

Optimal Mean Squared Error Imaging

Suman Chakravorty* Pierre T. Kabamba † David C. Hyland‡
Department of Aerospace Engineering
University of Michigan, Ann Arbor

May 31, 2002

Abstract

The problem of forming images that are optimal with respect to a Mean Square Error (MSE) criterion, based on finite data, is considered. First, it is shown that the MSE criterion is consistent with the general goal of classifying images, in that decreasing the MSE guarantees a decrease in the probability of misclassifying an image. The problem of choosing sampling locations for image formation that optimize the MSE is then formulated. It is shown that this MSE minimization problem has a solution under certain conditions and necessary conditions for a minimum are obtained. The results are illustrated on a simple image formation problem.

1 Introduction

This paper is motivated by the goal of taking high resolution images of exo-solar planets at distances of up to 15 parsecs. The images that are obtained would be used to answer true false questions such as ‘ Is the percentage of oxygen in the planet’s atmosphere greater than 10 ?’. We formulate an optimal imaging problem keeping the above objective in mind. We require that the imaging scheme be such that it minimizes the probability of misclassifying an image, given that the set of images has been partitioned into two non-empty classes. We show that, under certain assumptions, minimizing the mean-squared error in the estimated image results in the image being optimal with respect to the classification goals. We limit ourselves to the case when we can only take a finite number of measurements. In this case we postulate that we have a priori information that the

image belongs to a finite dimensional subspace of the original infinite dimensional space. This allows us to obtain an unbiased estimate of the image from a finite number of observations. Then we pose an optimization problem with respect to the mean squared error of the estimated images that result from different choices of the measurement/vantage points. We show that this optimization problem, under certain conditions, has a solution.

The process of light propagation between two parallel planes in three-dimensional space is described as a spatial convolution of the electromagnetic field of the object (source of the radiation) with the Huygens-Fresnel (H-F) operator^{1,2}. Though this is not explicitly stated, it is easily shown that the H-F operator is a unitary operator. Hence if we measured the field due to the object at some plane downstream from it, the image of the object is reconstructed by simply convolving the measured field with the inverse of the H-F operator (i.e. its adjoint). The problem is well-posed since unitary operators are all-pass filters the resulting deconvolution problem is trivial. However all optical instruments can only measure the field over a finite region. In this case the image formed by the optical instrument is represented by a spatial convolution of the object field with the point-spread function (psf) of the optical instrument³. The psf of an optical instrument is a low-pass filter and hence the problem of deconvolution in this case is ill-posed³. This problem has been widely studied in the image processing community and is solved using regularization methods³⁻⁵. In some applications, the point-spread function of the optical instrument being used might not be known perfectly. These problems are treated as blind-deconvolution problems where the psf of the instrument and the image are estimated simultaneously^{6,7}. Thus the problem is under-determined in the sense that the measure of the support of the image is much larger than the measure of the measurement region. This problem

*Graduate Student Research Assistant

†Professor

‡Professor

¹ Copyright © 2002 The American Institute of Aeronautics and Astronautics, Inc. All rights reserved.

can be overcome by having apriori knowledge that allows us to constrain the image to be in a class of functions with the property that any function in the class can be determined perfectly (in the absence of noise) by taking measurements in a region of arbitrarily small measure. Under conditions of bandlimitedness/ analyticity on the image, the image can be reconstructed perfectly over the entire infinite plane by taking measurements on a countable set/set of arbitrarily small measure^{1,8}. If we have further apriori information that the image belongs to a certain finite dimensional subspace, then the image can be reconstructed perfectly by taking a finite number of observations. The problem of finding the best orthonormal basis to represent such a class of images has been studied extensively and has a counterpart in finding best rational orthonormal basis in controls⁹⁻¹².

In medical imaging research, the Mean Squared Error (MSE) has been used as the criterion for posing a variety of static and sequential observation selection problems¹³⁻¹⁵. The problem in the case of medical imaging very closely resembles the MSE minimization that we consider in this paper. However our problem is generalized to the case of infinite dimensional spaces. Also we justify the use of the MSE as an optimization criterion by showing that it is consistent with the goal of classifying an image correctly. Also note that the problem of optimal imaging is similar to the optimal sensor placement problem in various other applications such as large flexible structures, chemical processes and power systems¹⁶⁻¹⁹.

The rest of the paper is organized as follows. In section 2 we formulate the optimal imaging problem according to the classification goals. In section 3, we show that the MSE is consistent with the classification goals. In section 4 we formulate the problem of obtaining optimal vantage points with respect to the MSE. In section 5 we give a simple example to illustrate the concepts presented in the paper.

2 Imaging Problem Formulation

In three-dimensional space, consider two parallel planes as shown in Figure(1). The process of light transmission between these planes is represented mathematically as a spatial convolution $i_m = h * i_o$ where i_o is the electromagnetic field emanating from the object in the object plane, i_m is the field on the

measurement plane, and the operator h is unitary and known. The imaging problem is then to reconstruct i_o based on the noise corrupted knowledge of i_m over a finite region of the measurement plane (i.e. we don't have measurements of i_m over the whole infinite plane) in an 'optimal fashion' i.e. solve the following equation for i_o

$$i_m = A(h * i_o) + n, \quad (2.1)$$

where A is defined as follows:

$$A(\bar{P}) = \begin{cases} 1 & \text{if } \bar{P} \in S \\ 0 & \text{otherwise} \end{cases}$$

S denotes the finite region over which the measurements are taken and \bar{P} represents some point on the measurement plane.

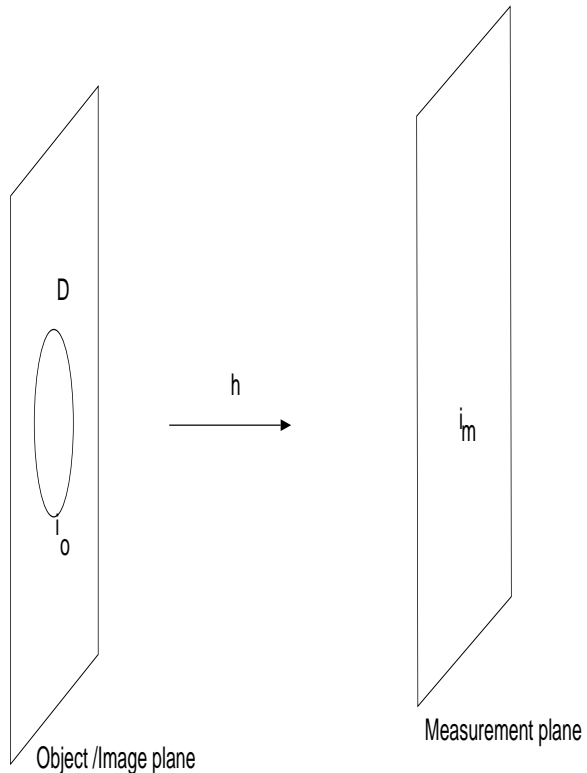


Figure 1: Process of Image Formation

The solution of eq.(2.1) is to be optimal in the sense of minimizing the probability of misclassifying the image as will be defined next. Assume that the set of images \mathcal{I} has been partitioned into two non-empty subsets W_1, W_2 i.e.

$$\begin{aligned} \mathcal{I} &= W_1 \cup W_2 \\ W_1 \cap W_2 &= \phi \end{aligned}$$

Let \hat{i}_o be an estimate of i_o from eq.(2.1). Then

Definition 2.1 *The misclassification error $\epsilon(\hat{i}_o)$ for an estimate \hat{i}_o of object i_o is defined to be*

$$\epsilon(\hat{i}_o, i_o) = \text{prob}(\hat{i}_o \notin W_k / i_o \in W_k) \quad (2.2)$$

where W_k denotes class k .

Hence the estimate \hat{i}_o of the object should be such that it minimizes the misclassification error of the object into one of the two classes I/II , i.e.,

$$\hat{i}_o^* = \arg \min_{\hat{i}_o} \epsilon(\hat{i}_o, i_o) \quad (2.3)$$

In general, the misclassification error would be extremely difficult to evaluate from the estimated image and hence we need to find an error metric that could easily be evaluated from the estimated image and that is consistent with the classification goal as formulated above. We could frame this mathematically as finding an error metric between the actual(i) and estimated image(\hat{i}), $\|i - \hat{i}\|$ such that

$$\arg \min_{\hat{i}} \|i - \hat{i}\| = \arg \min_{\hat{i}} \epsilon(\hat{i}, i) \quad (2.4)$$

Given this error metric, the problem of finding an optimal estimate would be to find \hat{i}_o^* such that

$$\hat{i}_o^* = \arg \min_{\hat{i}_o} \|i - \hat{i}_o\|$$

Thus the imaging problem posed is two-fold:

P 1 *Find an error metric that is consistent with goal of classifying a given image into one of several possible classes.*

P 2 *Find the optimal estimate of the image w.r.t the above metric*

In Section3 we answer Problem P1 and show that the Mean Squared Error criterion is an error metric that is consistent with the goals of image classification. In section4 we formulate the problem of finding the optimal vantage points with respect to the MSE of the estimated image.

3 Optimal Imaging : The Case of Linear / Isotropic features

In this section we show that the mean squared error/weighted-mean squared error satisfies the goal of image classification as previously defined,

under certain conditions. The problem of optimal imaging is stated as follows :

Let i denote a fixed but unknown image. Let $\hat{i}(\theta)$ denote a family of estimates of i parametrized by $\theta \in S$. Find

$$\theta^* = \arg \min_{\theta \in S} \epsilon_p(\hat{i}(\theta), i).$$

Consider the following assumptions,

A 3.1 *The noise corrupting the measurements, e , is a zero-mean, and quadratic-mean-continuous random process*

A 3.2 *We are given a feature extraction function $F : L_2(D) \rightarrow \mathfrak{R}$ and $x_0 \in \mathfrak{R}$ such that*

$$\begin{aligned} F(i) \leq x_0 &\iff i \in \mathcal{I}_1, \\ F(i) > x_0 &\iff i \in \mathcal{I}_2. \end{aligned} \quad (3.1)$$

A 3.3 *The noise process corrupting the measurements is Gaussian.*

Next we present the following result.

Proposition 3.1 *Suppose assumptions A3.1, A 3.2, A 3.3 hold. Let \hat{i}_1 and \hat{i}_2 be noisy estimates of i such that $E|F(i) - F(\hat{i}_1)|^2 \leq E|F(i) - F(\hat{i}_2)|^2$. Let F be a linear map. Then $\epsilon_p(\hat{i}_1, i) \leq \epsilon_p(\hat{i}_2, i)$.*

Proof:

By Assumption A3.2, there exists $x_0 \in \mathfrak{R}$ such that $F(i) < x_0 \Rightarrow F(i) \in \mathcal{I}_2$ and $F(i) > x_0 \Rightarrow F(i) \in \mathcal{I}_1$. Let $\delta i_1 = i - \hat{i}_1$ and $\delta i_2 = i - \hat{i}_2$. Since $\delta i_1, \delta i_2$ satisfy A3.3 and F is linear, $(F(i) - F(\hat{i}_{1,2}))$ are gaussian random variables. Note that by Assumption A3.1, the estimates \hat{i}_1 and \hat{i}_2 are unbiased and hence we can set

$$F(\hat{i}_1) \equiv N(F(i), \sigma_1), \quad (3.2)$$

$$F(\hat{i}_2) \equiv N(F(i), \sigma_2), \quad (3.3)$$

where by hypothesis we have that $\sigma_1 \leq \sigma_2$. Let $i \in \mathcal{I}_2$.

$$\begin{aligned} \epsilon_p(\hat{i}_1, i) &= \text{prob}(F(\hat{i}_1) \notin \mathcal{I}_2) = \text{prob}(F(\hat{i}_1) > x_0) \\ &= \text{prob}\left(\frac{F(\hat{i}_1) - F(i)}{\sigma_1} > \frac{x_0 - F(i)}{\sigma_1}\right). \end{aligned} \quad (3.4)$$

Noting that $\frac{F(\hat{i}_1) - F(i)}{\sigma_1} \equiv N(0, 1)$,

$$\epsilon_p(\hat{i}_1, i) = \text{prob}(N(0, 1) > \frac{x_0 - F(i)}{\sigma_1}). \quad (3.5)$$

$$(3.6)$$

Similarly,

$$\epsilon_p(\hat{i}_1, i) = \text{prob}(N(0, 1) > \frac{x_0 - F(i)}{\sigma_2}). \quad (3.7)$$

Noting that $x_0 > F(i)$ and $\sigma_1 \leq \sigma_2$, it follows from above that $\epsilon_p(\hat{i}_1, i) \leq \epsilon_p(\hat{i}_2, i)$.

The case when $i \in \mathcal{I}_1$ can be treated similarly.

Q.E.D

Hence, from the above proposition we can conclude that if A3.1-A3.3 are satisfied by the estimates, then the optimal imaging problem reduces to finding θ^* where

$$\theta^* = \arg \min_{\theta \in \mathcal{S}} E|F(i) - F(\hat{i}(\theta))|^2.$$

Hence, to choose the optimal estimate we need to evaluate the quantity $E|F(i) - F(\hat{i}(\theta))|^2$ for the various estimates. It is possible to evaluate the above for the case of a linear map. In fact,

$$E|F(i) - F(\hat{i}(\theta))|^2 = \int_D \int_D f(t)R_{\delta i(\theta)}(t, s)\bar{f}(s)dt ds, \quad (3.8)$$

where $f \in L_2$ is the unique representation of the linear operator F^{20} . Thus, in the case of a linear feature map, evaluating

$$\theta^* = \arg \min_{\theta \in \mathcal{S}} \int_D \int_D f(t)R_{\delta i(\theta)}(t, s)\bar{f}(s)dt ds,$$

results in an optimal estimate in the sense that the misclassification error is minimized.

In the following we show that if a feature map satisfies the property of "local isotropy" then minimizing the MSE of the estimates results in an optimal estimate.

Definition 3.1 A functional $g : X \rightarrow R$ is called 'locally isotropic' if it is differentiable throughout X and, given $x \in X$, $|\delta g(x; h)|$ is a constant for all $h \in X$ where $|\delta g(x; h)|$ represents the frechet deruvative of the functional g at x with increment h .

A 3.4 The feature map F is locally isotropic and linear.

Then we have the following proposition

Proposition 3.2 Suppose assumptions A3.1, A3.2, A3.4 hold. If $E\|i - \hat{i}_1\|^2 \leq E\|i - \hat{i}_2\|^2$, then it follows that $E|F(i) - F(\hat{i}_1)|^2 \leq E|F(i) - F(\hat{i}_2)|^2$, the terms $F(\hat{i}_1)$ and $F(\hat{i}_2)$ being interpreted as quadratic mean integrals.

Proof: Let us expand the various terms involved in the above expression.

$$E\|i - \hat{i}_1\|^2 = E \int_0^T |\delta i_1(t)|^2 dt = \int_0^T R_1(t, t) dt. \quad (3.9)$$

By Assumption A3.1, δi_1 is quadratic mean continuous, $R_1(t, t)$ is continuous and hence the integral $\int_0^T R_1(t, t) dt$ exists. Hence

$$E\|i - \hat{i}_1\|^2 = \int_0^T R_1(t, t) dt, \quad (3.10)$$

$$E\|i - \hat{i}_2\|^2 = \int_0^T R_2(t, t) dt. \quad (3.11)$$

By Mercer's theorem ²¹, any covariance function $R(t, s)$ can be represented as

$$R(t, s) = \sum_{n=1}^{\infty} \lambda_n \varphi_n(t) \bar{\varphi}_n(s), \quad (3.12)$$

where $\{\varphi_n(t)\}_{n=1}^{\infty}$ is orthonormal in $L_2[0, T]$ and $\{\lambda_n\}$ are real and positive. Hence it follows from (3.10), (3.12) that

$$\int_0^T R_1(t, t) dt = \sum_{n=1}^{\infty} \lambda_n^{(1)}, \quad (3.13)$$

$$\int_0^T R_2(t, t) dt = \sum_{n=1}^{\infty} \lambda_n^{(2)}. \quad (3.14)$$

By hypothesis, we have

$$\sum_{n=1}^{\infty} \lambda_n^{(1)} \leq \sum_{n=1}^{\infty} \lambda_n^{(2)}. \quad (3.15)$$

Also, we have

$$\begin{aligned} E|F(i) - F(\hat{i}_1)|^2 &= E \left(\int_0^T f(t) \delta i_1(t) dt \int_0^T f(s) \delta i_1(s) ds \right) \\ &= \int_0^T \int_0^T f(t) R_1(t, s) \bar{f}(s) dt ds \end{aligned} \quad (3.16)$$

Again by Mercer's theorem it follows that

$$\begin{aligned} E|F(i) - F(\hat{i}_1)|^2 &= \int_0^T \int_0^T f(t) \left(\sum_{n=1}^{\infty} \lambda_n^{(1)} \varphi_n(t) \bar{\varphi}_n(s) \right) \bar{f}(s) dt ds \\ &= \sum_{n=1}^{\infty} \lambda_n^{(1)} \left(\int_0^T \int_0^T f(t) \varphi_n^{(1)}(t) \bar{\varphi}_n^{(1)}(s) \bar{f}(s) dt ds \right) \\ &= \sum_{n=1}^{\infty} \lambda_n^{(1)} \left(\int_0^T f(t) \varphi_n^{(1)}(t) dt \right)^2 \end{aligned} \quad (3.17)$$

i.e.,

$$E|F(i) - F(\hat{i}_1)|^2 = \sum_{n=1}^{\infty} \lambda_n^{(1)} |\delta F(i; \varphi_n^{(1)})|^2, \quad (3.18)$$

$$E|F(i) - F(\hat{i}_2)|^2 = \sum_{n=1}^{\infty} \lambda_n^{(2)} |\delta F(i; \varphi_n^{(2)})|^2. \quad (3.19)$$

By the local isotropy of F and (3.15) it follows that

$$\begin{aligned} E|F(i) - F(\hat{i}_1)|^2 &= \sum_n \lambda_n^{(1)} K \\ &\leq \sum_n \lambda_n^{(2)} K = E|F(i) - f(\hat{i}_2)|^2, \end{aligned} \quad (3.20)$$

where $K = \delta F(i; h)$.

Q.E.D

In view of the above result, we have that if the feature map F was "linear-isotropic", then finding

$$\theta^* = \arg \min_{\theta \in S} \|i - \hat{i}(\theta)\|_2$$

results in an optimal estimate in that the misclassification error is minimized.

4 Optimal Image Acquisition : Finite Number of Observations

In this section we formulate the problem of optimal observation selection for imaging, given that only a finite number of observations can be made. We utilize the results of the previous section in order to show optimality of the vantage points. In this section we shall be assuming that the feature involved is linear or isotropic. We require that the image be perfectly determined through a finite number of observations in the absence of noise. In the presence of noise, the optimal choice of observations would be those that minimize the effect of noise on the synthesized image.

Definition 4.1 We define an N -sampling domain as a set $S = \{t_1, t_2, \dots, t_N\}$ where $\forall i, t_i \in \mathcal{R}$.

We denote by i/S the set of the sample values of the function i on the set S . We make the following assumption.

A 4.1 Let $i \in \text{span}[\phi_1, \dots, \phi_N]$ where $\{\phi_1, \dots, \phi_N\}$ is a known orthonormal set of vectors.

Suppose we take noise corrupted measurements of Hi at $S = \{t_1, \dots, t_N\}$. We need to estimate i from

these observations (note that the problem here is just a particular case of the one defined in eq(2.1)). By the linearity of H and the A4.1, it follows that $Hi \in \text{span}[H\phi_1, \dots, H\phi_N]$. Let $H\phi_k = \psi_k$. We obtain the estimate of the image \hat{i} , in the following steps:

Step 1 First we form an estimate of $Hi, \hat{H}i_S$, as:

$$\hat{H}i_S(t) = [\psi_1(t), \dots, \psi_N(t)] \Psi_S^{-1} (Hi/S + \epsilon), \quad (4.1)$$

where

$$\Psi_S \equiv \begin{bmatrix} \psi_1(t_1) & \psi_2(t_1) & \dots & \psi_n(t_1) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(t_N) & \psi_2(t_N) & \dots & \psi_n(t_N) \end{bmatrix},$$

assuming that Ψ_S is invertible and

$$\epsilon = \begin{bmatrix} \epsilon(t_1) \\ \vdots \\ \epsilon(t_N) \end{bmatrix},$$

$\epsilon(t_i)$ being the noise corrupting the observation at t_i .

Step 2 Then we obtain the estimate of i, \hat{i} , by setting $\hat{i} = H^{-1} \hat{H}i_S$.

It can be shown that since H is a unitary operator,

$$E\|i - H^{-1} \hat{H}i_S\|^2 = E\|Hi - \hat{H}i_S\|^2, \quad (4.2)$$

We assume that the noise corrupting the measurement at t_k is zero-mean with variance $\sigma^2(t_k)$ and that it is independent of the noise corrupting the measurements at any other point.

It can be shown that the value of the MSE in taking measurements on the sampling set $S = \{t_1, \dots, t_n\}$ is given by

$$e_2(S) = \text{trace}(\Psi_S^{-1} R_\epsilon^S \Psi_S^{-1*}), \quad (4.3)$$

where Ψ_S is as previously defined ,

$$R_\epsilon^S = \text{diag}(\sigma^2(t_1), \dots, \sigma^2(t_N)), \quad (4.4)$$

and $(.)^*$ denotes the complex conjugate transpose of a matrix.

Similarly it can be shown that the value of $E|F(i) - F(\hat{i})|^2$ is given by

$$e_f(S) = \text{trace}(\Psi_S^{-1} R_\epsilon^S \Psi_S^{-1*} G),$$

where

$$G = FF^*,$$

and

$$F = [\langle f, \phi_1 \rangle, \dots, \langle f, \phi_N \rangle].$$

Let S be constrained to remain in some $D \subset \mathfrak{R}^N$. Then the problem of optimal imaging can be stated as find

$$S^* = \arg \min_{S \in D} e_2(S), \quad (4.5)$$

for the case of an isotropic feature, and as find

$$S^* = \arg \min_{S \in D} e_f(S), \quad (4.6)$$

for the case of a linear feature.

In the following proposition we show the existence of a solution to the above optimization problem

Proposition 4.1 *If the basis functions $\{\psi_1, \dots, \psi_N\}$ are continuous and the constraint set D is compact then the functionals e_2 and e_f attain their minimum on D .*

Proof:

By the assumption that the functions $\{\psi_1, \dots, \psi_n\}$ are orthonormal, it follows that there exists some $S_0 \in D$ s.t. $e_2(S_0) < \infty$. Consider the set $F = \{S | e_2(S) \leq \epsilon(S_0)\}$. The function e_2 is continuous at all points in F since the only points of discontinuity of the functional e_2 is where it is unbounded. Hence by the definition of continuous functions it follows that F is closed and also compact (since it is a subset of D). Thus e_2 attains its minimum on F and hence in D . The case for e_f can be treated similarly.

Q.E.D.

It can be shown that the above optimization problems can be framed as constrained optimization problems with inequality constraints (the set constraint is treated as the inequality constraint). The following are first order necessary conditions for a relative minimum ²²

$$\begin{aligned} \nabla \epsilon(S^*) + \mu^t \nabla \bar{g}(S^*) &= 0, \\ \mu^t \bar{g}(S^*) &= 0, \mu \geq 0, \\ &\text{where} \\ \frac{\partial \epsilon}{\partial t_i} &= 2\sigma^2 \text{trace}(\Psi^{-1} \frac{\partial \Psi}{\partial t_i} \Psi^{-1} \Psi^*), \end{aligned} \quad (4.7)$$

$$\bar{g}(S) = \begin{bmatrix} a_1 & t_1 \\ \vdots & \vdots \\ a_n & t_N \end{bmatrix},$$

assuming that the noise is independent identically distributed, i.e., $R_\epsilon^S = \sigma^2 I$ and $t_i \geq a_i \forall i = 1, \dots, N$.

Remarks:

1) It can be shown that the choice of the optimal vantage points is independent of the choice of the basis vectors.

2) If we consider the case when the basis consists of only one vector, the optimization problem is to choose one point such that the MSE is minimized. It turns out that this optimal vantage point is precisely that point where the ‘ signal to noise ratio ’ of the function to the noise is maximum.

3) The error criterion looks like a generalized signal to noise ratio .

5 Illustrative Example

In this section we present a simple numerical example that illustrates the problem formulation that was presented in the previous section. The example presented here is a finite-dimensional case. In this finite dimensional setting it can be seen that a feature cannot be isotropic. (However note that it can be shown that isotropic maps do exist in general). In this example we consider a linear feature.

In Fig.2, the topmost plot represents the image at the object plane. It is a discretized image with four pixels on each side. The bottom plot in the same figure represents the image at the measurement plane. The feature considered in this example is the sum of the pixel values of the image. It is assumed that the image lies in class I if the sum is less than a threshold value of 4.1 and it belongs to class II otherwise. The sum of the pixel values of the object as shown in Fig.7 is 4 and hence the object belongs to class I.

We assume that it is known that the support of the object is constrained to the central four pixels of the image. Note that the image can be represented as a vector if we stack the rows of the image lexicographically . In doing this, the image is represented as an element in \mathfrak{R}^{16} . It then follows that the knowledge that the support is constrained to the central 4 pixels of the picture translates into saying that the basis of the image is $\{e_6, e_7, e_{10}, e_{11}\}$ where e_i denotes the i^{th} co-ordinate vector in \mathfrak{R}^{16} . Hence, the basis for the image at the measurement plane is $\{He_6, He_7, He_{10}, He_{11}\}$. These basis images are shown in Fig.3.

The problem of optimal imaging with respect

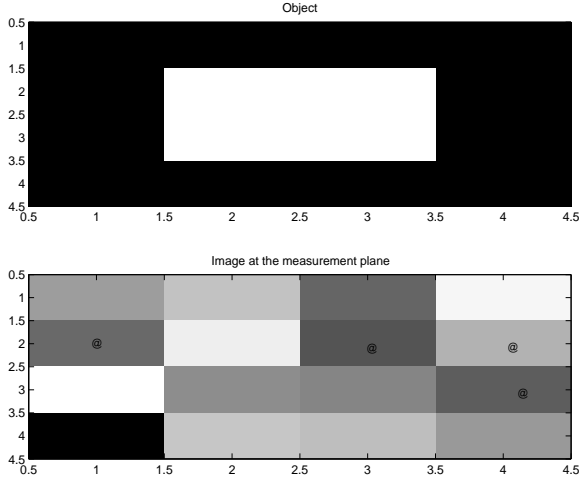


Figure 2:

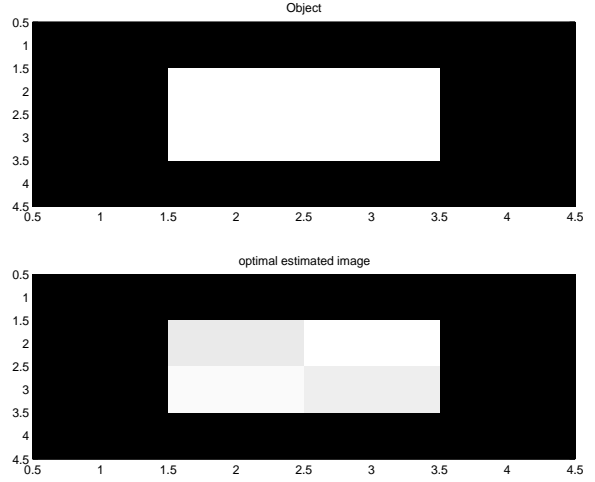


Figure 4:

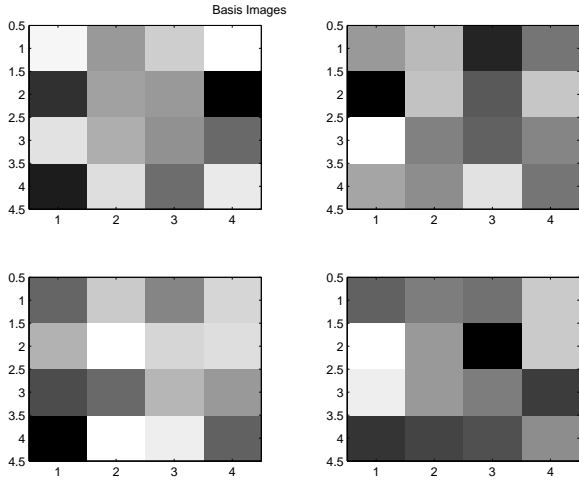


Figure 3:

to the weighted-MSE is to choose the four pixels of the image at the measurement plane at which to make measurements so that the weighted-MSE in estimating the image is minimized, assuming that the statistics of the noise corrupting the measurements is known. We assume that the noise corrupting the measurements is i.i.d with Gaussian statistics. The variance of the noise is assumed to be 0.05. In Fig.9 we show the object and the optimal estimated image. The optimal locations of the measurements on the measurement plane are represented by the '@' in Fig.4.

In Fig.5 we show a comparison plot of the weighted-MSE and the misclassification error . Every point on the X-axes in the plots represents a

particular choice for the measurement points on the measurement plane. We obtained that the optimal weighted-MSE estimate and the optimal ME estimate were the same. However there were cases for which the above did not hold true in the simulations . (The different cases were obtained by different choices of the transformation between the image and the measurement plane). The misclassification errors for this example were obtained through Monte-Carlo simulations and would account for the discrepancy between the minimum weighted-MSE estimate and the minimum ME estimate in these cases. In closing we would like to say that in a significant number of cases the min weighted-MSE estimate exactly matches the minimum ME estimate.

6 Conclusions

An optimal imaging problem was formulated such that the solution minimizes the probability of misclassifying the image. It was shown that if the classification was done using a given feature map and if the feature map was isotropic, the MSE criterion is consistent with the optimal imaging problem. The problem of obtaining optimal vantage points was then formulated with respect to the MSE criterion. It was shown that the optimization problem has a solution under the condition that the observations be constrained to a compact set. The results were illustrated through a simple finite dimensional example. It was noted that the isotropy of the feature map may be a restrictive condition. It will be our future endeavour to relax this condition. We note that the MSE criterion is one particular topology on

the image estimates under which the goal of optimal imaging is satisfied. However there might be other error criteria that could be defined under which the above goal could still be satisfied. The relationship of the MSE topology with the other topologies, if they exist, remains to be explored.

References

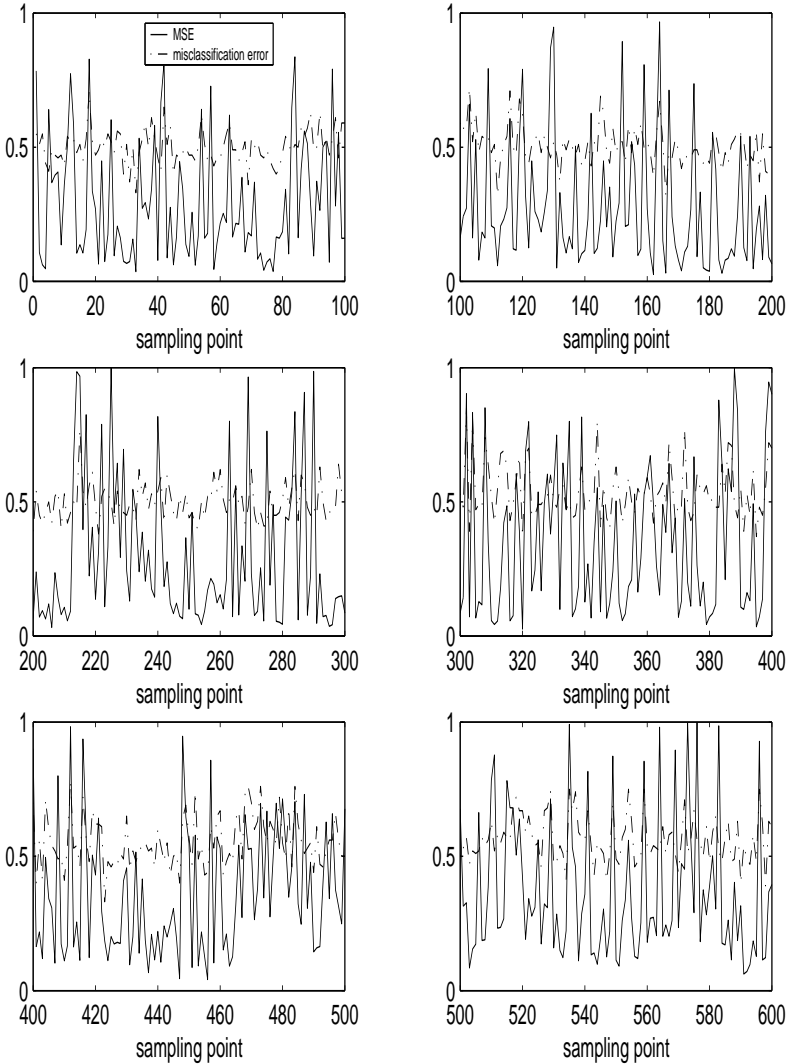


Figure 5:

- [1] J.W.Goodman, Introduction to Fourier Optics, Boston,MA, Mcgraw Hill,1996
- [2] J.W.Goodman, Statistical Optics, Wiley Classics Library,2000
- [3] G Demoment, Image Reconstruction and Restoration: Overview of Common Estimation Structures and Problems, IEEE Transactions on Acoustics,Speech and Signal Processing, vol.37,pp2024-2036,1989
- [4] A.K.Katsagellos, Iterative Image Restoration Algorithms, Optical Engineering, vol.28 n.7, pp735-748,1989
- [5] A.K.Katsagellos(ed.), Digital Image Restoration, Springer-Verlag,New York, 1991
- [6] M.Cannon, Blind Deconvolution of Spatially Invariant Image Blurs with Phase, IEEE Transactions on Acoustics,Speech and Signal Processing, vol.24,pp58-63,1976
- [7] D. Kundur, D. Hatzinakos, A Novel Blind Deconvolution Scheme for Image Restoration using Recursive Filtering, IEEE Transactions Signal Processing, vol.46 n.2, 1998
- [8] R.J.Marks, Introduction to Shannon Sampling and Interpolation theory, Springer-Verlag, New York, 1991
- [9] N.Saito, Image Approximation and Modeling via Least Statistically Dependent Bases, Pattern Recognition, vol.34, pp1765-1784, 2001
- [10] I.F.Gorodnitsky, B.D.Rao, Sparse Signal Reconstruction from Limited Data using FO-CUSS : A Re-weighted Minimum Norm Algorithm, IEEE Transactions on Signal Processing, vol.45 n.3, 1997
- [11] P. Bodin, L.F.Villemoes, B.Wahlberg, Selection of Best Orthonormal Rational Basis, SIAM J. Control Optim., vol.38 n4, pp995-1032,2000

- [12] Y.Cao, D.N.Levin, Using an Image Database to Constrain the Acquisition and Reconstruction of MR Images of the Human Head, *IEEE Transactions on Medical Imaging*, vol.14 n.2,pp.350-361, 1995
- [13] S.J.Reeves, Sequential Algorithms for Observation Selection, *IEEE Transactions on Signal Processing*, vol.47 n.1,pp.123-131,1999
- [14] Y.Gao, S.J.Reeves, Optimal K-Space Sampling in MRSI for Images with a Limited Region of Support, *IEEE Transactions on Medical Imaging*, vol.19 n.12,pp.1168-1178,2000
- [15] S.K.Nagle, D.N.Levin, Multiple Region MRI, *Magnetic Resonance in Medicine*, vol.41, pp.774-786,1999
- [16] R.E.Skelton, D.Chiu, Optimal Selection of Inputs and Outputs in Linear Stochastic Systems, *The Journal of Astronautical Sciences*, vol.XXXI n.3, pp.399-414,1983
- [17] P.G.Maghani, S.M.Joshi, Sensor/Actuator Placement for Flexible Space Structures, *IEEE Transactions on Aerospace and Electronic Systems*, vol.29 n.2,pp.345-351,1993
- [18] W.H.Chen, J.H.Seinfeld, Optimal Location of Process Measurements, *Optimal Location of Process Measurements, Int. J. Control*, vol.21 n.6, pp. 1003-1014, 1975
- [19] J.E.Farach, W.M.Grady, A.Arapostathis, An Optimal Procedure for Placing Sensors and Estimating the Locations of Harmonic Sources in Power Systems, *IEEE Transactions on Power Delivery*, vol.8 n.3,pp.1303-1310,1993
- [20] D.G.Luenberger, *Optimization by Vector Space Methods*, Wiley, New York, 1968
- [21] E.Wong, B.Hajek, *Stochastic Processes in Engineering Systems*, Springer-Verlag, New York, 1985
- [22] D.G.Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley,Reading,MA, 1984