf{y}}(\theta)$ is the N-vector of model predictions for the measurements, and $\mathbf{y}$ is the N-vector of actual measurements. Stated mathematically, the vector of parameter estimates is specified as

$$\widehat{\theta} = \arg\min_{\theta} V(\theta) \,, \tag{2}$$

where the minimization criterion for least squares is defined by

$$V(\theta) = \frac{1}{2}\|\mathbf{r}(\theta)\|^2 = \frac{1}{2}\sum_{\ell=1}^{N} r_\ell^2(\theta) \,, \tag{3}$$

with $r_\ell(\theta)$ denoting the ℓth component of the error vector, and the factor of $\frac{1}{2}$ being simply for convenience in some later expressions. Minimization of the above criterion by the Gauss-Newton method involves iterated *linearization* of the nonlinear problem around the current best guess of the parameter estimates, and therefore requires the $N \times n$ *Jacobian* or gradient or matrix of first partial derivatives of the error vector with respect to the parameter vector:

$$\mathbf{J}(\theta) = \frac{\partial \mathbf{r}(\theta)}{\partial \theta} \,. \tag{4}$$

The Gauss-Newton method begins with an initial guess of the parameter estimate, say $\widehat{\theta}_0$. The next guess is then computed as

$$\widehat{\theta}_1 = \widehat{\theta}_0 + \alpha_1 \mathbf{p}_1 \,, \tag{5}$$

where α_1 is a scalar (of the order of unity) that fixes the step size in the Gauss-Newton *direction* $\mathbf{p}_1$. This direction is computed by solving a *linear* least squares problem associated with the linearization of the original problem around the initial guess. Specifically, $\mathbf{p}_1$ satisfies the so-called "normal" equations:

$$(\mathbf{J}'\mathbf{J})\,\mathbf{p}_1 = \mathbf{J}'\,\mathbf{r} \,, \tag{6}$$

where $'$ denotes matrix transposition, and both $\mathbf{J}$ and $\mathbf{r}$ in (6) are evaluated at the current estimate $\widehat{\theta}_0$. In principle, $\mathbf{p}_1$ could be found by inverting the matrix $(\mathbf{J}'\mathbf{J})$ that pre-multiplies it in the above equation, but from a numerical point of view there are better methods of actually computing $\mathbf{p}_1$, see [2]. Note that $(\mathbf{J}'\mathbf{J})$ is invertible, and correspondingly $\mathbf{p}_1$ is uniquely determinable, if and only if the n columns of $\mathbf{J}$ are independent. From the definition in (4), we see that this condition is equivalent to requiring that increments in the various parameters should perturb the error vector in n independent directions (in N-space).

The step size α_1 in (5) may be picked so as to obtain (close to) the greatest possible decrease of the criterion $V(\cdot)$ by movement in the specified direction. Once $\widehat{\theta}_1$ has been found, the entire procedure is repeated, but now linearizing around $\widehat{\theta}_1$. This iteration is continued until the desired degree of convergence has been achieved. A more detailed algorithm description may be found in [4]; a clear summary in the context of generator identification is given in [12].

Other Newton-type approaches to solving the minimization problem (2) involve the *Hessian*, which is the $n \times n$ matrix of second partial derivatives of the error criterion $V(\theta)$ with respect to the parameter vector θ, and is easily seen to be given by

$$\mathbf{H}(\theta) = \mathbf{J}'(\theta)\mathbf{J}(\theta) + \sum_{\ell=1}^{N} r_\ell(\theta)\,\frac{\partial r_\ell(\theta)}{\partial \theta\,\partial \theta'} \,. \tag{7}$$

For small residuals, the Hessian can evidently be approximated by

$$\mathbf{H}(\theta) \approx \mathbf{J}'(\theta)\mathbf{J}(\theta) \,, \tag{8}$$

which is the matrix on the left side of (6). In the remainder of this paper, we shall use the term "Hessian" and the symbol $\mathbf{H}$ to refer to this approximation of the strict Hessian, namely $\mathbf{J}'\mathbf{J}$.

Efficient methods exist for computing the Jacobian when the data is modeled as comprising time samples of the output of a state-space model, which is the case with our test system. For a detailed description of these methods, we refer the reader to [4]. Explicit treatments of the application of least squares parameter estimation in this manner to the identification of synchronous machines, including the calculation of gradient functions, can be found in [9, 12, 14].

B. Parameter Conditioning

The Hessian matrix $\mathbf{H} = \mathbf{J}'\mathbf{J}$ on the left of the normal equations (6) is symmetric and positive semidefinite, so all its eigenvalues are real and non-negative. Suppose $\mathbf{H}$ is actually singular, with just one eigenvalue at 0 and some associated eigenvector; this happens if and only if the n columns of $\mathbf{J}$ actually contain only $n-1$ independent vectors. An immediate implication of the singularity is that the step direction computed from the normal equations (6) can be varied in the direction of this eigenvector of $\mathbf{H}$ without affecting the error criterion (at least to first order). Such indeterminacy would be highly undesirable in a physical parameter estimation problem, because it would indicate that the parameters cannot be unambiguously estimated from the given measurements. Note that such parameter indeterminacy can exist even if the model predictions fit the measurements exactly, i.e. even if