SatNav & ATCALS

Naval Aviation PAX River

Technical Exchange Meeting on GPS Technology

The Trans-Kalman Filter (TKF), "Rhobust" Synthesis of GNC & Estimation Systems (Rho-Syn), and Disturbance Accommodating Control (DAC), together with Rhobust Adaptive Nonlinear System Identification (RANSID) as applicable to GPS, JPALS, and other ARINC activities

Robert W. Bass
ARINC Engineering Services

August 16, 2005

The Trans-Kalman Filter (TKF)

It is insufficiently widely appreciated that in designing Kalman Filters (or their generalization to BLUE [Best Linear Unbiased Estimator] filters) it is impossible to improve the estimate of one particular state variable by deliberately allowing degraded estimates of the other state variables. Indeed, if one attempts to replace the Euclidean norm (*RSS*) of the error-vector by a quadratic form based upon ANY other positive-definite weighting matrix ***W*** (e.g. in attempting to minimize a "weighted" *Root Sum Square* rather than standard *RSS*) the result turns out to be identical to what one obtains if ***W*** is taken to be the Identity matrix, i.e. the standard *RSS*!! In an unpublished paper [completed when I was self-employed and before I had joined ARINC] I have shown how to derive the "optimal" Filter Gain Factors when only certain *specified* error variables are considered and the others ignored, resulting e.g. in an alpha-beta tracker for range & azimuth which gives smaller azimuth errors than any possible tracker designed by KF theory, but at the expense of not improving the measured ranges at all. In my paper I have derived the general theory of stable Trans-Kalman Filters (**TKF**s) but so far worked out in detail just this one special case, which has been endorsed by Kalman himself.

The Trans-Kalman Filter (TKF)

- Suppose that the problem is considered only in Discrete Time, wherein the time-interval $dT \equiv: t_k - t_{k-1}$ between epochs $\{t_k\}$ of process-output observations, or sensor measurements, need not be constant, but is considered to be *known retroactively* at each new observation's epoch t_k , for $k = 1, 2, 3, \ldots$. Then, letting $x^k = x(t_k)$, ($k = 1, 2, 3, \ldots$), one considers the LTI stochastic process

$$x^{k+1} = \Phi \cdot x^k + \Gamma \cdot v^k \ ,$$
$$y^k = H \cdot x^k + w^k \ ,$$

- where Φ is a constant $n \times n$ *state-transition* matrix and Γ is a constant $n \times m$ *exogenous input-coupling* matrix or *actuator* matrix (with $m \leq n$), and H is a constant $l \times n$ *output coupling* matrix or *sensor* matrix (with $l \leq n$), and where the m-vector v^k and the l-vector w^k are, respectively, zero-mean Gaussian stochastic white-noise processes of given *covariance* matrices Q & R, *i.e.*

$$\boldsymbol{E}\{v^k\} \equiv 0 \ , \quad \boldsymbol{E}\{v^j \cdot (v^k)'\} = \delta_{jk} \cdot Q,$$
$$\boldsymbol{E}\{w^k\} \equiv 0 \ , \quad \boldsymbol{E}\{w^j \cdot (w^k)'\} = \delta_{jk} \cdot R,$$

- where ' denotes vector-matrix *transposition*; $\boldsymbol{E}$ denotes mathematical *expectation*; δ_{jk} denotes the Kronecker delta; and $Q = Q' \geq 0$ is a non-negative definite $m \times m$ *disturbance-covariance* matrix, while $R = R' > 0$ is a positive-definite $l \times l$ *noise-covariance* matrix.

The Trans-Kalman Filter (TKF)

We postulate the *information-theoretic architecture* of the KF, namely the *linear* estimator

$$x_e^{k+1} = \Phi\cdot x_e^{k} + K\cdot\{ y^{k+1} - H\cdot\Phi\cdot x_e^{k} \} \equiv$$

$$\equiv (I_n - K\cdot H)\cdot\Phi\cdot x_e^{k} + K\cdot H\cdot\Phi\cdot x^{k} + K\cdot H\cdot\Gamma\cdot v^{k} + K\cdot w^{k+1},$$

where K is the celebrated $n \times m$ *Kalman gain matrix* and which *estimator*, when subtracted from the above to provide the dynamics of the *estimation-error* $e \equiv: x - x_e$, yields, after substituting new stochastic processes $\{\xi^k\}$ and $\{\eta^k\}$ by the *definitions*

$$\xi^k \equiv: (I_n - K\cdot H)\cdot\Gamma\cdot v^k, \qquad \eta^k \equiv: -K\cdot w^{k+1},$$

$$e^{k+1} = \Psi\cdot e^k + \xi^k + \eta^k, \qquad \boldsymbol{E}\{\xi^k\cdot(\eta^k)'\} \equiv 0,$$

where

$$\Psi \equiv: (I_n - K\cdot H)\cdot\Phi, \qquad \boldsymbol{E}\{\xi^k\} \equiv 0, \qquad \boldsymbol{E}\{\eta^k\} \equiv 0,$$

$$\boldsymbol{E}\{\xi^j\cdot(\xi^k)'\} = \delta_{jk}\cdot\boldsymbol{Q}, \qquad \boldsymbol{Q} \equiv: (I_n - K\cdot H)\cdot\Gamma\cdot Q\cdot\Gamma'\cdot(I_n - K\cdot H)',$$

$$\boldsymbol{E}\{\eta^j\cdot(\eta^k)'\} = \delta_{jk}\cdot\boldsymbol{R}, \qquad \boldsymbol{R} \equiv: K\cdot R\cdot K',$$

and where I_n denotes the $n \times n$ *identity matrix*.

The Trans-Kalman Filter (TKF)

Now define P_k, the $n \times n$ *estimation-error covariance* matrix, by

$$P_k \;\equiv:\; \boldsymbol{E}\{e^k\cdot(e^k)'\} \equiv (P_k)' \;\ge\; 0,$$

and as usual find that, for $k = 1, 2, 3, \ldots$,

$$P_{k+1} = \Psi\cdot P_k\cdot\Psi' + \boldsymbol{Q} + \boldsymbol{R}\,.$$

In the LTI case it is convenient to take K to be a constant, so we postulate that P_k has a steady-state limit $P = P' \ge 0$, i.e. $P_k \to P$ as $k \to +\infty$. Thus, finally, we have

$$P = \Psi\cdot P\cdot\Psi' + \boldsymbol{Q} + \boldsymbol{R}\,.$$

Using tensor algebra, we may define uniquely scalar functions φ, ψ such that

$$\varphi = (K,Q\,;\,\Phi,\Gamma,H,dT),\quad \psi = (K,R\,;\,\Phi,\Gamma,H,dT),$$

$$P_{11} = \varphi(K,Q) + \psi(K,R).$$

For example, if $l = m = 1$, then

$$Q = Q_0\cdot\sigma_d^2 > 0,\qquad R = R_0\cdot\sigma_s^2 > 0,$$

where $Q_0 = Q_0(\Phi,\Gamma,H,dT)$, $R_0 = R_0(\Phi,\Gamma,H,dT)$, and where σ_d^2 is the continuous-time process-disturbance spectral density and σ_s^2 is the continuous-time sensor-noise spectral density, and we may therefore find scalar functions ϕ and ψ such that, by repetition of the preceding arguments,

$$\phi = \phi(K\,;\,\Phi,\Gamma,H,dT),\qquad \psi = \psi(K\,;\,\Phi,\Gamma,H,dT),$$

$$P_{11} = \phi(K)\cdot\sigma_d^2 + \psi(K)\cdot\sigma_s^2\,,$$

where by inspection it may be verified that P_{11} is truly *linear* in the disturbance and sensor variances σ_d^2 & σ_s^2.

The Trans-Kalman Filter (TKF)

Consider rectilinear constant-speed motion defined by a state-vector $x = [r,v]'$, where r denotes range and $v = dr/dt$ denotes velocity, and let r_m denote the measured range. Then the α-β **Tracker Algorithm** may be stated as

$$r_{new} = (1-\alpha)\cdot[r_{old} + v_{old}\cdot dT] + \alpha\cdot r_m, \qquad (0 < \alpha < 2),$$

$$v_{new} = (1-\beta)\cdot v_{old} + \beta\cdot[(r_m - r_{old})/dT], \qquad (0 < \beta < 2\cdot(2-\alpha)).$$

Note that even if α & β are taken to be constants, the algorithm automatically ***adapts*** to whatever the most recent update time-interval $dT \equiv: t_{new} - t_{old}$ may have been. Moreover, for fixed dT, it may be regarded as a steady-state or limiting KF by setting

$$\Phi = [1, dT\,;\, 0, 1], \quad \Gamma\cdot Q\cdot\Gamma' = [dT^3/3, dT^2/2\,;\, dT^2/2, dT]\cdot\sigma_a^2,$$

$$H = [1, 0], \qquad K = [\alpha, \beta/dT]', \qquad R = \sigma_s^2/dT,$$

where σ_a^2 is the continuous-time variance of white-noise-like *acceleration disturbance* and σ_s^2 is the continuous-time variance of the white-noise-like *sensor noise*. In this $n = 2$ case, straightforward application of the preceding theory yields, after eliminating the redundancy in the $n^2 = 4$ equations for P_{ij}, just $(1/2)\cdot n\cdot(n+1) = 3$ independent linear equations in 3 unknowns, to solve for (P_{11}, P_{12}, P_{22}). After a truly enormous amount of algebra, best done by a symbolic-algebra engine, the final result simplifies greatly down to

$$\begin{bmatrix} 1-\alpha & (1-\alpha)\cdot dT \\ -\beta/dT & (1-\beta) \end{bmatrix}$$

The Trans-Kalman Filter (TKF)

$$P_{11} \;=\; \phi(\alpha,\beta)\cdot\sigma_a^{\,2}\cdot(dT^3/3) + \psi(\alpha,\beta)\cdot\sigma_s^{\,2}/\mathrm{dT},$$

$$\phi \quad \equiv: \; [(1-\alpha)^2/\alpha]\cdot\{[3\cdot(2-\alpha)-\beta]/[\beta\cdot(2\cdot(2-\alpha)-\beta)]\},$$

$$\psi \quad \equiv: \; [2\cdot\alpha^2 + (2 - 3\cdot\alpha)\cdot\beta]/[\alpha\cdot(2\cdot(2-\alpha)-\beta)],$$

Defining the *disturbance-to-noise ratio* ρ by $\rho \equiv: \sigma_a^{\,2}/\sigma_s^{\,2}$, we wish to minimize

$$P_{11} \;=\; \phi(\alpha,\beta)\cdot\rho\cdot(dT^3/3) + \psi(\alpha,\beta)/\mathrm{dT}.$$

The *maximally robust* (or, in complete generality, “least worst”) α-β Tracker is given by

$$\alpha_{\mathrm{OPT}} \;=\; 0.886554, \qquad \beta_{\mathrm{OPT}} \;=\; 1.411793\,,$$

$$P_{11} \;=\; (0.0243289)\cdot\sigma_a^{\,2}\cdot(dT^3/3) \;+\; \alpha_{\mathrm{OPT}}\cdot\sigma_s^{\,2}/\mathrm{dT},$$

In this case the error-state *closed-loop* state-transition matrix becomes one whose characteristic polynomial turns out to be independent of dT and is

$$z^2 - (2 - \alpha - \beta)\cdot z \;+ (1 - \alpha) \;\equiv\; (z - z_1)\cdot(z - z_2) \;=\; 0,$$

whose complex roots $z_{1,2} \equiv: -0.149 \pm 0.302\cdot\boldsymbol{i}$ satisfy $|\,z_1\,| = |\,z_2\,| = 0.336818$ and therefore define poles *inside* the unit circle of the complex z-plane, ensuring *tracker stability* for arbitrary but known dT and *arbitrary/unknown* $(\sigma_a^{\,2}, \sigma_s^{\,2})$. Indeed, the state-vector estimate’s error-norm $\|e^k\|$ decreases by a *factor* of at least (1/3) [e.g. < 0.04 for $k \geq 3$] after each new update at the same time-increment dT !

“Rhobustness” (Bass, analog-time 1956; discrete-time 2002)

A STABILITY-RHOBUSTNESS MARGIN IN DISCRETE TIME

Excerpt from a letter to Prof. Wilson J. Rugh of the Dept. of Electrical & Computer Engineering at the Johns Hopkins University dated August 13, 2002 in which I told him about my having generalized from analog time to discrete time (in connection with an evening course on DSP for the local branch of Florida Inst. of Technology that I was then teaching) my 'Rhobustness' Criterion that I had first published for analog-time systems in 1956 [**10**] and then improved in an IEEE Conference paper [**8**] in 1991):

==

I have just derived what I believe is an elegant generalization from continuous-time systems to discrete-time systems of my theory of "**rhobustification**" of control and observer systems, as follows:

In your book [2] you prove that if the system

$$x^{k+1} \;=\; A\cdot x^{k} + B\cdot u^{k}, \qquad u \;=\; -\,K\cdot x,$$

is stabilized by state-feedback, to become [closed-loop]

$$x^{k+1} \;=\; A_{cl}\cdot x^{k}, \qquad x^{0} = x^{o}, \qquad A_{cl} \;\equiv:\; A - B\cdot K,$$

then there must be positive numbers $\gamma \;\geq\; 1$ and $\lambda \;<\; 1$ $[\lambda \equiv: \max\{|\mathrm{eig}(A_{cl})|\}]$ such that

$$\|x^{k}\| \;\leq\; \gamma\cdot\|x^{o}\|\cdot\lambda^{k}, \qquad (k = 1,2,3, \ldots).$$

“Rhobustness” (Bass, analog-time 1956; discrete-time 2002)

I now define a <u>*stability “rhobustness” margin*</u> ρ, $(0 < \rho < 1)$, by

$$\rho \equiv: (1 - \lambda)/\gamma,$$

and prove that if in
<u>NONSTATIONARY/NONLINEAR & EXTERNALLY-FORCED</u> **actuality** the system is better-modeled by the 'perturbed' system

$$x^{k+1} = A_{cl} \cdot x^k + f(k, x^k) + g(k), \qquad (k = 1,2,3, \ldots),$$

where there exist (κ,δ) such that, for all x and k,

$$\|f(k, x)\| \le \kappa \cdot \|x\|, \ \|g(k)\| \le \delta, \text{ and where } \kappa < \rho, \text{ then}$$

$$\|x^k\| \le \gamma \cdot \|x^o\| \cdot \Lambda^k + \delta/\rho \ ,$$

$$\Lambda \equiv \{ \lambda + [\kappa/\rho](1 - \lambda)\} < 1, \qquad (k = 1,2,3, \ldots).$$

Hence the ***<u>larger</u>*** is ρ, the ***<u>LESS SENSITIVE</u>*** is the actual system to *UNMODELLED* **cross-couplings** & other *NONSTATIONARY* and/or *NONLINEAR* effects, as well as *EXTERNAL FORCING!*

“Rhobust” Tuning of Controls/Filters (Rho-Syn)

The classical KF works only when the assumed process-disturbance covariance matrix ***Q*** and assumed measurement-noise covariance matrix ***R*** [not to mention the usually unwisely ignored cross-variance matrix ***S*** which was present in Kalman’s original theory!] correspond **VERY** closely to physical reality. But in many practical aerospace applications (e.g. aircraft stability and control, and aerospace navigation) it has turned out that better results may be obtained by “tuning the KF” by empirical adjustment of ***Q*** and ***R***. This has led me to publish in various professional-meeting Proceedings a theory of “Rhobust” Tuning of Control & Filter Systems by fixing ***R*** and letting ***S*** vary linearly and ***Q*** vary quadratically in a fictitious design parameter η [that is a monotonic function of my innovataive “n-dimensional bandwidth” ω_{BW} of a stable Multiple-Input Multiple-Output (MIMO) system $\mathcal{S}$ of the Linear Time-Invariant (LTI) type] and which has the property that as η is varied between zero and infinity the robustness of $\mathcal{S}$, as measured by my 1956 criterion $\rho = \rho(\mathcal{S})$, which is a lower bound to the reciprocal of Doyle’s famous 1982 criterion $\mu(\mathbf{S})$, increases to its maximum and then decreases. Choosing the “optimal η” can be presented graphically, by eliminating η in favor of ω_{BW}, and plotting ω_{BW} horizontally and ρ vertically, as “choosing the bandwidth ω_{BW} which maximizes insensitivity to unmodeled dynamics (such as cross-couplings), to mistaken parameter assumptions, to neglected nonlinearities and external forcings, and to ignorance of the physically correct values of (***Q***,***R***,***S***).” I have developed and circulated to a few academic researchers a public-domain MATLAB^{TM} Toolkit to perform such a Rho-Synthesis (including a trade-off between Rhobustification and a half-dozen other desirable performance characteristics), that can be used in comparison with Doyle’s Mu-Syn Toolkit and Savonov’s K_m Toolkit to give the designer a *range* of possibilities: by analogy to flight-control systems, I have called the results “comfortable, tolerable, and survivable” in terms of how close the system gets to catastrophic instability conditions. When in a simulation Rho-Syn was applied to one published VTOL flight-control system, the rapidity of response was tripled, and the size of overshoots was divided by three, with no increase in the required bandwidth or “cost of control action.” There is no reason to doubt that similar improvements could be obtained in GPS-related technology by exploitation of this nearly-perfected but ignored technology.

“Rhobust” Tuning of Controls/Filters (Rho-Syn)

STABILITY/FIDELITY ROBUSTNESS TRADEOFFS IN LINEAR MULTIVARIABLE SYSTEMS DESIGN

Consider a Linear Time Invariant (LTI) Multiple-Input Multiple-Output (MIMO) system, which according to the Ho-Kalman Lemma can always be defined by three matrices (A,B,C) of appropriate dimensions such that if the input and output are time-history vectors u(t) & y(t) then the system’s evolution in time is governed by the differential equations dx/dt = Ax + Bu(t), y = Cx, or, letting s denote the complex frequency in Laplace transforms, the system’s transfer matrix $\Phi = C(sI_n - A)^{-1}B$, where I_n denotes the identity matrix, satisfies $y(s) = \Phi(s)u(s)$. (It is assumed that the design has been completed -- e.g. by use of the Control/Filter algebraic separation theorem [to be rehearsed] so that A is a stability matrix [Hurwitz matrix].) Techniques for evaluation of performance when u(t) is an undesired process disturbance can of course be applied before A has been fixed (e.g. when $A = \text{diag}(F_c, F_f)$, $F_c = F - GK$, $F_f = F - LH$, and the original system’s transfer matrix is $\Phi = H(sI_n - F)^{-1}G$ and the design problem is to choose the control gain matrix K and the filter gain matrix L). Note **six principal types of disturbances**: [1] impulsive, [2] internal or [3] external bounded, [4] stochastic, [5] constant and [6] harmonic; it is possible to define a quantitative characteristic of A which characterizes each: [1] response time; [2, 3] stability robustness margin; [4] dispersion attenuation factor; [5] [static] accuracy; [6a] [normalized] peak magnification & [6b] absolute peak resonance. In all cases, there is an rms power **cost of control/filtering**. If a one-parameter family of the control/filter gains (K,L) is varied, then each of these 7 performance criteria can be plotted as well as an 8th : a novel definition of multivariable bandwidth ω_{BW} as the geometric mean of the least and largest singular values of A (which agrees perfectly with the classical definition in the case n = 2 and is only a slight perturbation for n = 3). Better still, the 7 primary criteria can be plotted as functions of ω_{BW}, which displays clearly the classical result that cost of control varies like the cube of the bandwidth! **Stability robustness** and **fidelity robustness** peak at different bandwidths; the designer must decide how to weight each of the 6 criteria and can then plot a single composite tradeoff which exhibits as a rational design optimum the choice of “best” multivariable bandwidth, thereby fixing the optimal gains (K,L).

STABILITY/FIDELITY ROBUSTNESS TRADEOFFS

IN LINEAR MULTIVARIABLE SYSTEMS DESIGN

by Robert W. Bass

Linear Time Invariant (**LTI**)
Multiple-Input Multiple-Output (**MIMO**)
Control Systems can be modeled as:

x = n-vector **state**, u = r-vector **control**, y = m-vector **output**,
w = p-vector **disturbance**, v = m-vector **noise**

A, B, C, F appropriately dimensioned
(n × n, n × r, m × n, n × p) **constant matrices**

Plant Model

$$dx/dt = Ax + Bu + Fw(t),$$

$$y = Cx + v(t),$$

Practical State-Feedback Control Law

$u = Kx_e$, x_e = observer/filter **estimate** of x,

K = **control gain** matrix

Observer/Filter Dynamics

$dx_e/dt = Ax_e + Bu – L(y – Cx_e)$,

L = **observer gain** matrix

Estimation Error Dynamics

$de/dt = (A + LC)e + Fw(t) + Lv(t)$, $e = (x - x_e)$.

Dynamics of Overall Closed-Loop System

$$dx/dt = (A + BK)x - BKe + Fw(t),$$

$$de/dt = (A + LC)e + Fw(t) + Lv(t).$$

Introduction of 2n-vector Composite-State $\mathbf{x} = [x',e']'$

and (p + m)-vector Composite Disturbance $\mathbf{w} = [w',v']'$ yields

$$d\mathbf{x}/dt = \mathbf{A}\mathbf{x} + \mathbf{F}\mathbf{w}(t),$$

where the 2n × 2n matrix **A** is a function of the design matrices (K,L).

What quantifiable characteristics of $\mathbf{A} = \mathbf{A}(K,L)$ may be computed for each choice of (K,L) and whose optimizations provide rational design criteria?

Transition of Format/Notation to a more general Problem Format

In order to use previously prepared material,
the preceding notation will now be changed.

Replace the 2n-vector **x** by x and
replace the 2n × 2n matrix **A** by A and
replace the 2n-vector **Fw** by a more general
2n-vector v(t) **unrelated** to the previous noise m-vector v(t).

In other words,

$$d\mathbf{x}/dt = \mathbf{A}\mathbf{x} + \mathbf{F}\mathbf{w}(t) \quad \rightarrow \quad dx/dt = Ax + v(t).$$

Henceforth, we shall refer to the second system as an n-vector
system but in this new context the dimension n means
what was formerly denoted by 2n.

New General Problem Formulation

Thus, start over and consider the following completely general n-vector problem.

Given an LTI n-vector system

$$dx/dt = Ax + v(t), \qquad x(0) = 0,$$

what possible classes of disturbances/uncertainties should be considered?

Short of "singular perturbations" which change the effective dimension n, there are only 6 general kinds of disturbance/uncertainty types:

Possible Types of Disturbances

$$v(t) = v_\delta + v_i + v_e + v_s + v_h + v^o ,$$

IMPULSIVE $v_\delta = x^o \delta(t)$,

[bounded] **INTERNAL** $v_i = \Delta A(t,x)x$,

[bounded] **EXTERNAL** $v_e = g(t)$,

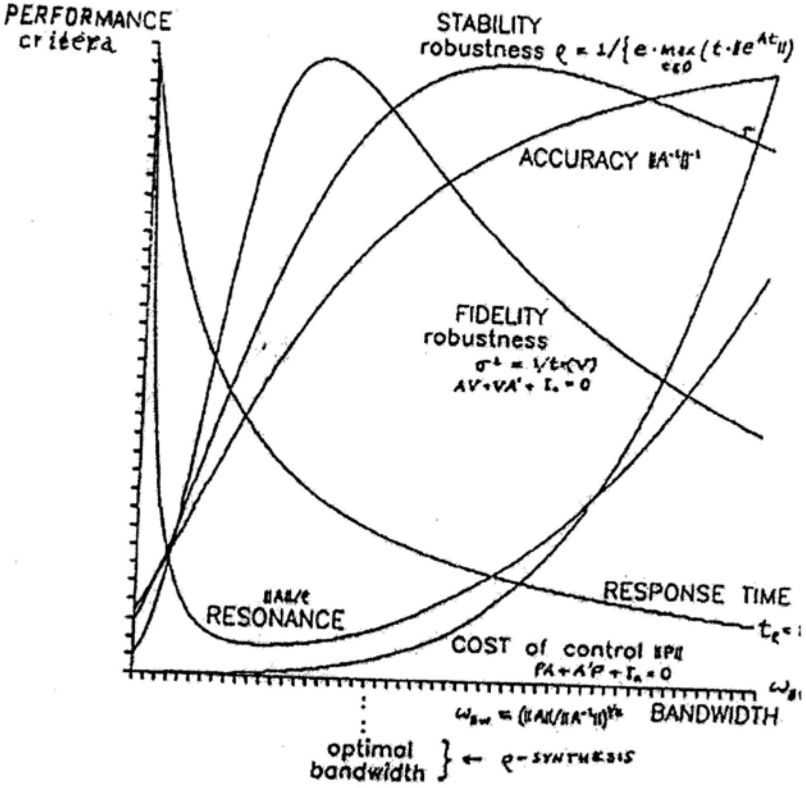
STOCHASTIC $\mathbf{E}\{v_s(t)\} \equiv 0,$

$\mathbf{E}\{v_s(t)\, v_s(\tau)'\} \equiv Q\delta(t - \tau),\quad Q = Q' \geq 0,$

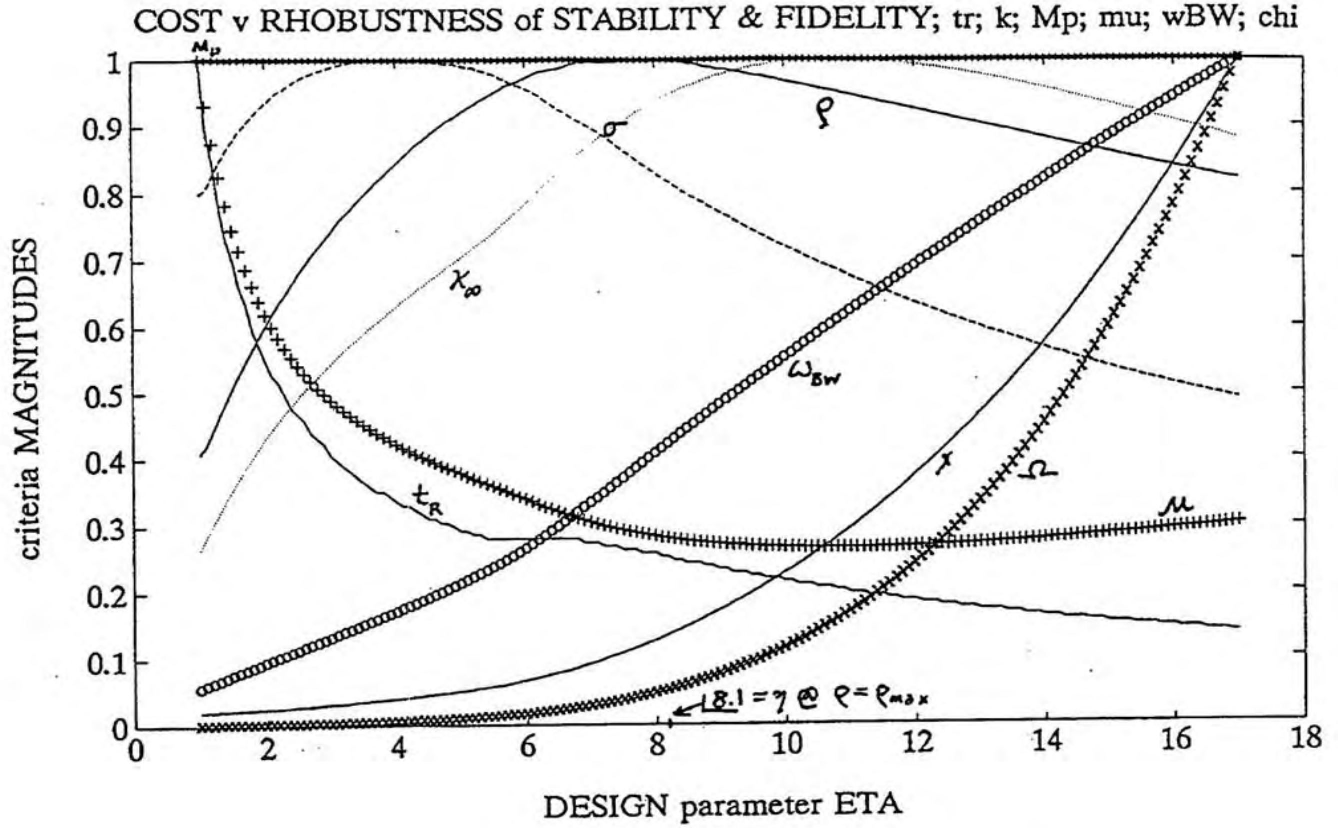
HARMONIC $v_h = v^o e^{j\,\omega\, t}$,

CONSTANT v^o .

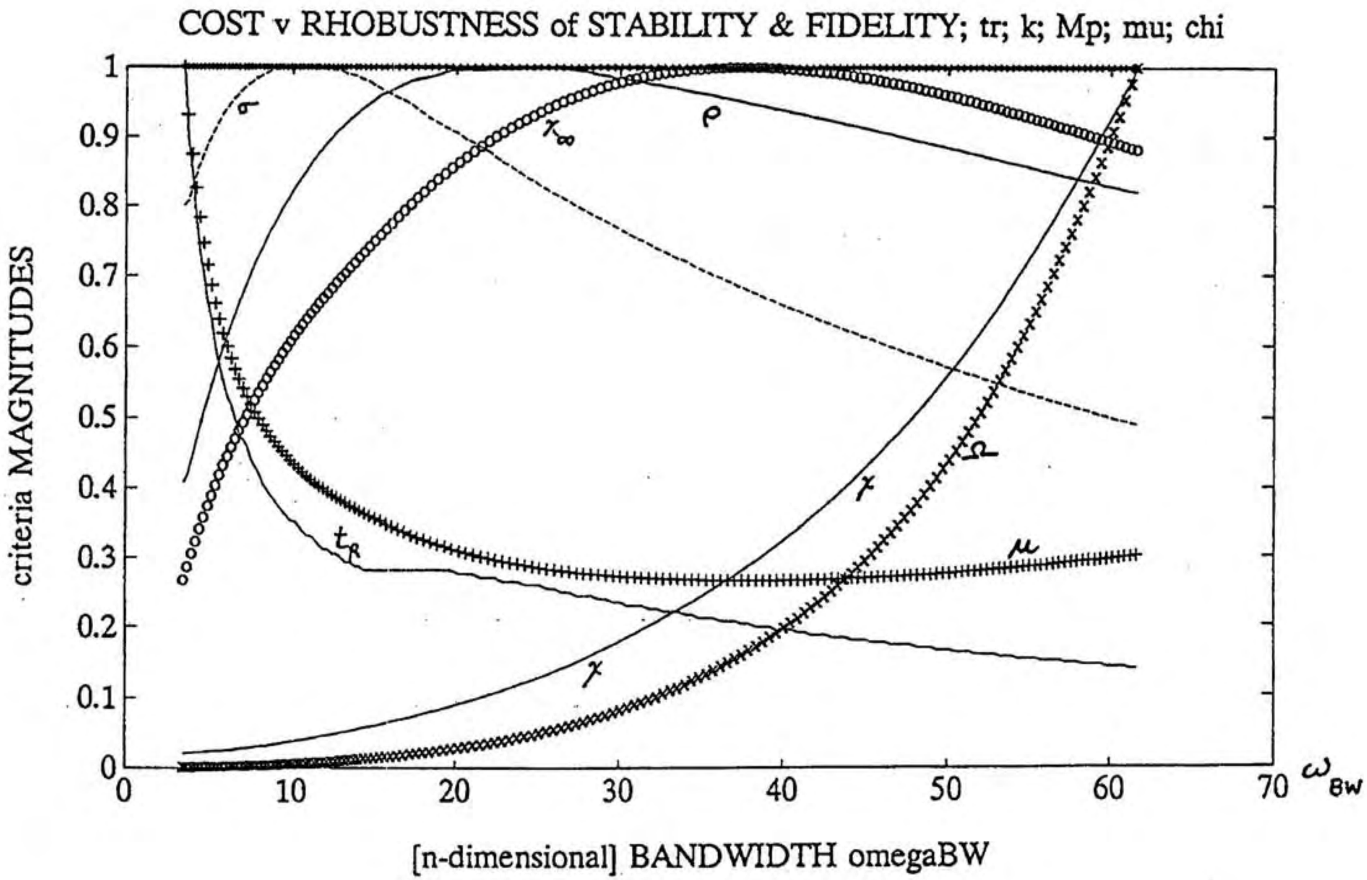
“Rhobust” Tuning of Controls/Filters (Rho-Syn)



“Rhobust” Tuning of Controls/Filters (Rho-Syn)



“Rhobust” Tuning of Controls/Filters (Rho-Syn)



Disturbance Accommodating Control (RhoSynDAC)

Incorporated in my public-domain RhoSyn Toolkit, but ignored (because of advertising propaganda in favor of µ and K_m) is C.D. Johnson’s technique of **Disturbance Accommodating Control** (**DAC**) which provides an adaptive online real-time capability that could be compared to continually updating the initially-assumed covariances ($\boldsymbol{Q}$,$\boldsymbol{R}$,$\boldsymbol{S}$) by a methodology which is formally the same as applying the KF in the case of “colored noise” but can be interpreted physically as continuously updating the “best-fitting waveform” to the disturbances and noises that are usually modeled by stochastic-process theory. When Dan Hill and I simulated the application of **RhoSynDAC** methodology to the redesign of the Electrostatic Gyro (ESG) used to navigate the Trident Submarine, the published result was that this Inertial Measurement Unit (IMU) showed rock-solid stability and virtually no overshoot whatsoever in the presence of nearby depth-charges! Furthermore, in the early days of GPS, some promising applications of DAC were promoted by reputable engineers, but somehow this has been overlooked or forgotten, though I would strongly advocate its revival.

Nonlinear (Quadratic & Cubic) Kalman-Bucy-Wiberg Filters

Another published but neglected technology is the rigorous generalization of the Kalman-Bucy Filter (**KBF**) to those nonlinear systems adequately approximable by retaining quadratic terms in the Taylor series expansions of its nonlinearities, both in its dynamics and in the kinematics of its measurements. In continuous time this is known (e.g. in the books of Bucy and of Jazwinski) as the **Bucy/Jazwinsky/Bass-Norum-Schwartz Filter** and in digital time is known (e.g. in Sage's book) as the **Masters-Sage Filter**. Both are special cases of the cubic-order **Wiberg Estimator** (**WE**) but Wiberg himself admits that his filter is "just too damned complicated" to be useful as other than a guide to *ad hoc* "fixes" of the continuously-relinearized **Extended Kalman Filter** (**EKF**), which is a special case of the quadratic filters just cited. In this connection Bass & Masters (prior to Bass's employment by ARINC) had circulated written funding proposals for combining nonlinear filtering with known input-output-measurement MIMO-LTI-System Identification (ID) technologies based upon the **Ho-Kalman Theorem** to provide a real-time capability for

Rhobust Adaptive Nonlinear System Identification (**RANSID**).

RANSID

Consider an arbitrary **“black box”** containing m input **ports** and l [lower case L] output **ports**, through which signals can be entered and received.

Initially consider analog **input** signals $u(t)$ & **output** signals $y(t)$, in **continuous time** t, $0 \leq t < +\infty$, where $u \in E^m$ and $y \in E^l$ are, respectively, real m-vectors and l-vectors, regarded as column-vectors in real k-dimensional Euclidean vector-space E^k , $k = (l,m)$. By ‘ we shall denote the operation of row-column **transposition**, i.e.

$$u' = [u_1 ,u_2, \ldots,u_m], \quad y' = [y_1, y_2, \ldots, y_l],$$

wherein the columns are transposed to be displayed as rows. In this notation, the *norm* $\| x \|$ of an n-vector x is given by the positive square-root of $x'.x$, and the Euclidean *norm* $\| A \|$ of a non-null n × n matrix A is the smallest number $\| A \| > 0$ such that $\| A.x \| \leq \| A \|.\| x \|$ for all x.

In the case of theoretically perfect analog instrumentation, and no noise or process disturbances, we can conceive of keeping an **input record** $\{ u(t) \mid 0 \leq t < +\infty \}$, or at least a record $\{ u(t) \mid 0 \leq t \leq T \}$ for a finite length of time, where the duration time T is indefinitely large, and an **output record** $\{u(t) \mid 0 \leq t < +\infty \}$.

In the sampled-data or **discrete-time** case, we assume that there is a fixed interval of time T_s, called the **sampling time**, and keep the input and output records only at times $t_n = n.T_s$, $(n = 0, 1, 2, \ldots , N)$. In this case one often suppresses the sampling time and defines the input-output records in terms of sequences of l-vectors and m-vectors called

$$y^k = y(t_k), \quad u^k = u(t_k) , \qquad (k = 0, 1, 2, \ldots , N).$$

According to the profound ***Ho-Kalman Lemma***, under very broadly plausible assumptions pertaining to the **LTI** dynamics inside the black box, there must exist constant matrices (A,B,C) of dimensions respectively $n \times n$, $n \times m$, & $l \times n$, such that for (discrete) times $k = 1, 2, 3,\ldots,N\text{-}1$,

$$x^{k+1} = A.x^k + B.u^k,$$

$$y^k = C.x^k .$$

Here the n-dimensional **state-vector** x^k is **not** known, but its existence is merely inferred! All that is actually measured and recorded are the sequences $\{ u^k \}$ and $\{ y^k \}$. The problem is to derive an algorithm which will enable computation of the matrices (A,B,C) from the input-output records alone. Such a **System Identification** algorithm in the **LTI** case being considered is called an **ID algorithm**.

$\sum_{j=0}^{k-1} A^{k-j-1} B u^j$

RANSID

The most difficult problem is to determine the ***state-dimension*** *n*. One approach is to try a sequence of dimensions, say $n = 2, 3, 4, \ldots, 100$, and see which provides the smallest residual errors after application of the ID procedure. But (Akaike, 1977) there must be a penalty for increasing *n*, lest we fit noise! If the initial state x^0 is given, then one has an *initial condition* problem. It is easy to verify by an inductive proof that the unique solution to this problem is given by

$$x^k = A^{k-1}.x^0 + \Sigma_{j=0}^{k-1} A^{k-j-1}.B.u^j , \quad (k = 1, 2, 3, \ldots, N)$$

$$y^k = C.x^k .$$

The system's ***poles*** are the **roots** z_k , ($k = 1, 2, 3, \ldots, n$), of the *n*th degree polynomial determinantal equation $\Delta(z) = \det(z.I_n - A) = 0$, where I_n denotes the $n \times n$ identity matrix. If the poles are distinct, it is easy to show that there are positive numbers $\gamma \geq 1$ and $\lambda = max\{ |z_k| \}$ such that

$$\|A^k\| \leq \gamma.\lambda^k , \quad (k = 1, 2, 3, \ldots, N).$$

Accordingly a *NASC* (Necessary & Sufficient Condition) for **BIBO** (Bounded-Input, Bounded-Output) **Stability** of the system is that all of the poles lie inside of the unit- circle $|z| = 1$ in the complex *z*-plane. (Sufficiency is obvious when $\lambda < 1$, and necessity is shown by constructing a counter-example wherein for a particular input $\{ u^k \}$ the magnitudes of the outputs, $\{ \|y^k\| \}$, can be driven to unbounded growth if any pole lies on or outside of the unit-circle, i.e. if $\lambda \geq 1$.)

Usually the only systems of engineering interest are those which are stable, and therefore the various system ID procedures are most readily applicable in the case wherein $\lambda < 1$.

Notice that if the system is started from a *relaxed* state (i.e. $x^0 = 0$), then

$$y^k = \Sigma_{j=0}^{k-1} M_{k-j+1}.u^j , \qquad (k = 1, 2, 3, \ldots, N),$$

where the ***Markov Parameters*** $\{ M_j \}$ are defined as

$$M_j = C.A^{j-1}.B , \qquad (j = 1, 2, 3, \ldots, n).$$

When there is no ***process disturbance*** (i.e. errors in recording of $\{ u^k \}$) and no ***measurement noise*** (i.e. errors in recording of $\{ y^k \}$), then by clever matrix-algebra algorithms one can compute the Markov parameters from the input/output records, and use them to define a much larger matrix called the **Hankel Matrix**. The principal discovery of the Ho-Kalman Lemma is that the ***rank*** of the Hankel Matrix is just the system's state-vector ***dimension*** n! Once this has been determined, it is not difficult to use another matrix algorithm to "pull the Markov Parameters apart" and find out the matrices (A,B,C) which define the system.

The Ho-Kalman Lemma has profound theoretical significance, in that it limits the possible structural nature of arbitrary unknown LTI dynamical systems. But its application to real-life problems is not always easy.

RANSID

The chief difficulty in application of the Ho-Kalman system ID methodology is that real data is always to some extent corrupted by disturbances or noise, whence the rank of the Hankel Matrix is not an exact quantity, and the value of n may be incorrect; also the resultant values of (A,B,C) may be erroneous, both because of a mistaken choice of the dimension n, and, even if n is correct, because of numerical imperfections in the values of the empirically-determined Markov Parameters$\{ M_j \}$. Consequently there are available many MATLAB toolkits which purport to determine (n, A, B, C) from experimental measurements of $\{ u^k \}$and $\{ y^k \}$, but each has different strengths and weaknesses, and each performs better on certain classes of problems than on others.

I have an unpublished manuscript on the **BLUE** algorithm for estimating the Markov Parameters in the presence of process disturbances and measurement noises, but this long-standing unrequited goal is not quite finished yet.

Conclusion:

- Many valuable results have been obtained in this System ID Technology field, but there does not seem to have been developed as yet any universally superior approach or uniquely best methodology; consequently it is prudent to attempt to apply several different approaches to ID of the same system and then use pragmatic *ad hoc* reasoning to decide which is probably the most useful result in any given case.