PART 2

Stochastic Systems

Problem 2.1

On error of estimation and minimum of cost for wide band noise driven systems

Agamirza E. Bashirov
Department of Mathematics
Eastern Mediterranean University
Mersin 10
Turkey
agamirza.bashirov@emu.edu.tr

1 DESCRIPTION OF THE PROBLEM

The suggested open problem concerns the error of estimation and the minimum of the cost in the filtering and optimal control problems for a partially observable linear system corrupted by wide band noise processes.

Recent results allow to construct a wide band noise process in a certain integral form on the basis of its autocovariance function and design the optimal filter and the optimal control for a partially observable linear system corrupted by such wide band noise processes. Moreover, explicit formulae for the error of estimation and for the minimum of the cost have been obtained. But, the information about wide band noise contained in its autocovariance function is incomplete. Hence, every autocovariance function generates infinitely many wide band noise processes represented in the integral form. Consequently, the error of estimation and the minimum of the cost mentioned above are for a sample wide band noise process corresponding to the given autocovariance function.

The following problem arises: given an autocovariance function, what are the least upper and greatest lower bounds of the respective error of estimation and the respective minimum of the cost? What are the distributions of the error of estimation and the minimum of the cost? What are the parameters of the wide band noise process producing the average error and the average minimum of the cost?

2 MOTIVATION AND HISTORY OF THE PROBLEM

Modern stochastic optimal control and filtering theories use white noise driven systems. Results such as the separation principle and the Kalman-Bucy filtering are based on the white noise model. In fact, white noise, being a mathematical idealization, gives only an approximate description of real noise. In some fields the parameters of real noise are near to the parameters of white noise and, so, the mathematical methods of control and filtering for white noise driven systems can be satisfactorily applied to them. But in many fields white noise is a crude approximation to real noise. Consequently, the theoretical optimal controls and the theoretical optimal filters for white noise driven systems become not optimal and, indeed, might be quite far from being optimal. It becomes important to develop the control and estimation theories for the systems driven by noise models that describe real noise more adequately. Such a noise model is the wide band noise model.

The importance of wide band noise processes was mentioned by Fleming and Rishel [1]. An approach to wide band noise based on approximations by white noise was used in Kushner [2]. Another approach to wide band noise based on representation in a certain integral form was suggested in [3] and its applications to space engineering and gravimetry was discussed in [4, 5]. Filtering, smoothing, and prediction results for wide band noise driven linear systems are obtained in [3, 6]. The proofs in [3, 6] are given through the duality principle and, technically, they are routine, making further developments in the theory difficult. A more handle technique based on the reduction of a wide band noise driven system to a white noise driven system was developed in [7, 8, 9]. This technique allows to find the explicit formulae for the optimal filter and for the optimal control, as well as for the error of estimation and for the minimum of the cost in the filtering and optimal control problems for a wide band noise driven linear system. In particular the open problem described here was originally formulated in [9]. A complete discussion of the subject can be found in the recent book [10].

3 AVAILABLE RESULTS AND DISCUSSION

The random process φ with the property $\operatorname{cov}(\varphi(t+s), \varphi(t)) = \lambda(t,s)$ if $0 \le s < \varepsilon$ and $\operatorname{cov}(\varphi(t+s), \varphi(t)) = 0$ if $s \ge \varepsilon$, where $\varepsilon > 0$ is a small value and λ is a nonzero function, is called a *wide band noise process* and it is said to be stationary (in wide sense) if the function λ (called the *autocovatiance function* of φ) depends only on s (see Fleming and Rishel [8]).

Starting from the autocovariance function λ, one can construct the respective wide band noise process φ in the integral form

$$\varphi(t) = \int_{-\min(t,\varepsilon)}^{0} \phi(\theta) w(t+\theta)\, d\theta, \ \ t \ge 0, \tag{1}$$

where w is a white noise process with $\operatorname{cov}(w(t), w(s)) = \delta(t-s)$, δ is the Dirac's delta-function, $\varepsilon > 0$ and ϕ is a solution of the equation

$$\int_{-\varepsilon}^{-s} \phi(\theta)\phi(\theta+s)\,d\theta = \lambda(s),\ 0 \le s \le \varepsilon. \tag{2}$$

The solution φ of (2) is called a *relaxing function.* Since in (2) ϕ has only one variable the process φ from (1) is stationary in wide sense (except small time interval $[0, \varepsilon]$). The following theorem from [8, 9] is crucial for the proposed problem.

Theorem: *Let* $\varepsilon > 0$ *and let* λ *be a continuous real-valued function on* $[0, \varepsilon]$*. Define the function* λ_0 *as the even extension of* λ *to the real line vanishing outside of* $[-\varepsilon, \varepsilon]$ *and assume that* λ_0 *is a positive definite function with* $\mathcal{F}(\lambda_0)^{1/2} \in L_2(-\infty, \infty)$ *where* $\mathcal{F}(\lambda_0)$ *is the Fourier transformation of* λ_0*. Then there exists an infinite number of solutions of the equation* (2) *in the space* $L_2(-\varepsilon, 0)$ *if* λ *is a nonzero function a.e. on* $[-\varepsilon, 0]$*.*

The nonuniqueness of the solution of equation (2) demonstrates that the covariance function λ does not provide complete information about the respective wide band noise process φ. Therefore, for given λ, a sample solution ϕ of (2) generates the random process φ in the form (1) that could be considered as a less or more adequate model of real noise. In order to make a reasonable decision about the relaxing function, one of the ways is studying the distributions of the error of estimation and the minimum of the cost in filtering and control problems, finding the average error and the average minimum and identifying the relaxing function $\bar{\phi}$ producing these average values. For this, the explicit formulae from [7, 8, 9] (they are not given here because of the length) can be used to investigate the problem analytically or numerically. Also, the proof of the theorem from [8, 9] can be useful for construction different solutions of equation (2).

Finally, note that in a partially observable system both the state (signal) and the observations may be disturbed by wide band noise processes. Hence, the suggested problem concerns both these cases and their combination as well.

BIBLIOGRAPHY

[1] W. H. Fleming and R. W. Rishel, *Deterministic and Stochastic Optimal Control*, New York, Springer Verlag, 1975, p. 126.

[2] H. J. Kushner, *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, Boston, Birkhäuser, 1990.

[3] A. E. Bashirov, "On linear filtering under dependent wide band noises", *Stochastics*, 23, pp. 413-437, 1988.

[4] A. E. Bashirov, L. V. Eppelbaum and L. R. Mishne, “Improving Eötvös corrections by wide band noise Kalman filtering,” *Geophys. J. Int.*, 108, pp. 193-127, 1992.

[5] A. E. Bashirov, “Control and filtering for wide band noise driven linear systems”, *Journal on Guidance, Control and Dynamics*, 16, pp. 983-985, 1993.

[6] A. E. Bashirov, H. Etikan and N. Şemi, “Filtering, smoothing and prediction of wide band noise driven systems”, *J. Franklin Inst., Eng. Appl. Math.*, 334B, pp. 667-683, 1997.

[7] A. E. Bashirov, “On linear systems disturbed by wide band noise”, *Proceedings of the 14th International Conference on Mathematical Theory of Networks and Systems*, Perpignan, France, June 19-23, 7 p., 2000.

[8] A. E. Bashirov, “Control and filtering of linear systems driven by wide band noise,” *1st IFAC Symposium on Systems Structure and Control*, Prague, Czech Republic, August 29-31, 6 p.,2001.

[9] A. E. Bashirov and S. Uğural, “Analyzing wide band noise with application to control and filtering”, *IEEE Trans. Automatic Control*, 47, pp. 323-327, 2002.

[10] A. E. Bashirov, *Partially Observable Linear Systems Under Dependent Noises*, Basel, Birkhäuser, 2003.

Problem 2.2

On the stability of random matrices

Giuseppe Calafiore*, Fabrizio Dabbene**
*Dip. di Automatica e Informatica
**IEIIT-CNR
Politecnico di Torino
C.so Duca degli Abruzzi, 24 Torino,
Italy
{giuseppe.calafiore, fabrizio.dabbene}@polito.it

1 INTRODUCTION AND MOTIVATION

In the theory of linear systems, the problem of assessing whether the omogeneous system $\dot{x} = Ax$, $A \in \mathbb{R}^{n,n}$ is asymptotically stable is a well understood (and fundamental) one. Of course, the system (and we shall say also the matrix A) is stable if and only if $\mathrm{Re}\lambda_i < 0$, $i = 1, \ldots, n$, being λ_i the eigenvalues of A.

Evolving from this basic notion, much research effort has been devoted in recent years to the study of *robust* stability of a system. Without entering in the details of more than thirty years of fruitful research, we could condense the essence of the robust stability problem as follows: given a bounded set $\mathbf{\Delta}$ and a stable matrix $A \in \mathbb{R}^{n,n}$, state whether $A_\Delta = A + \Delta$ is stable for all $\Delta \in \mathbf{\Delta}$. Since the above deterministic problem may be computationally hard in some cases, a recent line of study proposes to introduce a probability distribution over $\mathbf{\Delta}$, and then to assess the *probability* of stability of the *random matrix* $A + \Delta$. Actually, in the probabilistic approach to robust stability, this probability is *not* analytically computed but only *estimated* by means of randomized algorithms, which makes the problem feasible from a computational point of view, see, for instance, [3] and the references therein.

Leaving apart the randomized approach, which circumvents the problem of analytical computations, there is a clear disparity between the abundance of results available for the deterministic problem (both positive and negative results, in the form of computational "hardness," [2]) and their deficiency in the probabilistic one. In this latter case, almost no analytical result is known among control researchers.

The objective of this note is to encourage research on random matrices in the control community. The one who adventures in this field will encounter unexpected and exciting connections among different fields of science and beautiful branches of mathematics.

In the next section, we resume some of the known results on random matrices, and state a simple new (to the best of our knowledge) closed form result on the probability of stability of a certain class of random matrices. Then, in section 3 we propose three open problems related to the analytical assessment of the probability of stability of random matrices. The problems are presented in what we believe is their order of difficulty.

2 AVAILABLE RESULTS

Notation : A real random matrix $\mathbf{X}$ is a matrix whose elements are real random variables. The probability density (pdf) of $\mathbf{X}$, $f_{\mathbf{X}}(X)$ is defined as the joint pdf of its elements. The notation $\mathbf{X} \sim \mathbf{Y}$ means that $\mathbf{X}, \mathbf{Y}$ are random quantities with the same pdf. The Gaussian density with mean μ and variance σ^2 is denoted as $N(\mu, \sigma^2)$. For a matrix X, $\rho(X)$ denotes the spectral radius, and $\|X\|$ the Frobenius norm. The multivariate Gamma function is defined as $\Gamma_n(x) = \pi^{n(n-1)/4} \prod_{i=1}^{n} \Gamma(x - (i-1)/2)$, where $\Gamma(\cdot)$ is the standard Gamma function.

In this note, we consider the class of random matrices (a class of random matrices is often called an "ensemble" in the physics literature) whose density is invariant under orthogonal similarity. For a random matrix $\mathbf{X}$ in this class, we have that $\mathbf{X} \sim U\mathbf{X}U^T$, for any fixed orthogonal matrix U. For *symmetric* orthogonal invariant random matrices, it can be proved that the pdf of $\mathbf{X}$ is a function of only its eigenvalues $\Lambda \doteq \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, i.e.,

$$f_{\mathbf{X}}(X) = g_{\mathbf{X}}(\Lambda). \tag{1}$$

Orthogonal invariant random matrices may seem specialized, but we provide below some notable examples:

1. G_n: Gaussian matrices. It is the class of $n \times n$ real random matrices with independent identically distributed (iid) elements drawn from $N(0, 1)$.

2. W_n: Whishart matrices. Symmetric $n \times n$ random matrices of the form $\mathbf{X}\mathbf{X}^T$, where $\mathbf{X}$ is G_n.

3. GOE_n: Gaussian Orthogonal Ensemble. Symmetric $n \times n$ random matrices of the form $(\mathbf{X} + \mathbf{X}^T)/2$, where $\mathbf{X}$ is G_n.

4. S_n: Symmetric orthogonal invariant ensemble. Generic symmetric $n \times n$ random matrices whose density satisfies (1). W_n and GOE_n are special cases of these.

5. US_n^ρ: Symmetric $n \times n$ random matrices from S_n, which are uniform over the set $\{X \in \mathbb{R}^{n,n} : \rho(X) \leq 1\}$.

6. US_n^F: Symmetric $n \times n$ random matrices from S_n, which are uniform over the set $\{X \in \mathbb{R}^{n,n} : \|X\| \leq 1\}$.

Whishart matrices have a long history, and are well studied in the statistics literature, see [1] for an early reference. The Gaussian Orthogonal Ensemble is a fundamental model used to study the theory of energy levels in nuclear physics, and it has been originally introduced by Wigner [9, 8]. A thorough account of its statistical properties is presented in [7].

A fundamental result for the S_n ensemble is that the joint pdf of the eigenvalues of random matrices belonging to S_n is known analytically. In particular, if $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ are the eigenvalues of a random matrix $\mathbf{X}$ belonging to S_n, then their pdf $f_{\mathbf{\Lambda}}(\Lambda)$ is

$$f_{\mathbf{\Lambda}}(\Lambda) = \frac{\pi^{n^2/2}}{\Gamma_n(n/2)} g_{\mathbf{X}}(\Lambda) \prod_{1 \leq i < j \leq n} (\lambda_i - \lambda_j). \tag{2}$$

This result can be deduced from [7], and it is also presented in [4]. For some of the ensembles listed above, this specializes to:

$$W_n : \tfrac{\pi^{n^2}}{\Gamma_n^2(n/2)} \exp(-\tfrac{1}{2}\textstyle\sum_i \lambda_i) \prod_i \lambda_i^{-1/2} \prod_{1 \leq i < j \leq n} (\lambda_i - \lambda_j) \tag{3}$$

$$\mathrm{GOE}_n : \tfrac{1}{2^{n/2}\prod_i \Gamma(i/2)} \exp(-\tfrac{1}{2}\textstyle\sum_i \lambda_i^2) \prod_{1 \leq i < j \leq n} (\lambda_i - \lambda_j) \tag{4}$$

$$\mathrm{US}_n^\rho : K_u \prod_{1 \leq i < j \leq n} (\lambda_i - \lambda_j),\ 1 \geq \lambda_1 \geq \ldots \geq \lambda_n \geq -1. \tag{5}$$

The normalization constant K_u in the last expression can be determined in closed form solving a Legendre integral, see eq. (17.6.3) of [7]

$$K_u = n! 2^{\frac{n}{2}(n+1)} \prod_{j=0}^{n-1} \frac{\Gamma(3/2 + j/2)\Gamma^2(1 + j/2)}{\Gamma(3/2)\Gamma((n+j+3)/2)}. \tag{6}$$

Clearly, knowing the joint density of the eigenvalues is a key step in the direction of computing the probability of stability of a random matrix. We remark that the above results all refer to the symmetric case, which has the advantage of having all *real* eigenvalues. Very little is known for instance about the distribution of the eigenvalues of generic Gaussian matrices G_n. By consequence, to the best of our knowledge, nothing is known about the probability of stability of Gaussian random matrices (i.e., matrices drawn using Matlab `randn` command). Famous asymptotic results (i.e., for $n \to \infty$) go under the name of "circular laws" and are presented in [6]. An exact formula for the distribution of the *real* eigenvalues may be found in [5]. We show below a (seemingly new) result regarding the probability of stability for the US_n^ρ ensemble.

2.1 Probability of stability for the US_n^ρ ensemble

Given an $n \times n$ real random matrix $\mathbf{X}$, let $f_{\mathbf{\Lambda}}(\Lambda)$ be the marginal density of the eigenvalues of $\mathbf{X}$. The *probability of stability* of $\mathbf{X}$ is defined as

$$P \doteq \int \cdots \int_{\mathrm{Re}\Lambda<0} f_{\mathbf{\Lambda}}(\Lambda)\mathrm{d}\Lambda. \tag{7}$$

We now compute this probability for matrices in the US_n^ρ ensemble, whose pdf is given in (5). To this end, we first remove the ordering of the eigenvalues, and therefore divide by $n!$ the pdf (5). Then, the probability of stability is

$$P_{US} = \frac{K_u}{n!} \int_{-1}^{0} \cdots \int_{-1}^{0} \prod_{1\le i<j\le n} |\lambda_i - \lambda_j| \, \mathrm{d}\lambda_1 \cdots \mathrm{d}\lambda_n. \tag{8}$$

This multiple integral is a Selberg type integral whose solution is reported for instance in [7], p. 339. The above probability results to be

$$P_{US} = 2^{-\frac{1}{2}n(n+1)}.$$

3 OPEN PROBLEMS

The probability of stability can be computed also for the GOE_n ensemble and the US_n^F ensemble, using a technique of integration over alternate variables. We pose this as the first open problem (of medium difficulty):

Problem 1: *Determine the probability of stability for the GOE_n and the US_n^F ensembles.*

A much harder problem would be to determine an analytic expression for the density of the eigenvalues (which are now both real and complex) of Gaussian matrices G_n, and then integrate it to obtain the probability of stability for the G_n ensemble:

Problem 2: *Determine the probability of stability for the G_n ensemble.*

A numerical estimate of the probability of (Hurwitz) stability for G_n matrices is reported in table 2.2.1, as a function of dimension n.

n	1	2	3	4	5	6
Prob.	0.500	0.250	0.104	0.037	0.011	0.003

Table 2.2.1 Estimated probability of stability for G_n matrices.

As the reader may have noticed, all the problems treated so far relate to random matrices with zero mean. From the point of view of robustness analysis it would be much more interesting to consider the case of *biased* random matrices. This motivates our last (and most difficult) open problem:

Problem 3: *Let $A \in \mathbb{R}^{n,n}$ be a given stable matrix. Determine the probability of stability of the random matrix $A + \mathbf{X}$, where $\mathbf{X}$ belongs to one of the ensembles listed in section 2.*

ACKNOWLEDGEMENTS

The authors wish to thank Professor Madan Lal Mehta for reading and giving his precious advice on this manuscript.

BIBLIOGRAPHY

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York, 1958.

[2] V. D. Blondel and J. N. Tsitsiklis, "A survey of computational complexity results in systems and control," *Automatica*, 36:1249–1274, 2000.

[3] G. Calafiore, F. Dabbene, and R. Tempo, "Randomized algorithms for probabilistic robustness with real and complex structured uncertainty," *IEEE Trans. Aut. Control*, 45(12):2218–2235, December 2000.

[4] A. Edelman, *Eigenvalues and Condition Numbers of Random Matrices*, Ph.D. thesis, Massachusetts Institute of Technology, Boston, 1989.

[5] A. Edelman, "How many eigenvalues of a random matrix are real?", *J. Amer. Math. Soc.*, 7:247–267, 1994.

[6] V. L. Girko, *Theory of Random Determinants*, Kluwer, Boston, 1990.

[7] M. L. Mehta, *Random Matrices*, Academic Press, Boston, 1991.

[8] E. P. Wigner, "Distribution laws for the roots of a random Hermitian matrix," In C. E. Porter, ed., *Statistical Theories of Spectra: Fluctuations*. Academic, New York, 1965.

[9] E. P. Wigner, "Statistical properties of real symmetric matrices with many dimensions," In C. E. Porter, ed., *Statistical Theories of Spectra: Fluctuations*. Academic, New York, 1965.

Problem 2.3

Aspects of Fisher geometry for stochastic linear systems

Bernard Hanzon
Institute for Econometrics, Operations Research and System Theory
Technical University of Vienna
Argentinierstr. 8/119
A-1040 Vienna
Austria
bhanzon@eos.tuwien.ac.at

Ralf Peeters
Department of Mathematics
Maastricht University
P.O. Box 616, 6200 MD Maastricht
The Netherlands
ralf.peeters@math.unimaas.nl

1 DESCRIPTION OF THE PROBLEM

Consider the space S of stable minimum phase systems in discrete-time, of order (McMillan degree) n, having m inputs and m outputs, driven by a stationary Gaussian white noise (innovations) process of zero mean and covariance Ω. This space is often considered, for instance in system identification, to characterize stochastic processes by means of linear time-invariant dynamical systems (see [8, 18]). The space S is well known to exhibit a differentiable manifold structure (cf. [5]), which can be endowed with a notion of distance between systems, for instance by means of a Riemannian metric, in various meaningful ways.

One particular Riemannian metric of interest on S is provided by the so-called *Fisher metric*. Here the Riemannian metric tensor is defined in terms of local coordinates (i.e., in terms of an actual parametrization at hand) by the Fisher information matrix associated with a given system. The open question raised in this paper reads as follows:

> *Does there exist a uniform upper bound on the distance induced by the Fisher metric for a fixed $\Omega > 0$, between any two systems in S?*

In case the answer is affirmative, a natural follow-up question from the differential geometric point of view would be whether it is possible to construct a finite atlas of charts for the manifold S, such that the charts as subsets of Euclidean space are bounded (i.e., contained in an open ball in Euclidean space), while the distortion of each chart remains finite.

2 MOTIVATION AND BACKGROUND OF THE PROBLEM

An important and well-studied problem in linear systems identification is the construction of parametrizations for various classes of linear systems. In the literature a great number of parametrizations for linear systems have been proposed and used. From the geometric point of view the question arises whether one can qualify various parametrizations as good or bad. A parametrization is a way to (locally) describe a geometric object. Intuitively, a parametrization is better the more it reflects the (local) structure of the geometric object. An important consideration in this respect is the *scale* of the parametrization, or rather the *spectrum of scales*, see [4]. To explain this, consider the tangent space of a differential manifold of systems, such as S. The differentiable manifold can be supplied with a Riemannian geometry, for example, by smoothly embedding the differentiable manifold in an appropriate Hilbert space: then the tangent spaces to the manifold are linear subspaces of the Hilbert space, which induces an inner product on each of the tangent spaces and a Riemannian metric structure on the manifold. If such a Riemannian metric is defined, then any sufficiently smooth parametrization will have an associated Riemannian metric tensor. In local coordinates (i.e., in terms of the parameters used) it is represented by a symmetric, positive definite matrix at each point. The eigenvalues of this matrix reflect the local scales of the parametrization: the scale of any infinitesimal movement starting from a given point, will vary between the largest and the smallest eigenvalue of the Riemannian metric tensor at the point involved. Over a set of points the scale will clearly vary between the largest eigenvalue to be found in the spectra of the corresponding set of Riemannian metric matrices and the smallest eigenvalue to be found in that same set of spectra. Following Milnor (see [12]), who considered the question of finding good charts for the earth, we define the distortion of a parametrization, which we will call the *Milnor distortion*, as the quotient of the largest scale and the smallest scale of the parametrization.

Note that this concept of Milnor distortion is a generalization of the concept of the condition number of a matrix. However it is (in general) *not* the maximum of the condition numbers of the set of Riemannian metric matrices.

Indeed, the largest eigenvalue and the smallest eigenvalue that enter into the definition of the Milnor distortion do not have to correspond to the Riemannian metric tensor at the same point.

If one has an atlas of overlapping charts, one can calculate the Milnor distortion in each of the charts and consider the largest distortion in any of the charts of the atlas. One could now be tempted to define this number as the distortion of the atlas and look for atlases with relatively small distortion. However, in this case, the problem shows up that it is always possible to take a large number of small charts, each one displaying very little distortion (i.e., distortion close to one), while such an atlas may still not be desirable as it may require a huge number of charts. The difficulty here is to trade off the number of charts in an atlas against the Milnor distortion in each of those charts. At this point, we have no clear natural candidate for this trade-off. But at least for atlases with an equal *finite* number of charts the concept of maximal Milnor distortion could be used to compare the atlases.

3 AVAILABLE RESULTS

In trying to apply these ideas to the question of parametrization of linear systems, the problem arises that many parametrizations turn out to have in fact an infinite Milnor distortion. Consider for example the case of real SISO discrete-time strictly proper stable systems of order one. (See also [9] and [13, section 4.7].) This set can be described by two real parameters, e.g., by writing the associated transfer function into the form $h(z) = b/(z-a)$. Here, the parameter a denotes the pole of the system and the parameter b is associated with the gain. The Riemannian metric tensor induced by the H_2 norm of this parametrization can be computed as $\begin{pmatrix} b^2(1+a^2)/(1-a^2)^3 & ab/(1-a^2)^2 \\ ab/(1-a^2)^2 & 1/(1-a^2) \end{pmatrix}$, see [9]. Therefore it tends to infinity when a approaches the stability boundary $|a| = 1$, whence the Milnor distortion of this parametrization becomes infinity. In this example the geometry is that of a flat double infinite-sheeted Riemann surface. Locally, it is isometric with Euclidean space and therefore one can construct charts that have the identity matrix as their Riemannian metric tensor (see [13]). However, in this case, this means that close to the stability boundary the distances between points become arbitrarily large. Therefore, although it is possible to construct charts with optimal Milnor distortion, this can only be done at the price of having to work with infinitely large (i.e., unbounded) charts. If one wants to work with charts in which the distances remain bounded then one will need infinitely many of them on such occasions.

In the case of stochastic Gaussian time-invariant linear dynamical systems without observed inputs, the class of stable minimum-phase systems plays an important role. For such stochastic systems the (asymptotic) Fisher information matrix is well-defined. This matrix is dependent on the parametrization used and admits the interpretation of a Riemannian metric tensor (see [15]). There is an extensive literature on the computation of Fisher information, especially for AR and ARMA systems. See, e.g., [6, 7, 11]. Much of this interest derives from the many applications in practical settings: it can be used to establish local parameter identifiability, it is used for parameter estimation in the method of scoring, and it is also known to determine the local convergence properties of the popular Gauss-Newton method for least-squares identification of linear systems based on the maximum likelihood principle (see [10]).

In the case of stable AR systems, the Fisher metric tensor can, for instance, be calculated using the parametrization with Schur parameters. From the formulas in [14] it follows that the Fisher information for scalar AR systems of order one driven by zero mean Gaussian white noise of unit variance is equal to $1/(1-\gamma_1^2)$. Here γ_1 is required to range between -1 and 1 (to impose stability) and to be nonzero (to impose minimality). Although this again implies an infinite Milnor distortion, the situation here is structurally different from the situation in the previous case: the length of the curve of systems obtained by letting γ_1 range from 0 to 1 is finite! Indeed, the (Fisher) length of this curve is computed as $\int_0^1 \frac{1}{\sqrt{1-\gamma_1^2}} d\gamma_1 = \pi/2$.

Let the *inner geometry* of a connected Riemannian manifold of systems be defined by the shortest path distance: $d(\Sigma_1, \Sigma_2)$ is the Riemannian length of the shortest curve connecting the two systems Σ_1 and Σ_2. Then, in this simple case, the Fisher geometry has the property that the corresponding inner geometry has a uniform upper bound. Therefore, this example provides an instance of a *subset* of the manifold S for which the answer to the question raised is affirmative.

As a matter of fact, if one now reparametrizes the set of systems as in [17] by θ defined through $\gamma_1 = \sin(\theta)$, then the resulting Fisher information quantity becomes equal to 1 everywhere. Thus, it is bounded and the Milnor distortion of this reparametrization is finite. But at the same time the parameter chart itself remains bounded! Hence, also the “follow-up question” of the previous section is answered affirmative here.

If one considers SISO stable minimum-phase systems of order 1, it can be shown likewise that also here the Fisher distance between two systems is uniformly bounded and that a finite atlas with bounded charts and finite Milnor distortion can be designed. Whether this also occurs for larger state-space dimensions is still unknown (to the best of the authors’ knowledge) and this is precisely the open problem presented above.

To conclude, we note that the role played by the covariance matrix Ω of the driving white noise is rather limited. It is well known that if the system equations and the covariance matrix are parametrized independently of each other, then the Fisher information matrix attains a block-diagonal structure (see, e.g., [18, Ch. 7]. The covariance matrix Ω then appears as a weighting matrix for the block of the Fisher information matrix associated with the parameters involved in the system equations. Therefore, if Ω is known, or rather if an upper bound on Ω is known (which is likely to be the case in any practical situation!), its role with respect to the questions raised can be largely disregarded. This allows to restrict attention to the situation where Ω is fixed to the identity matrix I_m.

BIBLIOGRAPHY

[1] S.-I. Amari, *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics 28, Springer Verlag, Berlin, 1985.

[2] S.-I. Amari, "Differential geometry of a parametric family of invertible linear systems Riemannian metric, dual affine connections, and divergence," *Mathematical Systems Theory*, 20, 53–82, 1987

[3] C. Atkinson and A. F. S. Mitchell, "Rao's distance measure. *Sankhyā: The Indian Journal of Statistics*, Series A, 43(3), 345–365, 1981.

[4] R. W. Brockett and P. S. Krishnaprasad "A scaling theory for linear systems," *IEEE Trans. Aut. Contr.*, AC-25, 197–206, 1980.

[5] J. M. C. Clark, "The consistent selection of parametrizations in system identification," *Proc. Joint Automatic Control Conference*, 576–580. Purdue University, Lafayette, Indiana, 1976.

[6] B. Friedlander, "On the Computation of the Cramer-Rao Bound for ARMA Parameter Estimation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-32 (4), 721–727.

[7] E. J. Godolphin and J. M. Unwin, "Evaluation of the covariance matrix for the maximum likelihood estimator of a Gaussian autoregressive-moving average process," *Biometrika*, 70 (1), 279–284, 1983.

[8] E. J. Hannan and M. Deistler, *The Statistical Theory of Linear Systems.* John Wiley & Sons, New York, 1988.

[9] B. Hanzon, *Identifiability, Recursive Identification and Spaces of Linear Dynamical Systems*, CWI Tracts 63 and 64, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, 1989.

[10] B. Hanzon and R. L. M. Peeters, "On the Riemannian Interpretation of the Gauss-Newton Algorithm," In: M. Kárný and K. Warwick (eds.), *Mutual Impact of Computing Power and Control Theory*, 111–121. Plenum Press, New York, 1993.

[11] A. Klein and G. Mélard,"On Algorithms for Computing the Covariance Matrix of Estimates in Autoregressive Moving Average Processes," *Computational Statistics Quarterly*, 1, 1–9, 1989.

[12] J. Milnor, "A problem in cartography," *American Math. Monthly*, 76, 1101–1112, 1969.

[13] R. L. M. Peeters, *System Identification Based on Riemannian Geometry: Theory and Algorithms*. Tinbergen Institute Research Series, vol. 64, Thesis Publishers, Amsterdam, 1994.

[14] R. L. M. Peeters and B. Hanzon, "Symbolic computation of Fisher information matrices for parametrized state-space systems," *Automatica*, 35, 1059–1071, 1999.

[15] C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, 37, 81–91, 1945.

[16] N. Ravishanker, E. L. Melnick and C.-L. Tsai, "Differential geometry of ARMA models," *Journal of Time Series Analysis*, **11**, 259–274, .

[17] A. L. Rijkeboer, "Fisher optimal approximation of an AR(n)-process by an AR(n-1)-process," In: J. W. Nieuwenhuis, C. Praagman and H. L. Trentelman eds., *Proceedings of the 2nd European Control Conference ECC '93*, 1225–1229, Groningen, 1993.

[18] T. Söderström and P. Stoica, *System Identification*, Prentice-Hall, New York, 1989.

Problem 2.4

On the convergence of normal forms for analytic control systems

Wei Kang
Department of Mathematics,
Naval Postgraduate School
Monterey, CA 93943
USA
wkang@nps.navy.mil

Arthur J. Krener
Department of Mathematics,
University of California
Davis, CA 95616
USA
ajkrener@ucdavis.edu

1 BACKGROUND

A fruitful technique for the local analysis of a dynamical system consists of using a local change of coordinates to transform the system to a simpler form, which is called a normal form. The qualitative behavior of the original system is equivalent to that of its normal form which may be easier to analyze. A bifurcation of a parameterized dynamics occurs when a change in the parameter leads to a change in its qualitative properties. Therefore, normal forms are useful in analyzing when and how a bifurcation will occur. In his dissertation, Poincaré studied the problem of linearizing a dynamical system around an equilibrium point, linear dynamics being the simplest normal form. Poincaré's idea is to simplify the linear part of a system first, using a linear change of coordinates. Then the quadratic terms in the system are simplified, using a quadratic change of coordinates, then the cubic terms, and so on. For systems that are not linearizable, the Poincaré-Dulac theorem provides the normal form.

Given a C^∞ dynamical system in its Taylor expansion around $x = 0$,

$$\dot{x} = f(x) = Fx + f^{[2]}(x) + f^{[3]}(x) + \cdots \tag{1}$$

where $x \in \Re^n$, F is a diagonal matrix with eigenvalues $\lambda = (\lambda_1, \ldots, \lambda_n)$, and $f^{[d]}(x)$ is a vector field of homogeneous polynomial of degree d. The dots $+\cdots$ represent the rest of the formal power series expansion of f. Let $\mathbf{e}_k$ be the k-th unit vector in $\Re^n$. Let $m = (m_1, \ldots, m_n)$ be a vector of non-negative integers. In the following, we define $|x|$ and x^m by $|m| = \sum |m_i|$ and $x^m = x_1^{m_1} x_2^{m_2} \ldots x_n^{m_n}$. A nonlinear term $x^m \mathbf{e}_k$ is said to be resonant if $m \cdot \lambda = \lambda_k$ for some nonzero vector of non-negative integers m and some $1 \leq k \leq n$.

Definition 1 The eigenvalues of F are in the Poincaré domain if their convex hull does not contain zero, otherwise they are in the Siegel domain.

Definition 2: The eigenvalues of F are of type (C, ν) for some $C > 0, \nu > 0$ if

$$|m \cdot \lambda - \lambda_k| \geq \frac{C}{|m|^\nu}$$

For eigenvalues in the Poincaré domain, there are at most a finite number of resonances. For eigenvalues of type (C, ν), there are no resonances and as $|m| \to \infty$ the rate at which resonances are approached is controlled.

A formal change of coordinates is a formal power series

$$z = Tx + \theta^{[2]}(x) + \theta^{[3]}(x) + \cdots \tag{2}$$

where T is invertible. If $T = I$, then it is called a near identity change of coordinates. If the power series converges locally, then it defines a real analytic change of coordinates.

Theorem 1: *(Poincaré-Dulac) If the system (1) is C^∞ then there exists a formal change of coordinates (2) transforming it to*

$$\dot{z} = Az + w(z)$$

where A is in Jordan form and $w(z)$ consists solely of resonant terms. (If some of the eigenvalues of F are complex then the change of coordinates will also be complex). In this normal form, $w(z)$ need not be unique.

If the system (1) is real analytic and its eigenvalues lie in the Poincaré domain (2), then $w(z)$ is a polynomial vector field and the change of coordinates(2) is real analytic.

Theorem 2: (Siegel) *If the system (1) is real analytic and its eigenvalues are of type (C, ν) for some $C > 0, \nu > 0$, then $w(z) = 0$ and the change of coordinates (2) is real analytic.*

As is pointed out in [1], even in cases where the formal series are divergent, the method of normal forms turns out to be a powerful device in the study of nonlinear dynamical systems. A few low degree terms in the normal form often give significant information on the local behavior of the dynamics.

2 THE OPEN PROBLEM

In [3], [4], [5], [10], and [8], Poincaré's idea is applied to nonlinear control systems. A normal form is derived for nonlinear control systems under change of state coordinates and invertible state feedback. Consider a C^∞ control system

$$\dot{x} = f(x,u) = Fx + Gu + f^{[2]}(x,u) + f^{[3]}(x,u) + \cdots \tag{3}$$

where $x \in \Re^n$ is the state variable, $u \in \Re$ is a control input. We only discuss scalar input systems, but the problem can be generalized to vector input systems. Such a system is called *linearly controllable* at the origin if the linearization (F, G) is controllable.

In contrast with Poincaré's theory, a homogeneous transformation for (3) consists of both change of coordinates and invertible state feedback,

$$z = x + \theta^{[d]}(x), \quad v = u + \kappa^{[d]}(x,u) \tag{4}$$

where $\theta^{[d]}(x)$ represents a vector field whose components are homogeneous polynomials of degree d. Similarly, $\kappa^{[d]}(x)$ is a polynomial of degree d. A formal transformation is defined by

$$z = Tx + \textstyle\sum_{d=2}^{\infty} \theta^{[d]}(x), \quad v = Ku + \sum_{d=2}^{\infty} \kappa^{[d]}(x,u) \tag{5}$$

where T and K are invertible. If T and K are identity matrices then this is called a near identity transformation.

The following theorem for the normal form of control systems is a slight generalization of that proved in [3], see also [8] and [10].

Theorem 3: *Suppose (F, G) in (3) is a controllable pair. Under a suitable transformation (5), (3) can be transformed into the following normal form*

$$\begin{array}{rcl} \dot{z}_i & = & z_{i+1} + \sum_{j=i+2}^{n+1} p_{i,j}(\bar{z}_j) z_j^2 \quad 1 \leq i \leq n-1 \\ \dot{z}_n & = & v \end{array} \tag{6}$$

where $z_{n+1} = v$, $\bar{z}_j = (z_1, z_2, \cdots, z_j)$, and $p_{i,j}(\bar{z}_j)$ is a formal series of $\bar{z}_j$.

Once again, the convergence of the formal series $p_{i,j}$ in (6) is not guaranteed, hence nothing is known about the convergence of the normal form.

Open Problem (The Convergence of Normal Form): *Suppose the controlled vector field $f(x,u)$ in (3) is real analytic and F, G is a controllable pair. Find verifiable necessary and sufficient conditions for the existence of a real analytic transformation (5) that transforms the system to the normal form (6).*

Normal forms of control systems have proven to be a powerful tool in the analysis of local bifurcations and local qualitative performance of control systems. A convergent normal form will make it possible to study a control system over the entire region in which the normal form converges. Global or semi-global results on control systems and feedback design can be proved by studying analytic normal forms.

3 RELATED RESULTS

The convergence of the Poincaré normal form was an active research topic in dynamical systems. According to Poincaré's Theorem and Siegel's theorem, the location of eigenvalues determines the convergence. If the eigenvalues are located in the Poincaré domain with no resonances, or if the eigenvalues are located in the Siegel domain and are of type (C, ν), then the analytic vector field that defines the system is biholomorphically equivalent to a linear vector field. Clearly, the normal form converges because it has only a linear part. The Poincaré-Dulac theorem deals with a more complicated case. It states that if the eigenvalues of an analytic vector field belong to the Poincaré domain, then the field is biholomorphically equivalent to a polynomial vector field. Therefore, the Poincaré normal form has only finite many terms, and hence is convergent.

As for control systems, it is proved in [5] that if an analytic control system is linearizable by a formal transformation, than it is linearizable by an analytic transformation. It is also proved in [5] that a class of three-dimensional analytic control systems, which are not necessarily linearizable, can be transformed to their normal forms by analytic transformations. No other results on the convergence of control system normal forms are known to us.

The convergence problem for control systems has a fundamental difference from the convergence results of Poincaré-Dulac. For the latter, the location of the eigenvalues are crucial and the eigenvalues are invariant under change of coordinates. However, the eigenvalues of a control system can be changed by linear state feedback. It is unknown what intrinsic factor in control systems determines the convergence of their normal form or if the normal form is always convergent.

The convergence of normal forms is an important problem to be addressed. Applications of normal forms for control systems are proved to be successful. In [6] the normal forms are used to classify the bifurcation of equilibrium sets and controllability for uncontrollable systems. In [7] the control of bifurcations using state feedback is introduced based on normal forms. For discrete-time systems, normal form and the stabilization of Naimark-Sacker bifurcation are addressed in [2]. In [10] a complete characterization for the symmetry of nonlinear systems is found for linearly controllable systems.

In addition to linearly controllable systems, the normal form theory has been generalized to larger family of control systems. Normal forms for systems with uncontrollable linearization are derived in several papers ([6], [7], [8], and [10]). Normal forms of discrete-time systems can be found in [9] and [2]. The convergence of these normal forms is also an open problem.

BIBLIOGRAPHY

[1] V. I. Arnold, *Geometrical Method in the Theory of Ordinary Differential Equations*, Springer-Verlag, New York, Berlin, 1988.

[2] B. Hamzi, J.-P. Barbot, S. Monaco, and D. Normand-Cyrot, "Nonlinear discrete-time control of systems with a Naimark-Sacker bifurcation," *Systems and Control Letters*, 44, 4, pp. 245-258, 2001.

[3] W. Kang, *Extended controller normal form, invariants and dynamical feedback linearization of nonlinear control systems*, Ph.D. dissertation, University of California at Davis, 1991.

[4] W. Kang and A. J. Krener, "Extended quadratic controller normal form and dynamic feedback linearization of nonlinear systems," *SIAM J. Control and Optimization*, 30 (1992), 1319-1337.

[5] W. Kang, "Extended controller form and invariants of nonlinear control systems with a single input," *J. of Mathematical Systems, Estimation and Control*, 6 (1996), 27-51.

[6] W. Kang, "Bifurcation and normal form of nonlinear control systems - Part I and II, *SIAM J. Control and Optimization*, 36 (1998), 193-212 and 213-232.

[7] W. Kang, "Bifurcation control via state feedback for systems with a single uncontrollable mode," *SIAM J. Control and Optimization*, 38, (2000), 1428-1452.

[8] A. J. Krener, W. Kang, and D. E. Chang, "Control bifurcations," *IEEE Transactions on Automatic Control*, forthcoming.

[9] A. J. Krener and L. Li, "Normal forms and bifurcations of discrete time nonlinear control systems," *Proc. of* 5^{th} *IFAC NOLCOS Symposium*, Saint-Petersburg, 2001.

[10] I. Tall and W. Respondek, "Feedback classification of nonlinear single-input controls systems with controllable linearization: Normal forms, canonical forms, and invariants," preprint.