Proceedings of the
American Control Conference
Baltimore, Maryland • June 1994

Smoothly Time Varying Systems and Toeplitz Least Squares Problems

Michael Stewart and Paul Van Dooren
Coordinated Science Lab

University of Illinois
Urbana, IL 61801

This work was partially supported by
NSF(grant CCR 9209349) and DARPA(grant 60NANB2D1272)

Abstract

This paper explores the implications of assuming a system to be smoothly time-varying for least squares based system identification, as well as conditions under which least squares solutions are smoothly time-varying. By requiring persistent excitation and that the order of the model be chosen appropriately, using a standard singular value based scheme, it is shown that the subspace tracking, least squares and total least squares problems all yield smooth solutions. Specific tracking bounds are given, which show that any smooth system which realizes the input/output relation with small error must be close to the least squares solution. This indicates that if smoothness is desired, the least squares estimate is a reasonable choice. The underlying matrix problem has Toeplitz structure which can be exploited in the algorithmic implementation.

1. Introduction

There are many applications of system identification in the fields of adaptive control and signal processing. When on-line estimates of changing system parameters are desired, it is usually assumed that these parameters change slowly enough for the identification algorithm to track them with a small, bounded error. This paper considers significance of smoothness from a somewhat broader perspective. Three issues are involved. The first is the ability of the identification algorithm to track a smooth model when the order is estimated appropriately. The second is to find circumstances under which the estimates themselves will be smooth. The third is to find circumstances under which a smooth model of a given order can be fit to a set of input/output data.

Before the identification algorithm can even be applied, it is necessary to have an estimate of the system order. However, a unique order for a time-varying system is not a well defined concept. In fact, any pair of input and output sequences, u_t and y_t can be realized by a zero order time-varying system

$$y_t = \frac{y_t}{u_t} u_t = \theta_0(t) u_t,$$

and if $|\theta_0(t) - \theta_0(t-1)| < \Delta$ for some small Δ, then this would seem to be a reasonable model. Even data taken from a higher order time-invariant model will have such a zero order time-varying representation. It seems that it is necessary to choose the order appropriately to fit a smooth model to the data. This paper suggests that choosing the order to give a smooth model is a reasonable approach to order estimation and shows that this can be achieved for models derived from least squares problems by using a standard singular value based order estimation scheme.

Given the scalar input and output sequences u_t and y_t we define the vector

$$\phi_t^T = [\ y_t \quad u_t \quad \cdots \quad y_{t-n+1} \quad u_{t-n+1}\ ],$$

where n is the presumed system order, and the block Toeplitz matrix

$$A_t = W_t \begin{bmatrix} \phi_0^T \\ \phi_1^T \\ \vdots \\ \phi_t^T \end{bmatrix}$$

where

$$W_t = \text{diag}(\lambda^t, \lambda^{t-1}, \ldots, 1)$$

for $0 < \lambda < 1$ is a weighting matrix. A_t has the structure of a block Toeplitz matrix multiplied by the diagonal weighting matrix W_t. Such a matrix allows fast updating of the least squares solution as rows are added. We also define

$$\hat{\phi}_t^T = \begin{bmatrix} \phi_t^T & y_{t+d} & y_{t+d-1} & \cdots & y_{t+1} \end{bmatrix}$$

and

$$B_t = W_t \begin{bmatrix} y_d & y_{d-1} & \cdots & y_1 \\ y_{d+1} & y_d & \cdots & y_2 \\ \vdots & \vdots & & \vdots \\ y_{d+t} & y_{d+t-1} & \cdots & y_{1+t} \end{bmatrix}.$$

Generally for system identification $d = 1$, but for much of this paper, the more general case will be considered.

In a least squares scheme we are interested in finding X_t to minimize

$$\left\| \begin{bmatrix} A_t & B_t \end{bmatrix} \begin{bmatrix} X_t \\ -I_d \end{bmatrix} \right\|_F^2.$$

This is a standard approach for system identification and can be easily implemented recursively. Because of the block Toeplitz structure of the data matrix A_t, the problem of finding X_t recursively from X_{t-1} can be done quickly, in $O(n)$ time. The algorithms to do this take several forms. Examples are the fast transversal filters algorithm, [1], the RLS lattice algorithm, [4], and the QR-based fast least squares algorithm, [5]. Some of these algorithms do not explicitly compute the least squares solution, but instead compute residuals. This sometimes limits the range of applicability of the algorithm to adaptive filtering rather than system identification, where an explicit solution is needed.

For total least squares we wish to find X_t so that

$$\begin{bmatrix} A_t + \hat{A}_t & B_t + \hat{B}_t \end{bmatrix} \begin{bmatrix} X_t \\ -I_d \end{bmatrix} = 0$$

where $\hat{A}_t$ and $\hat{B}_t$ are chosen to satisfy

$$\min_{\text{range}(B_t+\hat{B}_t)\subseteq\text{range}(A_t+\hat{A}_t)} \left\| \begin{bmatrix} \hat{A}_t & \hat{B}_t \end{bmatrix} \right\|_F^2.$$

Both the LS and TLS approaches give explicit estimates of the system. In the case $d = 1$, they can be viewed as an attempt to estimate a vector v_t for which

$$y_t = \phi_{t-1}^T v_t.$$

The subspace tracking problem involves finding a basis for the right singular vectors associated with the d smallest singular values, $\sigma_{2n+1}, \ldots, \sigma_{2n+d}$. In the case $d = 1$, the $(2n + 1)$ right singular vector provides an estimate of an implicit form of the system with norm one. An implicit form of the system is a vector v_t for which

$$\hat{\phi}_t^T v_t = 0.$$

The total least squares and subspace tracking approaches are not as commonly used in system identification and cannot be implemented as conveniently in recursive form as the least squares approach. Their presence in this paper is primarily to provide intermediate results which can be used to derive smoothness properties for least squares. The order estimation scheme considered here uses singular values of the matrix $\begin{bmatrix} A_t & B_t \end{bmatrix}$. The subspace tracking and total least squares problems are characterized in terms of the singular vectors of the data matrix. This makes it easy to derive smoothness results for these two schemes using constraints imposed through order estimation. The results are then extended to the least squares approach.

We make the following assumptions about the singular values of various matrices

1. $\sigma_{2n}(A_i) \geq \underline{\rho}$ and $\sigma_d(B_i) \geq \underline{\rho}$ for $i = t, t-1$.
2. $\sigma_1(A_i) \leq \overline{\rho}$ and $\sigma_1(B_i) \leq \overline{\rho}$ for $i = t, t-1$.
3. $\sigma_{2n+1}(\begin{bmatrix} A_i & B_i \end{bmatrix}) \leq \nu$ for $i = t, t-1$.
4. $\sigma_{2n}(\begin{bmatrix} A_i & B_i \end{bmatrix}) \geq \sigma$ for $i = t, t-1$.

The first two assumptions are persistency of excitation conditions and are equivalent to assumptions on the conditioning of the least squares problem. These conditions are common in the system identification literature. Since the squared singular values of A_i correspond to the eigenvalues of $A_i^T A_i$, the inequalities applied to A_i are equivalent to

$$\underline{\rho}^2 I \leq \sum_{j=0}^{i} \lambda^{i-j} \phi_j \phi_j^T \leq \overline{\rho}^2 I.$$

The existence of such $\underline{\rho}$ and $\overline{\rho}$ is implied if there exists s, α and β such that

$$\alpha I \leq \sum_{j=t}^{i+s} \phi_j \phi_j^T \leq \beta I$$

for all t. This a standard persistence of excitation condition, and the proof that this implies the exponentially weighted condition can be found in [3]. Similar conditions can be stated for B_i.

The third and fourth assumptions are the ones which determine the fitness of the order of the model. Taking $d = 1$, if ν/σ, the noise-to-signal ratio in a subspace tracking context, is small, then the data can be closely approximated by a time-invariant system of order n over the time period of interest (defined by the exponential window). This approach without exponential weighting is standard for estimating the order of a time invariant system. In the exponentially windowed case, if the same sort of gap holds for all time, then it seems reasonable to assume that the data can be well approximated by a slowly time varying system. This is in fact the case, and such systems can be generated by standard least squares system identification schemes.

The assumptions on the singular values will be used to prove smoothness for the identification approaches discussed previously, as well as bounds on the errors each algorithm achieves in reconstructing the data. In a practical scheme, the bounds on the singular values, and hence the smoothness bounds, would hold only for sufficiently large t. The derivations will make no mention of this because to derive a maximum change in going from time $t-1$ to time t, it will only be necessary to assume that the singular value bounds hold at $t-1$ and t.

The results proceed in stages with each depending on the previous one. The first is for subspace tracking. The result for that algorithm is used to derive smoothness for total least squares. Finally, the total least squares result is used to derive a bound for the least squares solution.

2. Smoothness for Subspace Tracking

We wish to find smoothness bounds for the right singular subspace of $\begin{bmatrix} A_t & B_t \end{bmatrix}$ associated with the singular values $\sigma_{2n+1}, \ldots, \sigma_{2n+d}$. Let the SVD of $\begin{bmatrix} A_t & B_t \end{bmatrix}$ be

$$\begin{bmatrix} U_1(t) & U_2(t) \end{bmatrix} \begin{bmatrix} \Sigma_1(t) & 0 \\ 0 & \Sigma_2(t) \end{bmatrix} \begin{bmatrix} V_1^T(t) \\ V_2^T(t) \end{bmatrix}$$

where $\Sigma_2(t)$ is $d \times d$ and $\Sigma_1(t)$ is $2n \times 2n$.

We wish to find a bound on the distance between the small singular subspaces, and hence the projectors, from one time step to the next:

$$\mathrm{d}(t) = \|V_2(t)V_2^T(t) - V_2(t-1)V_2^T(t-1)\|_2.$$

The matrix norm used here is the one induced by the Euclidean norm.

The process starts off by showing that $\begin{bmatrix} A_{t-1} & B_{t-1} \end{bmatrix} V_2(t)$ is small. The bound is

$$\left\| \begin{bmatrix} A_{t-1} & B_{t-1} \end{bmatrix} V_2(t) \right\|_2 \leq \frac{\nu}{\lambda}. \qquad (1)$$

The next step is to use (1) to bound the component of the columns of $V_2(t)$ which is in range($V_1(t-1)$). To do this, let

$$V_2(t) = C + D$$

where $V_2^T(t-1)D = 0$. Every column of C is in range($V_2(t-1)$) and every column of D is in its orthogonal complement. This implies that $C^T D = 0$ and $I = C^T C + D^T D$. The inequality (1) can be used to show

$$d(t) = \|D\|_2 \leq \frac{\nu}{\lambda\sigma}. \qquad (2)$$

The result shows that when the gap between the singular values is large, the right singular subspace associated with the small singular values changes slowly. Here neither of the persistence of excitation conditions were used. Since they are equivalent to assumptions about the conditioning of the least squares problem, they will be useful in translating the bound in (2) to a bound for the least squares and total least squares problems.

3. Smoothness for Total Least Squares

For the case of $d = 1$, the right singular vector $v_2(t)$ associated with $\sigma_{2n+1}(t)$ provides an estimate of an implicit system giving y_t from u_t, in the sense that

$$\hat{\phi}_t^T v_2(t)$$

is small. One approach to finding an explicit model is to divide the first $2n$ components of $v_2(t)$ by the last component. For the more general case when $d \neq 1$, if the matrix of right singular vectors is further partitioned as

$$\begin{bmatrix} V_1(t) & V_2(t) \end{bmatrix} = \begin{bmatrix} V_{11}(t) & V_{12}(t) \\ V_{21}(t) & V_{22}(t) \end{bmatrix}$$

where $V_{22}(t)$ is $d \times d$. The total least squares solution is given by $-V_{12}(t)V_{22}^{-1}(t)$. To insure invertibility of $V_{22}(t)$ we assume that $\underline{\rho} > \nu$. [2]

In fact, more than invertibility of $V_{22}(t)$ will be required. Since can be shown that

$$d_{TLS}(t) \leq \|V_{22}^{-1}\|_2^2(1 + \|V_{22}^{-1}\|_2^2) \cdot \\ \|V_2(t)V_2^T(t) - V_2(t-1)V_2^T(t-1)\|_2 \;,$$

we derive a bound on $\|V_{22}^{-1}(t)\|$

$$\|V_{22}^{-1}(t)\|_2 \leq \frac{2\bar{\rho}^2}{\underline{\rho}^2 - \nu^2}.$$

For simplicity the bound on $\|V_{22}^{-1}(t)\|$ will be represented by a single constant

$$\|V_{22}^{-1}(t)\|_2 \leq K,$$

giving a final form to the bound of

$$d_{TLS} \leq K^2(1+K)\frac{\nu}{\lambda\sigma}. \qquad (3)$$

There are no projections involved in the expression. Unlike the subspace tracking problem, this result does not show a bound for the subspace spanned by the solution, but for the solution itself. It is the first result in this paper which applies to an explicit solution. The second is the least squares solution and it will be shown to be smooth through the use of (3).

4. Smoothness for Least Squares

The least squares solution is given by choosing X_t to minimize

$$\min_{X_t} \|A_t X_t - B_t\|_F^2.$$

The approach to proving its smoothness is to show that it is close to the total least squares solution. If ν is small then it follows that the set of equations $A_t X_t = B_t$ is nearly consistent, in the sense that,

$$\begin{aligned} \|A_t X_t - B_t\|_2 &\leq \left\| \begin{bmatrix} A_t & B_t \end{bmatrix} V_2(t)(V_{22}^{-1}(t)) \right\| \\ &\leq K\nu. \end{aligned}$$

If the equations have an exact solution, then the total least squares and least squares solutions will both equal the exact solution. It seems plausible that if it is possible to solve the equations with small residual, then the two solutions will be close. This is in fact the approach that is used here. For additional discussion of closeness of least squares and total least squares residuals, see [6]. As in the proof of smoothness for the small right singular subspace, we decompose the solution into its components in range($V_1(t)$) and range($V_2(t)$),

$$\begin{bmatrix} X_t \\ -I \end{bmatrix} = V_2(t)C + V_1(t)D.$$

It can be shown that

$$\|D\|_2 \leq \frac{K\nu}{\sigma}.$$

Using this fact we find that,

$$\left\|X_t + V_{12}(t)(V_{22}^{-1}(t))\right\|_2 \leq \frac{(K+1)K\nu}{\sigma}$$

Putting together the bound for the difference between least squares and total least squares and the smoothness result for total least squares gives

$$\|X_t - X_{t-1}\|_2 \leq K(1+K)\left(2 + \frac{K}{\lambda}\right)\frac{\nu}{\sigma} \qquad (4)$$

Since the bound has ν/σ as a factor, it goes to zero as ν becomes small compared to σ. In the case $\nu = 0$, the data can be modeled by a time-invariant system, and this result shows that the identification algorithm will produce a time-invariant solution.

5. Tracking Error for the Implicit Solution

In this section we are concerned with the special case $d = 1$. Given a sequence of vectors $\hat{v}_t$ for which

$$\hat{\phi}_t^T \hat{v}(t) = 0 \qquad (5)$$

for all t, where in this case,

$$\hat{\phi}_t^T = [\ y(t) \quad \phi_{t-1}\ ],$$

We are interested in determining how good an estimate $v_2(t)$ is of $\hat{v}_t$. We make a smoothness assumption about $\hat{v}_t$, requiring that

$$||\hat{v}_t\hat{v}_t^T - \hat{v}_{t-1}\hat{v}_{t-1}^T|| \leq \Delta,$$

for all t. We also require that $||\hat{v}_t|| = 1$ for all t, so that the matrices in the smoothness assumption are projections.

Using arguments similar to those in preceeding sections (bounding an component to show that one vector is almost in the range of another), it can be shown that that

$$||\hat{v}_t\hat{v}_t^T - v_2(t)v_2^T(t)||^2 = ||d||^2 \leq \frac{2\Delta\overline{\rho}^2\lambda^2(1+\lambda^2)}{\sigma^2(1-\lambda^2)^3}$$

It follows from this that *any smooth, implicitly defined system which realizes the input/output relation (5), must be close to the singular vector corresponding to* $\sigma_{2n+1}\left([\ A_t \quad b_t\ ]\right)$.

6. Tracking Error for the Explicit Least Squares Solution

This section shows that any model $\bar{\theta}_t$ with output $\bar{y}_t$ which achieves the input/output relation with a bounded error

$$|\bar{y}_t - y_t| \leq K_1,$$

where $\bar{y}_t$ is defined by the recursion

$$\bar{y}_t = \bar{\phi}_{t-1}^T\bar{\theta}_t.$$

and

$$\bar{\phi}_t^T = [\ \bar{y}_t \quad u_t \cdots \quad \bar{y}_{t-n+1} \quad u_{t-n+1}\ ].$$

must be close to the least squares estimate. Here θ_t is taken to be the regular least squares estimate for the data y_t and u_t. Assume that $\bar{\theta}_t$ is bounded

$$||\bar{\theta}_t|| \leq K_2$$

and that it is smooth

$$||\bar{\theta}_t - \bar{\theta}_{t-1}|| \leq \Delta.$$

Using the triangle inequality results in a tracking error bound of,

$$||\theta_t - \bar{\theta}_t|| \leq \frac{\overline{\rho}^2\Delta\lambda}{\underline{\rho}^2(1-\lambda)^2} + \frac{K_1(nK_2+1)\overline{\rho}}{(1-\lambda)\underline{\rho}^2}.$$

Clearly, if Δ and K_1 are both small, then this error will be small. From this it follows that *any smooth system with bounded parameters which realizes the input/output data with bounded (small) error must be close to the least squares solution.*

It should be noted that unlike bounds in previous sections, the bounds given here require that $||A_i||_2 < \overline{\rho}$ for all $0 \leq i \leq t$. Previous results only required assumptions on the singular values to hold at times t and $t-1$.

7. Concluding Remarks

The results given in this paper show that looking for a gap in the singular values is a reasonable approach to the order estimation problem for smoothly time varying systems. This is a fact which is accepted implicitly in many recursive identification schemes, and this paper gives a quantitative justification by proving rigorous bounds.

References

[1] J. M. Cioffi and T. Kailath, "Fast, Recursive-Least-Squares Transversal Filters for Adaptive Filtering," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 32, pp. 304-337.

[2] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Baltimore: Johns Hopkins Press, 1989.

[3] R. M. Johnstone, C. R. Johnson, R. R. Bitmead and B. D. O. Anderson, "Exponential Convergence of Recursive Least Squares with Exponential Forgetting Factor," *Systems and Control Letters*, vol. 2, pp. 77-82, August 1982.

[4] D. T. L. Lee, M. Morf and B. Friedlander, "Recursive Least Squares Ladder Estimation Algorithms," *IEEE Transactions on Circuits and Systems*, vol. 28, pp. 467-481.

[5] P. A. Regalia and M. G. Bellanger, "On the Duality Between Fast QR Methods and Lattice Methods in Least Squares Adaptive Filtering," *IEEE Transactions on Signal Processing,* vol. 39, pp. 879-891.

[6] S. Van Huffel and J. Vandewalle, "The Total Least Squares Problem: Computational Aspects and Analysis," Frontiers in Applied Mathematics Series, SIAM, Philadelphia, 1991.