Convergence rates for eigenstructure identification using subspace methods [1]

Albert Benveniste
IRISA, Campus de Beaulieu,
35042 Rennes cedex, France,
benveniste@irisa.fr

Michèle Basseville
IRISA, Campus de Beaulieu,
35042 Rennes cedex, France,
basseville@irisa.fr

Laurent Mével
IRISA, Campus de Beaulieu,
35042 Rennes cedex, France,
laurent.mevel@irisa.fr

Abstract

We apply the general results of companion paper [1] on the relationship between identification and local tests, to the estimation of convergence rates for MIMO system eigenstructure identification using subspace algorithms. We provide a new and practical estimator for such convergence rates.

1 Introduction

This paper is a follow-up of [1]. In the latter reference we examine relationships between the consistency of so-called M-estimators, the local asymptotic normality of statistics for change detection, and the asymptotic normality of estimators. And we use this relationship for proposing an alternative approach to estimate the convergence rate of identification procedures for general M-estimators.

In this paper we apply this approach to the interesting case of subspace methods for eigenstructure identification. We first briefly review the results of [1]. Then we state the problem of eigenstructure identification. In section 2 we discuss system theoretic issues related to eigenstructure identification. Then in section 3 we derive our main results by exploiting Central Limit Theorems for local test statistics in order to derive estimators for the convergence rates of subspace algorithms for eigenstructure identification.

1.1 On the relationship between identification and testing

Here we briefly review the results of [1], and cast them under the notations of the present paper. We consider an observation process $(Y_1, Y_2, ...)$ and try to estimate parameter θ of its distribution P_θ as a solution $\widehat{\theta}_n$ of $\sum_{i=1}^n H(\theta, Y_i) = 0$:

$$\widehat{\theta}_n \quad = \quad \arg_\theta \quad : \quad \sum_{i=1}^n H(\theta, Y_i) = 0 \ . \tag{1}$$

Of course this is based on our knowledge that the true parameter $\theta_\star$ satisfies [1]:

$$E_{\theta_\star}[H(\theta_\star, Y_i)] = 0. \tag{2}$$

The obtained estimator is called an M-estimator [9]. This estimator corresponds sometimes to a *quasi-likelihood* estimate $\widehat{\theta}_n = \arg\min_\theta \sum_{i=1}^n G(\theta, Y_i)$, where G is the quasi-likelihood [2]. However, in some cases equations like (1) are solved and H is not the gradient of any function G. An interesting case is that of the *subspace methods,* which we investigate in detail in this paper.

In [1] we study the distribution under $P_{\theta+\widetilde{\theta}/\sqrt{n}}$ of the following statistics, which is directly related to M-estimator (1):

$$U_n \quad = \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n H(\theta, Y_i) \tag{3}$$

We show that, for n large,

$$\begin{aligned} \text{under } P_\theta \quad &: \quad U_n \sim \mathcal{N}(\,0\,,\,\Sigma(\theta)\,) &(4)\\ \text{under } P_{\theta+\widetilde{\theta}/\sqrt{n}} \quad &: \quad U_n \sim \mathcal{N}\left(\,-h'(\theta)\,\widetilde{\theta}\,,\,\Sigma(\theta)\right) &(5)\end{aligned}$$

where

$$\begin{aligned} h'(\theta) \quad &= \quad +E_{\theta_\star}\left(\nabla_\theta H(\theta, Y)\big|_{\,\theta\,=\,\theta_\star}\right) \\ &= \quad -\nabla_{\theta_\star}\left(E_{\theta_\star}[H(\theta, Y)]\right)\big|_{\,\theta_\star\,=\,\theta} \end{aligned} \tag{6}$$

where the two different, but equivalent, expressions for $h'(\theta)$ result from the identity

$$E_{\theta_\star}[H(\theta, Y_i)] = 0 \quad \text{iff} \quad \theta = \theta_\star\,.$$

The important point in (4,5) is that covariance matrix $\Sigma(\theta)$ is the same under both distributions.

Next, assuming that the (Y_i) are drawn from distribution $P_{\theta_\star}$ (i.e., $\theta_\star$ is the "true" underlying system parameter), then, for $\theta - \theta_\star$ small, one has

$$U_n(\theta) \tag{7}$$

[1] This work was supported by Eureka project no 1562 Sinopsys (Model based **S**tructural monitoring using **in-op**eration **sys**tem identification) coordinated by Lms, Leuven, Belgium.

[1] Notation E_θ denotes expectation under distribution P_θ.
[2] For $f(x, y)$ a function, we shall denote by $\nabla_x f$ the partial derivative of f w.r.t. x, when it exists.

$$\approx \quad U_n(\theta_\star) + \frac{1}{n}\left(\sum_1^n \nabla_\theta H(\theta_\star, Y_i)\right)\ (\theta - \theta_\star)\sqrt{n}$$

Since $\widehat{\theta}_n$ is characterized by the equation $U_n(\widehat{\theta}_n) = 0$, we get, using (7) and the law of large numbers,

$$\sqrt{n}(\widehat{\theta}_n - \theta_\star) \quad \approx \quad -h'(\theta_\star)^{-1}\ U_n(\theta_\star) \tag{8}$$

and we deduce immediately the Central Limit Theorem for $\widehat{\theta}_n$ from the asymptotic behavior of the statistics U_n.

Besides its interest for its own, this relationship between $\widehat{\theta}_n$ and $U_n(\theta_\star)$ has a practical interest. The Central Limit Theorem for M-estimators states (see [1]) :

$$\sqrt{n}(\widehat{\theta}_n - \theta_\star) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, h'(\theta_\star)^{-T}\Sigma(\theta_\star)h'(\theta_\star)^{-1}\right) \tag{9}$$

where $h'(\theta)$ is defined in (6) and

$$\Sigma \quad = \quad R(0) + \sum_{i>0} R(i) + R(i)^T \tag{10}$$

$$R(i) \quad = \quad \mathrm{Cov}(H(\theta_\star, Y_1), H(\theta_\star, Y_{i+1}))\ , \tag{11}$$

While estimating or calculating the sensitivity matrix $h'(\theta)$ is most of the time easy, estimating or calculating matrix Σ often leads to illconditioned estimates. In contrast formula (8) suggests that it is enough to estimate the covariance matrix of $U_n(\theta_\star)$, which can be safely achieved from drawing samples of this statistics.

In fact, a good estimate for Σ is obtained by taking an empirical estimate for the covariance matrix of $U_n(\theta')$, for θ' close enough to the true system $\theta_\star$. In particular, one can bootstrap this estimate by taking for θ' the actual estimate $\widehat{\theta}$ for $\theta_\star$. One straightforward implementation of this idea is as follows. Given estimate $\widehat{\theta}$ and a validation data sample $Y_1, \ldots, Y_N$ where N is large enough,

1. divide $Y_1, \ldots, Y_N$ into K blocks of length n
2. construct $U_n^k(\widehat{\theta})$ for the kth block (12)
3. estimate empirically the covariance of $(U_n^k(\widehat{\theta}))_{k=1,\ldots,K}$

In this paper, we apply the above general method to the interesting case of *subspace methods for output-only eigenstructure identification.*

1.2 The problem of eigenstructure identification

We consider linear multi-variable systems described by a discrete-time state space model :

$$\left\{ \begin{array}{rcl} X_{k+1} & = & FX_k + \varepsilon_k \\ Y_k & = & HX_k + \nu_k \end{array} \right. \tag{13}$$

where state X and observed output Y have dimensions m and r respectively. The state noise process $(\varepsilon_k)_k$ is an unmeasured Gaussian white noise sequence with zero mean. We assume noise ε_k to be stationary, that is of constant covariance matrix. The measurement noise process $(\nu_k)_k$ is assumed to be an unmeasured MA(ι) Gaussian sequence with zero mean. In the sequel, we use the notational convention that $\iota = -1$ for no measurement noise, and $\iota = 0$ for white (i.i.d.) measurement noise. Note that, with this MA assumption for its structure, measurement noise does *not* affect[3] the eigenstructure of system (13).

Let $G \stackrel{\text{def}}{=} \mathbf{E}\left(X_k Y_k^T\right)$ be the cross-correlation between state X_k and observation Y_k, and let :

$$\mathcal{O}_p \quad = \quad \begin{pmatrix} H \\ HF \\ \vdots \\ HF^{p-1} \end{pmatrix} \quad \text{and} \tag{14}$$

$$\mathcal{C}_p \quad = \quad \begin{pmatrix} G & FG & \ldots & F^{p-1}G \end{pmatrix}$$

be the pth-order observability matrix of system (13) and controllability matrix of pair (F, G), respectively. We assume that, for p large enough, both observability and controllability matrices have full rank m. The problem we consider is

> to identify the observed system eigenstructure, that is the collection of m pairs (λ, ϕ_λ), where λ ranges over the set of eigenvalues of state transition matrix F, $\phi_\lambda = H\varphi_\lambda$ and φ_λ is the corresponding eigenvector.

In all what follows, we assume that the system has no multiple eigenvalues, and thus that the λ's and φ_λ's are pairwise complex conjugate. In particular, 0 is *not* an eigenvalue of state transition matrix F. The collection of (λ, ϕ_λ) provides us with a canonical parameterization of the pole part of system (13). In the sequel, referring to vibration monitoring [5], such a pair (λ, ϕ_λ) is called a mode. The set of the m modes is considered as the system parameter θ :

$$\theta \stackrel{\text{def}}{=} \begin{pmatrix} \Lambda \\ \mathrm{vec}\,\Phi \end{pmatrix} \tag{15}$$

In (15), Λ is the vector whose elements are the eigenvalues λ, Φ is the matrix whose columns are the ϕ_λ's, and vec is the column stacking operator. Parameter θ has size $(r+1)m$. The problem is to identify parameter vector θ.

Note that we take the important assumption that ***exact model order m is known.*** Clearly this is not a

[3] The same would not hold true with an AR assumption for the measurement noise structure.

realistic assumption in practice. But this is an acceptable assumption for the theoretical task of providing convergence rates for subspace algorithms for MIMO system eigenstructure identification.

2 System theoretic issues and eigenstructure identification

In this section, we present a variant of subspace algorithms for output only, eigenstructure identification for MIMO systems. For general references on subspace algorithms, the reader is referred to [10, 11, 12, 13]. The variant we consider here is not using raw data as input, but rather covariance matrices from data, but it has been recognized that this is a minor modification. Also, compared with [10], we do *not* consider weighting matrices. The reason is that is has been shown in [5] that weighting matrices play no role when exact model order is known, which we assume here.

The arguments presented here are well known [12], we just state them here for the sake of clarity and completeness. We are given a sequence of covariances : $R_j \stackrel{\text{def}}{=} \mathbf{E}\left(Y_{k+j}\, Y_k^T\right)$ of output Y_k of a state space model (13). For $q \geq p+1$, let $\mathcal{H}_{p+1,q}$ be the block-Hankel matrix :

$$\mathcal{H}_{p+1,q} = \begin{pmatrix} R_{\iota+1} & R_{\iota+2} & \dots & R_{\iota+q} \\ R_{\iota+2} & R_{\iota+3} & \dots & R_{\iota+q+1} \\ \vdots & \vdots & \ddots & \vdots \\ R_{\iota+p+1} & \dots & \dots & R_{\iota+p+q} \end{pmatrix} \tag{16}$$

As mentioned above, integer ι reflects the assumed correlation in the measurement noise sequence $(\nu_k)_k$. It should be considered as a design parameter for the proposed algorithms.

Choosing the eigenvectors of F as a basis[4] for the state space of model (13) yields the following particular representation of the observability matrix introduced in (14) :

$$\mathcal{O}_{p+1}(\theta) = \begin{pmatrix} \Phi \\ \Phi\Delta \\ \vdots \\ \Phi\Delta^p \end{pmatrix} \tag{17}$$

where diagonal matrix Δ is defined as $\Delta = \text{diag}(\Lambda)$, and Λ and Φ are as in (15). For any other state basis, the observability matrix $\mathcal{O}_{p+1}$ can be written as :

$$\mathcal{O}_{p+1} = \mathcal{O}_{p+1}(\theta)\, T \tag{18}$$

for a suitable $m \times m$ invertible matrix T. Because of the definition of $\mathcal{H}_{p+1,q}$, $\mathcal{O}_p$ and $\mathcal{C}_q$ in (16) and (14), a direct computation of the R_j's from the model equations : $R_{\iota+j+1} = H\, F^{\iota+j+1}\, G$, for $j \geq 0$, leads to the following well known factorization property :

$$\mathcal{H}_{p+1,q} = \mathcal{O}_{p+1}\, \left(F^{\iota+1}\, \mathcal{C}_q\right) \tag{19}$$

From (19), (17) and (18), we get that the following property characterizes whether some parameter $\theta_\star$ agrees with a given output covariance sequence $(R_j)_j$ [12] :

$$\begin{array}{l} \mathcal{O}_{p+1}(\theta_\star) \text{ and } \mathcal{H}_{p+1,q} \\ \text{have the same left kernel space,} \end{array} \tag{20}$$

where we recall that the left kernel space of matrix M is the kernel space of matrix M^T. Property (20) can be checked as follows :

1. From θ as in (15), form $\mathcal{O}_{p+1}(\theta)$.
2. Pick an orthonormal basis of the left kernel space of matrix $\mathcal{O}_{p+1}(\theta)$, in terms of the columns of some matrix S of co-rank m such that :

$$S^T\, S = I_s \tag{21}$$
$$S^T\, \mathcal{O}_{p+1}(\theta) = 0 \tag{22}$$

   Matrix S has dimensions $(p+1)r \times s$, where $s = (p+1)r - m$, and it is not unique; it can be obtained, for example, by the SVD-factorization of $\mathcal{O}_{p+1}(\theta)$. Even though S is not unique, in the sequel we regard it sometimes as a function of parameter θ and denote it by $S(\theta)$.
3. Then, (20) rephrazes as follows :

$$S^T(\theta_\star)\, \mathcal{H}_{p+1,q} = 0\,, \tag{23}$$

   compare with (2).

When only sample data are available, not the exact covariance matrices R_j, the idea is to replace, in the step 3 above, the exact Hankel matrix $\mathcal{H}_{p+1,q}$ by its estimate $\widehat{\mathcal{H}}_{p+1,q}$ based on sample covariances from data. Therefore we get the M–estimator

$$\widehat{\theta} \quad = \quad \arg_\theta \left[S^T(\theta)\, \widehat{\mathcal{H}}_{p+1,q} = 0\right] \tag{24}$$

compare with (1).

In (24), the empirical Hankel matrix $\widehat{\mathcal{H}}_{p+1,q}$ has not rank m but generically full rank, therefore equation $S^T(\theta)\, \widehat{\mathcal{H}}_{p+1,q} = 0$ is to be understood as follows : $S(\theta)$ shall be orthogonal to the subspace generated by the m principal left singular vectors of $\widehat{\mathcal{H}}_{p+1,q}$. Therefore the map

$$\widehat{\mathcal{H}}_{p+1,q} \longmapsto \widehat{S} \longmapsto \widehat{\mathcal{O}}_{p+1} \longmapsto \widehat{\theta}$$

is implemented via

$$\widehat{\mathcal{H}}_{p+1,q} \longmapsto (\,\widehat{\mathcal{O}}_{p+1}\,,\,\widehat{\mathcal{C}}_q\,) \longmapsto \widehat{\theta}$$

[4] This is called the modal basis in the vibration monitoring application [3].

where $\widehat{\mathcal{H}}_{p+1,q} = \widehat{\mathcal{O}}_{p+1}\,\widehat{\mathcal{C}}_q^{\iota\,T}$ ($\mathcal{C}_q^\iota = F^{\iota+1}\,\mathcal{C}_q$), is obtained via the truncated SVD of $\widehat{\mathcal{H}}_{p+1,q}$. And this is (one version of) the wellknown subspace algorithm for eigenstructure identification.

3 Deriving convergence rates for eigenstructure identification from subspace tests

Here we show how the general procedure of the introduction applies to this case. This is based on (1,2,3) on the one hand, and (23,24) on the other hand.

3.1 Subspace-based tests

We first explain how subspace-based tests are designed. Assume we have at hand a nominal model θ, and newly collected data $Y_1, \ldots, Y_n$. Compute the empirical covariance sequence : $\widehat{R}_j \stackrel{\text{def}}{=} 1/(n-j)\ \sum_{k=1}^{n-j} Y_{k+j}\ Y_k^T$. Then, perform steps 1 and 2 of section 2, and replace step 3 by :

3. Define the residual vector :

$$U_n(\theta) \stackrel{\text{def}}{=} \sqrt{n}\ \text{vec}\left(\ S^T(\theta)\ \widehat{\mathcal{H}}_{p+1,q}\ \right) \qquad (25)$$

   where $\widehat{\mathcal{H}}_{p+1,q}$ is the empirical block-Hankel matrix obtained by substituting $\widehat{R}_j$ for R_j in (16).

3.2 Estimating the convergence rate

We apply the general method described in the introduction. Referring to (9) we need to estimate the pair (h', Σ). Covariance Σ is estimated using the procedure (12). Therefore we only need to estimate h', this is detailed next.

IEstimation of h'. n this section, we investigate the estimation and the rank of the Jacobian matrix $h'(\theta)$ of a residual of the form (25).

From (6), second expression for $h'(\theta)$, and denoting by $\theta_\star$ the true system parameter corresponding to Hankel matrix $\mathcal{H}$, we get :

$$h'(\theta) \quad = \quad \nabla_{\theta_\star}\ \text{vec}\left(S^T(\theta)\ \mathcal{H}_{p+1,q}\right)\big|_{\theta=\theta_\star}\ ,$$

which also writes :

$$h'(\theta) \quad = \quad \nabla_{\theta_\star}\ \text{vec}\left(S^T(\theta)\ \mathcal{O}_{p+1}(\theta_\star)\ F^{\iota+1}\ \mathcal{C}_q\right)\big|_{\theta=\theta_\star}$$

thanks to factorization (19) of the Hankel matrix. Hence, denoting by $\mathcal{O}'_{p+1}(\theta)$ the derivative of $\mathcal{O}_{p+1}(\theta)$ wrt. θ, we get :

$$\begin{aligned} & h'(\theta) \qquad (26) \\ = \ & \left(\mathcal{C}_q^T\ F^{\iota+1\ T} \otimes I\ S^T(\theta)\right)\ \mathcal{O}'_{p+1}(\theta_\star) \\ = \ & S^T(\theta)\ \left(\mathcal{C}_q^T\ F^{\iota+1\ T} \otimes I_{(p+1)r}\right)\ \mathcal{O}'_{p+1}(\theta_\star) \\ = \ & S^T(\theta)\ \left(\mathcal{H}_{p+1,q}^T\ \mathcal{O}_{p+1}^{\dagger\,T}(\theta_\star) \otimes I_{(p+1)r}\right)\ \mathcal{O}'_{p+1}(\theta_\star) \end{aligned}$$

The last equality, where $\mathcal{O}_{p+1}^\dagger(\theta)$ is the pseudo-inverse of $\mathcal{O}_{p+1}(\theta)$, is obtained using factorization (19) again. A consistent estimate $\widehat{M}$, based on a data sample, is obtained by substituting $\widehat{\mathcal{H}}$ for $\mathcal{H}$ in (26). We refer the reader to [5] for an analytical expression for the derivative $\mathcal{O}'_{p+1}(\theta)$, as well as considerations about the rank of matrix $h'(\theta)$.

4 Conclusion

In this paper we have proposed an estimator for the convergence rate of output-only, MIMO system, eigenstructure identification, using subspace algorithms. We have assume full knowledge of model order. In addition, the estimate we propose is well conditioned, practical, and of reasonable cost. This was made possible thanks to general results on the relationahip between identification and local tests.

References

[1] A. Benveniste and B. Delyon (2000). Using local tests to estimate convergence rates for identification. Submitted to CDC'2000.

[2] Basseville, M. and I.V. Nikiforov (1993). *Detection of Abrupt Changes — Theory and Applications*. Prentice Hall, N.J. http://www.irisa.fr/sigma2/kniga.

[3] Basseville, M., A. Benveniste, G. Moustakides and A. Rougée (1987). Detection and diagnosis of changes in the eigenstructure of non-stationary multivariable systems. *Automatica*, **23**, 479–489.

[4] Basseville, M., A. Benveniste, B. Gach-Devauchelle, M. Goursat, D. Bonnecase, P. Dorey, M. Prevosto and M. Olagnon (1993). Damage monitoring in vibration mechanics : issues in diagnostics and predictive maintenance. *Mechanical Systems and Signal Processing*, **7**(5), 401–423.

[5] M. Basseville, M. Abdelghani, A. Benveniste (2000). Subspace-based fault detection algorithms for vibration monitoring. *Automatica,* vol.36(1), 101–109.

[6] Benveniste, A., M. Basseville and G. Moustakides (1987). The asymptotic local approach to change detection and model validation. *IEEE Trans. Autom. Control*, **AC-32**(7), 583–592.

[7] Benveniste, A., M. Métivier and P. Priouret (1990). *Adaptive Algorithms and Stochastic Approximations*. Springer, New York.

[8] Delyon, B., A. Juditsky and A. Benveniste (1997). On the relationship between identification and local tests. *Research Report* IRISA no 1104. ftp://ftp.irisa.fr/techreports/1997/PI-1104.ps.gz.

[9] P. J. Huber (1981). *Robust statistics*, Wiley.

[10] Van Overschee, P. and B. De Moor (1996). *Subspace Identification for Linear Systems : Theory – Implementation – Methods.* Kluwer.

[11] Viberg, M. (1995). Subspace-based methods for the identification of linear time-invariant systems. *Automatica*, **31**(12), 1835–1853.

[12] Viberg, M., B. Wahlberg and B. Ottersten (1997). Analysis of state space system identification methods based on instrumental variables and subspace fitting. *Automatica*, **33**(9), 1603–1616.

[13] Gustafsson, T. (1996). A 4SID-based change detection algorithm. *IFAC'96 World Conf.*, San Francisco, CA.