

Optimal Measurement under Cost Constraints for Estimation of Propagating Wave Fields

Ayça Özçelikkale, Haldun M. Ozaktas, and Erdal Arıkan

Department of Electrical Engineering, Bilkent University, TR-06800, Ankara, Turkey

Email: ayca, haldun, arikan@ee.bilkent.edu.tr

Abstract—We give a precise mathematical formulation of some measurement problems arising in optics, which is also applicable in a wide variety of other contexts. In essence the measurement problem is an estimation problem in which data collected by a number of noisy measurement probes are combined to reconstruct an unknown realization of a random process $f(x)$ indexed by a spatial variable $x \in \mathbb{R}^k$ for some $k \geq 1$. We wish to optimally choose and position the probes given the statistical characterization of the process $f(x)$ and of the measurement noise processes. We use a model in which we define a cost function for measurement probes depending on their resolving power. The estimation problem is then set up as an optimization problem in which we wish to minimize the mean-square estimation error summed over the entire domain of f subject to a total cost constraint for the probes. The decision variables are the number of probes, their positions and qualities. We are unable to offer a solution to this problem in such generality; however, for the *metrical problem* in which the number and locations of the probes are fixed, we give complete solutions for some special cases and an efficient numerical algorithm for computing the best trade-off between measurement cost and mean-square estimation error. A novel aspect of our formulation is its close connection with information theory; as we argue in the paper, the mutual information function is the natural cost function for a measurement device. The use of information as a cost measure for noisy measurements opens up several direct analogies between the measurement problem and classical problems of information theory, which are pointed out in the paper.

I. INTRODUCTION

The problems addressed in this work were motivated mostly by measurement problems in optics, although the results are applicable in a wide variety of other contexts. The linear wave equation is of fundamental importance in many areas of science and engineering. It governs the propagation of electromagnetic, acoustic, and other kinds of fields. Its solutions in free space can be expressed in many forms. One of these is to express the field over one plane in terms of that on another, through a diffraction integral, a convenient approximation of which is well-known as the Fresnel diffraction integral [1]. Generalizations of such so-called quadratic-phase integrals allow one to characterize a broad class of optical systems involving arbitrary concatenations of lenses and sections of free space [2]. Such integral transforms are related to the fractional Fourier transform [3], which provides an elegant and pure description of these systems.

In this paper we consider a very general measurement scenario. Say we have an optical field propagating through some system characterized by any of the above integral transforms,

or indeed, any linear input-output relationship. We wish to recover the wave field as economically as possible. We are concerned with accuracy both in the sense of spatial resolution and in the sense of the accuracy of point-wise measurements. We are also concerned with the cost of performing the measurements and the trade-offs between cost and accuracy. For a given measurement cost (or error), we would like to know how to best make the measurements so as to minimize the measurement error (or cost), leading to a Pareto-optimal trade-off. In particular, we are interested in questions such as how many measurements we should make, where we should best situate our detectors, how the sensitivity of each detector should be chosen, and so forth, in order to obtain the best trade-off.

These questions are not merely of interest for practical purposes. A study of these issues also leads us to an understanding of the information-theoretic relationships inherent in the wave equation and what happens to the information carried by a wave field as it propagates through a system. Specifically we wish to develop a better understanding of what information certain parts of a wave field contain about the other parts, and characterize the dependence and redundancy inherent in different parts of a propagating wave. The present work aims to propose a unified framework within which such issues can be systematically studied.

Propagation of information in optical fields has been studied at least since the 1950s. The concept of the number of degrees of freedom (DOF) is central to several works including [4]–[9]. A different approach is pursued in [10], where the concepts of structural information and metrical information are introduced; these concepts find use in [11] and [12]. A sampling theory approach is taken in [13]. Most works under the name of “optics and information theory” have dealt with issues of sampling, degrees of freedom, and the like, rather than concepts involving Shannon entropy. A number of works utilizing Shannon entropy in different optical contexts have appeared [14]–[17]; nevertheless, we are not aware of any works which try to address measurement problems of the kind dealt with in this paper from an information-theoretical perspective.

II. PROBLEM FORMULATION

In the specific measurement scenario under consideration in this paper, noisy measurements are done at the output of a linear system in order to estimate the input of the system.

More precisely, the measured signal is of the form

$$g(\mathbf{x}) = L\{f(\mathbf{x})\} + n(\mathbf{x}), \quad (1)$$

where $\mathbf{x} \in \mathbb{R}^k$, $f : \mathbb{R}^k \rightarrow \mathbb{R}$ is the unknown input random process, $n : \mathbb{R}^k \rightarrow \mathbb{R}$ is the random process denoting the inherent system noise which is independent of the input f , and $g : \mathbb{R}^k \rightarrow \mathbb{R}$ is the output of the linear system. The dimension k is typically 1 or 2.

Measurements are done at various locations $\xi_1, \dots, \xi_M \in \mathbb{R}^k$ to obtain the observed variables $s_i \in \mathbb{R}$ according to the measurement model

$$s_i = g(\xi_i) + m_i, \quad (2)$$

for $i = 1, \dots, M$. Here m_i denotes the measurement noise introduced by the measurement device used at location ξ_i . Each measurement indexed by i is done with a possibly different noise variance $\sigma_{m_i}^2$. We allow repeated measurements so that more than one ξ_i may equal the same \mathbf{x} ; however, we assume that different measurements are statistically independent even if performed at the same site.

By putting the measured values in vector form, the measurement vector $\mathbf{s} = [s_1, \dots, s_M]^T$ is obtained

$$\mathbf{s} = \mathbf{g} + \mathbf{m}, \quad (3)$$

where $\mathbf{g} = [g(\xi_1), \dots, g(\xi_M)]^T$, $\mathbf{m} = [m_1, \dots, m_M]^T$. We assume each of these measurements given in (2) are done with a corresponding cost $C_i = (1/2) \log(\sigma_{s_i}^2 / \sigma_{m_i}^2)$ where $\sigma_{s_i}^2$ denotes the variance of s_i . The plausibility of this cost function is discussed in section III. Total cost, the cost of the scheme given in (3), is defined as the sum of the cost of each measurement. The objective is to minimize the mean-square error (MSE) between $f(\mathbf{x})$ and $\hat{f}(\mathbf{x}|\mathbf{s})$, the estimate of $f(\mathbf{x})$ given \mathbf{s} . We consider only linear minimum mean-square error (LMMSE) estimators. The problem in its most general form can be stated as one of determining

$$\varepsilon(\beta) = \inf_{\xi^M, M, \mathbf{C}^M} E \left\{ \int \|f(\mathbf{x}) - \hat{f}(\mathbf{x}|\mathbf{s})\|^2 d\mathbf{x} \right\} \quad (4)$$

subject to $\sum_{i=1}^M C_i \leq \beta$, where $\xi^M = [\xi_1, \dots, \xi_M]^T$ is the vector of sampling points, $\mathbf{C}^M = [C_1, \dots, C_M]^T$ is the associated cost vector, β is the total allowed cost, E denotes statistical expectation w.r.t. the statistics of the random field vector f and the measurement noise for the chosen measurement configuration, and $\|\cdot\|$ denotes Euclidean norm. The estimator $\hat{f}(\mathbf{x}|\mathbf{s})$ is the LMMSE estimator.

A. The metrical problem

In this paper, we study a discretized version of the above problem by assuming that (i) the space variable \mathbf{x} is quantized to a fixed finite set of points $\mathbf{x}_1, \dots, \mathbf{x}_N$ and we wish to estimate only the values $f(\mathbf{x}_i)$, $i = 1, \dots, N$, and (ii) the number M and locations ξ_1, \dots, ξ_M of the measurements are frozen and are not part of the optimization problem. (Note that we do not assume that the measurement locations are a subset of the points $\mathbf{x}_1, \dots, \mathbf{x}_N$. In fact, it may be that these two sets may be quite removed from each other, e.g., we may

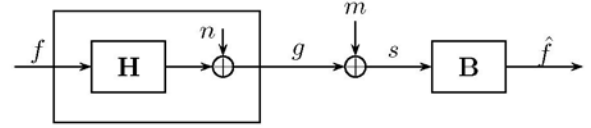


Fig. 1. Measurement system model block diagram.

wish to estimate the field at the outer edges of a region with measurements done in the center.) The problem is then reduced to computing

$$\varepsilon(\beta, \xi^M) = \min_{\mathbf{C}^M} E \left\{ \sum_{i=1}^N \|f(\mathbf{x}_i) - \hat{f}(\mathbf{x}_i|\mathbf{s})\|^2 \right\}, \quad (5)$$

where the cost vector is to be chosen subject to $\sum_{i=1}^M C_i \leq \beta$. We refer to this problem in which the emphasis is on understanding the relationship between the estimation error and the allotment of cost to the measurement devices at specified positions as the *metrical* problem, as opposed to the *structural* problem in which the emphasis is on how to best choose the number and locations of the measurement devices.

In this simplified metrical framework, where the coordinates $\mathbf{x}_1, \dots, \mathbf{x}_N$ and ξ_1, \dots, ξ_M are fixed, the measured vector \mathbf{g} can be related to the target vector \mathbf{f} by a matrix equation

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}, \quad (6)$$

where \mathbf{f} is a column vector with N elements, \mathbf{H} is an $M \times N$ matrix, and \mathbf{g} and \mathbf{n} are M dimensional column vectors. The coordinates of these vectors are defined as $f_i = f(\mathbf{x}_i)$, $g_j = g(\xi_j)$, and $n_j = n(\xi_j)$, $i = 1, \dots, N$, $j = 1, \dots, M$. The measured vector \mathbf{s} is given by (3) as before. We consider this problem under the assumptions that \mathbf{f} , \mathbf{n} , and \mathbf{m} are independent random vectors, with zero mean and known covariance matrices \mathbf{K}_f , \mathbf{K}_n , $\mathbf{K}_m = \text{diag}(\sigma_{m_1}^2, \dots, \sigma_{m_M}^2)$. We consider only linear estimators of the form $\hat{\mathbf{f}}(\mathbf{s}) = \mathbf{B}\mathbf{s}$ where \mathbf{B} is an N by M matrix. The MSE for such an estimator equals $E \{ \text{tr}[(\mathbf{f} - \mathbf{B}\mathbf{s})(\mathbf{f} - \mathbf{B}\mathbf{s})^T] \}$ or $\text{tr}(\mathbf{K}_f - 2\mathbf{B}\mathbf{H}\mathbf{K}_f + \mathbf{B}\mathbf{K}_s\mathbf{B}^T)$ where tr denotes the trace operator and $\mathbf{K}_s = \mathbf{H}\mathbf{K}_f\mathbf{H}^T + \mathbf{K}_n + \mathbf{K}_m$ is the covariance of \mathbf{s} . Concisely the *metrical measurement problem* is:

Given covariances $\mathbf{K}_f \in \mathbb{R}^{N \times N}$, $\mathbf{K}_n \in \mathbb{R}^{M \times M}$, a system matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$, and a budget $\beta \geq 0$, compute

$$\varepsilon(\beta) = \min_{\mathbf{K}_m, \mathbf{B}} \text{tr}(\mathbf{K}_f - 2\mathbf{B}\mathbf{H}\mathbf{K}_f + \mathbf{B}\mathbf{K}_s\mathbf{B}^T), \quad (7)$$

where the minimization is over all $\mathbf{B} \in \mathbb{R}^{N \times M}$ and all $\mathbf{K}_m = \text{diag}(\sigma_{m_1}^2, \dots, \sigma_{m_M}^2)$ subject to

$$\sum_{i=1}^M \frac{1}{2} \log \left(\frac{\sigma_{s_i}^2}{\sigma_{m_i}^2} \right) \leq \beta, \quad (8)$$

where $\sigma_{s_i}^2$ is the i th diagonal element of \mathbf{K}_s . \square

A block diagram illustrating this problem is given in Fig. 1. For any fixed \mathbf{K}_m , the optimization over \mathbf{B} is a standard LMMSE problem with solution

$$\mathbf{B} = \mathbf{K}_f\mathbf{H}^T\mathbf{K}_s^{-1}. \quad (9)$$

So, we could reduce the optimization problem to one over \mathbf{K}_m only, by substituting the optimal value of \mathbf{B} , but the form (7) is better suited to numerical optimization, as discussed in Section V.

The metrical problem above differs from a standard LMMSE estimation problem in that the covariance \mathbf{K}_m of measurement noise is subject to optimization. We are allowed to “design” the noise levels of the measurement devices subject to a cost constraint so as to minimize the overall estimation error. To our knowledge this problem is novel.

B. Relation to Rate-Distortion Theory

It is clear from Fig. 1 that, by the data-processing theorem [18, p. 80], we have $I(\mathbf{f}; \hat{\mathbf{f}}) \leq I(\mathbf{g}; \mathbf{s})$; i.e., the estimate $\hat{\mathbf{f}}$ can only provide as much information about \mathbf{f} as the measurement devices extract from the observable \mathbf{g} . In turn, by standard arguments, we have $I(\mathbf{g}; \mathbf{s}) \leq \sum_{i=1}^M I(g_i; s_i)$. The cost function $1/2 \log(1 + \sigma_{g_i}^2 / \sigma_{m_i}^2)$ that we use upper-bounds $I(g_i; s_i)$ whenever the measurement noise is Gaussian with a given variance $\sigma_{m_i}^2$ and the variance of the measured quantity is fixed as $\sigma_{g_i}^2$. Thus, for Gaussian measurement noise, we have $I(\mathbf{f}; \hat{\mathbf{f}}) \leq \beta$ where β is the total measurement budget.

The goal of measurements is the minimization of the MSE $\varepsilon(\beta) = E[d(\mathbf{f}, \hat{\mathbf{f}})]$ within a budget β where d denotes the summation on the right side of (5). From a rate-distortion theory viewpoint, interpreting d as a *distortion measure*, this is similar to minimizing the average distortion in the reconstruction of \mathbf{f} from a representation $\hat{\mathbf{f}}$ subject to a *rate constraint* $I(\mathbf{f}; \hat{\mathbf{f}}) \leq \beta$. This viewpoint immediately gives the bound $\varepsilon(\beta) \geq D(\beta)$ where $D(\beta)$ is the *distortion-rate function* applicable to this situation.

In the rate-distortion framework one is given complete freedom in forming the reconstruction vectors $\hat{\mathbf{f}}$ subject only to a rate constraint, which in measurement terminology would mean the ability to apply arbitrary transformations on the observable \mathbf{g} before performing a measurement (so as to carry out the measurement in the most favorable coordinate system) and not being constrained to linear measurements or linear estimators. Thus, the measurement problem can be seen as a special instance of the rate-distortion problem in which the formation of the reconstruction vector is restricted by various measurement constraints.

III. PROPOSED COST FUNCTION

Which properties of a measurement device should figure in its price (i.e., the fee for using it once)? A measurement device has basically two properties: range and resolution. We conceptualize measurement devices as instruments whose range can be adjusted freely to any interval $[-a + b, a + b]$ for any desired $a > 0$ and b before each measurement; so range is not a factor in determining the price of measurements in our model. Thus, the only remaining characteristic of a measurement device relevant to the cost issue is its resolution, which is a vague notion referring to the number of signal levels that the device can reliably distinguish. It may be

argued heuristically that resolution in a measurement process $s = g + m$ can be quantified by

$$\rho = \varrho \sqrt{\frac{\sigma_s^2}{\sigma_m^2}} = \varrho \sqrt{\left(1 + \frac{\sigma_g^2}{\sigma_m^2}\right)}, \quad (10)$$

where $\varrho >$ is a scaling constant that depends on how reliably the levels need to be distinguished. The heuristic argument we refer to here is precisely the same as the one given by [19], [20], [21] in defining the number of distinguishable signal levels at the receiver of an additive noise channel. The square-root in the expression keeps the resolution invariant under scaling of the input signal by any constant. Clearly, in the limit of very noisy measurements, ϱ should be 1; so, we set $\varrho = 1$ henceforth. Next, we list some properties that any plausible cost function must possess.

- 1) $C(\rho)$ must be a non-negative, monotonically increasing function of ρ , with $C(1) = 0$ since a device with one measurement level gives no useful information.
- 2) For any integer $L \geq 1$, we must have $LC(\rho) \geq C(\rho^L)$. This is because a measurement device with ρ levels can be used L times in succession with range adjustments between measurements to distinguish ρ^L levels.

Omitting further details due to space limitations, these are the basic arguments for using the cost function

$$C(\rho) = \log \rho = \frac{1}{2} \log \left(\frac{\sigma_s^2}{\sigma_m^2} \right) = \frac{1}{2} \log \left(1 + \frac{\sigma_g^2}{\sigma_m^2} \right). \quad (11)$$

$C(\rho)$ has the same form as Shannon’s formula for the capacity of a Gaussian noise channel; this has a minimax economic interpretation omitted here due to lack of space. Broadly, the mutual information $I(s; g) = h(s) - h(m)$ is the natural cost of a measurement $s = g + m$, since such a measurement is analogous to sending g across an additive noise channel.

IV. SPECIAL CASES

1) Single-Point Estimation with Repeated Measurements:

In the notation of Section II-A, we consider the case in which the space variable \mathbf{x} has $N = 1$ possible value, namely, \mathbf{x}_1 . One is allowed to make M measurements $s_i = g(\xi_i) + m_i$ subject to the usual cost constraint and the added restriction that $\xi_i = \mathbf{x}_1$, $i = 1, \dots, M$. By studying this case, we wish to see which measurement alternative is better: (i) to make one high quality measurement by renting the best device within budget limits, or (ii) to split the budget among multiple lower quality devices. Simple LMMSE analysis shows that the first alternative is better. We omit details due to space limitations.

2) *Diagonal Case:* When the matrices \mathbf{H} , \mathbf{K}_f , \mathbf{K}_n are diagonal, we refer to this case as the diagonal case. In this case $N = M$ and we further restrict the example by taking $\mathbf{K}_n = \mathbf{0}$, and $\mathbf{x}_i = \xi_i$, all i . The resulting problem is one where there are N separate LMMSE problems tied together by a total cost constraint. By standard techniques, we obtain the optimal solution as

$$\sigma_{m_i}^2 = \begin{cases} \frac{\nu \sigma_{f_i}^2}{\sigma_{f_i}^2 - \nu} & \text{if } \sigma_{f_i}^2 - \nu > 0 \\ \infty & \text{if } \sigma_{f_i}^2 - \nu \leq 0 \end{cases} \quad (12)$$

where the parameter ν is selected so that the total cost is β . Notice that for coordinates i where there is a non-trivial measurement ($\sigma_{m_i}^2 < \infty$), we have $1/\sigma_{m_i}^2 + 1/\sigma_{f_i}^2 = 1/\nu$, which is reminiscent of the “water-filling” solutions common in such information-theoretic problems (e.g., [18, p. 485]).

V. CALCULATION OF ESTIMATION ERROR

This section gives here a “double descent” method for solving the optimization problem (7), where we take turns in fixing \mathbf{B} and \mathbf{K}_m and minimizing over the other variable. This algorithm is known not to converge to the optimal solution for some hand-crafted examples where it starts from carefully chosen initial conditions; however, we have found no example where the algorithm failed to reach the optimal solution when the problematic initial state was slightly perturbed. For fixed \mathbf{K}_m , the \mathbf{B} that minimizes (7) is given by (9). On the other hand, if we fix \mathbf{B} , minimizing (13) over \mathbf{K}_m is equivalent to

$$\min_{\mathbf{K}_m} \text{tr}(\mathbf{B}\mathbf{K}_m\mathbf{B}^T) \quad (13)$$

subject to (8), after we eliminate \mathbf{K}_m -independent terms. In turn, this problem is equivalent to

$$\min_{\sigma_{m_1}^2, \dots, \sigma_{m_M}^2} \sum_{i=1}^M a_i \sigma_{m_i}^2 \quad (14)$$

subject to (8), where $a_i = \sum_{j=1}^M b_{j,i}^2$. By standard techniques of convex programming, the solution is given by

$$\sigma_{m_i}^2 = \frac{-\sigma_{g_i}^2 + \sqrt{\sigma_{g_i}^4 + \frac{4\sigma_{g_i}^2}{\nu a_i}}}{2}, \quad (15)$$

where $\nu > 0$ is a parameter chosen so that the total cost is β . The resulting algorithm is the following.

- 1) Initialization: Set $\mathbf{K}_m^{(0)} = \mathbf{0}$; $t = 0$.
- 2) Minimize over \mathbf{B} : Set $\mathbf{B}^{(t+1)} = \mathbf{K}_f \mathbf{H}^T (\mathbf{K}_s^{(t)})^{-1}$ where $\mathbf{K}_s^{(t)} = \mathbf{H}\mathbf{K}_f\mathbf{H}^T + \mathbf{K}_n + \mathbf{K}_m^{(t)}$.
- 3) Minimize over \mathbf{K}_m : Obtain $\mathbf{K}_m^{(t+1)}$ by solving the equations (15) with $a_i = \sum_{j=1}^M (b_{j,i}^{(t+1)})^2$.
- 4) Stoppage: If the percentage error, $100 \varepsilon(\beta) / \text{tr}(\mathbf{K}_f)$, over 10 consecutive iterations does not change by more than 0.01, then stop; otherwise, increment t and go to step 2.

VI. NUMERICAL RESULTS

We consider measurements of an optical wave field where all measurement probes are placed uniformly on a reference surface perpendicular to the axis of propagation and at a certain distance z from the source of the optical field. Based on its close relationship to the Fresnel diffraction integral, the propagation of light can be interpreted as an act of continual fractional Fourier transformation (FRT), with the fractional order a starting from 0 and increasing with the distance of propagation z , asymptotically reaching 1 for very large values of z [22]. Thus the measurement process in question can be modeled by taking the system matrix \mathbf{H} as the FRT matrix.

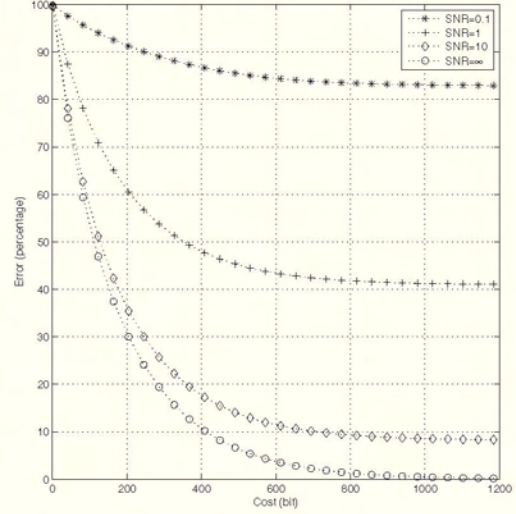


Fig. 2. Experiment 1: Error vs. cost for different values of SNR. $N = 256$, $M = 256$, $a = 0.5$, SNR variable. The vertical axis is percentage error $100 \varepsilon(\beta) / \text{tr}(\mathbf{K}_f)$.

Further information on the FRT and its computation may be found in [3], [23], [24]. The system matrix \mathbf{H} is taken to be the N by N real equivalent of the $N/2$ by $N/2$ complex FRT matrix [23]. In experiments, we have used randomly chosen covariance matrices \mathbf{K}_f and \mathbf{K}_n , and defined a parameter

$$\text{SNR} \triangleq \frac{\text{tr}(\mathbf{H}\mathbf{K}_f\mathbf{H}^T)}{\text{tr}(\mathbf{K}_n)} = \frac{\text{tr}(\mathbf{K}_f)}{\text{tr}(\mathbf{K}_n)}. \quad (16)$$

(Equivalence of the two forms is due to orthonormality of real FRT matrices, $\mathbf{H}^T\mathbf{H} = \mathbf{I}$.) SNR measures the ratio of signal power to inherent system noise power, before measurements.

To obtain the trade-off curve between the MSE error $\varepsilon(\beta)$ and measurement cost β , the algorithm of Section V has been used. The cost β is measured in bits by taking logarithms to base 2.

Experiment 1: In this experiment, the FRT order was fixed as $a = 0.5$, with $N = M = 256$. SNR was variable, ranging over 0.1, 1, 10 and ∞ . The computed $\varepsilon(\beta)$ vs. β curves are presented in Fig. 2. We notice that $\varepsilon(\beta)$ is very sensitive to increases in β for small enough β , then it becomes less responsive and eventually saturates at the asymptote for infinite measurement accuracy.

We have repeated the above experiment for several different values of the FRT order a . The resulting trade-off curves were nearly indistinguishable from the ones in Fig. 2, although the distribution of cost among the measurement devices changes with a . This indicates that the MSE is not critically dependent on how far the measurement devices are placed along the propagation axis, which is not surprising since the field that we are trying to reconstruct undergoes a reversible unitary transformation as it propagates, without picking up additional noise due to propagation effects such as turbulence in the atmosphere. Had the model included such effects, we would

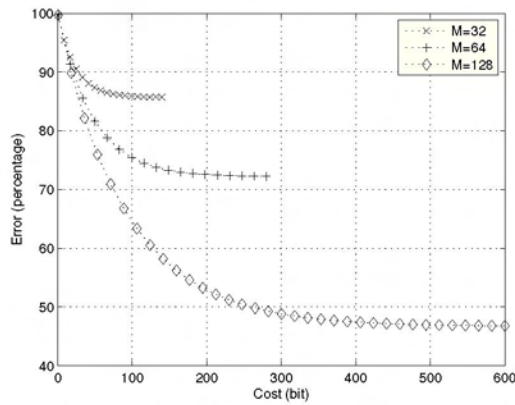


Fig. 3. Experiment 2: Error vs. cost for $N = 256$, $a = 0.5$, $\text{SNR}=10$, $M = 32, 64, 128$ variable.

expect a degradation of the trade-off curves as measurements are made at increasing distances. This subject is left for future work.

Experiment 2: This experiment investigates the dependence of $\varepsilon(\beta)$ for a fixed β on M . Fig. 3 shows the result for $a = 0.5$, $N = 256$, $\text{SNR} = 10$, and $M = 32, 64, 128$. The measurement locations were chosen from a uniform grid and the grid for $M = 32$ was a sub-grid of that for $M = 64$ which was a sub-grid of that for $M = 128$. We see that for low values of cost, added degrees of freedom for measurement locations does not significantly improve performance. However, as the allowed cost is increased, the benefit of spreading the measurement devices to more locations becomes significant. This example illustrates the importance of the *structural problem* in which the main task is to determine the number and locations of the measurement devices. Clearly, one may attack the structural problem by solving multiple carefully chosen instances of the metrical problem; however, a more methodical approach is desired and left as a subject for future study.

VII. CONCLUSION

We have given a precise formulation of the measurement problem for estimation of scalar wave fields in an optical context; the formulation has information-theoretic as well as estimation-theoretic elements. The formulation is also general enough to be applicable in other areas where one wishes to estimate the input of a linear system using noisy measurements. The novelty of the estimation problem here stems from two elements specific to this problem: (i) allowing the number and location of measurement devices to be variable, (ii) the assignment of costs to measurements. We have pointed out the close connection of the measurement problem to the rate-distortion problem, and also the essential differences between them. A numerical algorithm has been given which rapidly determines the trade-off between measurement cost and mean-square error. The numerical examples showed the applicability of the method to problems of interest in optics and have revealed quantitative trade-offs that were readily interpretable

in an optical context. The structural problem, in which the number and locations of the probes are subject to optimization, is an interesting open problem left for future study.

ACKNOWLEDGEMENT

This work was supported in part by TÜBİTAK grant EEEAG-105E065. Additionally, H. M. Ozaktas was supported in part by the Turkish Academy of Sciences.

REFERENCES

- [1] L. Onural and H. M. Ozaktas, "Signal processing issues in diffraction and holographic 3dtv," *Signal Processing: Image Communication*, in print.
- [2] M. J. Bastiaans, "Applications of the Wigner distribution function in optics," in *The Wigner Distribution: Theory and Applications in Signal Processing*, W. Mecklenbrauker and F. Hlawatsch, Eds. Amsterdam: Elsevier, 1997, pp. 375–426.
- [3] H. M. Ozaktas, Z. Zalevsky, and M. A. Kutay, *The Fractional Fourier Transform with Applications in Optics and Signal Processing*. New York: Wiley, 2001.
- [4] W. Lukozs, "Optical systems with resolving powers exceeding the classical limit," *Journal of the Optical Society of America*, vol. 56, no. 11, pp. 1463–1472, November 1966.
- [5] G. Toraldo Di Francia, "Resolving power and information," *Journal of the Optical Society of America*, vol. 45, no. 7, pp. 497–501, July 1955.
- [6] —, "Degrees of freedom of an image," *Journal of the Optical Society of America*, vol. 59, no. 7, pp. 799–804, July 1969.
- [7] F. Gori and G. Guattari, "Effects of coherence on the degrees of freedom of an image," *Journal of the Optical Society of America*, vol. 61, no. 1, pp. 36–39, January 1971.
- [8] —, "Degrees of freedom of images from point-like-element pupils," *JOSA*, vol. 64, no. 4, pp. 453–458, April 1974.
- [9] R. Piestun and D. A. Miller, "Electromagnetic degrees of freedom of an optical system," *Journal of the Optical Society of America A*, vol. 17, no. 5, pp. 892–902, May 2000.
- [10] D. MacKay, "Quantal aspects of scientific information," *IEEE Transactions on Information Theory*, vol. 1, no. 1, pp. 60–80, February 1953.
- [11] J. T. Winthrop, "Propagation of structural information in optical wave fields," *Journal of the Optical Society of America*, vol. 61, no. 1, pp. 15–30, January 1971.
- [12] D. Gabor, "Light and information," in *Progress In Optics*, E. Wolf, Ed. Elsevier, 1961, vol. I, ch. 4, pp. 109–153.
- [13] L. Onural, "Sampling of the diffraction field," *Applied Optics*, vol. 39, no. 32, pp. 5929–5935, November 2000.
- [14] F. T. Yu, *Entropy and Information Optics*. New York: Marcel Dekker, 2000.
- [15] R. Barakat, "Some entropic aspects of optical diffraction imagery," *Optics Communications*, vol. 156, no. 6, pp. 235–239, November 1998.
- [16] P. Réfrégier and J. Morio, "Shannon entropy of partially polarized and partially coherent light with Gaussian fluctuations," *JOSA A*, vol. 23, no. 12, pp. 3036–3044, December 2006.
- [17] A. Stern and B. Javidi, "Shannon number and information capacity of three-dimensional integral imaging," *Journal of the Optical Society of America A*, vol. 21, no. 9, pp. 1602–1612, September 2004.
- [18] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [19] R. V. L. Hartley, "Transmission of information," *Bell System Technical Journal*, vol. 7, pp. 535–563, July 1928.
- [20] C. E. Shannon, "A mathematical theory of communication," *BSTJ*, vol. 27, pp. 379–423 and 623–656, July and October 1948.
- [21] B. M. Oliver, J. R. Pierce, and C. E. Shannon, "The philosophy of PCM," in *Proceedings of the I.R.E.*, vol. 36, November 1948, pp. 1324–1331.
- [22] H. M. Ozaktas and D. Mendlovic, "Fractional Fourier optics," *Journal of the Optical Society of America A*, vol. 12, no. 4, pp. 743–751, April 1995.
- [23] Ç. Candan, M. A. Kutay, and H. M. Ozaktas, "The discrete fractional Fourier transform," *IEEE Transactions on Signal Processing*, vol. 48, no. 5, pp. 1329–1337, May 2000.
- [24] H. M. Ozaktas, O. Arkan, M. A. Kutay, and G. Bozdağı, "Digital computation of the fractional Fourier transform," *IEEE Transactions on Signal Processing*, vol. 44, no. 9, pp. 2141–2150, September 1996.