International Institute for Applied Systems Analysis
IIASA
www.iiasa.ac.at

Optimization of Measurement Schedules and Sensor Designs for Linear Dynamic Systems

Mehra, R.K.

IIASA Research Memorandum
March 1976

RM-76-15

OPTIMIZATION OF MEASUREMENT SCHEDULES AND SENSOR DESIGNS FOR LINEAR DYNAMIC SYSTEMS

Raman K. Mehra

March 1976

Optimization of Measurement Schedules and Sensor Designs for Linear Dynamic Systems[1]

Raman K. Mehra

Abstract

This paper presents new results on the problem of measurement scheduling, sensor location and design for linear dynamic systems. Both time-invariant and time-varying systems are considered and different norms of the Observability and Information matrices are maximized with respect to the structural parameters of the system. A close connection is established between these problems and the Kiefer-Wolfowitz Theory of Experimental Design for Regression problems. Both randomized and nonrandomized designs are considered. It is shown that the optimal designs obey certain minmax properties that lead to rapidly convergent algorithms. The results are illustrated by an analytical and a numerical example.

I. Introduction

The importance of measurement system design has been recognized for a long time, but very few attempts have been made at solving this problem. Meier, Peschon and Dressler [1] considered the problem of measurement scheduling and obtained a computational solution using dynamic programming. Johnson [2] defined measures of the quality of controllability and observability and attempted to maximize these measures with respect to structural parameters in the control distribution matrix and the measurement system matrix. (These parameters, in certain systems, depend on controller and sensor locations.) The optimization procedure leads to a nonlinear eigenvalue problem. Müller and Weber [3] considered other measures of the quality of controllability and observability and obtained

[1]This research was supported by the U.S. Office of Naval Research under the Joint Services Electronics Program through Contract N00014-67-A-0298-0006 extended to the Division of Engineering and Applied Physics, Harvard University, and partly by IIASA.

optimum locations of thrusters for satellite attitude control problems.

An extensive literature on the problem of experimental design exists in the field of statistics. Especially significant is the work of Kiefer and Wolfowitz [4] and Fedorov [5] on the theory of optimal experiments for regression. The systems considered are static, and only recently have their results been extended to dynamic systems for input design [6,7,8]. In this paper, we show how similar results can be obtained for a whole range of problems in measurement system design for dynamic systems (both time-invariant and time-varying). The interesting thing about these results is that in addition to revealing interesting minmax properties of optimal designs, they also provide simple algorithms for the computation of such designs.

The organization of the paper is as follows. In section 2, we motivate and formulate mathematically different optimization problems and the associated performance criteria. The results on optimal measurement schedules are presented in section 3. The optimization of sensor and controller structures is considered in section 4. Examples presented in section 5 illustrate the design procedures and elaborate the differences between randomized and nonrandomized designs.

2. Optimization Problems and Criteria

Consider a finite-dimensional time-varying linear dynamic system with state equations

$$\dot{x}(t) = Fx(t) + Gu(t) + \Gamma w(t) \tag{1}$$

$$y(t) = Hx(t) + v(t) \tag{2}$$

$$t_0 \leq t \leq t_1$$

where state $x \in \mathbb{R}^n$, control $u \in \mathbb{R}^m$, measurement $y \in \mathbb{R}^p$, process noise $w \in \mathbb{R}^q$ and measurement noise $v \in \mathbb{R}^p$. Processes w(t) and v(t) are uncorrelated zero mean Gaussian white noise (ZMGWN) processes with intensities Q(t) and R(t). The initial state

$x(0)$ is normally distributed with mean x_0 and covariance P_0 and is independent of the noise processes $\{w(t), v(t), t_0 \leq t \leq t_1\}$.

It is well known that for the above system, the Kalman-Bucy filter [9] provides a minimum-variance unbiased estimate of the state conditional on all past data. The minimal covariance matrix $P(t)$ of the filtered state estimate obeys the following Riccati Differential Equation

$$\dot{P} = FP + PF^T + \Gamma Q \Gamma^T - PH^T R^{-1} HP \tag{3}$$

$$P(0) = P_0 \quad .$$

For a system with fixed structure, i.e. specified P_0, F, Γ, Q and R matrices, matrix $P(t)$ is fixed. But in many practical problems, it is possible to vary H and R under certain constraints. For example, on a flight vehicle, the location of accelerometers influences H and measurement scheduling affects $R(t)$. In the latter case, $R^{-1}(t)$ denotes the precision of the measurements and zero precision ($R^{-1}(t) = 0$) corresponds to no measurement at time t. In certain applications, e.g. Inertial Navigation, it is possible to control the level of precision $R^{-1}(t)$ indirectly by making multiple measurements close together and averaging them. We, therefore, formulate a <u>measurement scheduling problem</u> as follows:

<u>Problem 1'</u>: Select a measurement precision matrix $R^{-1}(t)$ subject to a constraint on the total precision $\int_{t_0}^{t_1} \mathrm{tr}\, R^{-1}(t)dt \leq C$ to minimize a suitable norm of the error covariance matrix $P(t_0,t_1)$.

Problem 1' is similar to the problem considered by Meier, Peschon and Dressler [1], if $R^{-1}(t)$ is allowed to take only two values, one of which is zero (no measurement)*.

The nonlinear nature of the Riccati Eq. (3) makes it very difficult to obtain analytical results and to gain any insight into the nature of the optimal measurement schedules. This

---

*The question of practical implementation of the solution to Problem 1' will be considered in section 3.

difficulty is circumvented here by working with the following well-known bounds on $P(t_0,t_1)$ [10]:

$$\left\{M(t_0,t_1) + W^{-1}(t_0,t_1)\right\}^{-1} \leq P(t_0,t_1)$$

$$\leq M^{-1}(t_0,t_1) + W(t_0,t_1) \quad , \qquad (4)$$

where $M(t_0,t_1)$ is the Fisher Information Matrix for estimating $x(t_1)$ from observations $\{y(t), t_0 \leq t \leq t_1\}$ and $W(t_0,t_1)$ is a generalized controllability matrix, the two satisfying the following equations:

$$M(t_0,t_1) = \int_{t_0}^{t_1} \Phi^T(t,t_1)H^T(t)R^{-1}(t)H(t)\Phi(t,t_1)dt \qquad (5)$$

$$W(t_0,t_1) = \int_{t_0}^{t_1} \Phi(t_1,t)\Gamma(t)Q(t)\Gamma^T(t)\Phi^T(t_1,t)dt \quad . \qquad (6)$$

In Eq. (5) and (6), $\tau(t,\tau)$ is the transition matrix of the system (Eq. (1))

$$\frac{d}{dt}\Phi(t,\tau) = F(t)\Phi(t,\tau) \qquad (7)$$

$$\Phi(\tau,\tau) = I \quad . \qquad (8)$$

Since $W(t_0,t_1)$ does not depend on H and R, Eq. (4) shows that the key quantity for minimizing $P(t_0,t_1)$ is $M^{-1}(t_0,t_1)$. For example, minimization of tr $P(t_0,t_1)$ can be achieved indirectly by minimizing the upper bound, i.e. tr $M^{-1}(t_0,t_1)$. Similarly, the minimization of $|P(t_0,t_1)|$ can be achieved indirectly by either minimizing $|M^{-1} + W|$, the upper bound, or maximizing $|M + W^{-1}|$ (inverse of lower bound). For simplicity, we will first consider criteria involving only the Information Matrix $M(t_0,t_1)$. Notice that for the case of no process noise, i.e. $Q(t) = 0$, and no prior information on $x(0)$, i.e. $P^{-1}(0) = 0$, the solution of the Riccati Eq. (3) is

$$P(t,t_0) + M^{-1}(t,t_0) \quad . \tag{9}$$

For this case, the results presented in sections 3, 4 and 5 are exact. In particular, the maximization of $|M(t_0,t_1)|$ that is to be considered in the following problem formulations is equivalent to the minimization of the volume of the uncertainty ellipsoid for estimating $x(t_1)$ based on the observations $\{y(t), t_0 \leq t \leq t_1\}$ [8].

Based on the above discussion, we state the following problems:

<u>Problem 1</u>: Same as Problem 1' with $P(t_0,t_1)$ replaced by $M^{-1}(t_0,t_1)$.

<u>Problem 2</u>: Minimize a suitable norm of $M^{-1}(t_0,t_1)$ with respect to free parameters in H subject to a norm constraint on H. (Nonrandomized Design, a control version of this problem, is the location of jets on a satellite for attitude control; see section 5.2.)

<u>Problem 3</u>: Minimize a suitable norm of $M^{-1}(t_0,t_1)$ with respect to a probability distribution on the free parameters b in F and H, b being constrained to the set Ω_b. (Randomized Design, see sections 5.1 and 5.2 for examples).

In the next section, we provide a solution to Problem 1 and in sections 4A and 4B, we provide solutions to problems 3 and 2 respectively. Problem 2 is similar to the one considered in Refs. [2, 3], but problem 3 is new. It is motivated by the work of Kiefer and Wolfowitz [4] and Fedorov [5] on Experimental Design in Regression. It is known that randomization can improve performance greatly in experimental design problems. The reader may also refer to Ref. [8] for a brief summary of the results of Ref. [4, 5] and for solutions to a closely related problem, viz. Input Design for parameter estimation.

3. Measurement Scheduling (Problem 1)

3.1 Scalar Measurement Case:

Define normalized precision information

$$\sigma(t) = R^{-1}(t)/C \quad . \tag{10}$$

Since

$$\int_{t_0}^{t_1} R^{-1}(t)dt \leq C \quad ,$$

it follows that

$$\int_{t_0}^{t_1} \sigma(t)dt \leq 1 \quad . \tag{11}$$

Let

$$M(t_0,t_1) = C\hat{M}(t_0,t_1)$$

where

$$\hat{M}(t_0,t_1) = \int_{t_0}^{t_1} \Phi^T(t,t_1)H^TH\Phi(t,t_1)\sigma(t)dt \quad . \tag{12}$$

$\hat{M}(t_0,t_1)$ will be called a Normalized Information Matrix.

It is obvious that the maximization of a suitable norm of $\hat{M}(t_0,t_1)$ under the constraint (11) is equivalent to Problem 1.

Definition:

If $\sigma(\cdot)$ is a continuous function, the design will be called continuous. If $\sigma(\cdot)$ consists only of delta functions, the design will be called discrete. We first prove the following result.

Theorem 1: For any continuous design, there exists a discrete design with no more than $[n(n + 1)/2 + 1]$ points such that the two designs have the same information matrix $\hat{M}(t_0,t_1)$.

Proof: The proof follows from a classical theorem of Caratheodory which states that a point in the convex hull, S^*, of a set S in an m-dimensional Euclidean space can be expressed as a linear combination of $(m + 1)$ or less points of S; i.e. if $s^* \in S^*$, one can find α_i such that

$$s^* = \sum_{i=1}^{m+1} \alpha_i s_i \tag{13}$$

where

$$\sum_{i=1}^{m+1} \alpha_i = 1 \quad , \qquad \alpha_i \geq 0 \quad .$$

We now identify S* with the set of all information matrices $\hat{M}$ in an $n(n + 1)/2$ dimensional Euclidean space[2]. This set is convex since if $\hat{M}(\sigma_1)$ and $M(\sigma_2)$ belong to S* and

$$\sigma = (1 - \alpha)\sigma_1 + \alpha\sigma_2 \quad ; \qquad 0 \leq \alpha \leq 1 \quad ,$$

then

$$\hat{M}(\sigma) = (1 - \alpha)\hat{M}(\sigma_1) + \alpha\hat{M}(\sigma_2) \quad . \tag{14}$$

Therefore, $\hat{M}(\sigma)$ belongs to S*. Furthermore S* is the convex hull of point measurement (i.e. $\sigma(t) = \delta(t - t_1)$) information matrices

$$\hat{M}(t_i) = \Phi^T(t_i,t_1)H^T(t_i)H(t_i)\Phi(t_i,t_1) \quad . \tag{15}$$

It now follows from Catathéodory's Theorem that any $\hat{M}(t_0,t_1)$ be written as

$$\hat{M}(t_0,t_1) = \sum_{i=1}^{n(n+1)/2+1} \hat{M}(t_i) \;\; {}_i \quad . \tag{16}$$[3]

The right hand side of Eq. (16) corresponds to the discrete design

$$\sigma(t) = \sum_{i=1}^{n(n+1)/2+1} \sigma_i\delta(t - t_i) \quad . \tag{17}$$

We now prove a theorem characterizing the solution of problem 1.

<u>Theorem 2</u>: Let * be the optimal measurement schedule. Then $\int_{t_0}^{t_1} \sigma^*(t)dt = 1$ and $\sigma^*(t)$ may be chosen to consist of $\leq \frac{n(n + 1)}{2}$

---

[2] An $n \times n$ symmetric matrix can be represented by a point $R^{n(n+1)/2}$.

[3] If $\hat{M}(t_0,t_1)$ lies on the boundary of the set S*, e.g. when $|M|$ is maximum, then only $n(n+1)/2$ points are required to represent $\hat{M}(t_0,t_1)$.

points, i.e. measurements are made at only k different time point. Furthermore, the following are equivalent:

(i) $\sigma^*(t)$ maximizes $|\hat{M}(t_0,t_1,\sigma)|$

(ii) $\sigma^*(t)$ minimizes $\underset{t}{\text{Max}}\ d(t_0,t_1,t,\sigma)$,

where

$$d(t_0,t_1,t,\sigma) = H(t)\Phi(t,t_1)\hat{M}^{-1}(t_0,t_1,\sigma)\Phi^T(t,t_1)H^T(t) \tag{18}$$

(iii) $$\underset{t}{\text{Max}}\ d(t_0,t_1,t,\sigma^*) = n \quad . \tag{19}$$

All designs satisfying (i)-(iii) and their linear combinations are optimal and have the same information matrix.

Remark: To illustrate the nature of the results in Theorem 2, we show that the quantity $d(t_0,t_1,t,\sigma)$ is the variance of the estimate $\hat{y}(t|t_0,t_1) = H(t)\hat{x}(t|t_0,t_1)$, i.e. the conditional estimate of the output based on all observations over the period $[t_0,t_1]$. This is easily seen from the relations

$$y(t) = H(t)\Phi(t,t_1)x(t_1) \tag{20}$$

and

$$\hat{y}(t|t_0,t_1) = H(t)\Phi(t,t_1)\hat{x}(t_1|t_0,t_1) \quad . \tag{21}$$

(It is assumed that Q = O.) Thus,

$$E\left\{ y(t) - \hat{y}(t|t_0,t_1)^2 \right\} = H(t)\Phi(t,t_1)M^{-1}(t_0,t_1)\Phi^T(t,t_1)H^T(t)$$

$$= d(t_0,t_1,t,\sigma) \quad . \tag{22}$$

Part (ii) of Theorem 2 implies that the design σ^* is also optimal in a minmax sense, i.e. it minimizes the maximum over time of the output prediction variance. From part (iii), Eq. (19), the minmax value of the prediction variance is n, the dimension of the state vector. In fact, it will be shown in Corrollary 1 that the function $d(t_0,t_1,t,\sigma^*)$ looks like a Chebychev function with all local maxima of $d(t_0,t_1,t,\ )$ (Fig.1). From an Information Theoretic viewpoint, measurements are made atthose times when the output entropy is maximum. This appears to be a fairly general principle.

Proof: We prove the theorem in four parts:

(1) The condition $\int_{t_0}^{t_1} \sigma^*(t)dt = 1$ follows trivially from the fact that a scaling of $\sigma^*(\cdot)$ by c scales $|M(t_0,t_1,\sigma^*)|$ by c^n. Heuristically, it is optimal to use all the precision that is available.

(2) We now show that for any normalized[4] design σ,

$$\max_t d(t_0,t_1,t,\sigma) \geq n \quad . \tag{23}$$

Consider

$$\int_{t_0}^{t_1} d(t_0,t_1,t,\sigma)\sigma(t)dt$$

$$= \mathrm{tr}\left|\hat{M}^{-1}(t_0,t_1,\sigma) \int_{t_0}^{t_1} \Phi^T(t,t_1)H^T(t)H(t)\Phi(t,t_1)\sigma(t)dt\right|$$

$$= \mathrm{tr}\left|\hat{M}^{-1}(t_0,t_1,\sigma)M(t_0,t_1,\sigma)\right| = n \quad . \tag{24}$$

Since $1 \geq \sigma(t) \geq 0$ and $\int_{t_0}^{t_1} \sigma(t)dt = 1$, Eq.(23) follows from Eq.(24).

We now show that for an optimal design σ^* satisfying property (i), the inequality (23) is reversed, i.e.

$$\max_t d(t_0,t_1,t,\sigma^*) \leq n \quad . \tag{25}$$

Then parts (ii) and (iii) of Theorem 2 follow from Eq.(23) and

---

[4] A normalized design $\sigma(.)$ is such that $\int_{t_0}^{t_1} (t)dt = 1$.

Then parts (ii) and (iii) of Theorem 2 follow from Eq. (23) and (25).

Consider a perturbation from the optimal design of the following type.

$$\sigma(t) = (1 - \alpha)\sigma^*(t) + \alpha\delta(t - t_i) \quad , \qquad t_0 \leq t \leq t_1 \qquad (26)$$

where $\delta(t - t_i)$ is a delta function at t_i and $0 \leq \alpha < 1$. The information matrix for $\sigma(t)$ design is (for simplicity, the subscripts (t_0, t_1) are omitted from $\hat{M}$)

$$\hat{M}(\sigma) = (1 - \alpha)\hat{M}(\sigma^*) + \alpha\Phi^T(t_i,t_1)H^T(t_i)\Phi(t_i,t_1) \quad . \qquad (27)$$

Since σ^* maximizes $|\hat{M}|$ or equivalently $\log |\hat{M}|$, any deviation from δ^* such as Eq.(26) should result in a decrease of $\log |\hat{M}|$, i.e.

$$\frac{\partial}{\partial\alpha} \log|\hat{M}(\sigma)|_{\alpha=0} \leq 0$$

or

$$\mathrm{tr}\ \hat{M}^{-1}(\sigma)\ \frac{\partial\hat{M}(\sigma)}{\partial\alpha}\Bigg|_{\alpha=0} \leq 0$$

or

$$- n + d(t_0,t_1,t_i,\sigma^*) \leq 0 \quad . \qquad (28)$$

Since Eq.(28) holds for any t_i, Eq.(25) follows if t_i is chosen to maximize $d(t_0,t_1,t,\sigma^*)$ with respect to t.

(3) We now show that (i) and (iii) follow from (ii). Indeed (iii) follows directly from (ii) and Eq.(23). To show that (i) also follows, assume the contrary, i.e. σ^* satisfying (ii) does not maximize $|\hat{M}|$. Consider a perturbed normalized design.

$$\begin{aligned} \bar{\sigma}(t) &= (1 - \alpha)\sigma^*(t) + \alpha\sigma(t) \\ \hat{M}(\sigma) &= (1 - \alpha)\hat{M}(\sigma^*) + \alpha M(\sigma) \end{aligned} \quad , \quad t_0 \leq t \leq t_1 \quad , \quad 0 \leq \alpha \leq 1 \quad .$$

It follows that σ can be chosen such that

$$\frac{\partial}{\partial\alpha} \log|\hat{M}(\bar{\sigma})| \Big|_{\alpha=0} > 0$$

or

$$\mathrm{tr}\left\{- I + \hat{M}^{-1}(\sigma^*)\hat{M}(\sigma)\right\} > 0$$

or

$$\mathrm{tr}\left\{\hat{M}^{-1}(\sigma^*)\hat{M}(\sigma)\right\} > n \quad . \tag{29}$$

Now using Theorem 1, $\hat{M}(\sigma)$ can be expressed as (cf. Eq.(16))

$$\hat{M}(\sigma) = \sum_{i=1}^{k} \hat{M}(t_i)\sigma_i$$

where

$$k \leq n(n+1)/2 + 1 \quad , \quad \sum_{i=1}^{k} \sigma_i = 1 \quad , \quad 0 \leq \sigma_i \leq 1$$

and

$$\hat{M}(t_i) = \Phi^T(t_i,t_1)H^T(t_i)H(t_i)\Phi(t_i,t_1) \quad .$$

Thus

$$\sum_{i=1}^{k} d(t_0,t_1,t_i,\sigma^*)\sigma_i > n \quad . \tag{30}$$

But from (iii), it easily follows that

$$\sum_{i=1}^{k} d(t_0,t_1,t_i,\sigma^*)\sigma_i \leq n \quad . \tag{31}$$

Eqs. (30) and (31) are in contradiction unless (i) holds.

(4) To prove the last part of Theorem 2, consider two designs σ_1^* and σ_2^* both satisfying (i)-(iii). Construct a new design

$$\sigma^* = (1-\alpha)\sigma_1^* + \alpha\sigma_2^* \quad , \qquad 0 < \alpha < 1 \quad .$$

Then

$$\hat{M}(\sigma^*) = (1 - \alpha)M(\sigma_1^*) + \alpha M(\sigma_2^*) \quad .$$

From the concavity of log $|M|$,

$$\log|\hat{M}(\sigma^*)| \geq (1 - \alpha)\log|\hat{M}(\sigma_1^*)| + \alpha\log|\hat{M}(\sigma_2^*)| \quad .$$

But by definition $\log|\hat{M}(\sigma^*)| \not< \log|\hat{M}(\sigma_i)|$, for i = 1,2, so that $\hat{M}(\sigma^*) = \hat{M}(\sigma_1^*) = \hat{M}(\sigma_2^*)$.

Corollary: At all points t_i of the optimal design, $d(t_0,t_1,t_i,\sigma^*)$ attains it maximum value n.

Proof: Assume the contrary, that for some t_i,

$$d(t_0,t_1,t_i,\sigma^*) < n \quad .$$

Then

$$\sum_{i=1}^{k} d(t_0,t_1,t_i,\sigma^*)\sigma_i < n$$

which contradicts Eq. (24). Thus

$$d(t_0,t_1,t_i,\sigma^*) = n$$

for all points t_i of the optimal design (see Fig.1).

3.2 Algorithm for Computation of σ^*

Algorithm 1:

(a) Start with a discrete normalized design $\sigma^0 = \{\sigma^0(t_i), i = 1,\ldots,k\}$ such that $\hat{M}(\sigma^0)$ is nonsingular. The number of points k in the design must be more than n to have a nonsingular $M(\sigma^0)$. Let j = 0.

(b) Compute $d(t_0,t_1,t, {}^j)$ and determine its maximum with respect to t. Denote the maximizing t as t_{k+j+1}.

(c) If $\max_t d(t_0,t_1,t,\sigma^0) = n$, stop, since σ^j is then the optimal design. Otherwise proceed to (d).

(d) Update σ^j to $\sigma^{j+1} = \left\{ \sigma^{j+1}(t_i),\ i = 1,\dots,k+j+1 \right\}$ as follows:

$$\sigma^{j+1}(t_i) = (1 - \alpha^j)\sigma^j(t_i) + \alpha^j \delta(t - t_{k+j+1}) \qquad (32)$$

$$i = 1,\dots,k+j+1 \quad ,$$

where $0 < \alpha^j \leq 1$ is chosen by either a one-dimensional search to maximize $|M(\sigma^{j+1})|$ or any sequence such that

$$|M(\sigma^{j+1})| > |M(\sigma^j)| \quad , \quad \sum_{j=0}^{\infty} \alpha^j = \infty \ \lim_{j\to\infty} \alpha^j = 0.$$

(e) Go to (b).

Remarks:

1. The above algorithm is very similar to an Input Design algorithm in Frequency Domain described in Ref. [6]. Identical proofs of convergence of the algorithm hold in both cases. These methods are related to the work of Kiefer, Wolfowitz and Fedorov [4,5], the last reference being an extensive treatment of the subject.

2. A practical implementation of the design would consist of clusters of measurements at the design points t_i; i.e. multiple measurements will be made close together, but without violating the assumption of independence of measurement noise to achieve a specified precision level $\sigma(t_i)$. The closeness of any given design σ to the optimal design σ^* may be tested using the following bound which is derived analogously to a similar bound in Input design [7]:

$$\frac{|\hat{M}(\sigma)|}{|M(\sigma^*)|} \geq \exp n - \max_t d(t_0,t_1,t,\sigma) \quad . \qquad (33)$$

In practice, the optimal design σ^* may not be achieved precisely, but one can check its closeness to the optimal design by using Eq.(33).

3.3 Vector Measurement Case

This case, though technically more difficult, leads to results essentially similar to the scalar measurement case. Since the proofs and derivations are similar to those for Input Design [7], we only state the main results without proofs. Notice, however, a special case in which the measurement noise matrix may be written as $R^{-1}(t) = (t)\bar{R}^{-1}(t)$ where $\bar{R}(t)$ is fixed and only (t) is subject to choice. By appropriate scaling, we can convert the constraint $\int_{t_0}^{t_1} \mathrm{tr}\left(R^{-1}(t)\right)dt = C$ to the constraint $\int_{t_0}^{t_1} \sigma(t)dt = 1$. Then the problem of selecting $\sigma(t)$ is similar to that of the scalar measurement case and all the results of Theorem 2 apply with $M(t_0,t_1,\sigma)$ and $d(t_0,t_1,t,\sigma)$ modified appropriately.

We now consider the case in which complete $R^{-1}(t)$ processes can be selected subject to $\mathrm{tr} \int_{t_0}^{t_1} R^{-1}(t)dt \leq 1$ and show that $R^{-1}(t)$ may be chosen to be zero everywhere (i.e. no measurements) except at a finite number of time points. In actual practice, it may be difficult to make measurements in such a way as to achieve the optimal design since it would involve controlling correlations between measurements. For this reason, we present the following results more for theoretical completeness rather than for immediate practical applications.

Theorem 3: For the optimal measurement policy $\sum^* = (R^*)^{-1}$, $\int_{t_0}^{t_1} \mathrm{tr} \sum^*(t)dt = 1$ and the following are equivalent:

(i) $\sum^*$ maximizes $|\hat{M}|$,

(ii) $\sum^*$ minimizes $\max_t \lambda_{max}\left(D(t,\sum)\right)$ where

$$D(t,\textstyle\sum) = H(t)\Phi(t,t_1)\hat{M}^{-1}(t_0,t_1,\textstyle\sum)\Phi^T(t,t_1)H^T(t) \quad , \tag{34}$$

and $\lambda_{max}(B)$ denotes the maximum eigenvalue of $D(t,\sum)$,

(iii) $\max_t \lambda_{max}\left(D(t,\sum^*)\right) = n$. (35)

The information matrices of all normalized designs satisfying (i)-(iii) are identical and their linear combinations also satisfy (i)-(iii).

3.4 Other Criteria

Theorem 2 is easily extended to more general criteria such as $tr(M^{-k})$, $k > 0$. We state here the results for the scalar measurement case only.

Theorem 4: For an optimal design σ^*, the following are equivalent:

(i) σ^* minimizes $tr(M^{-k})$, $k > 0$

(ii) σ^* minimizes $\max_t d(t,\sigma)$ where

$$d(t,\sigma) = H(t)\Phi(t,t_1)\hat{M}^{-(k+1)}(\sigma)\Phi^T(t,t_1)H^T(t) \quad (36)$$

(iii) $\max_t d(t,\sigma^*) = tr\left(\hat{M}^{-k}(\sigma^*)\right)$. (37)

Remark:

If the determinant of the lower bound, i.e. $\left|M + W^{-1}\right|$, in the Riccati solution (4) is to be minimized, then the relevant quantity $d(t,\sigma)$ is

$$d(t,\sigma) = H(t)\Phi(t,t_1)\left(M(\sigma) + W^{-1}\right)^{-1}\Phi^T(t,t_1)H^T(t) \quad (38)$$

and

$$\max_t d(t,\sigma^*) = tr\left[\left(M(\sigma^*) + W^{-1}\right)^{-1}M(\sigma^*)\right] \quad . \quad (39)$$

Similarly, for minimizing the determinant of the upper bound, viz. $\left|M^{-1} + W\right|$, the relevant $d(t,\sigma)$ is

$$d(t,\sigma) = H(t)\Phi(t,t_1)\left(M^{-1}(\sigma) + W\right)^{-1}\Phi^T(t,t_1)H^T(t)$$

and

$$\max_t d(t,\sigma^*) = \mathrm{tr}\left[\left(M^{-1}(\sigma^*) + W\right)^{-1} M(\sigma^*)\right] \quad . \tag{40}$$

4. Optimization of Sensor Designs

In this section, we consider solutions to Problems 2 and 3 posed in section 2. First, Problem 3 involving Randomized Designs will be considered since its solution resembles closely the solutions to the measurement scheduling problem.

A. Randomized Designs:

Let b denote the vector of design parameters in H and F and let Ω_b be the set of allowable b values. It will be assumed that Ω_b is a compact set and that randomization is permitted, i.e. different values of b may be chosen with different probabilities during the experiment. We, therefore, define a probability measure ξ for all Borel sets of Ω_b including single points and search for the optimal design ξ^* that maximizes a suitable norm of $M(\xi)$. It will be shown that ξ^* may be chosen to have a finite support, i.e. a discrete probability distribution with mass at only a finite number of points in Ω_b.

Information Matrix:

For a randomized design ξ, the expected information matrix is

$$M(\xi) = \int_{\Omega_b} M(b)\,\xi(db) \tag{41}$$

where

$$\int_{\Omega_b} \xi(db) = 1 \quad , \qquad 0 \le \xi(db) \le 1$$

and $M(b)$ is the information matrix corresponding to the single-point design $\xi(b) = 1$.

It is clear from Eq.(41) that the set of all $M(\xi)$ is a convex and closed set in $\mathbb{R}^{n(n+1)/2}$. It is also the convex hull of single-point information matrices $M(b)$ so that using Caratheodory's Theorem, one may write (see proof of Theorem 1 and Ref. [4-8]).

$$M(\xi) = \sum_{i=1}^{k} \xi_i M(b_i) \tag{42}$$

where

$$k \leq n(n + 1)/2 + 1$$

$$0 < \xi_i \leq 1 \quad , \qquad \sum_{i=1}^{k} \xi_i = 1 \quad .$$

If $M(b)$ is linear in b, then $M(\xi) = M(\bar{b})$, where $\bar{b}$ is the mean value of b. Thus, nothing is gained by randomization on those parameters which effect M linearly, i.e. $\sigma(t)$ in section 3.

We now derive the following theorem for D-optimal designs[5].

Theorem 5: Let ξ^* be the D-optimal design. Then the followingare equivalent.

(i) ξ^* maximizes $|M(\xi)|$

(ii) ξ^* minimizes $\max_{b \in \Omega_b} \mathrm{tr}\left\{M^{-1}(\xi) M(b)\right\}$

(iii) $\max_{b \in \Omega_b} \mathrm{tr}\left\{M^{-1}(\xi^*) M(b)\right\} = n \quad .$ (43)

All designs satisfying (i)-(iii) have the same information matrix M and their linear combinations also satisfy (i)-(iii).

Proof: Since the proof of this theorem is similar to that of Theorem 1 we only give an outline. Consider

$$n = \mathrm{tr}\left(M^{-1}(\xi) M(\xi)\right) = \int_{\Omega_b} \mathrm{tr}\left(M^{-1}(\xi) M(b)\right) \xi(db) \leq \max_{b \in \Omega_b} \mathrm{tr}\left(M^{-1}(\xi) M(b)\right)$$

or

$$\max_{b \in \Omega_b} \left(\mathrm{tr}\ M^{-1}(\xi) M(b)\right) \geq n \quad . \tag{44}$$

[5] A D-optimal design is one that maximizes the determinant of the information matrix.

However, for a design $\xi = (1 - \alpha)\xi^* + \alpha\xi_0$, $0 < \alpha \leq 1$ it follows from (i) that

$$\frac{\partial}{\partial\alpha} \log\left|M(\xi)\right|\Big|_{\alpha=0} \leq 0$$

or

$$\mathrm{tr}\left[M^{-1}(\xi^*)\Big(M(\xi_0) - M(\xi^*)\Big)\right] \leq 0$$

or

$$\mathrm{tr}\left[M^{-1}(\xi^*)M(\xi_0)\right] \leq n \quad . \tag{45}$$

Choose ξ_0 to be a design that assigns probability 1 to the value $b_0 \varepsilon \Omega_b$ maximizing $\mathrm{tr}\left|M^{-1}(\xi)M(b)\right|$. Then

$$\max_{b\varepsilon\Omega_b} \mathrm{tr}\left[M^{-1}(\xi^*)M(b)\right] \leq n \quad . \tag{46}$$

Eqs. (44) and (46) are contradictory unless (ii) and (iii) hold. To show that (ii) implies (i), we assume the contrary, i.e. ξ^* minimizes $\max_{b\varepsilon\Omega_b} \left[\mathrm{tr}\ M^{-1}(\xi)M(b)\right]$ so that (43) holds, but ξ^* does not maximize $\left|M(\xi)\right|$ or there exists a design $\bar{\xi} = (1-\alpha)\xi^* + \alpha\xi$ such that

$$\frac{\partial}{\partial\alpha} \log\left|M(\bar{\xi})\right|\Big|_{\alpha=0} > 0$$

or

$$\mathrm{tr}\left[M^{-1}(\xi^*)M(\xi)\right] > n \quad . \tag{47}$$

Using Eq. (42), this implies

$$\sum_{i=1}^{k} \xi_i\ \mathrm{tr}\left[M^{-1}(\xi^*)M(b)\right] > n$$

or

$$\max_{b \varepsilon \Omega_b} \operatorname{tr}\left[M^{-1}(\xi^*)M(b)\right] > n \quad . \qquad (48)$$

This contradicts (ii) and (iii) so that ξ^* must maximize $|M(\xi)|$. The rest of the theorem follows easily from the concavity of $|M(\xi)|$.

The following lemmas are easily proved from Theorem 5.

Lemma 1: At all points b_i, $i = 1,\ldots,k$ of the design ξ^*,

$$\operatorname{tr}\left\{M^{-1}(\xi^*)M(b_i)\right\} = n \quad . \qquad (49)$$

Lemma 2: Let $b \equiv H = h^T$ and consider a scalar measurement case with $\Omega_h = h^T h \leq 1$. Then

$$\max_{h \varepsilon \Omega_h} \operatorname{tr}\left\{M^{-1}(\xi^*)M(h)\right\} = \lambda_{max}\left(C'(\xi^*)\right) = n \qquad (50)$$

where

$$C'(\xi^*) = \int_{t_0}^{t_1} \Phi(t,t_1)M^{-1}(\xi^*)\Phi^T(t,t_1)\sigma(t)dt \quad . \qquad (51)$$

The points of the design ξ^* are eigenvectors of unit length of $C'(\xi^*)$ corresponding to $\lambda_{max} = n$ and are equal in number to the multiplicity of this eigenvalue.

The computation of ξ^* is easily performed by an iteration similar to that for computing σ^*.

Algorithm 2:

(a) Start with a design ξ_0 such that $M(\xi_0)$ is nonsingular. Let $j = 0$.

(b) Compute $\operatorname{tr}\left\{M^{-1}(\xi_j)M(b_j)\right\}$ and find its maximum over $b_j \varepsilon \Omega_b$. If the maximum is equal to n, stop. Otherwise go to (c).

(c) Let $\xi_{j+1} = (1 - \alpha)\xi_j + \alpha_j \xi(b_j)$ (52)

where

$$0 \leq \alpha_j \leq 1 \quad , \qquad b_j \in \Omega_b$$

is such that

$$\mathrm{tr}\left\{M^{-1}(\xi_j)M(b_j)\right\} \geq \mathrm{tr}\left\{M^{-1}(\xi_j)M(b) \quad , \quad \forall b \in \Omega_b\right\} .$$

The choice of α_j is done by a search to maximize $|M(\xi_{j+1})|$ or according to the rule in Algorithm 1.

(d) Go to step (b).

B. Nonrandomized Designs (Prob. 2):

In this section, we discuss the problem considered by Johnson [2] and Müller and Weber [3]. The motivation for considering nonrandomized designs is that in many practical situations, it is not possible to employ randomized designs. The price paid for nonrandomization can be quite high as will be shown by a numerical example in the next section. Furthermore, the optimization problem becomes nonconvex leading to the appearance of local optima and severe difficulties in computation.

Since the examples given in the next section are for SISO (single-input single-output) time variant systems, we present results here only for these systems. It was shown by Johnson [2] that for this case, the optimization of the Information Matrix may be replaced by the optimization of the Observability Matrix of the following kind:

$$Q = \sum_{i=0}^{n-1} (F^T)^i H^T H F^i \quad . \tag{53}$$

We give the results here for the maximization of a general norm $m_s = \left(\frac{1}{n}\mathrm{tr}Q^s\right)^{1/s}$, $s \leq 0$ subject to the norm constraint $HH^T \leq 1$ (our results are similar to those of Müller and Weber [3] except for the correction of an error in the results of Ref.[13]). Notice that

$$\lim_{s\to 0} m_s = |Q|^{\frac{1}{n}} \tag{54}$$

$$m_{-1} = n/\mathrm{tr}(Q) \tag{55}$$

$$m_{-\infty} = \lambda_{min}(Q) \quad . \tag{56}$$

Theorem 6: Let H* be a locally optimal nonrandomized sensor design. Then $H^*H^{*T} = 1$ and the following conditions are necessary for local optimality:

(i) H* is a normalized eigenvector of

$$A(H) = \sum_{i=0}^{n-1} F^i Q^{s-1}(H)(F^T)^i \tag{57}$$

corresponding to the eigenvalue

$$\lambda^* = \mathrm{tr}\left(Q^s(H^*)\right) \quad . \tag{58}$$

(This is a nonlinear eigenvalue problem).

(ii) If λ^* is the maximum eigenvalue of A(H*), then H* minimizes the maximum eigenvalue of A(H). If λ^* is the minimum eigenvalue of A(H*), then H* maximizes the minimum eigenvalue[6] of A(H).

Proof: The property $H^*H^{*T} = 1$ follows trivially since m_s is a homogeneous function of H of degree 2. Let $h = H^T$ and define the Lagrangian function

$$L(h,\rho) = m_s(h) - \rho(h^T h-1) \tag{59}$$

where ρ is a Lagrange multiplier.

The stationary condition $\left.\frac{\partial L}{\partial h}\right|_{h^*} = 0$ and the fact that

$\frac{\partial m_s}{\partial h} = \frac{2}{nm_s^{s-1}} A(h)h$ imply

$$\left[A(h^*)-nm_s^{s-1}(h^*)\rho\right] h^* = 0 \tag{60}$$

[6] It is stated incorrectly in Ref.[3] that λ^* is always the minimum eigenvalue of A(H*). A counterexample is given in 5.1.

Premultiplying Eq. (62) by h^{*T}, we get

$$h^{*T}A(h^*)h^*-nm_s^{s-1}(h^*)\rho = 0$$

or

$$\rho = m_s(h^*) \quad . \tag{61}$$

From Eqs. (60)-(61), part (i) of Theorem 6 follows. To prove part (ii), consider

$$\begin{aligned} tr(Q^s) &= tr(Q^{s-1}Q) \\ &= tr \sum_{i=0}^{n-1} Q^{s-1}\left(F^T\right)^i hh^T F^i \\ &= tr(AB) \end{aligned} \tag{62}$$

where

$$B = hh^T \quad .$$

Since A and B are symmetric and $B \geq 0$, it is easily shown that (see Appendix A):

$$\lambda_{min}\left(A(h)\right) \leq tr(AB) = tr\left(Q^s(h)\right) \leq \lambda_{max}\left(A(h)\right) \quad . \tag{63}$$

From Eq. (65), it is clear that <u>if</u>

$$\lambda^*(A) = \lambda_{max}(A) = tr\left(Q^s(h^*)\right)$$

<u>then</u> h^* minimizes the value of $\lambda_{max}\left(A(h)\right)$. Similarly if $\lambda^*(A) = \lambda_{min}(A)$, then h^* maximizes the value of $\lambda_{min}\left(A(h)\right)$.

<u>Remark</u>: For maximizing $|Q|$, we consider $\lim s \to 0$. This gives

$$A(h) = \sum_{i=1}^{n-1} F^i Q^{-1}\left(F^T\right)^i \tag{64}$$

and

$$\lambda^*\left(A(h^*)\right) = n \quad .$$

The following algorithm may be used to compute h^*.

<u>Algorithm 3</u>:

(a) Start with an h_0 such that $Q(h_0)$ is nonsingular. Let $k = 0$.

(b) Compute $A(h_k) = \sum_{i=1}^{n-1} F^i Q^{-1}(h_k)\left(F^T\right)^i$ and find all eigenvalue-eigenvector pairs $\left(\lambda_j^k, \phi_j^k,\ j = 1,\ldots,n\right)$ of $A(h_k)$. If for some j, $\lambda_j^k = n$ and $\phi_j^k = h_k$ within specified tolerances, then h_k is a candidate for local optimality. Evaluate (numerically or analytically) $\left.\frac{\partial^2 L}{\partial h^2}\right|_{h_k}$ and check for negative- definiteness. Stop if the above conditions are satisfied. Otherwise proceed to (c). (Additional computational steps will have to be added to search for the Global Optimum).

(c) Order the eigenvalues such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and compute successively the scalar products $\left|\left(\lambda_j^k - n\right)h_k^T \phi_j^k\right|$, $j = 1,\ldots,n$. Let $j = i$ be the eigenvalue for which this product is maximized[7]. Then set

$$h_{k+1} = \frac{1}{\sqrt{c}}\left[(1 - \beta)h_k + \beta\phi_i^k\right] \quad , \tag{65}$$

where $0 < \beta \leq 1$ and

$$c = (1 - \beta)^2 + \beta^2 + 2\beta(1 - \beta)h_k^T\phi_i^k \tag{66}$$

is a normalizing factor.

---

[7] It is shown in Appendix B that the directional gradient is

$$\frac{\partial}{\partial\beta} \log|Q(h_{k+1})|\,\Big|_{\beta=0} = 2\left(\lambda_i^k - n\right)h_k^T\phi_i^k$$

where h_{k+1} satisfies Eqs. (65)-(66).

Choose β by either a one-dimensional search to maximize $Q(h_{k+1})$ or any sequence satisfying the conditions given in Algorithm 1.

(d) Go to step (b).

5. Examples

5.1 Second Order Integrator[8]

Consider the system with states $x_1 = \dot{x}$, $x_2 = x$ and the system equations $\ddot{x} = u$, or in state-vector form,

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u \quad . \tag{67}$$

Only one measurement is made continuously in time,

$$y(t) = \begin{bmatrix} h_1, h_2 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + v(t) \quad . \tag{68}$$

We consider the problem of selecting h_1 and h_2 under the constraint $(h_1^2 + h_2^2) \_ 1$. Both randomized and nonrandomized solutions will be considered and it will be shown that they are the same in this case. This is also true for the control version of the N^{th} order integrator problem considered in Refs. [2,3]. The observability matrix is

$$Q = hh^T + F^T hh^T F$$

$$= \begin{bmatrix} 1 & h_2\sqrt{1 - h_2^2} \\ h_2\sqrt{1 - h_2^2} & h_2^2 \end{bmatrix} ,$$

where $|Q| = h_2^4$;

clearly $h_2^* = \pm 1$ and $h_1^* = 0$.

---

[8]This example is given here mainly for illustrative purposes.

We first verify Theorem 6,

$$A(h) = \frac{1}{h_2^4}\begin{bmatrix} h_2^2 & -h_2\sqrt{1 - h_2^2} \\ -h_2^2\sqrt{1 - h_2^2} & 1 + h_2^2 \end{bmatrix}$$

The eigenvalues of A(h) are λ/h_2^2 where

$$\lambda = \frac{1}{2}\left[(2h_2^2 + 1) \pm \sqrt{-4h_2^4 + 4h_2^2 + 1}\right] .$$

Setting $\lambda/h_2^4 = 2$, a nonlinear equation for h_2 is obtained. The second condition to be satisfied by h_2 is

$$A(h)\begin{bmatrix} \sqrt{1 - h_2^2} \\ h_2 \end{bmatrix} = 2\begin{bmatrix} \sqrt{1 - h_2^2} \\ h_2 \end{bmatrix} .$$

It is easily shown that these conditions are satisfied only when $h_2 = \pm 1$. For this h_2

$$A(h^*) = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

so that $\lambda^* = n = 2$ is the maximum eigenvalue of $A(h^*)$.

We now consider a randomized solution in which two values of h, viz. h and h', are selected with probabilities p and (1 - p). Then

$$Q(p) = \begin{bmatrix} 1 \quad , & ph_2\sqrt{1 - h_2)^2} + (1 - p)h_2'\sqrt{1 - (h'_2)^2} \\ Q_{12} & , \quad p(h_2)^2 + (1 - p)(h_2')^2 \end{bmatrix}$$

$$= \begin{bmatrix} Q_{11}(p) & , \quad Q_{12}(p) \\ Q_{12}(p) & , \quad Q_{22}(p) \end{bmatrix}$$

$$
\begin{aligned}
|Q(p)| &= Q_{11}Q_{22} - Q_{12}^2 \\
&= \left[p(h_2) \quad + (1 - p)(h_2')^2 \right] \\
&\quad - \left[ph_2\sqrt{1 - (h_2)^2} + (1 - p)h_2' \cdot \sqrt{1 - (h_2')^2} \right]^2 .
\end{aligned}
$$

It is clear that the maximum of $|Q(p)|$ is attained for a p making the negative term zero, i.e. $h_2 = h_2' = 0, \pm 1$. To maximize the terms in the first bracket, $h_2 = h_2' = \pm 1$. Thus the optimal solution is nonrandomized.

A graphical illustration which sheds further light on the effect of randomization is shown in Fig.2. The matrix Q(p) can be represented in $\mathbb{R}^3$, but since $Q_{11}(p) = 1$, the set of all Q(p) lies on a two-dimensional plane which is depicted in Fig.2. Without randomization, the attainable values of Q_{12} and Q_{22} lie on a parabola and its mirror image, i.e. curve ABCD. As h_2 is varied from 0 to 1, the point (Q_{12},Q_{22}) moves along the parabolas from A to C. However, by randomization, all (Q_{12},Q_{22}) values inside the parabolas are attainable. Thus the set of attainable values is expanded and made convex.

The contours of constant $|Q| = Q_{11}Q_{22} - Q_{12}$ are also parabolas and it is clear from Fig.2 that the maximum of $|Q|$ is attained at point C. Since this is a boundary point which is attainable by a nonrandomized design, there is no advantage in randomization on the present case. However, we will show in the next example that in more general situations this is not the case, since the optimum of $|Q|$ may be attained in the interior of the set which is only attaibable via randomization.

5.2 Satellite Attitude Control of Optimally Located Thrust Jets

This example is taken from Müller and Weber [3], where the nonrandomized solution is computed using a nonlinear programming technique. The results presented in Section 4 are easily applied to this problem using the Duality Principle. The state vector dimension is 6, consisting of three angular positions (ϕ,ψ,θ) and three angular velocities $(\dot{\phi},\dot{\psi},\dot{\theta})$. The equations of motion are linearized for small angle deviations. The resulting F and G matrices are

$$F = \begin{bmatrix} 0, & -.1517, & 0, & -3.33, & 0, & 0 \\ .5, & 0, & 0, & 0, & -.5, & 0 \\ 0, & 0, & 0, & 0, & 0, & -1.7 \\ 1.0, & 0, & 0, & 0, & 0, & 0 \\ 0, & 1.0, & 0, & 0, & 0, & 0 \\ 0, & 0, & 1.0, & 0, & 0, & 0 \end{bmatrix}$$

$$G = \begin{bmatrix} \cos\alpha / .857 \\ \cos\beta / .2857 \\ \cos\gamma \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where α,β,γ define the direction of thrust.

By using a state transformation T = Diag.[.857,.2857,1,1,1,1], we can convert the above system to (F,G) such that $(G')^T G' = \cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$.

Table 1 summarizes the results of maximizing $|Q_C|$ where Q_C is related to the controllability matrix by

$$Q_C = \sum_{i=0}^{n-1} F^i G G^T \left(F^T\right)^i .$$

Algorithms 2 and 3 of section 4 were used to compute the randomized and nonrandomized optimal designs. It is seen that using randomization, $|Q_C|$ is increased from 627 to 2.43×10^6 and $m_0 = |Q_C|^{1/6}$ is increased from 2.93 to 11.59. A randomized design which is nearly optimal consists of four different locations to be used in the proportion (.1652,.2348,.2968,.3032).

The implementation of the randomized solution would require controllable thrust direction, but would involve no extra fuel consumption. For the nonrandomized design, the eigenvalue in contrast to the previous example where it was the largest eigenvalue of A.

Nonrandomized solution

α^* 56.91

β^* 52.48

γ^* 54.70

$$m_O^{NR} = |M|^{1/6} = 2.93$$

$$|M|_{NR} = 626.81 \quad .$$

Randomized solution:

$$m_O^R = 11.59$$

$$|M|_R = 2.4274 \times 10^6$$

$$\frac{|M|_R}{|M|_{NR}} = 3872 \quad , \quad \frac{m_O^R}{m_O^{NR}} = 3.96 \quad .$$

Table 1: Comparison of Nonrandomized and Randomized solutions to the Satellite Attitude Control Problem.

	p_1	p_2	p_3	p_4
	.1652	.2348	.2968	.3032
α^*	84.66	118.0	87.4	45.5
β^*	73.69	90.7	106.6	90.4
γ^*	98.45	146.6	89.1	125.1

6. Conclusions

A large number of different problems arising in the design of measurement systems for estimating states in linear systems are solved by using an Experimental Design approach based on the work of Kiefer, Wolfowitz and Fedorov [4,5]. The techniques are easily extended to parameter estimation using Fisher Information Matrix criteria derived in Refs.[6,7.8] on Input Design. Three computational algorithms based on certain minimax properties of the designs are given. By duality, similar results hold for the control problem. This is illustrated by a problem in Satellite Attitude Control for which both randomized and nonrandomized solutions are computed numerically.

References

[1] Meier, L., Peschon, J. and Dressler, R.M., "Optimal Control of Measurement Subsystems," IEEE Transactions on Automatic Control, October 1967.

[2] Johnson, C.D., "Optimization of a Certain Quality of Complete Controllability and Observability for Linear Dynamic Systems," ASME Transactions Journal Basic Engineering, 91, Series D, 228-23 (1969).

[3] Müller, P.C. and Weber, H.I., "Analysis and Optimization of Certain Qualities of Controllability and Observability for Linear Dynamic Systems," Automatica, Vol. 8 237-246, 1972.

[4] Kiefer, J. and Wolfowitz, K., "The Equivalence of Two Extremum Problems," Canadian Journal Mathematics, 12, 363 (1960).

[5] Fedorov, V.V., Theory of Optimal Experiments, Academic Press, New York, 1972.

[6] Mehra, R.K., "Frequency-Domain Synthesis of Optimal Inputs for Linear System Parameter Estimation." TR 645, Harvard University Division of Engineering and Applied Physics, July 1973.

[7] Mehra, R.K., "Frequency-Domain Synthesis of Optimal Inputs for MIMO Systems with Process Noise," Preprints, IEEE Decision and Control Conference, Phoenix, Arizona, 1974.

[8] Mehra, R.K., "Optimal Input Signals for Parameter Estimation in Dynamic Systems - Survey and New Results," IEEE Transactions on Automatic Control, December 1974.

[9] Kalman, R.E. and Bucy, R.S., "New Results in Linear Filtering and Prediction Theory," J. Basic Engineering, 83, 95-108, 1961.

[10] Bucy, R.S. and Joseph, P.D., Filtering for Stochastic Processes with Applications to Guidance, Interscience Publishers, John Wiley and Sons, New York, 1968.

Appendix A

Lemma: Let A and B be n×n symmetric matrices $B \geq 0$. Then

$$\lambda_{min} tr(B) \leq tr(AB) \leq \lambda_{max}(A) tr(B) \quad .$$

Proof: Let P b e an orthogonal matrix with eigenvectors of A as its columns. Then

$$PAP^T = Diag\left[\lambda_1, \ldots, \lambda_n\right]$$

$$tr(AB) = tr(PAP^T PBP^T) = \sum_{i=1}^{n} \lambda_i (PBP^T)_{ii} \quad .$$

Since $(PBP^T)_{ii} - 0$,

$$\lambda_{min} tr(B) \leq tr(AB) \leq \lambda_{max}(A) TR(PBP^T) = \lambda_{max}(A) tr(B) \quad ,$$

where λ_{max} is the maximum eigenvalue of A, and λ_{min} is the minimum eigenvalue of A.

Appendix B: Computation of the Directional Gradient (Algorithm 3)

In this appendix, we calculate the gradient of $|Q|$ with respect to β when h is updated according to the iterative scheme of Eq.(67):

$$h_{k+1} = \left[(1 - \beta) h_k + \beta \phi_j^k \right] / \sqrt{c} \quad ,$$

where $0 \leq \beta \leq 1$ and ϕ_j^k is a normalized eigenvector of $A(h_k) = \sum_{i=0}^{p-1} F^i Q^{-1}(h_k)(F^T)^i$ corresponding to the eigenvalue λ_j^k.

From Eq.(3)

$$Q(h_{k+1}) = \frac{1}{c} \sum_{i=0}^{p-1} (F^T)^i \Big[(1 - \beta^2) h_k h_k^T + \beta^2 \phi_j^k \phi_j^{k^T}$$

$$+ \beta(1 - \beta)(h_k \phi_j^{k^T} + \phi_j^k h_k^T) \Big] F^i$$

$$\frac{\partial}{\partial \beta} \log|Q(_{k+1})| = \operatorname{tr} \left[Q^{-1}(h_{k+1}) \frac{\partial Q(h_{k+1})}{\partial \beta} \right]$$

$$= \operatorname{tr} \Big\{ Q^{-1}(h_{k+1}) \Big[\frac{1}{c} \sum_{i=1}^{p-1} (F^T)^i \, -2(1 - \beta) h_k h_k^T$$

$$+ 2\beta \phi_j^k \phi_j^{k^T} + (1 - 2\beta)(h_k \phi_j^{k^T} + \phi_j^k h_k^T) \; F^i \Big]$$

$$- \frac{1}{c} \frac{\partial c}{\partial \beta} Q(h_{k+1}) \Big\} \quad .$$

For $\beta = 0$, $h_{k+1} = h_k$, $c = 1$ and $\frac{\partial c}{\partial \beta} = -2 + 2\phi_j^{k^T} h_k$

$$\frac{\partial}{\partial \beta} \log|Q(h_{k+1})|_{\beta=0} = 2(\lambda_j^k - n)\phi_j^{k^T} h_k \quad .$$

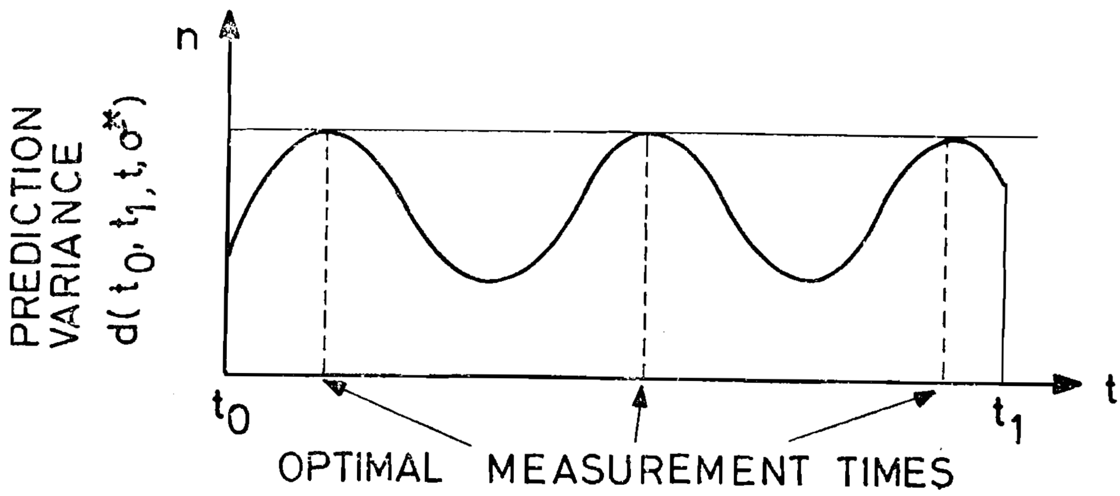

FIG.1: PLOT OF PREDICTION VARIANCE.

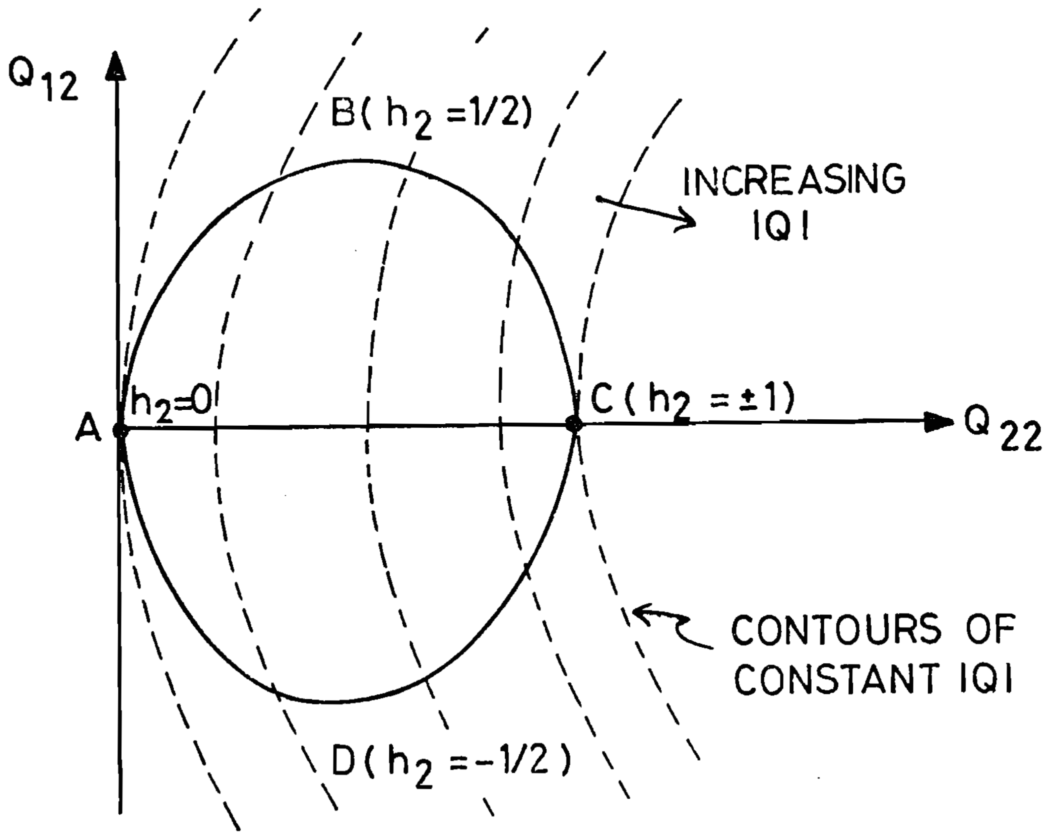

FIG. 2 : GRAPHICAL ILLUSTRATION OF THE SOLUTION TO THE SECOND ORDER INTEGRATOR PROBLEM.