CONTRIBUTIONS TO INFORMATION-BASED COMPLEXITY, IMAGE UNDERSTANDING, AND LOGIC CIRCUIT DESIGN

DAVID LEE

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY
1985

CUCS-182-85

ABSTRACT

Contributions to Information-based Complexity, Image Understanding, and Logic Circuit Design

David Lee

This work consists of three parts. The first two describe new results in information-based complexity and applications of the general theory to the field of computer vision. The last presents an average case result of a different sort: the design of a binary comparator.

Part I is joint work with G. W. Wasilkowski. In this part, we study approximation of linear functionals on a separable Banach space equipped with a Gaussian measure. We study optimal information and optimal algorithms in average case, probabilistic and asymptotic settings, for a general error functional. We prove that adaptive information is not more powerful than nonadaptive information and that μ-spline algorithms, which are linear, are optimal in all three settings. We specialize our results to spaces of functions with continuous rth derivatives, equipped with a Wiener measure. In particular, we show that the interpolation by the natural splines of degree $r+1$ yields the optimal algorithms. We apply the general results to the problem of integration.

Part II of this work studies the following image understanding problems: 2 & 1/2 D sketch, shape from shading, and optical flow. We point out how known general optimality results in the worst case model may be applied to these problems. We indicate some preliminary results and work in progress, concerning the numerical solution of these problems. Algorithms which differ from those currently used in practice are proposed.

Part III provides a design of a binary comparator with completion signal, for the purpose of optimizing the average processing time. The average propagation delay is a constant, independent of n, the number of inputs, while the logic complexity is a linear function of n.

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

I am grateful to Professors J. R. Kender, J. F. Traub, S. H. Unger, G. W. Wasilkowski and H. Wozniakowski, for their encouragement, advice and help during this work.

1. Chapter 1

Approximation of Linear Functionals on a Banach Space with a Gaussian Measure

(Joint Work with G. W. Wasilkowski)

Abstract

We study approximation of linear functionals on separable Banach spaces equipped with a Gaussian measure. We study optimal information and optimal algorithms in average case, probabilistic and asymptotic settings, for a general error criterion. We prove that adaptive information is not more powerful than nonadaptive information and that μ-spline algorithms, which are linear, are optimal in all three settings. Some of these results hold for approximation of linear operators. We specialize our results to space of functions with continuous rth derivatives, equipped with a Wiener measure. In particular, we show that the natural splines of degree $2r+1$ yields the optimal algorithms. We apply the general results to the problem of integration.

1 Introduction

We illustrate the concepts and results of this paper by an integration example. We discuss this example at considerable length in the introduction, because it gives a simple illustration of the more general results obtained in this work. We state results for this example here, and defer the proofs until Section 5.

Suppose we want to approximate the integral of f, $Sf = \int_0^1 f(t)\,dt$, where f is a function with regularity r, i.e., $f \in C^r[0,1]$. We assume that instead of knowing f, we have information $N(f)$, which consists of n function or derivative values at adaptively chosen points t_i. The problem is to find the optimal

information and the optimal algorithms to minimize the approximation error. Much is known about this problem for the worst case setting. For instance, see [37], Ch.6, for a list of references. In that setting, the error of an algorithm is measured by its performance for the worst integrand. It was shown that information consisting of function values at equi-spaced points is almost optimal. Furthermore, for this information the algorithm based on perfect splines of degree r is optimal, with error $\Theta(n^{-r})$.

In this paper, we study three settings: average case, probabilistic and asymptotic. We assume that the space $C^r[0,1]$ is equipped with a probability measure, which reflects a belief of how often a function may occur as an integrand. We choose a Wiener measure, which is an example of a Gaussian measure. We study Wiener measures because they are among the most widely used and studied measures on function spaces and because they are of interest to physicists and statisticians. We seek optimal information and optimal algorithms. The following results are obtained for all three settings:

(i) The same algorithm based on interpolation by natural splines of degree $2r+1$ is optimal.

(ii) The same information, function values at equi-spaced points, is almost optimal.

(iii) Adaptive information is not more powerful than nonadaptive information.

We now comment on these results. We first elaborate on result (i), and compare it with the corresponding result for the worst case setting.

The optimal algorithm has the following properties:

- in the average case setting, it has the minimal average error,

- in the probabilistic setting, it has the maximal probability of not exceeding a prescribed error bound;
- in the asymptotic setting, the sequence of optimal algorithms has the best rate of convergence for *almost all* functions.

The minimal average error is equal to the square of the error of the l_2 -approximation of a certain function by splines of degree r. This l_2 -approximation error is equal to $\Theta(n^{-(r+1)})$, as proven in [31]. Hence the average error is $\Theta(n^{-(r+1)})$ and the best rate of convergence is $n^{-(r+1)}$.

There is a number of statistical papers addressing relations between splines or smoothing splines to the problem of Bayesian estimation, see e.g., [4, 19, 20, 45]. Our results specialized to the integration problem may be viewed as a continuation of this study; see also Remark 5.1.

For the average case, probabilistic and asymptotic settings, the probability measure supplies additional information to that provided by the integrand values. Therefore, we expect different optimality results than in the worst case setting. Indeed, in the worst case setting, the optimal algorithm is based on perfect splines of degree r, where r is the regularity of the underlying function space $C^r[0,1]$. In the three settings of this paper, the optimal algorithm is based on natural splines of degree $2r+1$. In terms of approximation error, the worst case error of the optimal algorithm is $\Theta(n^{-r})$, and the best possible rate of convergence is roughly n^{-r}. Therefore, the average case error of the optimal algorithm is an order smaller and the best possible rate of convergence in the average case setting is an order faster than that in the worst case setting. Furthermore, the resulting

optimal μ-spline algorithm is linear. Hence, if precomputation is used, we have a very simple and easy to implement optimal algorithm, which is desirable from the practical point of view.

Result (ii), which is a conclusion from [31] and the general results of Sections 2, 3 and 4, states that evaluating the integrand at equi-spaced points is almost optimal in all three settings. Result (iii) states that one can use information with a priori chosen points, which cannot be improved by any adaptively chosen points. Notice that this is a very desirable property from the practical point of view, since:

- nonadaptive information has much simpler structure than adaptive information, and in seeking optimal information, we can confine ourselves to nonadaptive information only;
- nonadaptive information can be computed very efficiently in parallel, whereas adaptive information is ill-suited for parallel computation.

The above reported results for integration are consequences of a more general approach, which is studied in the first part of this paper. In the general setting, we study approximation of a linear functional defined on a separable Banach space, equipped with a Gaussian measure. Information consists of adaptive continuous linear functionals. We study optimal information and optimal algorithms in average case, probabilistic and asymptotic settings. The following results are obtained in all three settings:

(i) the μ-spline algorithm is uniquely optimal;

(ii) the same information is optimal;

(iii) adaptive information is not more powerful than nonadaptive information.

Comments similar to that of the integration problem with a Wiener measure, which we used for illustration, hold in general. Some of these results hold even for approximation of linear operators which need not be continuous; we point it out when it occurs.

We want to add that there exists a statistical literature addressing the approximation of linear functionals defined on function spaces with information consisting of function and derivative evaluations, see e.g., [19, 20, 22, 31, 34, 45]. There are also papers dealing with the approximation of linear operators defined on separable Hilbert spaces with arbitrary adaptive linear functional evaluations as information, see e.g., [16, 46, 49, 50]. We know no literature addressing approximation problems defined on general separable Banach spaces.

2 Average Case Setting

In this section we study optimal algorithms and optimal information on the average. We begin with the definitions of basic concepts.

Let S be a continuous linear functional defined on a real separable Banach space F_1 ,

$$S : F_1 \to R. \tag{1}$$

We wish to construct $x = x(f) \in R$, which approximates Sf with a possibly small error. The error between Sf and x is measured by $E(Sf - x)$, where

$$E : R \to R_+ , \quad R_+ = [0,+\infty),$$

is called an *error functional*. The error between Sf and x is often measured by the absolute value of the difference, $Sf - x$. This corresponds to the error functional $E(\cdot) = |\cdot|$, which is convex and symmetric. These two properties of E are crucial for our analysis. We assume that the error functional E is convex and symmetric. (Some results hold even when E is neither convex nor symmetric and when S is a linear operator. In fact, it is enough to assume that E is an arbitrary function so that $E(S(\cdot) - g)$ is measurable for any fixed $g \in S(F_1)$. We will point it out when it occurs.)

We assume that the element f is unknown. Instead, we know Nf, where N is called an *information operator* (or *information*). In this paper, we assume that $N : F_1 \to R^n$ has the form

$$Nf = [L_1(f), \ldots, L_n(f)], \quad \forall\, f \in F_1, \tag{2}$$

where $L_1, \ldots L_n$ are continuous linear functionals. Without loss of generality, we assume that $L_1, \ldots, L_n$ are linearly independent. The number n is called the *cardinality* of N and is denoted by $card(N) = n$.

Knowing Nf, we construct an approximation x to Sf by an algorithm ϕ, $x = \phi(Nf)$. By an *algorithm* ϕ that uses N, we mean any mapping

$$\phi : N(F_1) = R^n \to R. \tag{3}$$

In the *average case setting* the error of ϕ is measured by the average value of $E(Sf - \phi(Nf))$, i.e.,

$$e^{avg}(\phi,N) = \int_{F_1} E(Sf - \phi(Nf))\, \mu(df), \tag{4}$$

where μ is a probability measure defined on the σ-field $\mathcal{B}(F_1)$ of Borel sets from F_1. To guarantee that the error of ϕ is well defined, we assume that ϕ is measurable. However, this assumption is not restrictive, as is discussed in [48].

We seek algorithms whose average error is as small as possible; such algorithms are called *optimal*. Note that the error of an algorithm ϕ depends on the error functional E and on the probability measure μ, $e^{avg}(\phi,N) = e^{avg}(\phi,N;E,\mu)$. Hence the optimal algorithm also depends on E and μ. In this paper we assume that the measure μ is Gaussian. We recall the definition and basic properties of Gaussian measures in Section 2.1.

2.1 Gaussian Measures

Let F_1^* be the dual space of F_1, i.e., F_1^* is the space of all continuous linear functionals $L : F_1 \to R$. Let μ be a probability measure defined on $\mathcal{B}(F_1)$. Then μ is a Gaussian measure iff the characteristic functional ψ_μ of μ, is of the form

$$\psi_\mu(L) = exp\{iL(a) - L(VL)/2\}, \quad \forall\ L \in F_1^*, \quad i = \sqrt{-1}, \tag{5}$$

for some element $a \in F_1$ and a linear operator $V : F_1^* \to F_1$. The characteristic functional ψ_μ of μ is defined by

$$\psi_\mu(L) = \int_{F_1} \exp\{iL(f)\}\ \mu(df), \quad \forall\ L \in F_1^*.$$

It is known, see [44], that for a Gaussian measure μ, with ψ_μ of the form (5), the *mean element* m_μ is equal to a, i.e.,

$$L(a) = L(m_\mu) = \int_{F_1} L(f)\, \mu(df), \qquad \forall\, L \in F_1^*, \tag{6}$$

and the *correlation operator* C_μ is equal to V, i.e.,

$$L_1(VL_2) = L_1(C_\mu L_2) = \int_{F_1} L_1(f - m_\mu)\, L_2(f - m_\mu)\, \mu(df), \quad \forall\, L_1, L_2 \in F_1^*. \tag{7}$$

The correlation operator C_μ of a Gaussian measure μ has a number of important properties (some of which will be discussed below). We remark that not every operator

$V : F_1^* \to F_1$ can be a correlation operator of some Gaussian measure. If F_1 is a separable Hilbert space, then V is a correlation operator of some Gaussian measure μ iff V is symmetric, nonnegative definite and has finite trace. If F_1 is, as in this paper, a separable Banach space, then the complete characterization of correlation operators of Gaussian measures remains an open problem.

Let C_μ be the correlation operator of a Gaussian measure μ. Then C_μ is symmetric,

$$L_1(C_\mu L_2) = L_2(C_\mu L_1), \quad \forall\, L_1, L_2 \in F_1^*,$$

and nonnegative definite,

$$L(C_\mu L) \geq 0, \quad \forall\, L \in F_1^*.$$

Hence, the correlation operator C_μ generates a semi-innerproduct on F_1^*,

$$\langle L_1, L_2 \rangle_\mu = L_1(C_\mu L_2), \quad \forall\, L_1, L_2 \in F_1^*,$$

and the corresponding semi-norm on F_1^*,

$$\|L\|_\mu = (<L,L>_\mu)^{1/2}, \quad \forall\, L \in F_1^*.$$

Here $<\cdot\,,\,\cdot>_\mu$ and $\|\cdot\|_\mu$ are an innerproduct and a norm, respectively, iff C_μ is one-to-one, or equivalently, iff C_μ is positive definite, i.e., $L(C_\mu L) > 0$, $\forall\, L \in F_1^*$, $L \neq 0$.

2.2 Spline Algorithms are Optimal on the Average

From now on, let μ be the Gaussian measure on $\mathbf{B}(F_1)$ with mean element zero and correlation operator C_μ as in (7), and let $N(f) = [L_1(f), \ldots, L_n(f)]$ be information. Without loss of generality we assume that the functionals $L_1, \ldots, L_n$ are orthonormalized such that

$$<L_i\,,\,L_j>_\mu = \delta_{i,j}\,, \quad \forall\, i, j = 1, 2, \ldots, n, \tag{8}$$

where $\delta_{i,j}$ is the Kronecker delta.

For $y = [y_1, y_2, \ldots, y_n] \in R^n$, let

$$\sigma(y,N) = \sum_{j=1}^{n} y_j\, C_\mu L_j\,. \tag{9}$$

Then $\sigma = \sigma(y,N) \in C_\mu(F_1^*)$ for every $y \in R^n$. The element σ has the following two properties:

(i) σ interpolates y,

$$N(\sigma) = y, \tag{10}$$

(ii) for every functional $L^\wedge \in F_1^*$ such that $C_\mu L^\wedge = \sigma$, $L^\wedge$ has the minimal semi-norm,

$$\|L^\wedge\|_\mu = \min\, \{\|L\|_\mu\colon L \in F_1^* \text{ and } N(C_\mu L) = y\}. \tag{11}$$

Indeed, $L_i(\sigma) = \sum_{j=1}^{n} y_j <L_i\,,\,L_j>_\mu = y_i$ which proves (10). To prove (11), take

$L \in F_1^*$ such that $N(C_\mu L) = y$. Let $f = C_\mu L$. Then $f = \sigma + h$, where $N(h) = 0$. Define $R = L - L^\wedge$. Then $R \in F_1^*$ and $C_\mu R = f - \sigma = h$. Thus, $N(C_\mu R) = 0$, i.e., $L_j(C_\mu R) = R(C_\mu L_j) = 0$, $j = 1, 2, \ldots, n$. Observe that $\|L\|_\mu^2 = \|L^\wedge\|_\mu^2 + \|R\|_\mu^2 + 2\langle L^\wedge, R\rangle_\mu$, where $\langle L^\wedge, R\rangle_\mu = R(C_\mu L^\wedge) = R\sigma = \sum_{j=1}^n y_j R(C_\mu L_j) = 0$. Thus $\|L\|_\mu \geq \|L^\wedge\|_\mu$ as claimed in (11).

Note that if C_μ is one-to-one, then $L^\wedge = C_\mu^{-1}\sigma$ and $\|L^\wedge\|_\mu^2 = (C_\mu^{-1}\sigma)\,\sigma = \sum_{j=1}^n y_j^2$. Then, (11) can be rewritten as

$$\sum_{j=1}^n y_j^2 = (C_\mu^{-1}\sigma)\,\sigma = \min\{(C_\mu^{-1}f)f : f \in C_\mu(F_1^*) \text{ and } Nf = y\}.$$

The element $\sigma(y,N)$ satisfying (i) and (ii) is called in the literature the *spline interpolating* y or, briefly, the *spline*. To stress the dependence of σ on the measure μ, we shall call $\sigma = \sigma(y,N)$ the *μ-spline*.

We define the *μ-spline algorithm* that uses N by

$$\phi^s(y) = S\sigma(y,N) = \sum_{j=1}^n y_j S(C_\mu L_j), \quad \forall\, y \in R^n. \tag{12}$$

Thus, the μ-spline algorithm is linear in y. We now prove that it is optimal.

Let the *average radius* of information N be defined as the minimal error among all algorithms that use N,

$$r^{avg}(N) = \inf_\phi e^{avg}(\phi,N). \tag{13}$$

By an *optimal algorithm* we mean any algorithm ϕ^* with

$$e^{avg}(\phi^*,N) = r^{avg}(N).$$

Theorem 2.1 Let N be of the form (8). Then the μ-spline algorithm ϕ^s is optimal and

$$e^{avg}(\phi^s,N) = r^{avg}(N) = \int_{F_1} E(Sf)\ \nu(df,N). \tag{14}$$

Here $\nu(\cdot,N)$ is the Gaussian measure on $\mathbf{B}(F_1)$ with mean element zero and correlation operator

$$C_\nu = (I - \sigma_N)\ C_\mu\ , \tag{15}$$

where the operator $\sigma_N : F_1 \to F_1$ is given by $\sigma_N(f) = \sum_{j=1}^{n} L_j(f)\ C_\mu L_j$, $\forall\ f \in F_1$. □

Proof.

The proof follows from the fact that the μ-spline is the mean of conditional (a posteriori) measure $\mu(\cdot \mid N(f) = y)$ which turns out to be equal to a translated measure ν (see the Appendix).

In Lemma 6.3 of the Appendix, we prove that for every ϕ,

$$e^{avg}(\phi,N) =$$

$$\int_{R^n} (2\pi)^{-n/2} \{ \int_{F_1} E(S(f + \sigma(y,N)) - \phi(y))\ \nu(df,N) \}\ exp\{ -\sum_{j=1}^{n} y_j^2/2 \}dy.$$

Hence, due to linearity of S,

$$e^{avg}(\phi,N) =$$

$$\int_{R^n} (2\pi)^{-n/2} \{ \int_{F_1} E(Sf-(\phi(y)-S\sigma(y,N)))\nu(df,N)\} exp\{ -\sum_{j=1}^{n} y_j^2/2\}dy. \tag{16}$$

In particular, we have

$$e^{avg}(\phi^s,N) = \int_{R^n} (2\pi)^{-n/2}\ \{ \int_{F_1} E(Sf)\ \nu(df,N) \}\ exp\{ -\sum_{j=1}^{n} y_j^2/2 \}\ dy$$

(17)

$$= \int_{F_1} E(Sf)\, \nu(df,N).$$

On the other hand, (16) yields

$$r^{avg}(N) =$$
$$\int_{R^n} (2\pi)^{-n/2} \{ \inf_{g\in\mathbf{R}} \int_{F_1} E(Sf - g)\, \nu(df,N)\} exp\{ - \sum_{j=1}^{n} y_j^2/2 \}\, dy$$

$$= \inf_{g\in\mathbf{R}} \int_{F_1} E(Sf - g)\, \nu(df,N).$$

Since $\nu(\cdot,N)$ is symmetric, i.e., $\nu(A,N) = \nu(-A,N)$, and E is symmetric, then, changing the variable f to $-f$, we get

$$r^{avg}(N) = \inf_{g\in\mathbf{R}} \int_{F_1} \{ \{E(Sf - g) + E(Sf + g)\}/2 \}\, \nu(df,N).$$

Convexity of E yields $E(Sf) \le (E(Sf - g) + E(Sf + g))/2$. Thus,

$$r^{avg}(N) = \int_{F_1} E(Sf)\, \nu(df,N) = e^{avg}(\phi^s,N).$$

Hence the μ-spline algorithm is optimal. []

Remark 2.1. It is possible to generalize Theorem 2.1 without assuming that E is convex and symmetric and that S is a functional. Let $S : F_1 \to F_2$ be a linear operator into a linear space F_2. Let E be an arbitrary functional from F_2 into $\mathbf{R}$, such that for every element $g \in F_2$ the functional $E(S(\cdot) - g)$ is measurable. The proof of Theorem 2.1 yields that

$$r^{avg}(N) = \inf_{g\in F_2} \int_{F_1} E(Sf - g)\, \nu(df,N).$$

Suppose that there exists an element $g^* \in R$ such that

$$r^{avg}(N) = \int_{F_1} E(Sf - g^*)\, \nu(df,N).$$

Define the *translated μ-spline algorithm* ϕ^* by

$$\phi^*(Nf) = \phi^s(Nf) + g^*.$$

Then

$$e^{avg}(\phi^*,N) =$$

$$\int_{R^n} (2\pi)^{-n/2} \{ \int_{F_1} E(Sf - (\phi^*(y) - \phi^s(y)))\, \nu(df,N) \}\, exp\{ -\sum_{j=1}^{n} y_j^2/2 \}\, dy$$

$$= \int_{F_1} E(Sf - g^*)\, \nu(df,N) = r^{avg}(N).$$

Thus, if g^* exists, then ϕ^* is optimal. On the other hand, if g^* does not exist, then there is no optimal algorithm, i.e., $r^{avg}(N) < e^{avg}(\phi,N)$, $\forall\ \phi$. However, taking g^* so that

$$\int_{F_1} E(Sf - g^*)\, \nu(df,N)$$

is close to $r^{avg}(N)$, we get an algorithm $\phi^*(Nf) = \phi^s(Nf) + g^*$ with almost minimal average error. []

We end this section by exhibiting an explicit formula for the radius of information N.

Theorem 2.2. Let N be of the form (8). Then

$$r^{avg}(N) = (2/\pi)^{1/2} \int_0^{\infty} E(t\ c(N)^{1/2})\, exp\{ -t^2/2 \}\, dt, \tag{18}$$

where

$$c(N) = S(C_\nu S) = \|S\|_\mu^2 - \sum_{j=1}^{n} \langle S, L_j \rangle_\mu^2. \qquad \square \qquad (19)$$

Proof.

Define the induced probability measure on $\mathbf{B}(R)$ by

$$\lambda(B,N) = \nu(S^{-1}(B),N), \quad \forall\ B \in \mathbf{B}(R). \qquad (20)$$

Since $\nu(\cdot,N)$ is Gaussian with mean element zero and correlation operator C_ν, $\lambda(\cdot,N)$ is the Gaussian (*normal*) measure on $\mathbf{B}(R)$ with mean element (*value*) zero and correlation operator (*value*) given by

$$c(N) = S(C_\nu S) = \langle S,S \rangle_\mu - \sum_{j=1}^{n} \langle S, L_j \rangle_\mu^2.$$

Changing variables in $r^{avg}(N) = \int_{F_1} E(Sf)\ \nu(df,N)$ by $x = Sf$, we get

$$r^{avg}(N) = \int_R E(x)\ \lambda(dx,N)$$
$$= (2\pi\ c(N))^{-1/2} \int_{-\infty}^{\infty} E(x)\ exp\{ - x^2/[2c(N)] \}dx.$$

From this and symmetry of E, we get (18). $\square$

Theorem 2.2 states that the average radius of N is fully characterized by $c(N)$. Note that due to (8), $c(N)$ can be rewritten as

$$c(N) = \|S - \sum_{j=1}^{n} \langle S, L_j \rangle_\mu L_j\ \|_\mu^2$$
$$= inf\ \{\|S - G\|_\mu^2 :\ G \in lin\{L_1, ..., L_n\}\}.$$

Thus, the correlation value $c(N)$ is equal to the error of the least square approximation of S by $lin\{L_1, ..., L_n\}$ with respect to $\langle \cdot , \cdot \rangle_\mu$.

2.3 Adaption Does not Help on the Average

Up to now we have considered information of the form $N(f) = [L_1(f), \ldots, L_n(f)]$. This means that N consists of n evaluations of continuous linear functionals L_i , $i = 1, 2, \ldots, n$, which are chosen independently of f. This is why N is called *nonadaptive*, denoted by N^{non}. In practice, a more general kind of information is often used. Namely, information consists of evaluations of n continuous linear functionals, which are chosen adaptively, depending on previously computed values. That is

$$N^a(f) = [L_1(f), L_2(f, y_1), \ldots, L_n(f, y_1, \ldots, y_{n-1})], \tag{21}$$

where $y_1 = y_1(f) = L_1(f)$, $y_i = y_i(f) = L_i(f, y_1, \ldots, y_{i-1})$, $i = 2, \ldots, n$. We assume that for a fixed $y = [y_1, \ldots, y_n] \in R^n$, $L_{i,y} = L_i(\cdot, y_1, \ldots, y_{i-1})$ is linear and continuous and that the mappings $G_i : R^n \to F_1^*$, $G_i(y) = L_{i,y}$, are Borel measurable. Such information N^a is called *adaptive*.

Since adaptive information has much richer structure than nonadaptive information, one might hope to achieve more by using adaptive information. We now prove that this hope is groundless. That is, *adaption does not help on the average*. (For similar results in Hilbert spaces, see [46, 49, 50]. This shows that the restriction to nonadaptive information, which we made in previous sections, is without loss of generality.

To prove that adaption does not help, we construct, for any adaptive N^a, nonadaptive information of a similar structure to N^a, which is not less powerful. The construction is obtained by fixing y_i in (21). More precisely, given N^a, $N^a(f) = [L_1(f), L_2(f, y_1), \ldots, L_n(f, y_1, \ldots, y_{n-1})$ with $y_i = y_i(f) = L_i(f, y_1(f), \ldots, y_{i-1}(f))]$, and given an element $y^* = [y_1^*, \ldots, y_n^*] \in R^n$, let

$$N^{non}{}_{y^*}(f) = [L_1(f),\ L_{2,y^*}(f),\ \ldots,\ L_{n,y^*}(f)],\quad \forall\ f \in F_1\ .$$

Of course, $N^{non}{}_{y^*}$ is nonadaptive.

Theorem 2.3 For every adaptive N^a, there exists an element $y^* \in R^n$, such that

$$r^{avg}(N^{non}{}_{y^*}) \le r^{avg}(N^a). \qquad \square$$

Proof.

We only sketch the proof, since it is similar to that of [46] Th.3.1. Without loss of generality, we can assume that for every $y = [y_1, \ldots, y_n] \in R^n$,

$$<L_{i,y}\ ,\ L_{j,y}>_\mu = \delta_{i,j}\ ,\ \forall\ i,j = 1, 2, \ldots, n.$$

It can be shown that for any algorithm ϕ, (16) holds with $\nu(\cdot,N)$ replaced by $\nu(\cdot,N^{non}{}_y)$. That is,

$$e^{avg}(\phi,N^a) =$$

$$\int_{R^n} (2\pi)^{n/2} \{ \int_{F_1} E(Sf-(\phi(y)-S\sigma(y,N^{non}{}_y)))\nu(df,N^{non}{}_y)\}\ exp\{-\sum_{j=1}^{n} y_j^2/2\}dy.$$

From this and (14) (see also Remark 2.1),

$$r^{avg}(N^a) =$$

$$\int_{R^n} (2\pi)^{-n/2}\ \{\ \inf_{g\in R}\ \int_{F_1} E(Sf - g)\ \nu(df,N^{non}{}_y)\}\ exp\{\ -\sum_{j=1}^{n} y_j^2/2\}\ dy$$

$$= \int_{R^n} (2\pi)^{-n/2}\ \ r^{avg}(N^{non}{}_y)\ exp\{\ -\sum_{j=1}^{n} y_j^2/2\}\ dy. \qquad (22)$$

Suppose it were true that for any element $y \in R^n$,

$$r^{avg}(N^a) <\ \ r^{avg}(N^{non}{}_y).$$

Integrating over y, we would have

$$r^{avg}(N^a) < \int_{R^n} (2\pi)^{-n/2} \; r^{avg}(N^{non}{}_y) \; exp\{ - \sum_{j=1}^{n} y_j^2/2\} \; dy,$$

which contradicts (22). Thus, there exists an element $y^* \in R^n$, such that

$$r^{avg}(N^{non}{}_{y^*}) \leq r^{avg}(N^a),$$

which completes the proof. []

Theorem 2.3 states that to find optimal information, it is sufficient to consider only nonadaptive information.

Remark 2.2 In the proof of Theorem 2.3 we do not use convexity or symmetry of E and the fact that S is a functional. Hence this theorem holds for S and E as in Remark 2.1. []

2.4 Optimal Information

Up to now we have assumed that the information operator $N = [L_1, ..., L_n]$ is fixed. Suppose now that we can choose functionals L_i , which form information N. Let **L** be a given class of permissible functionals L. We want to find n functionals L_i^* from **L** such that the average radius of N_n^*, $N_n^* = [L_1^*, ..., L_n^*]$ is minimal. Such information is called *nth optimal information*, and its radius is called the *nth minimal radius* (with respect to **L**),

$$r^{avg}(N_n^*) = r^{avg}(n,\mathbf{L}) = inf \; \{r^{avg}(N) : N = [L_1, ..., L_n], \; L_i \in \mathbf{L}\}.$$

We use Theorem 2.2 to exhibit optimal information. Due to that theorem, the radius of N is given by

$$r^{avg}(N) = (2/\pi)^{1/2} \int_0^{\infty} E(t \; c(N)^{1/2}) \; exp\{ - t^2/2 \} \; dt.$$

Since E is convex and symmetric, E is nondecreasing on R_+ . Thus to minimize $r^{avg}(N)$, it is enough to minimize $c(N)$. Let

$$c(n,\mathbf{L}) = inf\ \{c(N) :\ N = [L_1, \ldots, L_n],\ \ L_i \in \mathbf{L}\}$$
$$= inf\ \{\|S - G\|_\mu^{\,2} : G \in lin\{L_1, \ldots, L_n\},\ \ L_i \in \mathbf{L}\}. \tag{23}$$

Then

$$r^{avg}(n,\mathbf{L}) = (2/\pi)^{1/2} \int_0^\infty E(t\ c(n,\mathbf{L})^{1/2})\ exp\{\ -\ t^2/2\ \}\ dt. \tag{24}$$

This means that the nth minimal radius is fully characterized by the least square error of approximating S by a functional from n dimensional spaces spanned by elements from $\mathbf{L}$. Let $L_1^*, \ldots, L_n^* \in \mathbf{L}$ be such that

$$c(n,\mathbf{L}) = inf\ \{\|S - G\|_\mu^{\,2} : G \in lin\{L_1^*, \ldots, L_n^*\}\}. \tag{25}$$

Then N_n^*,

$$N_n^*(f) = [L_1^*(f), \ldots, L_n^*(f)] \tag{26}$$

is nth optimal.

3 Probabilistic Setting

In this section we study optimal algorithms and optimal information in the *probabilistic* setting. Let S, E, N and ϕ be defined as in Section 2. As for the average case setting we wish to approximate Sf by $\phi(Nf)$ so that $E(Sf - \phi(Nf))$ is small. In the probabilistic setting, the performance of ϕ is measured by the probability of success, i.e., by the probability that $E(Sf - \phi(Nf))$ is small. More precisely, given $\epsilon \geq 0$, let

$$prob(\phi,N,\epsilon) = \mu(\{f \in F_1 : E(Sf - \phi(Nf)) \leq \epsilon\}), \tag{27}$$

and let

$$prob(N,\epsilon) = \sup_{\phi} prob(\phi,N,\epsilon). \tag{28}$$

We seek an *optimal algorithm* ϕ^*, i.e., an algorithm with the largest probability of success,

$$prob(\phi^*,N,\epsilon) = prob(N,\epsilon).$$

3.1 Spline Algorithms are Optimal in the Probabilistic Setting

For nonadaptive N let ϕ^s be the μ-spline algorithm and let $c(N)$ be defined as in Section 2. For convex and symmetric E let

$$E^{-1}(\epsilon) = \sup \{t \in R : E(t) \leq \epsilon\}.$$

To avoid some unimportant complications, we assume that $E(0) = 0$, i.e., $E^{-1}(\epsilon)$ is well defined for any $\epsilon \in R_+$.

Theorem 3.1 For every nonadaptive N of the form (8), the μ-spline algorithm ϕ^s is optimal in the probabilistic setting,

$$prob(N,\epsilon) = prob(\phi^s,N,\epsilon) = (2/\pi)^{1/2} \int_0^b exp\{ - t^2/2 \} \, dt, \qquad (29)$$

with $b = b(N,\epsilon) = E^{-1}(\epsilon)/c(N)^{1/2}$. Here, by convention, $x/0 = +\infty$, $\forall\, x \in R_+$

.

Proof.

As in the proofs of Theorems 2.1 and 2.2. we conclude that

$$prob(\phi,N,\epsilon) =$$

$$(2\pi)^{-n/2} \int_{R^n} \nu(\{f \in F_1 : E(Sf-(\phi(y)-\phi^s(y))) \leq \epsilon\},N) exp\{-\sum_{j=1}^{n} y_j^2/2\} dy.$$

Therefore

$$prob(N,\epsilon) =$$

$$(2\pi)^{-n/2} \int_{R^n} \sup_{g \in \mathbf{R}} \nu(\{f \in F_1 : E(Sf - g) \leq \epsilon\},N) \, exp\{ - \sum_{j=1}^{n} y_j^2/2\} dy$$

$$= \sup_{g \in \mathbf{R}} \nu(\{f \in F_1 : E(Sf - g) \leq \epsilon\},N)$$

$$= \sup_{g \in \mathbf{R}} \lambda(\{t \in R : E(t - g) \leq \epsilon\},N),$$

where $\lambda(\cdot,N)$ is defined by (20). Since E is nondecreasing, $closure\{t \in R : E(t - g) \leq \epsilon\} = \{t \in R : |t - g| \leq E^{-1}(\epsilon)\}$, and therefore

$$prob(N,\epsilon) = \sup_{g \in \mathbf{R}} (2\pi)^{-1/2} \int_{a_1}^{a_2} exp\{ - t^2/2 \} \, dt,$$

with $a_1 = a_1(g) = (g - E^{-1}(\epsilon))/c(N)^{1/2}$ and $a_2 = a_2(g) = (g + E^{-1}(\epsilon))/c(N)^{1/2}$. It is easy to check that the supremum is attained for $g = 0$ and therefore

$$prob(N,\epsilon) = (2/\pi)^{1/2} \int_0^b exp\{ - t^2/2 \} \, dt.$$

For the μ-spline algorithm ϕ^s we get

$$prob(\phi^s,N,\epsilon) = \nu(\{f \in F_1 : E(Sf) \leq \epsilon\},N) = prob(N,\epsilon).$$

Hence ϕ^s is optimal. []

Remark 3.1 Similar to Remark 2.1 one can easily prove the optimality of a translated μ-spline algorithm for an arbitrary linear operator S and an arbitrary measurable E. More precisely, if there exists an element $g^* \in R$ for which $\nu(\{f \in F_1 : E(Sf - g^*) \leq \epsilon\}, N)$ attains its maximal value, then the algorithm $\phi^*(Nf) = \phi^s(Nf) + g^*$ is optimal. If such g^* does not exist, then there is no optimal algorithm. []

3.2 Adaption does not help in the Probabilistic Setting

Theorem 3.2 For every adaptive information N^a, there exists an element $y^* \in R^n$, such that

$$prob(N_{y^*}^{non},\epsilon) \geq prob(N^a,\epsilon). \qquad []$$

Since the proof is similar to that of Theorem 2.3, we omit it.

We want to add that Theorem 3.2 holds also for S and E as in Remark 2.1. In that case the element y^* might strongly depend on ϵ and in particular might be quite different than the corresponding element from Theorem 2.3. However, if S is a linear functional and E is convex and symmetric, as we assume in this paper, one can choose a common element y^* such that both

$$r^{avg}(N_y{}^{non}) \leq r^{avg}(N^a)$$

and

$$prob(N_y{}^{non},\epsilon) \geq prob(N^a,\epsilon),\ \forall\ \epsilon \geq 0$$

holds at the same time. That is, instead of N^a one can use $N_y{}^{non}$ which is at least as good as N^a in both the average case and the probabilistic settings for every ϵ. For instance, if $inf_y\ c(N_y{}^{non})$ is attained for $y = z$, one can set $y^* = z$. Otherwise, y^* should be so that $c(N_y{}^{non})$ is close enough to the above infimum.

3.3 Optimal Information in Probabilistic Setting

We want to find optimal information for the probabilistic setting. As in Section 2.4 let **L** be a given class of linear functionals. The *nth maximal probability* (with respect to **L**) is defined as

$$prob(n,\mathbf{L},\epsilon) = sup\ \{prob(N,\epsilon) : N = [L_1, ..., L_n],\quad L_i \in \mathbf{L}\},$$

and the *nth optimal information* $N_n{}^*$ is defined by

$$prob(N_n{}^*,\epsilon) = prob(n,\mathbf{L},\epsilon).$$

Similar to Section 2.4, we maximize $prob(N,\epsilon)$ iff the correlation value $c(N)$ is minimized. Thus,

$$prob(n,\mathbf{L},\epsilon) = (2/\pi)^{1/2} \int_0^{b^*} exp\{ - t^2/2 \}\ dt,$$

where $b^* = b^*(n,\epsilon) = E^{-1}(\epsilon)/c(n,\mathbf{L})^{1/2}$ and $c(n,\mathbf{L})$ is given by (23). Since for $N_n{}^*$ defined by (26), $c(N_n{}^*) = c(n,\mathbf{L})$, the information operator $N_n{}^*$ is also

optimal in the probabilistic setting. Note that N_n^* does *not* depend on ϵ. This and the optimality of the μ-spline algorithm tell us that using N_n^* and ϕ_n^*, we maximize the probability of success for *every* ϵ.

4 Asymptotic Setting with a Measure

In this section, we report very briefly about the optimality of μ-spline algorithms in the asymptotic setting, based on a recent paper [50]. Unlike in the average case and probabilistic settings, we do *not* fix the cardinality of information. Instead, we let $n = card(N_n)$ go to infinity, and we investigate how to approximate Sf by $\phi_n(N_n f)$ so that $|Sf - \phi_n(N_n f)|$ tends to zero as fast as possible for *almost all* $f \in F_1$. That is, we have a sequence of adaptive information operators $\{N_n\}$ as defined in Section 2.3, and we seek a sequence of algorithms which enjoys the best possible rate of convergence of $|Sf - \phi_n(N_n f)|$ for *almost all* $f \in F_1$. We assume that N_n is contained in N_{n+1}, i.e.,

$$N_{n+1}(f) = [N_n(f),\ L_{n+1}(f, y_1, \ldots, y_n)],$$

which is not a restrictive assumption in practice.

The asymptotic setting has been recently investigated in [50] for linear problems in Hilbert spaces with a measure; that is, for approximation of $S: F_1 \to F_2$, where F_1, F_2 are separable Hilbert spaces, F_1 is equipped with a Gaussian measure, and S is a continuous linear operator. Recall that here F_1 is a separable Banach space and S is a continuous linear functional. However, knowing that λ, the induced measure in (20), is Gaussian, one can obtain all the results from [50] for the case studied in this paper. We have

Theorem 4.1 For every sequence of adaptive information $\{N_n\}$, the μ-spline algorithms $\{\phi_n^{\,s}\}$ are optimal, i.e., for every sequence of algorithms $\{\phi_n\}$

$$\mu(\{f \in F_1 : \lim_{n\to\infty} |Sf - \phi_n(N_n(f))| / |Sf - \phi_n^{\,s}(N_n(f))| = 0\}) = 0. \qquad (30)$$

□

This theorem states that some algorithms might be better than the μ-spline algorithms *only* on a set of μ-measure zero. Hence the best rate of convergence is achieved by the μ-spline algorithms. This rate is characterized in

Theorem 4.2 For every sequence of adaptive information $\{N_n\}$,

$$\mu(\{f \in F_1 : \lim_{n\to\infty} |Sf - \phi_n^s(N_n f)| \,/\, c(N_n, f)^{1/2} = 0\}) = 0, \quad (31)$$

and

$$\mu(\{f \in F_1 : \lim_{n\to\infty} c(N_n, f)^{1/2} \,/\, |Sf - \phi_n^s(N_n f)| = 0\}) = 0, \quad (32)$$

where $c(N_n,f) = c(N^{non}{}_n, N_n f)$ is the correlation value of $\lambda(\cdot, N^{non}{}_{n,N_n f})$, see (19). □

Theorems 4.1 and 4.2 state close relations between the asymptotic setting with a Gaussian measure and the average case setting.

We now comment on optimal information. In the asymptotic setting information $\{N_n^{**}\}$ is optimal if it provides the best possible rate of convergence. Due to Theorem 4.2, this means that $c(N_n^{**})$ goes to zero with the best rate of convergence. Due to Theorems 4.1 and 4.2, no sequence of algorithms using any sequence of information operators has rate of convergence better than $c(n,\mathbf{L})$. Can this rate be achieved? Note that for N_n^*, as in Sections 2.4 and 3.4, $c(N_n^*) = c(n,\mathbf{L})$. However, for some specific $\mathbf{L}$, N_n^* need not contain N_{n-1}^*. In this case we use the following approach, suggested by [56]. Let N_n^{**} be a sequence of information operators, such that $N_n^{**} \subseteq N_{n+1}^{**}$, and for $n = 2^k - 1$,

$$N_n^{**}(f) = [N_1^*(f),\ N_2^*(f),\ N_4^*(f),\ \ldots,\ N_{2^{k-1}}^*(f)]. \tag{33}$$

Then for every n, N_n^{**} contains $N_{\lfloor n/2 \rfloor}^*$, and therefore,

$$c(N_n^{**}) \leq c(N^*_{\lfloor n/2 \rfloor}) = c(\lfloor n/2 \rfloor, \mathbf{L}).$$

Now, if $c(\lfloor n/2 \rfloor, \mathbf{L}) = \Theta(c(n,\mathbf{L}))$, which holds, for instance, for $c(n,\mathbf{L}) = \Theta(n^{-r})$, then

$$c(N_n^{**}) = \Theta(c(n,\mathbf{L})),$$

and therefore $\{N_n^{**}\}$ is optimal.

5 Applications

In this section we apply the results reported in the previous sections to linear problems in a space of regular functions equipped with a Wiener measure. Since the original work by Wiener in the 1920's, Wiener measures have received a great deal of attention. They are among the most widely used and studied measures on function spaces, and they are of interest to statisticians and physicists, see [4, 22, 31, 34, 45, 51, 52].

We begin with the definition and some properties of Wiener measures.

5.1 Wiener measures

Let $C^r[0,1]$ be the class of r times continuously differentiable functions defined on $[0,1]$. Let $C_0^r = \{f \in C^r[0,1] : f^{(j)}(0) = 0 , j = 0, 1, ..., r\}$, $r = 0, 1, ...$. Obviously, C_0^r equipped with the norm $\|f\| = sup \{|f^{(r)}(t)| : t \in [0,1]\}$, is a separable Banach space. Let D^r be the rth derivative operator, $D^r f = f^{(r)}$ for every f from $C^r[0,1]$. Note that D^r, restricted to C_0^r, is one-to-one and $D^r(C_0^r) = C_0^0$.

The Wiener measure μ on $\mathbf{B}(C_0^r)$ is defined by

$$\mu(B) = w(D^r(B)), \quad \forall\ B \in \mathbf{B}(C_0^r), \tag{34}$$

where w is the classical Wiener measure on $\mathbf{B}(C_0^0)$, uniquely defined by

$$w(\{f \in C_0^0 : f(t_j) - f(t_{j-1}) \in (a_j, b_j) , j = 1, ..., k\}) =$$

$$\prod_{j=1}^{k}(2\pi(t_j - t_{j-1}))^{-1/2} \int_{a_1}^{b_1}\cdots\int_{a_k}^{b_k} exp\{ - \sum_{j=1}^{k} y_j^2 /[2(t_j - t_{j-1})] \}\ dy, \tag{35}$$

for any $0 = t_0 < t_1 < ... < t_k \leq 1$ and $a_j < b_j$, see [21].

Take an operator T defined on the space of integrable functions on $[0,1]$

$$(Tf)(x) = \int_0^x f(t)\,dt. \tag{36}$$

Then for $m \geq 0$,

$$(T^m f)(x) = \int_0^x \int_0^{t_1} \cdots \int_0^{t_{m-1}} f(t_m)\,dt_m \cdots dt_1$$

$$= \int_0^x \{(x - t)^{m-1}/(m - 1)!\}\ f(t)\,dt \tag{37}$$

$$= \int_0^1 \{(x - t)_+^{m-1}/(m - 1)!\}\ f(t)\,dt,$$

where for $k \geq 1$

$$(x - t)_+^k = \begin{cases} (x - t)^k & \text{if } t \leq x, \\ 0 & \text{if } t > x, \end{cases}$$

and

$$(x - t)_+^0 = \begin{cases} 1 & \text{if } t \leq x, \\ 0 & \text{if } t > x. \end{cases}$$

Note that T^m is one-to-one and $T^m(C_0^j) = C_0^{j+m}$. Furthermore $D^m T^m = I$.

For every $L \in (C_0^r)^*$, take a function $s_L : [0,1] \to \mathbf{R}$, which is of bounded variation, continuous from the left and $s_L(1) = 0$, such that

$$L(f) = \int_0^1 f^{(r)}(t)\,d(s_L(t)), \quad \forall\ f \in C_0^r. \tag{38}$$

Such a function s_L exists due to Riesz Representation Theorem and the fact that $f = T^r f^{(r)}$.

For every $r \geq 0$, the Wiener measure μ on $\mathbf{B}(C_0^r)$ is the Gaussian measure with mean element zero and correlation operator C_μ ,

$$C_\mu L = - T^{r+1} s_L \ , \tag{39}$$

or equivalently,

$$(C_\mu L)(x) = L(T^r g_x),$$

where $g_x(t) = |x^{r+1} - (x - t)_+^{r+1}|/(r+1)!$, $\forall\ x,\ t \in [0,1]$. For completeness, we present a proof. For $r = 0$, $\mu = w$ and the characteristic functional of μ is equal to

$$\begin{aligned}
\psi_w(L) &= \int_{C_0^0} exp\{iL(f)\}\ w(df) \\
&= \int_{C_0^0} exp\{i \int_0^1 f(t)\ d(s_L(t))\}\ w(df) \\
&= exp\{ - \int_0^1 s_L^2(t)\ dt\ /2\ \} \\
&= exp\{ \int_0^1 (Ts_L)(t)\ d(s_L(t))/2\ \} \\
&= exp\{ - L(- Ts_L)/2\ \}.
\end{aligned}$$

For the third equality see [21] Th. 5.1, p.82. Thus, for $r = 0$, $C_w L = - Ts_L$. Hence w is Gaussian with mean element zero and correlation operator $C_w L = - Ts_L$. For $r \geq 1$, $\mu = wD^r$, and therefore

$$\psi_\mu(L) = \int_{C_0^r} exp\{iL(f)\}\ \mu(df) = \int_{C_0^0} exp\{iLT^r(f)\}\ w(df)$$

$$= \psi_w(LT^r), \qquad \forall L \in (C_0^r)^* .$$

Since $LT^r \in (C_0^0)^*$, we get

$$\Psi_\mu(L) = exp\{ - LT^r(C_w(LT^r)) /2 \} .$$

Note that $L(f) = \int_0^1 f^{(r)}(t)\, d(s_L(t))$ and $LT^r(f) = L(T^r f) = \int_0^1 f(t)\, d(s_L(t))$. Therefore $C_w(LT^r) = - Ts_L$, and $LT^r(C_w(LT^r)) = L(T^r(- Ts_L)) = L(- T^{r+1}s_L) = L(C_\mu L)$. Thus, $\psi_\mu(L) = exp\{ - L(C_\mu L)/2 \}$ which proves that μ is the Gaussian measure on $\mathbf{B}(C_0^r)$ with mean element zero and correlation operator C_μ, $C_\mu L = - T^{r+1}s_L$. We now prove that $(C_\mu L)(x) = L(T^r g_x)$. Indeed,

$$L(T^r g_x) = \int_0^1 (D^r T^r g_x)(t)\, d(s_L(t)) = \int_0^1 g_x(t)\, d(s_L(t))$$

$$= x^{r+1}/(r+1)! \int_0^1 1\, d(s_L(t)) - \int_0^1 (x - t)_+^{r+1}/(r+1)!\, d(s_L(t))$$

$$= - x^{r+1}/(r+1)!\, s_L(0) + \int_0^1 s_L(t)\, d((x - t)_+^{r+1}/(r+1)!) +$$

$$(x - 0)_+^{r+1}/(r+1)!\, s_L(0)$$

$$= (T^{r+1}s_L)(x)$$

as claimed. From (39) we see that for every $L \in (C_0^r)^*$, $C_\mu L$ is a function from C_0^r with $(r+1)st$ derivative of bounded variation.

5.2 Spline Algorithms and their Errors

In this section we apply results from Section 2 to 4 for $F_1 = C_0^r$, where μ is the Wiener measure. Without loss of generality we consider only nonadaptive information, since we know that adaption does not help. Let N, $N(f) = [L_1(f), L_2(f), \ldots, L_n(f)]$ be given. Then

$$L_i(f) = \int_0^1 f^{(r)}(t)\, d(s_i(t)) \tag{40}$$

for some s_i of bounded variation, continuous from the left and $s_i(1) = 0$. Note that

$$\begin{aligned} <L_i, L_j>_\mu &= L_i(C_\mu L_j) = L_i(-T^{r+1} s_j) \\ &= \int_0^1 s_j(t)\, s_i(t)\, d(t) = <s_i, s_j>_{l_2}. \end{aligned} \tag{41}$$

Thus, the assumption (8) that $<L_i, L_j>_\mu = \delta_{i,j}$ means that the functions s_i are orthonormal in l_2-norm. Then the μ-spline element interpolating $y = N(f)$ is of the following form

$$\sigma(y,N) = -\sum_{j=1}^{n} y_j T^{r+1} s_j . \tag{42}$$

For the μ-spline algorithm ϕ^s, $\phi^s(y) = S\sigma(y,N)$, we have

$$\begin{aligned} e^{avg}(\phi^s,N) &= (2/\pi)^{1/2} \int_0^\infty E(t\, c(N)^{1/2})\, exp\{ -t^2/2 \}\, dt \\ prob(\phi^s,N,\epsilon) &= (2/\pi)^{1/2} \int_0^b exp\{ -t^2/2 \}\, dt, \end{aligned} \tag{43}$$

where $b = b(N,\epsilon) = E^{-1}(\epsilon)/c(N)^{1/2}$ and

$$c(N) = inf \{\|S - G\|^2_\mu : G \in lin\{L_1, ..., L_n\}\}. \quad (44)$$

Since S belongs to $(C_0^r)^*$, there exists a function s of bounded variation, continuous from the left and $s(1) = 0$, such that

$$S(f) = \int_0^1 f^{(r)}(t)\, d(s(t)).$$

Due to (41), we get

$$c(N) = inf \{\int_0^1 (s(t) - g(t))^2\, dt : g \in lin\{s_1, ..., s_n\}\}. \quad (45)$$

Hence $c(N)$ equals to the square of the error of least square approximation of s by

$lin\{s_1, ..., s_n\}$ with respect to the l_2 -norm.

5.3 Standard Information

From now on we assume that N consists of function and derivatives evaluations,

$$N(f) = [f(t_1), ..., f^{(k_1)}(t_1), ..., f(t_p), ..., f^{(k_p)}(t_p)], \quad \forall\ f \in C_0^r, \quad (46)$$

where $0 = t_0 < t_1 < ... < t_p \leq 1$ and $k_1, ..., k_p \in \{0, ..., r\}$. Then the cardinality of N is given by

$$n = card(N) = \sum_{i=1}^{p} (k_i + 1).$$

Let $L_{i,j}(f) = f^{(j)}(t_i)$. Then

$$L_{i,j}(f) = \int_0^1 f^{(r)}(t)\, d(s_{i,j}(t)) \quad (47)$$

with

$$s_{i,j}(t) = \epsilon_j (t_i - t)_+^{r-j}/(r - j)! \ , \ i = 1, \ldots, p, \quad j = 0, \ldots, k_i \ ,$$

and $\epsilon_j = 1$ if $j = r$, and -1 otherwise. In order to construct the μ-spline algorithm one can take linear combinations of $s_{i,j}$ to normalize them as in (41). Then the μ-spline algorithm is given by (42) with y_i and s_i replaced by the corresponding linear combinations of $f^{(j)}(t_i)$ and $s_{i,j}$. Observe that the μ-spline element $\sigma = \sigma(N(f),N)$ is a linear combination of $s_{i,j}$, i.e., it is a piecewise polynomial interpolating $y = N(f)$,

$$\sigma^{(j)}(t_i) = f^{(j)}(t_i) \ , \ i = 1, \ldots, p, \ j = 0, \ldots, k_i \ ,$$

$$\text{and } \sigma^{(j)}(0) = 0, \ j = 0, \ldots, r.$$

Furthermore (11) and (39) yield that $\|C_\mu^{-1}\sigma\|^2_\mu = (C_\mu^{-1}\sigma)(\sigma) = \|\sigma^{(r+1)}\|^2_{l_2}$ is minimal among all functions g such that $g^{(j)}(t_i) = f^{(j)}(t_i)$, $i = 1, \ldots, p$, $j = 0, \ldots, k_i$, and $g^{(j)}(0) = 0$, $j = 0, \ldots, r$. Thus, $\sigma(\cdot) = \sigma(N(f),N)(\cdot)$ is the *natural spline of degree $2r+1$ interpolating f* at points $0 = t_0 , t_1 , \ldots, t_p \leq 1$, with multiplicities $m_0 = r + 1$, $m_1 = k_1 + 1 , \ldots, m_p = k_p + 1$, respectively. We have therefore the following

Theorem 5.1 For any information N of the form (46) and for any $r \geq 0$, the spline algorithm $\phi^s(y) = S\sigma(y,N)$ based on the natural spline $\sigma(y,N)(\cdot)$ of degree $2r+1$ is optimal for both the average case and probabilistic settings. Furthermore, the best rate of convergence in the asymptotic setting is achieved by the sequence of the spline algorithms. []

Remark 5.1 Theorem 5.1 states, in particular, that the natural spline of degree $2r+1$ is the mean of the conditional (a posteriori) measure. This fact is

mentioned in the statistical literature. For instance, it is written in [4], p.110, that this result is a well-known part of the folklore. We were unable to find the proof of this result. There is a number of very interesting papers, see e.g., [19, 20, 45], where similar problems are analyzed and relations between Bayesian statistics and spline functions are exhibited. In particular, [19] show that for a number of stochastic processes (excluding the Wiener case) spline functions (different than natural splines) are the means of conditional measures. □

5.4 Integration Problem

We now assume a specific form of the functional S. Let $Sf = \int_0^1 f(t)\,dt$. For any $f \in C_0^r$ we can rewrite this as

$$Sf = \int_0^1 f(t)\,dt = -\int_0^1 f^{(r)}(t)\,d\{(1-t)^{r+1}/(r+1)!\}.$$

We know that ϕ^s is optimal. We want to find optimal information of the form (46), i.e., with $\mathbf{L} = \{L : \exists\ x,\ \exists\ j \le r,\ L(f) = f^{(j)}(x)\ \forall\ f \in C_0^r\}$. Due to Sections 2.4 and 3.3, we need only to minimize $c(N)$. Note that for every N of the form (46),

$$c(N) = \inf\Big\{ \int_0^1 \{\,(1-t)^{r+1}/(r+1)! - q(t)\,\}^2\,dt : \ q \in \mathbf{P}(N)\Big\}$$

where

$$\mathbf{P}(N) = lin\{(t_i - \cdot\,)_+^{r-j}/(r-j)! : \ i = 1, \ldots, p,\ \ j = 0, \ldots, k_i\}.$$

This means that to find nth optimal information one needs to find optimal points and their multiplicity so that the l_2-error of approximating $s(x) = (1-t)^{r+1}/(r+1)!$ by spline functions of degree r is minimal. This minimization is a special

instance of a more general problem studied in [31]. In particular, they proved that

$$c(n,\mathbf{L}) = \Theta(n^{-2(r+1)}), \tag{48}$$

and, this is achieved by the following information N^*,

$$N^*(f) = [f(1/(n+1)),\ f(2/(n+1)),\ldots,\ f(n/(n+1))], \quad \forall\ f \in C_0^r. \tag{49}$$

This result and the results from Sections 2 and 3 imply that

Theorem 5.2 The information N^* defined by (49) is almost optimal for every $r \geq 0$ in the average case and probabilistic settings. []

For $n = 2^k - 1$, taking $\{N_n^{**}\}$ such that $N_n^{**} = [N_1^*, N_2^*, N_4^*, \ldots, N_{2^{k-1}}^*]$, as in Section 4, we conclude that

Theorem 5.3 The information $\{N_n^{**}\}$ is optimal for the asymptotic setting for every $r \geq 0$. The best possible rate of convergence $n^{-(r+1)}$ is achieved by the sequence of μ-spline algorithms that use $\{N_n^{**}\}$. []

5.5 Worst Case Setting and Average Case Setting - a Comparison

For the integration problem, we compare the optimality results in the average case setting with that in the worst case setting. In the latter setting, we approximate $Sf = \int_0^1 f(t)\,dt$ by using an algorithm ϕ, based on information N, so that $|Sf - \phi(Nf)| \ / \ \|f\|$ is small for every $f \in C_0^r$. By $\|\cdot\|$ we mean the norm on the space $F_1 = C_0^r$ as before, i.e.,
$\|f\| = \sup\ \{\ |f^{(r)}(t)| :\ t \in [0,1]\ \}$. The *worst case error* of ϕ is defined by

$$e^w(\phi,N) = \sup\ \{|Sf - \phi(N(f))| \ / \ \|f\| :\ f \in C_0^r\},$$

and the *worst case radius* of N is

$$r^w(N) = inf\ \{e^w(\phi,N) : \phi\}.$$

Let ϕ^* be an *optimal* algorithm, i.e.,

$$e^w(\phi^*,N) = r^w(N).$$

From [38] Appendix E, it follows that for an optimal algorithm ϕ^*

$$e^w(\phi^*,N) = sup\ \{|Sf - \phi^*(Nf)| :\ \|f\| \leq 1\} =$$

$$\inf_{\phi}\ sup\ \{|Sf - \phi(Nf)| :\ \|f\| \leq 1\}.$$

Gaffney and Powell [6] proved that for every f with $\|f\| \leq 1$,

$$p_1(x) \leq f(x) \leq p_2(x), \quad \forall\ x \in [0,1],$$

where p_1 and p_2 are lower and upper envelopes obtained from the *perfect splines* of degree r with n-r+1 knots interpolating f, $N(p_1) = N(p_2) = N(f)$. Thus, the algorithm ϕ^*,

$$\phi^*(Nf) = S(p_1 + p_2)/2,$$

provided by the perfect splines of degree r with n-r+1 knots interpolating f is optimal in the worst case setting. Information N^*_n remains almost optimal in the worst case setting, but its *worst case* radius is proportional to n^{-r}, instead of $n^{-(r+1)}$, as in the average case setting with $E(\cdot) = |\cdot|$.

We summarize and contrast the optimality results of these two settings in Table 1.

	worst case setting	average case setting
almost optimal information $N^{\diamond}_n$	function evaluations at equi-spaced points	function evaluations at equi-spaced points
optimal algorithm ϕ^* that uses $N^{\diamond}_n$	based on perfect splines of degree r with n-r+1 knots	based on natural splines of degree $2r+1$
radius of $N^{\diamond}_n$	$\Theta(n^{-r})$	$\Theta(n^{-(r+1)})$

Table 1

The optimality results of the two settings are quite different. In the worst case setting, the optimal algorithm is based on the *perfect* splines of degree r with n-r+1 knots. On the contrary, in the average case setting, the optimal algorithm is based on the *natural* splines of degree $2r+1$. Furthermore, the average error of the μ-spline algorithm is an order smaller that the worst case error of the optimal algorithm ϕ^*. This is due to the fact that the probability measure μ is concentrated on the set of functions with regularity almost $r + 1/2$, and therefore, supplies additional smoothness to the problem. This quantifies how much information is carried by the measure.

5.6 Complexity

We now discuss the above reported results from the complexity point of view. For the sake of simplicity we only consider the average case setting with $E(.) = |.|$.

Suppose that we want to approximate the integral of f with an average

(expected) error not exceeding a preassigned accuracy ϵ. Obviously, we would like to have information N and an algorithm ϕ with an average cost as small as possible. We assume that each function or derivative evaluation costs unity, and that the cost of each arithmetic operation is small compare to unity. Then the cost of each algorithm that uses information of cardinality n is at least n, and the cost of the spline algorithm is close to n, since it is a linear algorithm. On the other hand, no algorithm has average error less than or equal to ϵ, unless it uses information of cardinality at least $n^* = n^*(\epsilon)$, where

$$n^* = min \{ n : r^{avg}(n,\mathbf{L}) \leq \epsilon \} = \Theta(1/\epsilon^{1/(r+1)}).$$

This implies that the *complexity* (i.e., the minimal cost) of the integration problem is $\Theta(1/\epsilon^{1/(r+1)})$, and that the spline algorithm that uses $n^*(\epsilon)$ function values at equispaced points is *almost optimal from the complexity point of view*.

We stress that we have restricted the considered algorithms to those that use information of fixed cardinality, whereas, in practice, information of varying cardinality is commonly used. As will be reported in a forthcoming paper (see [47]), this restriction is without loss of generality. Hence, the spline algorithm that uses $n^*(\epsilon)$ function values at equispaced points remains almost optimal in the class of algorithms that use information of varying cardinality.

6 Appendix

In this appendix, we prove some properties of Gaussian measures, which were used in previous sections.

We begin with the following simple observation. Let μ be the Gaussian measure with mean element m_μ and correlation operator C_μ . Then for a continuous linear functional L, $L \in F_1^*$, the induced measure μL^{-1} is the Gaussian measure on $\mathbf{B}(R)$ with mean $L(m_\mu)$ and correlation $<L,L>_\mu$. That is, if $<L,L>_\mu$ is positive, then

$$\mu(L^{-1}(B)) = \tag{50}$$
$$(2\pi<L,L>_\mu)^{-1/2}\int_A exp\{-(t-L(m_\mu))^2/(2<L,L>_\mu)\}dt, \ \forall B\in\mathbf{B}(R),$$

and if $<L,L>_\mu = 0$, then

$$\mu(B) = \begin{cases} 1 & if\ L(m_\mu) \in B, \\ 0 & if\ L(m_\mu) \notin B. \end{cases}$$

Indeed, the characteristic functional ψ of μL^{-1} is given by $\psi(x) = \psi_\mu(xL)$, $\forall\ x \in R$, and therefore $\psi(x) = exp\{\ ixL(m_\mu) - x^2<L,L>_\mu/2\ \}$, as claimed.

Let now $N = [L_1, ..., L_n]$ be of the form (8), i.e., $<L_i, L_j>_\mu = \delta_{i,j}$, and let $m_\mu = 0$. For the induced probability measure μN^{-1} defined on $\mathbf{B}(R^n)$, we have

Lemma 6.1 The measure μN^{-1} is Gaussian with mean element zero and correlation matrix identity, i.e.,

$$\mu N^{-1}(A) = \mu(N^{-1}(A)) = (2\pi)^{-n/2}\int_A exp\{\ -\sum_{j=1}^{n} y_j^2/2\ \}\ dy. \qquad \square \tag{51}$$

Proof.

For the characteristic functional $\psi_{\mu N^{-1}}$ of μN^{-1}, we have

$$\psi_{\mu N^{-1}}(x) = \int_{R^n} exp\{i \sum_{j=1}^{n} x_j y_j \} \mu N^{-1}dy$$

$$= \int_{F_1} exp\{i \sum_{j=1}^{n} x_j L_j (f)\} \mu(df)$$

$$= \psi_{\mu}(\sum_{j=1}^{n} x_j L_j)$$

$$= exp\{ - \sum_{j,k=1}^{n} x_j x_k <L_j , L_k>_{\mu}/2 \}$$

$$= exp\{ - \sum_{j=1}^{n} x_j^2/2 \}$$

$$= exp\{ - <x,x>/2 \}, \qquad \forall\, x \in R^n.$$

This completes the proof. []

From [29] Th.8.1 p.147, we know that there exists a unique (up to a set of μN^{-1} measure zero in $\mathbf{B}(R^n)$) family of probability measures $\nu(\cdot \mid y,N)$ defined on $\mathbf{B}(F_1)$ and indexed by $y \in R^n$, such that for all $B \in \mathbf{B}(F_1)$,

$$\begin{aligned} &\nu(N^{-1}(y) \mid y,N) = 1, \quad \forall\, y \in R^n, \text{ a.e.,} \\ &\nu(B \mid \cdot, N) \text{ is } \mu N^{-1} \text{ - integrable,} \\ &\mu(B) = \int_{R^n} \nu(B \mid y,N)\, \mu N^{-1}(dy). \end{aligned} \tag{52}$$

This family, called *conditional measure*, is crucial for our study, since for every measurable mapping H, $H : F_1 \to R$,

$$\int_{F_1} H(f)\, \mu(df) = \int_{R^n} \{\int_{F_1} H(f)\, \nu(df \mid y,N)\}\, \mu N^{-1}(dy),$$

and due to Lemma 6.1,

$$\int_{F_1} H(f)\ \mu(df) =$$

$$(2\pi)^{-n/2} \int_{R^n} \{ \int_{N^{-1}\{y\}} H(f)\ \nu(df \mid y,N) \}\ exp\{ -\sum_{j=1}^{n} y_j^2/2 \}\ dy.$$

In particular, we have

$$e^{avg}(\phi,N) =$$

$$(2\pi)^{-n/2} \int_{R^n} \{ \int_{F_1} E(Sf - \phi(y))\ \nu(df \mid y,N) \}\ exp\{ -\sum_{j=1}^{n} y_j^2/2 \}\ dy,$$

and

$$prob(\phi,N,\epsilon) =$$

$$(2\pi)^{-n/2} \int_{R^n} \nu(\{f \in F_1 : E(Sf-\phi(y)) \leq \epsilon\} \mid y,N) exp\{-\sum_{j=1}^{n} y_j^2/2\} dy.$$

Lemma 6.2 For every N of the form (8), $\nu(\cdot \mid y,N)$ is the Gaussian measure on $\mathbf{B}(F_1)$ with mean element $m_y = \sigma(y,N)$ and correlation operator $C_\nu = (I - \sigma_N)\ C_\mu$. □

Proof.

We first prove that for every y there exists a Gaussian measure with mean element $\sigma(y,N)$ and correlation operator $C_\nu = (I - \sigma_N)\ C_\mu$. For this purpose, let χ be a functional from F_1^* into $\mathbf{C}$ defined by

$$\chi(L) = exp\{ -L(C_\nu L)/2 \}, \quad \forall\ L \in F_1^*.$$

Note that

$$\chi(L) =$$

$$exp\{-L(C_\mu L)/2\} exp\{\sum_{j=1}^{n} <L,L_j>_\mu^2/2\} \geq exp\{-L(C_\mu L)/2\} = \psi_\mu(L).$$

Thus, $0 \leq 1 - \chi(L) \leq 1 - \psi_\mu(L)$, $\forall\ L \in F_1^*$. Since ψ_μ is the characteristic functional of the measure μ, χ is also the characteristic functional of some probability measure defined on $\mathbf{B}(F_1)$, see [44] p.115. Therefore, $exp\{\ iL(\sigma(y,N)) - L(C_\nu L)/2\ \}$ is also a characteristic functional of a probability measure. Hence there exists a family of probability measures β_y such that

$$\psi_{\beta_y}(L) = exp\{\ iL(\sigma(y,N)) - L(C_\nu L)/2\ \}, \quad \forall\ L \in F_1^*.$$

Due to the definition of Gaussian measures, β_y is Gaussian with mean element $\sigma(y,N)$ and correlation operator $C_\nu = (I - \sigma_N)\ C_\mu$.

We now prove that $\nu(\ \cdot\ |\ y,N) = \beta_y(\cdot)$. To prove this equality, we only need to show that β_y satisfies (52). Since for all y's, β_y are Gaussian with a common correlation operator and mean elements $\sigma(y,N)$, respectively, they are translations of β_0 , i.e.,

$$\beta_y(B) = \beta_0(B - \sigma(y,N)), \quad \forall\ B \in \mathbf{B}(F_1).$$

Hence to prove that $\beta_y(N^{-1}(y)) = 1$, it is enough to show that $\beta_0(N^{-1}(0)) = 1$, since

$N^{-1}(y) = N^{-1}(0) + \sigma(y,N)$. Let $G(f) = \sum_{j=1}^{n} L_j^2(f)$. Then

$$\int_{F_1} G(f)\ \beta_0(df) = \sum_{j=1}^{n} \int_{F_1} L_j^2(f)\ \beta_0(df) = \sum_{j=1}^{n} L_j\ (C_\nu\ L_j),$$

since β_0 has correlation operator C_ν and mean element $\sigma(0,N) = 0$. A simple calculation yields that $L_j\ (C_\nu\ L_j) = 0$, $j = 1, 2, \ldots, n$. Thus

$$\int_{F_1} G(f)\ \beta_0(df) = 0.$$

Since $G(f) \geq 0$, and $G(f) > 0$ iff $f \notin N^{-1}(0)$, this proves that $\beta_0(F_1 - N^{-1}(0)) = 0$, and hence, $\beta_0(N^{-1}(0)) = 1$.

It is easy to observe that $\beta_y(B)$, as a function of y, is μN^{-1}-integrable for every $B \in \mathbf{B}(F_1)$. To complete the proof, we need only to show the last equality in (52). Let

$\mu^*(B) = \int_{R^n} \beta_y(B)\ \mu N^{-1}(dy)$, $\forall B \in \mathbf{B}(F_1)$. Of course, μ^* is a probability measure on $\mathbf{B}(F_1)$ whose characteristic functional ψ^* is given by

$$\psi^*(L) = (2\pi)^{-n/2} \int_{R^n} \{\int_{F_1} exp\{iL(f)\}\ \beta_y(df)\}\ exp\{ -\sum_{j=1}^{n} y_j^2/2 \}\ dy$$

$$= (2\pi)^{-n/2} \int_{R^n} exp\{\ iL(\sigma(y,N)) - L(C_\nu L)/2 \}\ exp\{ -\sum_{j=1}^{n} y_j^2/2 \}\ dy$$

$$= exp\{-L(C_\nu L)/2\}(2\pi)^{-n/2} \int_{R^n} exp\{iL(\sigma(y,N))\ exp\{-\sum_{j=1}^{n} y_j^2/2\} dy.$$

The last integral is equal to

$$(2\pi)^{-n/2} \int_{R^n} exp\{i \sum_{j=1}^{n} y_j <L,L_j>_\mu\}\ exp\{ -\sum_{j=1}^{n} y_j^2/2 \}\ dy$$

$$= exp\{ - <L,L>_\mu/2 + L(C_\nu L)/2 \}.$$

Thus we have

$$\psi^*(L) = exp\{ - <L,L>_\mu/2 \} = \psi_\mu(L).$$

Hence $\mu^* = \mu$, and the uniqueness of conditional measures implies that $\nu(\cdot \mid y,N) =$

$\beta_y(\cdot)$. This completes the proof. []

Lemma 6.3 For N of the form (8), let $\nu(\cdot,N)$ be the Gaussian measure on $\mathbf{B}(F_1)$ with mean element zero and correlation operator $C_\nu = (I - \sigma_N)C_\mu$. Then for every algorithm ϕ, we have

$$e^{avg}(\phi,N) =$$

(53)

$$(2\pi)^{-n/2}\int_{R^n}\{\int_{F_1} E(S(f+\sigma(y,N))-\phi(y))\nu(df,N)\}exp\{-\sum_{j=1}^{n} y_j^2/2\}dy,$$

and

$$prob(\phi,N,\epsilon) =$$

$$(2\pi)^{n/2}\int_{R^n}\nu(\{f\in F_1: E(S(f+\sigma(y,N))-\phi(y))\leq\epsilon\},N)exp\{-\sum_{j=1}^{n} y_j^2/2\}dy. \tag{54}$$

□

Proof.

From (52) and Lemma 6.2, we get

$$e^{avg}(\phi,N) = \tag{55}$$
$$(2\pi)^{-n/2}\int_{R^n}\int_{F_1} E(Sf - \phi(y))\,\beta_y(df)\,exp\{-\sum_{j=1}^{n} y_j^2/2\}\,dy,$$

where β_y is the Gaussian measure with mean element $\sigma(y,N)$ and correlation operator C_ν. Hence $\beta_0 = \nu(\cdot,N)$. Furthermore, since $\beta_y(B) = \beta_0(B - \sigma(y,N)) = \nu(B - \sigma(y,N),N)$, $\forall\ B \in \mathbf{B}(F_1)$, we can rewrite (55) as

$$e^{avg}(\phi,N) =$$

$$(2\pi)^{-n/2}\int_{R^n}\{\int_{F_1} E(S(f+\sigma(y,N))-\phi(y))\nu(df,N)\}exp\{-\sum_{j=1}^{n} y_j^2/2\}dy,$$

which proves (53). Since the proof of (54) is similar, we omit it. □

2. Chapter 2

Optimal Algorithms for Image Understanding

Abstract

We study the following three image understanding problems: 2 & 1/2 D sketch, shape from shading, and optical flow.

1 Introduction

One important application domain of information-based complexity is *Image Understanding* (IU). The characteristic feature of IU is the construction of a rich structure of a scene from *limited, contaminated* and *priced* information present in an image. In general, what is desired in the construction is some "best function" that fits the data derived from binocularity, projection, shading, motion, etc. Mathematically, this can be cast as an approximation (or interpolation) problem subject to some error criterion. For the purpose of approximation, one seeks appropriate algorithms, and much work has been done [7, 10, 11, 13, 15, 17, 18, 35].

In this part, we discuss three image understanding problems: *2 & 1/2 D sketch, shape from shading, and optical flow.* We point out how known general optimality results in information-based complexity theory may be applied to some of these problems. We indicate some preliminary results and work in progress, concerning the numerical solution of these problems. Algorithms which differ from those currently used in practice are proposed.

We first review briefly the relevant part of the theory, the *worst case model*, and then discuss the following problems in turn: 2 & 1/2 D sketch, shape from shading, and optical flow.

2 Some Basics of Information-based Complexity

We approximate a linear operator S: $F_1 \to F_2$, where F_1 is a linear space and F_2 is a normed space with norm $\|\cdot\|$. We recover Sf, for $f \in F_1$, based on *information* about f: $N(f) = [L_1(f), \cdots, L_n(f)]$, where L_i is a functional on F_1 , and N: $F_1 \to R^n$. An *algorithm* ϕ uses this information to construct an approximation $\phi(N(f)) \in F_2$, where ϕ: $R^n \to F_2$ is an arbitrary mapping. Since any Sf^, such that N(f^) = N(f), could be the element we want to approximate, and since in practice, we only consider f^ in a subset of F_1 , F_0 , we define the *worst case local algorithm error* as

$$e(\phi,N,f) = \sup_{\substack{f^\wedge \in F_0 \\ N(f^\wedge)=N(f)}} \|Sf^\wedge - \phi(N(f))\|, \quad f \in F_1 . \tag{1}$$

We seek a *strongly optimal algorithm*, denoted by ϕ^*, which minimizes the local algorithm error, for each f in F_1 , among all algorithms. If a strongly optimal algorithm exists, its algorithm error is the lower bound of all algorithm errors. We define the *radius of information*, [37], as

$$r(N) = \sup_{f \in F_1} e(\phi^*,N,f), \quad \text{where } \phi^* \text{ is a strongly optimal algorithm.} \tag{2}$$

In the following discussion of IU problems, we use the notation and terminology given above.

3 2 & 1/2 D Sketch in Shape from Stereo

The slight disparities in the images received from the left and right eyes enable humans to determine the shape and relative depth of visible surfaces. Marr-Poggio-Grimson's *stereo vision* algorithm is one such algorithm that intends to model the stereo processing of the human visual system. For the details, see [7].

We discuss only one stage of this algorithm: the 2 & 1/2 D sketch. After previous processing, disparity values are recorded. Based on these values, triangulations provide some depth values of the surface to be recovered. The problem of the derivation of the 2 & 1/2 D sketch is the determination of the best possible interpolation for smoothly completing the surface from the discrete set of depth values.

We address this issue from the point of view of information-based complexity. We first formulate the problem in the setting of the general theory. We then study the optimality of the spline algorithm and its implementation. Finally, we discuss possible decomposition of information for parallel or distributed computation.

3.1 Choosing the Class of Surfaces

Evidence from psychology indicates that human visual system is able to segment the surface into smooth patches of order two and to recover each patch. We assume that the part of the real world surface we want to recover is smooth and is viewed from a position free of accidental alignments. We further assume that it has been segmented from other nearby surfaces.

Formally, the surfaces can be represented as real valued functions of two variables defined on a region D,

$$f: D \to R,\ D \subseteq R^2. \tag{3}$$

We assume that f and its first and second order partial derivatives are all square integrable. That is, the class of surfaces are located with respect to a given background - the region D, and are smooth at least up to the local curvature: their curvatures are square-integrable. This class of surfaces is a linear space, denoted by F_1 , as in Section 2.

The surface we expect to "see" is part of the real world surface f, restricted to a bounded region $G \subseteq D$. This class of surfaces is also a linear space. Assume that this linear space is equipped with a norm for measuring the approximation error, and we denote this normed space by F_2 , as in Section 2.

Let S be the mapping which restricts a surface from D to G. Then S is a linear operator from F_1 to F_2 . If f is the original surface in F_1 , then, as described in Section 2, our goal is to recover Sf. We are interested in a strongly optimal algorithm, which minimizes the local algorithm error among all algorithms.

3.2 Quantify Surface Cues as Information

To recover a surface Sf, we need information about f, or *surface cues*, which is, in this case, the depth values of the surface f at n points. Formally, as in Section 2, information is given by

$$N(f) = [L_1(f), \cdots, L_n(f)] \tag{4}$$
$$= [f(x_1, y_1), \cdots, f(x_n, y_n)],\ (x_i, y_i) \in D,\ i = 1, \cdots, n.$$

3.3 Surface Consistence Constraint

A natural constraint from practice is the *surface consistency constraint*, see [7], which states that between known data, the surface cannot change in a radical manner. The change of a surface is quantified by its variation θ, defined as

$$\theta(f) = \{\int_D [\ (f_{xx})^2 + 2(f_{xy})^2 + (f_{yy})^2\ ]dx\ dy\ \}^{1/2}. \tag{5}$$

As in Section 2, we confine ourselves to the class of surfaces F_0 , which has uniformly bounded surface variation. For simplicity,, we assume that the bound is 1, i.e., $F_0 = \{f \in F_1 : \theta(f) \leq 1\}$.

3.4 Spline Algorithm is Strongly Optimal

Grimson [7] further explored the surface consistency constraint and proposed a *spline interpolation*, which interpolates the data and minimizes θ. This is the *spline algorithm.*

We are interested in strongly optimal algorithms, which minimize the local *algorithm error* as defined in (1). It is known (see e.g. [37] ch.4) that

Proposition 3.1 The spline algorithm ϕ^s is strongly optimal and linear. It has the form

$$\phi^s(N(f)) = S\sigma_{N(f)} = \sum_{i=1}^{n} f(x_i, y_i)\ S\sigma_i\ , \tag{6}$$

where the *spline*

$$\sigma_{N(f)} = \sum_{i=1}^{n} f(x_i\ ,\ y_i)\ \sigma_i\ , \tag{7}$$

and σ_i is a *basis spline*, i.e., the function of minimal surface variation such that $\sigma_i(x_j , y_j) = \delta_{i,j}$, and $\delta_{i,j}$ is the Kronecker delta.

3.5 Norms on F_2 and the Radius of Information

As in (1), a norm on F_2 measures the approximation error of an algorithm. There are infinitely many norms which can be assigned to F_2 , and one might expect that the optimality of algorithms depends on the norms. However, it turns out that the optimality of the spline algorithm is invariant with respect to the norms on F_2 , see e.g. [37] ch.4.

On the other hand, it is obvious that the algorithm error itself *does* depend on the norms on F_2 . For a fixed norm on F_2 , the radius of information is, [37] ch.4,

$$r(N) = \sup_{h \in \ker N} \frac{\| Sh \|}{\theta(h)}, \text{ where } \ker N = \{f \in F_1 : N(f) = 0\}. \qquad (8)$$

To provide an example, we assume that F_2 is equipped with a *supremum norm*. We also assume that information is depth data on a regular grid of a unit square region G. It is known, [1] p.46, that $r'(N) = \Theta(n^{-1})$, where n is the number of data points, for

$$\theta'(f) = \{ \int_D [(f_{xx})^2 + (f_{yy})^2] dx\, dy \}^{1/2}. \qquad (9)$$

Since $\theta'(f) \leq \theta(f)$ for all $f \in F_1$, from (8), it is obvious that $r(N) \leq r'(N)$. Therefore, $r(N) = O(n^{-1})$.

3.6 Implementing the Spline Algorithms

Implementing the spline algorithms is important in practice. From (6), we only have to construct the spline $\sigma_{N(f)}$, or σ for short. Much effort has been devoted to it by image understanding researchers, see [7, 35].

We now briefly discuss two different approaches. The first is to use the basis splines, and the second is to use an approach based on the reproducing kernel.

3.6.1 Basis Spline Approach

From (6), the spline is a linear combination of basis splines σ_i , and the coefficients are known depth values. The basis splines are data independent and can be precomputed. However, for large n, it may not be feasible to compute and store all basis splines because of time and space limitation. To implement the spline algorithm using this approach, one may have to explore efficient ways of storing and retrieving the basis splines.

3.6.2 Reproducing Kernel Approach

We discuss the construction of the spline using the reproducing kernel with the following further assumptions. Assume that $D = R^2$ and G is the unit square in R^2 and that $f \in F_1$ is a *Schwartz distribution* (see [24]), with all the partial derivatives interpreted in the distributional sense. It is known [5] that there exists a *reproducing kernel*, from the semi-Hilbert norm (5), of the form $K(x , y ; u , v) = (1/16\pi)^{-1}[(x - u)^2 + (y - v)^2]log\ [(x - u)^2 + (y - v)^2]$. The reproducing kernel is useful for representing the splines.

Let (x_α , y_α), (x_β , y_β) and (x_γ , y_γ) be arbitrary three non-colinear sampling points. Let p_j be the unique solution in $lin\{1,x,y\}$ for $p_j(x_i , y_i) = \delta_{i,j}$, where $i,j \in J = \{\alpha,\beta,\gamma\}$, and $\delta_{i,j}$ is the Kronecker delta.

It is known [25] that

$$\sigma(x\ ,\ y) - \sum_{j\in J} f(x_j\ ,\ y_j)p_j(x\ ,\ y) \tag{10}$$

$$= \sum_{i\in I-J} a_i\ \omega(x\ ,\ y\ ;\ x_i\ ,\ y_i)\ ,$$

where $I = \{1, \cdots, n\}$, n is the number of depth data and

$$\omega(x\ ,\ y\ ;\ u\ ,\ v) = K(x\ ,\ y\ ;\ u\ ,\ v) - \tag{11}$$

$$\sum_{j\in J} p_j(x\ ,\ y)K(x_j\ ,\ y_j\ ;\ u\ ,\ v) - \sum_{j\in J} p_j(u\ ,\ v)K(x\ ,\ y\ ;\ x_j\ ,\ y_j)$$

$$+ \sum_{j\in J}\sum_{k\in J} p_j(x\ ,\ y)p_k(u\ ,\ v)K(x_j\ ,\ y_j\ ;\ x_k\ ,\ y_k).$$

Therefore, to construct the spline σ, we only need to compute a_i . Since the spline interpolates the depth data, i.e., $\sigma(x_r\ ,\ y_r) = f(x_r\ ,\ y_r)$, to derive a_i , one has to solve a system of linear equations:

$$\sum_{i\in I-J} \omega(x_r\ ,\ y_r\ ;\ x_i\ ,\ y_i)\ a_i \tag{12}$$

$$= f(x_r\ ,\ y_r) - \sum_{j\in J} f(x_j\ ,\ y_j)p_j(x_r\ ,\ y_r), \qquad r \in I - J.$$

The coefficient matrix is *symmetric* and *positive definite*. To solve this system of linear equations, *Cholesky factorization* is studied in [24], with cost $O(n^3)$. It remains open if (12) can be solved in time essentially less than $O(n^3)$.

3.7 Adaption Does not Help

Information in (4) for 2 & 1/2 D sketch is depth data. This is *nonadaptive information*, since the sampling location of the *i*th depth value of N(f), $(x_i\ ,\ y_i)$, does not depend on the previously computed (i-1) depth values of N(f). If the sampling location of the *i*th depth value *does* depend on the (i-1) depth values

obtained, then we call it *adaptive information*. More precisely, adaptive information is defined as

$$N^a(f) = z = [z_1, \cdots, z_n], \tag{13}$$

where $z_i = f(x_i, y_i)$ and

$$x_i = x_i(z_1, \cdots, z_{i-1}),\ y_i = y_i(z_1, \cdots, z_{i-1}), \quad i = 2, \cdots, n.$$

Adaptive information has a richer structure than nonadaptive information. One might hope that the previously computed (i-1) depth values supply additional information for determining where to sample for the ith depth value. Counter-intuitively, adaptive information *cannot* aid 2 & 1/2 D sketch and some other image understanding problems, since for any adaptive information N^a, there exists nonadaptive information N^{non}, such that $r(N^{non}) \leq r(N^a)$, see [37] ch.2 and [38] ch.4. Therefore, in seeking the best places to sample, we can confine ourselves to nonadaptive information only, which is simple and can be collected in parallel, which is favorable for *parallel* or *distributed* computation.

4 Shape from Shading

Research in shape from shading explores the relationship between *image brightness* and object shape. A great deal of information is contained in the image brightness values, since image brightness is related to surface orientation. Information can also be obtained from *occluding boundaries* and other *boundary conditions*, see [15].

Algorithms are designed to determine shape from shading, including *characteristic strip expansion* [2, 9, 10, 11, 53, 54], *photometric stereo* [12, 13, 32, 55], and *numerical shape from shading and occluding boundaries* [15]. Characteristic strip expansion method resorts to solving nonlinear partial differential equations. Photometric stereo requires more than two images taken with different light sources. Numerical shape from shading and occluding boundaries results in a large system of nonlinear equations [15], and an iterative algorithm was proposed for solving it. The existence and uniqueness of the solution remain a problem. Furthermore, the convergence of the iterative method has not been established.

We will propose using a new iterative algorithm for solving the system of nonlinear equations derived in [15], and we will discuss its convergence. We study the existence and the uniqueness of the solution of the system as well. Finally, we discuss a preliminary approach based on the general theory of information-based complexity.

4.1 Numerical Shape from Shading and Occluding Boundaries

The goal of numerical shape from shading and occluding boundaries is to determine surface orientations from image brightness and boundary conditions. We discuss representation of surface orientations first, and then the algorithm proposed by Ikeuchi and Horn in [15].

4.1.1 Gaussian Sphere and Stereographic Projection

Surface orientation is quantified by the surface normal, a unit vector in R^3. A surface normal can be represented by a point on a unit sphere, called the *Gaussian sphere*. The part of the surface facing us corresponds to the northern hemisphere, while points on the occluding boundaries correspond to the points on the equator.

The northern hemisphere is then projected into a plane, the ξ-η plane, which is tangent to the sphere at the north pole. The projection center is the south pole. This is called *stereographic projection*. This is a conformal mapping, and the northern hemisphere is mapped onto a closed disc of radius 2 in the ξ-η plane. Therefore, points in this disc represent the surface orientations. Notice that orientations of the occluding boundaries correspond to the points on the circumference of that disc.

4.1.2 Image-Irradiance Equation and Boundary Conditions

The surface orientations are related to the image brightness by the following *image-irradiance equation*,

$$R(\xi,\eta) = E(x,y), \qquad (x,y) \in D, \tag{14}$$

where D is a unit square region, $\xi = \xi(x,y)$ and $\eta = \eta(x,y)$ represent the

surface orientation, $E(x,y) \in C$ is the brightness measured at the point (x,y), and $R(\cdot\ ,\ \cdot) \in C^1$ can be determined experimentally or theoretically if some information is available about the incident, emittant and phase angles, see [13, 27]. The image-irradiance equation provides information for determining the surface orientation from image brightness.

From (14) alone, one cannot determine the surface orientation (ξ,η) at each point (x,y). We need supplementary information from boundary conditions. The outline of the projection of an object in the image plane is called its *silhouette*. Some parts of it may correspond to sharp edges on the surface, and some parts to places where the surface curves around smoothly. The smooth parts of the surface correspond to the parts of the silhouette, called *occluding boundaries*, which supply important information about the shape of an object. Other information can also be obtained from *self-shadow boundaries, specular points* and *singular points*, see [15]. All these boundary conditions provide useful initial values for the iterative algorithm, which we will discuss next.

Without loss of generality, we assume that $R(\xi,\eta)$, $E(x,y) \geq 0$, and that $E(x,y) = 0$ if and only if (x,y) belongs to the occluding boundaries if and only if $\xi^2(x,y) + \eta^2(x,y) = 4$ if and only if $R(\xi(x,y),\eta(x,y)) = 0$. We further assume that the surface orientations on the boundaries of D is known.

4.1.3 Consistency Constraint

We assume that the surface we perceive is smooth. More specifically, we assume that the first order partial derivatives of $\xi(x,y)$ and $\eta(x,y)$ are square integrable. We also assume that real world surfaces tend to be stable, and the stability is measured by

$$\nu = \nu(\xi,\eta) = \int_D [\,(\xi_x)^2 + (\xi_y)^2 + (\eta_x)^2 + (\eta_y)^2\,]\,dxdy. \qquad (15)$$

where ξ_x , ξ_y , η_x and η_y denote the partial derivatives of ξ and η with respect to x and y. The *surface consistency constraint* is quantified as minimizing ν.

Thus, observing the image-irradiance equation (14) and boundary information, we are seeking functions $\xi(x,y)$ and $\eta(x,y)$, which tend to minimize (15). An approach used in [15] is *spline-smoothing*, see [23], i.e., to find ξ and η which minimize

$$\mu = \mu(\xi,\eta) = \qquad (16)$$
$$\int_D \{\,[(\xi_x)^2 + (\xi_y)^2 + (\eta_x)^2 + (\eta_y)^2] + \lambda[R(\xi,\eta) - E(x,y)]^2\,\}\,dxdy.$$

where the *penalty parameter* λ is set according to the accuracy of the measurement of the image brightness and the preciseness of the modeling of the lighting environment by R. The noisier the measurement and the less precise the modeling, the smaller the parameter λ. For example, in [15], λ is set, heuristically, in inverse proportion to the root-mean-square of the noise in the image brightness measurements.

4.1.4 An Iterative Algorithm

In the previous discussion, we described the image-irradiance equation, boundary conditions, the smoothness and consistency constraint, and arrived at spline-smoothing. All quantities involved are continuous functions.

We discretize the unit square region D in the xy-plane with mesh size h, and

discretize μ by using difference operators instead of differential operators, and summations instead of integrals. The corresponding *discrete smoothing-spline*, or DSS for short, minimizes

$$\mu = \sum_{i,j} (s_{i,j} + \lambda\, r_{i,j}), \tag{17}$$

where

$$s_{i,j} =$$

$$[(\xi_{i+1,j} - \xi_{i,j})^2 + (\xi_{i,j+1} - \xi_{i,j})^2 + (\eta_{i+1,j} - \eta_{i,j})^2 + (\eta_{i,j+1} - \eta_{i,j})^2]/h^2 ,$$

$$r_{i,j} = [R(\xi_{i,j} , \eta_{i,j}) - E_{i,j}]^2,$$

$\xi_{i,j}$ and $\eta_{i,j}$ represent the surface orientation at the regular grid point (ih,jh), and $E_{i,j}$ is the brightness measured at the grid point (ih,jh). The above minimization is subject to the boundary constraints, i.e., $\xi_{i,j}$ and $\eta_{i,j}$ are known if (ih,jh) belongs to the boundaries.

To minimize (17), we have to solve a large system of sparse nonlinear equations:

$$\begin{aligned} \xi_{i,j} &= \xi_{i,j}\hat{} - 4^{-1}\lambda h^2[R(\xi_{i,j} , \eta_{i,j}) - E_{i,j}]\, \partial R(\xi_{i,j} , \eta_{i,j})/\partial \xi, \\ \eta_{i,j} &= \eta_{i,j}\hat{} - 4^{-1}\lambda h^2[R(\xi_{i,j} , \eta_{i,j}) - E_{i,j}]\, \partial R(\xi_{i,j} , \eta_{i,j})/\partial \eta, \end{aligned} \tag{18}$$

where

$$\xi_{i,j}\hat{} = [\xi_{i+1,j} + \xi_{i,j+1} + \xi_{i-1,j} + \xi_{i,j-1}]/4, \text{ and}$$

$$\eta_{i,j}\hat{} = [\eta_{i+1,j} + \eta_{i,j+1} + \eta_{i-1,j} + \eta_{i,j-1}]/4.$$

To solve (18) for $\xi_{i,j}$ and $\eta_{i,j}$, Ikeuchi and Horn [15] proposed the following iterative algorithm:

$$\begin{aligned} &\xi_{i,j}^{(m+1)} = \\ &\xi_{i,j}\hat{}^{(m)} - 4^{-1}\lambda h^2[R(\xi_{i,j}^{(m)} , \eta_{i,j}^{(m)}) - E_{i,j}])]\, \partial R(\xi_{i,j}^{(m)} , \eta_{i,j}^{(m)})/\partial \xi, \\ &\eta_{i,j}^{(m+1)} = \end{aligned} \tag{19}$$

$$\eta_{i,j}^{-(m)} - 4^{-1}\lambda h^2[R(\xi_{i,j}^{(m)}, \eta_{i,j}^{(m)}) - E_{i,j}] \; \partial R(\xi_{i,j}^{(m)}, \eta_{i,j}^{(m)})/\partial \eta.$$

We can repeatedly use the values from the *m*th iteration on the right-hand side to compute the values for the *(m+1)*st iteration on the left-hand side. The initial values are supplied by the boundary conditions, i.e., $\xi_{i,j}$ and $\eta_{i,j}$ are known if (ih,jh) belongs to the boundaries.

The existence and uniqueness of the solution remain a problem, and the convergence of the iterative method has not been established. Furthermore, (18) is a *necessary condition* for minimizing (17), with the additional constraint that in a DSS, $\xi_{i,j}^{*2} + \eta_{i,j}^{*2} < 4$, if (ih,jh) does not belong to the boundary points.

4.2 A New iterative Algorithm

We propose a new iterative algorithm for solving (18), which, for a range of λ, converges to the unique solution of the system, which is the unique DSS, minimizing (17). For an arbitrary λ, the uniqueness of the solution and the convergence of the algorithm need further study.

4.2.1 Matrix A

Let $K + 1 = h^{-1}$ and $N = K^2$. In the rest of this part, we will deal with an NxN matrix

$$\mathbf{A} = \begin{pmatrix} B & -I & & & & \\ -I & B & -I & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & -I & B & -I \\ & & & & -I & B \end{pmatrix} \qquad (20)$$

where the KxK matrix

$$B = \begin{pmatrix} 4 & -1 & & & & \\ -1 & 4 & -1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & & -1 & 4 & -1 \\ & & & & -1 & 4 \end{pmatrix} \tag{21}$$

We state a few facts about matrix A, and for the details, see [33].

Matrix A is symmetric and positive definite, with eigenvalues $\{\lambda_{i,j}\}$, i,j = 1, ···, K, where

$$\lambda_{i,j} = 4\left[\sin^2\frac{\pi i}{2(K+1)} + \sin^2\frac{\pi j}{2(K+1)} \right]. \tag{22}$$

The inverse of A, A^{-1}, is also symmetric and positive definite, with eigenvalues $\{\mu_{i,j}\}$, i,j = 1, ···, K, where

$$\mu_{i,j} = \lambda_{i,j}^{-1}. \tag{23}$$

A^{-1} can be decomposed as

$$A^{-1} = H\Lambda H, \tag{24}$$

where the diagonal matrix

$$\Lambda = \mathrm{diag}\{\mu_{i,j}\}, \quad i,j = 1, \cdots, K, \tag{25}$$

and H is the tensor product

$$H = S \otimes S, \tag{26}$$

where the (i,j)th entry of the KxK matrix S is

$$s_{i,j} = [2/(K+1)]^{1/2} \sin[\pi ij/(K+1)]. \tag{27}$$

Multiplying A^{-1} by a vector costs O(N^2), using the conventional method. Since we can decompose $A^{-1} = S \otimes S \Lambda S \otimes S$ and Λ is a diagonal matrix, taking advantage of the structure of the entries of S, $s_{i,j}$, we can use Fast Fourier Transforms (FFT) for the multiplication, which costs O(N(*log* N)).

4.2.2 A New Iterative Algorithm

Equation (18) can be rewritten as

$$\mathbf{Mx} = -\lambda h^2 \mathbf{b}(\mathbf{x}), \tag{28}$$

where

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{O} \\ \mathbf{O} & \mathbf{A} \end{pmatrix} \tag{29}$$

where A is given in (20),

$$\mathbf{b} = \tag{30}$$

$$[\ \cdots,\ \{R(\xi_{i,j}\ ,\ \eta_{i,j}) - E_{i,j}\}\partial R(\xi_{i,j}\ ,\ \eta_{i,j})/\partial\xi,\ \cdots,\ \{R(\xi_{i,j}\ ,\ \eta_{i,j}) - E_{i,j}\}\partial R(\xi_{i,j}\ ,\ \eta_{i,j})/\partial\eta,\ \cdots\ ]^T,$$

and

$$\mathbf{x} = [\ \xi_{1,1}\ ,\ \cdots,\ \xi_{1,K}\ ,\ \cdots,\ \xi_{K,K},\ \eta_{1,1}\ ,\ \cdots,\ \eta_{1,K}\ ,\ \cdots,\ \eta_{K,K}\ ]^T. \tag{31}$$

Since A is non-degenerate, M is also non-degenerate, and therefore, (28) is equivalent to

$$\mathbf{x} = -\lambda h^2 \mathbf{M}^{-1}\mathbf{b}(\mathbf{x}), \tag{32}$$

We propose using the following iterative algorithm

$$\mathbf{x}^{(m+1)} = -\lambda h^2 \mathbf{M}^{-1}\mathbf{b}(\mathbf{x}^{(m)}), \tag{33}$$

where $\mathbf{x}^{(0)}$ is an arbitrary initial element.

We discuss the convergence of the iterative algorithm (33) and the uniqueness of the solution of (32).

4.2.3 Convergence of the Algorithm

Our goal is to find a DSS $x^* = [\ \xi_{1,1}^*\ ,\ \cdots,\ \xi_{1,K}^*\ ,\ \cdots,\ \xi_{K,K}^*\ ,\ \eta_{1,1}^*\ ,\ \cdots,\ \eta_{1,K}^*\ ,\ \cdots,\ \eta_{K,K}^*\ ]^T$, which minimizes (17), subject to the boundary constraints, i.e., the surface orientations are known on boundaries. We call a DSS *regular*, if $(\xi_{i,j}^*)^2 + (\eta_{i,j}^*)^2 < 4$, when (ih,jh) is not a boundary point. A regular DSS does not generate false occluding boundary points.

Since $(\xi_{i,j}, \eta_{i,j})$ is in the closed disc with radius two, denoted by S, x is defined on a compact set in R^{2N}, S^{2N}. Since μ in (17) is a continuous function of x, it obtains its minimum on S^{2N}, and therefore, DSS exists. We first show that a DSS is regular, and thus regular DSS exists. We then show that the DSS is unique and that algorithm (33) converges to this unique DSS.

We assume that there exist at least two boundary points, on which the surface orientations are different.

To prove that a DSS is regular, we need

Lemma 4.1 Let P_0 be a point on the circumference of a disc, and let $P_i \neq P_0$, i = 1,2,...,k, be points in the disc. Then for $\delta > 0$ and $K > 0$, there exists P_0^* in the disc, such that (i) $d(P_0\ ,\ P_0^*) < \delta$, and (ii)

$$\sum_{i=1}^{k} d(P_0\ ,\ P_i)^2 > \sum_{i=1}^{k} d(P_0^*\ ,\ P_i)^2 + Kd(P_0\ ,\ P_0^*)^2, \tag{34}$$

where d(P,Q) is the Euclidean distance between P and Q. []

Proof. Denote the angle spanned by the vectors P_0P_i and P_0P_j as $<P_i\ ,\ P_0\ ,\ P_j>$. Let $\alpha = \max\ \{\ <P_i\ ,\ P_0\ ,\ P_j>,\ i,j = 1,2,...,k\ \}$. Let $Q'P_0Q$ be the

bisector of α. Since P_i is in the disc, $\alpha < \pi$, $<P_i , P_0 , Q> < \pi/2$, and $<P_i , P_0 , Q'> = \pi - <P_i , P_0 , Q> > \pi/2$. Let P_0^* be a point on P_0Q such that $d(P_0 , P_0^*) < \delta$. Then in the triangle $P_0P_0^*P_i$, $d(P_0 , P_i)^2 = d(P_0^* , P_i)^2 + d(P_0 , P_0^*)^2 - 2d(P_0^* , P_i)d(P_0 , P_0^*)cos<P_0 , P_0^* , P_i>$. Since $lim_{P_0^* \to P_0} <P_0 , P_0^* , P_i> = <P_i , P_0 , Q'> > \pi/2$, and $lim_{P_0^* \to P_0} P_iP_0^* = P_iP_0$, for sufficiently small $d(P_0 , P_0^*)$, $- 2d(P_0^* , P_i)cos<P_0 , P_0^* , P_i> > Kd(P_0 , P_0^*)$. Therefore, $- 2d(P_0^* , P_i)d(P_0 , P_0^*)cos<P_0 , P_0^* , P_i> > Kd(P_0 , P_0^*)^2$, and $d(P_0 , P_i)^2 > d(P_0^* , P_i)^2 + Kd(P_0 , P_0^*)^2$. Taking summation over all i, we have (34). []

We are ready to prove

Lemma 4.2 A DSS is regular. []

Proof. We prove by contradiction. We assume, on the contrary, that there exists a DSS x^*, which is not regular, i.e.,

$$\mu(x^*) = \sum_{i,j} (s_{i,j}^* + \lambda\, r_{i,j}^*), \tag{35}$$

where

$$s_{i,j}^* = [(\xi_{i+1,j}^* - \xi_{i,j}^*)^2 + (\xi_{i,j+1}^* - \xi_{i,j}^*)^2 + (\eta_{i+1,j}^* - \eta_{i,j}^*)^2 + (\eta_{i,j+1}^* - \eta_{i,j}^*)^2]/h^2 ,$$

$$r_{i,j}^* = [R(\xi_{i,j}^* , \eta_{i,j}^*) - E_{i,j}]^2,$$

and there exists $(\xi_{i,j}^* , \eta_{i,j}^*)$, such that $\xi_{i,j}^{*2} + \eta_{i,j}^{*2} = 4$. Since (ih,jh) is an interior point of the region D, $E_{i,j} > 0$.

We have

$$\mu(x^*) = \tag{36}$$

$$F(x^*)+\lambda[R(\xi_{i,j}^*\ ,\ \eta_{i,j}^*)\ -\ E_{i,j}]^2+\{[(\xi_{i+1,j}^*\ -\ \xi_{i,j}^*)^2+(\eta_{i+1,j}^*\ -\ \eta_{i,j}^*)^2]+$$
$$[(\xi_{i,j+1}^*\ -\ \xi_{i,j}^*)^2+(\eta_{i,j+1}^*\ -\ \eta_{i,j}^*)^2]+[(\xi_{i-1,j}^*\ -\ \xi_{i,j}^*)^2+(\eta_{i-1,j}^*\ -\ \eta_{i,j}^*)^2]+$$
$$[(\xi_{i,j-1}^*\ -\ \xi_{i,j}^*)^2\ +\ (\eta_{i,j-1}^*\ -\ \eta_{i,j-1}^*)^2]\}/h^2,$$

where $F(x^*)$ does not contain $\xi_{i,j}^*$ and $\eta_{i,j}^*$.

Let $d(k,l;i,j)$ be the Euclidean distance between $(\xi_{k,l}^*\ ,\ \eta_{k,l}^*)$ and $(\xi_{i,j}^*\ ,\ \eta_{i,j}^*)$. Then

$$\mu\ (x^*) = F(x^*) + \lambda[R(\xi_{i,j}^*\ ,\ \eta_{i,j}^*)\ -\ E_{i,j}]^2 + \tag{37}$$
$$[d^2(i+1,j;i,j) + d^2(i,j+1;i,j) + d^2(i-1,j;i,j) + d^2(i,j-1;i,j)]/h^2.$$

Since $E_{i,j} > 0$, $R(\xi_{i,j}^*\ ,\ \eta_{i,j}^*) = 0$ and R is continuous, there exists a disc, centered at $(\xi_{i,j}^*\ ,\ \eta_{i,j}^*)$ with sufficiently small radius δ, such that for all (ξ,η) inside the disc, $[R(\xi,\eta)\ -\ E_{i,j}]^2 \le [R(\xi_{i,j}^*\ ,\ \eta_{i,j}^*)\ -\ E_{i,j}]^2 = E_{i,j}^2$.

Let $P = \{(i+1,j),\ (i-1,j),\ (i,j+1),\ (i,j-1)\}$, and let $A = \{(k,l) \in P:\ d(k,l;i,j) > 0\}$ and $K = |P\ -\ A|$. We analyse the following two cases, and arrive at a contradiction for each case.

Case 1. $K < 4$, i.e., there exists at least one (k,l), such that $d(k,l;i,j) > 0$. By applying Lemma 4.1, with $P_0 = (\xi_{i,j}^*\ ,\ \eta_{i,j}^*)$ and $\{P_i\} = A$, we know that there exists (ξ^*,η^*), inside the disc, centered at $(\xi_{i,j}^*\ ,\ \eta_{i,j}^*)$, with radius δ, such that

$$\sum_{(k,l)\in P} d(k,l;i,j)^2 = \sum_{(k,l)\in A} d(k,l;i,j)^2 > \sum_{(k,l)\in A} d^*(k,l;0,0)^2 + Kd^*(i,j;0,0)^2$$
$$= \sum_{(k,l)\in A} d^*(k,l;0,0)^2 + \sum_{(k,l)\in P-A} d^*(k,l;0,0)^2 = \sum_{(k,l)\in P} d^*(k,l;0,0)^2,$$

where $d^*(k,l;0,0)$ is the Euclidean distance between $(\xi_{k,l}^*, \eta_{k,l}^*)$ and (ξ^*,η^*).

Replacing $(\xi_{i,j}^* , \eta_{i,j}^*)$ by (ξ^*,η^*), from (37), we have x^{**}, such that $\mu(x^{**}) < \mu(x^*)$. Therefore, x^* is not a DSS, a contradiction.

Case 2. $K = 4$, i.e, $d(k,l;i,j) = 0$, $\forall$ $k,l \in P$. We separate out all the grid points, adjacent to any of the points in P. We denote this set of grid points by $P^{(1)}$. We analyse $P^{(1)}$, in a similar way as P, and we will arrive either at a contradiction or at the conclusion that $d(k,l;i,j) = 0$, $\forall$ $(k,l) \in P \cup P^{(1)}$, i.e., $(\xi_{k,l}^* , \eta_{k,l}^*) = (\xi^*,\eta^*)$, $\forall$ $(k,l) \in P \cup P^{(1)}$. We repeat the same arguments as we expand the region of grid points with identical (ξ^*,η^*). Since D is a unit square region and there exist at least two boundary points on which the (fixed) surface orientations are different, we will arrive at a contradiction no later than that the expanded region covers these two points.

Therefore, an irregular DSS does not exist. This completes the proof. []

We assume that $\{R(\xi,\eta) - E_{i,j}\}\partial R(\xi,\eta)/\partial\xi$ and $\{R(\xi,\eta) - E_{i,j}\}\partial R(\xi,\eta)/\partial\eta$ are *Lipschitz* functions, i.e.,

$$|\{R(\xi,\eta) - E_{i,j}\}\partial R(\xi,\eta)/\partial\xi - \{R(\xi',\eta') - E_{i,j}\}\partial R(\xi',\eta')/\partial\xi|$$

$$\leq L^{(1)}_{i,j}\{(\xi - \xi')^2 + (\eta - \eta')^2\}^{1/2}$$

and (38)

$$|\{R(\xi,\eta) - E_{i,j}\}\partial R(\xi,\eta)/\partial\eta - \{R(\xi',\eta') - E_{i,j}\}\partial R(\xi',\eta')/\partial\eta|$$

$$\leq L^{(2)}_{i,j}\{(\xi - \xi')^2 + (\eta - \eta')^2\}^{1/2}.$$

Let $\max \{L_{i,j}^{(k)}\}_{i,j,k} = \nu_0$. Then we have

Proposition 4.1 For $\lambda \in [0 , 2\pi^2\nu_0^{-1}[1 - \pi^2h^2/24]^2)$, algorithm (33) converges to the unique solution of (28), which is the unique DSS, minimizing (17). []

Proof. Let x^*be a regular DSS. Then $x^* = -\lambda h^2M^{-1}b(x^*)$. From (33) we have $x^{(m+1)} - x^* = -\lambda h^2M^{-1}[b(x^{(m)}) - b(x^*)]$, and so $||x^{(m+1)} - x^*||_2 \leq \lambda h^2||M^{-1}||_2\nu_0||x^{(m)} - x^*||_2$. Since $||M^{-1}||_2 = [8sin^2(\pi h/2)]^{-1} < [2\pi^2h^2(1 - \pi^2h^2/24)^2]^{-1}$, $||x^{(m+1)} - x^*||_2 \leq \lambda h^2[2\pi^2h^2(1 - \pi^2h^2/24)^2]^{-1}\nu_0||x^{(m)} - x^*||_2 = \lambda[2\pi^2(1 - \pi^2h^2/24)^2]^{-1}\nu_0||x^{(m)} - x^*||_2$. Since $\lambda < 2\pi^2\nu_0^{-1}(1 - \pi^2h^2/24)^2$, $\lambda[2\pi^2(1 - \pi^2h^2/24)^2]^{-1}\nu_0 < 1$. Therefore, $x^{(m)}$ converges to x^*.

Since $x^{(m)}$ has only one limit and (28) is a necessary condition satisfied by a regular DSS, $x^{(m)}$ converges to the unique solution of (28), which is the unique DSS. []

4.2.4 An Example

As an example, we estimate ν_0 and the range of λ for the case of a Lambertian surface with the incident rays coincident with the view direction. In this case, the image-irradiance equation [15]

$$R(\xi,\eta) = (4 - \xi^2 - \eta^2)/(4 + \xi^2 + \eta^2), \text{ where } \xi^2 + \eta^2 \leq 4. \tag{39}$$

Obviously, R is a Lipschitz function and

$$\nu_0 \leq \{ \{ \sup | \partial/\partial\xi \{[R(\xi,\eta) - E_{i,j}]\partial R(\xi,\eta)/\partial\xi\} | \}^2 + \tag{40}$$
$$\{\sup \{| \partial/\partial\eta \{[R(\xi,\eta) - E_{i,j}]\partial R(\xi,\eta)/\partial\eta\} |\}^2 \}^{1/2}.$$

Since $|\partial/\partial\xi \{[R(\xi,\eta) - E_{i,j}]\partial R(\xi,\eta)/\partial\xi\}| = |[\partial R/\partial\xi]^2 + (R - E_{i,j})\partial^2R(\xi,\eta)/\partial\xi^2| \leq |\partial R/\partial\xi|^2 + |R - E_{i,j}|\ |\partial^2R(\xi,\eta)/\partial\xi^2|$, we only need to estimate the bounds of the absolute values of $\partial R/\partial\xi$ and $\partial^2R(\xi,\eta)/\partial\xi^2$.

Since $\partial R/\partial \xi = -16\xi/(4 + \xi^2 + \eta^2)^2$. One checks that $|\partial R/\partial \xi| \leq 3\sqrt{3}/8$.

On the other hand, $\partial^2 R/\partial \xi^2 = 16(3\xi^2 - \eta^2 - 4)/(\xi^2 + \eta^2 + 4)^3$. One checks that $|\partial^2 R/\partial \xi^2| \leq 1/4$.

Since $0 \leq R, E_{i,j} \leq 1$, $|R - E_{i,j}| \leq 1$. Thus $|\partial/\partial \xi \{[R(\xi,\eta) - E_{i,j}]\partial R(\xi,\eta)/\partial \xi\}|$ $\leq 27/64 + 1/4 = 43/64$.

A similar analysis yields $|\partial/\partial \eta \{[R(\xi,\eta) - E_{i,j}]\partial R(\xi,\eta)/\partial \eta\}| \leq 27/64 + 1/4 = 43/64$.

From (40)

$$\nu_0 \leq 43\sqrt{2}/64. \tag{41}$$

Thus for $\lambda \in [0 , 64\sqrt{2}\pi^2(1 - \pi^2 h^2/24)^2/43)$, the algorithm (33) converges to the unique DSS.

4.2.5 Implementation of the Algorithm and its Complexity

From Subsection 4.2.1 and (29), M^{-1} is known, and therefore, to implement the algorithm (33), one has to multiply the 2Nx2N dense matrix M^{-1} by a vector, which costs $O(N^2)$, using the conventional matrix multiplication. However as discussed in Subsection 4.2.1, we can use FFT to reduce the cost to $O(N(\log N))$.

Let x^* be the unique solution of (28) and let $\lambda\nu_0[2\pi^2(1 - \pi^2 h^2/24)^2]^{-1} = \theta < 1$. Then by a similar argument as in the proof of Proposition 4.1, we have $||x^{(m)} - x^*||_2 \leq \theta\ ||x^{(m-1)} - x^*||_2$, and therefore,

$$||x^{(m)} - x^*||_2 \leq \theta^m\ ||x^{(0)} - x^*||_2 . \tag{42}$$

As an example, we derive the number of iterative steps to compute a solution of (28) with error bound O(h). Let k be the number of steps required, then we have

$$\theta^k = h,$$

$$k = log\ h\ /log\ \theta.$$

If $\theta = 1/2$, then $k = (log\ N)/2$. Thus

Proposition 4.2 For $\lambda \in [0\ ,\ \pi^2\nu_0^{-1}\ [1 - \pi^2h^2/24]^2)$, it takes $(log\ N)\ /2$ steps for $x^{(m)}$ to converge to the solution of (28) with error h. The total cost, using FFT for matrix multiplication, is thus $O(\ N(log\ N)^2\ )$.

4.3 Interpolating Spline and its Optimality

When the data are noisy, the spline-smoothing approach is appropriate. However, when the data are relatively precise, the *interpolating spline* approach is preferable. In that approach, one seeks a spline, which interpolates the data and minimizes (15). This is an *interpolatory algorithm* and is therefore *almost strongly optimal*, i.e., strongly optimal within a factor of 2, see [37] ch. 1. The uniqueness of the spline and its construction need further investigation.

5 Optical Flow

Biological systems typically move relatively continuously through the world, and the images projected on their retinas vary essentially continuously while they move. Such continuous flow of the imaged world across the retina is called *optical flow*. The optical flow assigns to every point on the visual field a two-dimensional "retinal velocity", at which it is moving across the visual field. We study the approximation of the optical flow, or *velocity field*, based on a sequence of images.

Assume that D is a bounded image domain of interest. Without loss of generality, we assume that D is a unit square. We denote the image brightness at $(x,y) \in D$, projected by a surface patch of a moving object at time t, by $E(x,y,t)$. After time Δt, the image of the same surface patch has moved to $(x + \Delta_1 , y + \Delta_2)$, with intensity $E(x + \Delta_1 , y + \Delta_2 , t + \Delta t)$. The average rate of image intensity change at image point (x,y) and within time interval $[t , t + \Delta]$, is

$$\frac{E(x + \Delta_1 , y + \Delta_2 , t + \Delta t) - E(x , y , t)}{\Delta t} =$$

$$\frac{E(x + \Delta_1 , y + \Delta_2 , t + \Delta t) - E(x , y + \Delta_2 , t + \Delta t)}{\Delta_1} \frac{\Delta_1}{\Delta t} +$$

$$\frac{E(x , y + \Delta_2 , t + \Delta t) - E(x , y , t + \Delta t)}{\Delta_2} \frac{\Delta_2}{\Delta t} +$$

$$\frac{E(x , y , t + \Delta t) - E(x , y , t)}{\Delta t} .$$

Let $\Delta t \to 0$. We assume that all the limits exist. We have

$$u(x,y,t) = \lim_{\Delta t \to 0} \frac{\Delta_1}{\Delta t} \quad \text{and} \quad v(x,y,t) = \lim_{\Delta t \to 0} \frac{\Delta_2}{\Delta t}, \tag{43}$$

where (u(x,y,t),v(x,y,t)) is the velocity field at image point (x,y) and time instance t. Thus we have [3, 14]:

$$- w(x,y,t) = p(x,y,t)\, u(x,y,t) + q(x,y,t)\, v(x,y,t) + r(x,y,t), \tag{44}$$

where $p = \partial E/\partial x$, $q = \partial E/\partial y$ and $r = \partial E/\partial t$ can be computed directly. The function - w is the rate of change of intensity, which is not known.

From (44) alone, one cannot determine u, v and w uniquely. Assume that the partial derivatives of u, v and w are square integrable. In addition to requiring that (44) be satisfied, a *consistency constraint* is imposed in [3] and [14], which is the minimization of the quadratic variations of u, v and w. Then u, v and w can be uniquely determined. In Subsections 5.1 and 5.2, we discuss two approaches for approximating u, v and w: *spline-smoothing* and the *interpolating spline*. We approximate the velocity field at an arbitrary time instance t, and we omit the time factor t in the following discussion.

5.1 Spline-smoothing

This approach was used by Horn and Schunck [14] (where $w \equiv 0$), and by Cornelius and Kanade [3]. They seek u,v and w, which minimize

$$\int_D \{ \lambda^{-1}[(u_x)^2 + (u_y)^2 + (v_x)^2 + (v_y)^2] + \mu^{-1}[(w_x)^2 + (w_y)^2] + [pu + qv + w + r]^2 \}\, dx\, dy, \tag{45}$$

where λ and μ are penalty parameters.

5.1.1 Gauss-Seidel Iterative Method

We discretize the unit square region D, with mesh size h, and discretize (45) by using difference operators instead of differential operators and summations instead of integrals. The corresponding *discrete spline-smoothing* is to minimize

$$\kappa = \sum_{i,j} [\ (\lambda^{-1}a_{i,j} + \mu^{-1}b_{i,j}) + c_{i,j}\ ], \tag{46}$$

where

$$a_{i,j} =$$

$$[(u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2 + (v_{i+1,j} - v_{i,j})^2 + (v_{i,j+1} - v_{i,j})^2]/h^2,$$

$$b_{i,j} =$$

$$[(w_{i+1,j} - w_{i,j})^2 + (w_{i,j+1} - w_{i,j})^2]/h^2, \quad \text{and}$$

$$c_{i,j} = (p_{i,j}u_{i,j} + q_{i,j}v_{i,j} + w_{i,j} + r_{i,j})^2,$$

where $u_{i,j}$, $v_{i,j}$, $w_{i,j}$, $p_{i,j}$, $q_{i,j}$ and $r_{i,j}$ are the function values of u, v, w, p, q and r at grid point (ih,jh), respectively, where $i,j = 1, \cdots, K, K + 1 = h^{-1}$.

Let $U = [u_{1,1}, \cdots, u_{1,K}, \cdots, u_{K,K}]^T$, $V = [v_{1,1}, \cdots, v_{1,K}, \cdots, v_{K,K}]^T$, and $W = [w_{1,1}, \cdots, w_{1,K}, \cdots, w_{K,K}]^T$. We want to find U, V and W, which minimize (46).

A *necessary condition* for minimizing (46) is

$$\frac{\partial \kappa}{\partial u_{i,j}} = \frac{\partial \kappa}{\partial v_{i,j}} = \frac{\partial \kappa}{\partial w_{i,j}} = 0, \quad i,j = 1, \cdots, K. \tag{47}$$

Therefore,

$$\begin{aligned}
&4u_{i,j} - u_{i+1,j} - u_{i,j+1} - u_{i-1,j} - u_{i,j-1} = \\
&-\lambda h^2 p_{i,j}(p_{i,j}u_{i,j} + q_{i,j}v_{i,j} + w_{i,j} + r_{i,j}) , \\
&4v_{i,j} - v_{i+1,j} - v_{i,j+1} - v_{i-1,j} - v_{i,j-1} = \\
&-\lambda h^2 q_{i,j}(p_{i,j}u_{i,j} + q_{i,j}v_{i,j} + w_{i,j} + r_{i,j}) ,
\end{aligned} \tag{48}$$

and

$$4w_{i,j} - w_{i+1,j} - w_{i,j+1} - w_{i-1,j} - w_{i,j-1} =$$
$$-\mu h^2(p_{i,j}u_{i,j} + q_{i,j}v_{i,j} + w_{i,j} + r_{i,j}) .$$

To solve this system of linear equations, the Gauss-Seidel iterative method is proposed in [3, 14]. The convergence of the algorithm remains to be analyzed. Furthermore, even if the Gauss-Seidel iterative method converges for this case, it is known to converge slowly [39].

5.1.2 Conjugate Gradient Iterative Method

Let $N = K^2$. Then the system of linear equations (48) can be rewritten as

$$M \begin{pmatrix} U \\ V \\ W \end{pmatrix} = - h^2 \begin{pmatrix} \lambda PR \\ \lambda QR \\ \mu R \end{pmatrix} , \tag{49}$$

where the diagonal NxN matrices

$$P = \mathrm{diag}(p_{1,1}, \cdots, p_{1,K}, \cdots, p_{K,K}) \quad \text{and}$$
$$Q = \mathrm{diag}(q_{1,1}, \cdots, q_{1,K}, \cdots, q_{K,K}), \tag{50}$$

the Nx1 vector $R = [r_{1,1}, \cdots, r_{1,K}, \cdots, r_{K,K}]^T$, and the coefficient matrix

$$M = \begin{pmatrix} A + \lambda h^2P^2 & \lambda h^2PQ & \lambda h^2P \\ \lambda h^2PQ & A + \lambda h^2Q^2 & \lambda h^2Q \\ \mu h^2P & \mu h^2Q & A + \mu h^2I \end{pmatrix} , \tag{51}$$

where the NxN matrix A is given in (20).

The coefficient matrix M is not symmetric for $\lambda = \mu$. To symmetrize M, we multiply the last equation of (48) by λ/μ, and we have

$$M^* \begin{pmatrix} U \\ V \\ W \end{pmatrix} = - \lambda h^2 \begin{pmatrix} PR \\ QR \\ R \end{pmatrix}, \tag{52}$$

where

$$M^* = \begin{pmatrix} A + \lambda h^2 P^2 & \lambda h^2 PQ & \lambda h^2 P \\ \lambda h^2 PQ & A + \lambda h^2 Q^2 & \lambda h^2 Q \\ \lambda h^2 P & \lambda h^2 Q & (\lambda/\mu)A + \lambda h^2 I \end{pmatrix} . \tag{53}$$

We have

Proposition 5.1 The coefficient matrix M^* is sparse, symmetric and positive definite. []

Proof. Matrix M^* is sparse, since A is sparse, and P and Q are diagonal matrices. We now show that M^* is positive definite. Let $x = [U,V,W]^T$. Then we have

$$<M^* x , x> = \tag{54}$$

$$< \begin{pmatrix} A & & \\ & A & \\ & & (\lambda/\mu)A \end{pmatrix} x , x > + \lambda h^2 <PU+QV+W, PU+QV+W>.$$

Since $<PU + QV + W , PU + QV + W> = \|PU + QV + W\|_2^2 \geq 0$ and A is positive definite, see Subsection 4.2.1, and therefore $(\lambda/\mu)A$ is also positive definite, we have

$<M^* x , x> > 0$ for $x \neq 0$. []

Since M^* is non-degenerate, we have

Corollary 5.1 The systems of linear equations (52) and hence (49) have one and only one solution. []

To solve for it, we propose using the *conjugate gradient* iterative method, see [8] ch.7. It converges much faster than Gauss-Seidel, and is *optimal* in terms of computational complexity, see [40].

On the other hand, algorithms with *simple, local and parallel* operations are more preferable in image understanding, since they are suitable for parallel or distributed computation and feasible for a biological system. The conjugate gradient method requires global interaction, which might not be desirable. We propose using the *Chebyshev* iterative method, which favors simple, local and parallel operations, and converges much faster than Gauss-Seidel.

5.1.3 Chebyshev Iterative Method

We estimate the minimum and the maximum eigenvalues of the matrix M^* in (53), since this is crucial for applying the Chebyshev iterative method. The minimum and the maximum eigenvalues of M^* are

$$\lambda_{min}(M^*) = \inf \{ \langle M^* x , x\rangle : \|x\|_2 = 1 \}$$
$$\text{and} \qquad (55)$$
$$\lambda_{max}(M^*) = \sup \{ \langle M^* x , x\rangle : \|x\|_2 = 1 \} .$$

Let

$$F = \begin{pmatrix} A & & \\ & A & \\ & & (\lambda/\mu)A \end{pmatrix} . \qquad (56)$$

Then

$$\langle M^* x , x\rangle = \langle F x , x\rangle + \lambda h^2 \langle PU+QV+W, PU+QV+W\rangle. \qquad (57)$$

$$\inf\{\langle F x , x\rangle : \|x\|_2=1\} = \lambda_{min}(F) = \min\{\lambda_{min}(A), (\lambda/\mu)\lambda_{min}(A)\}$$
$$= 8\theta \sin^2[\pi/2(K+1)] = 8\theta \sin^2(\pi h/2), \quad \text{where } \theta = \min \{ 1 , \lambda/\mu \}.$$

$$\sup\{<F\ x\ ,\ x>:\ ||x||_2=1\} = \lambda_{max}(F)=\max\{\lambda_{max}(A),(\lambda/\mu)\lambda_{max}(A)\}$$

$$= 8\eta cos^2[\pi/2(K+1)] = 8\eta cos^2(\pi h/2),\quad \text{where } \eta = \max\ \{\ 1\ ,\ \lambda/\mu\ \}.$$

$$\sup\ \{\ \lambda h^2<PU + QV + W\ ,\ PU + QV + W>:\ ||x||_2 = 1\ \} =$$

$$\sup\ \{\ \lambda h^2||PU + QV + W||_2^{\,2} :\ ||x||_2 = 1\ \}$$

$$\leq \lambda h^2 C^2,$$

where $C = \max\ \{|p_{i,j}|\} + \max\ \{|q_{i,j}|\} + 1.$

We have

Proposition 5.2 The minimum and the maximum eigenvalues of the matrix M^*

$$8\theta sin^2(\pi h/2) \leq \lambda_{min}(M^*) \leq 8\theta sin^2(\pi h/2) + \lambda h^2 C^2,$$

and (58)

$$8\eta cos^2(\pi h/2) \leq \lambda_{max}(M^*) \leq 8\eta cos^2(\pi h/2) + \lambda h^2 C^2,$$

where

$$\theta = \min\ \{\ 1,\ \lambda/\mu\ \},$$

$$\eta = \max\ \{\ 1,\ \lambda/\mu\ \} \text{ and}$$

$$C = \max\ \{|p_{i,j}|\} + \max\ \{|q_{i,j}|\} + 1.$$

The condition number of M^* is

$$\lambda_{max}/\lambda_{min} \leq \{8\eta cos^2(\pi h/2) + \lambda h^2 C^2\}/\{8\theta sin^2(\pi h/2)\}. \quad (59)$$

□

For small h, the condition number of M^* is of order h^{-2}.

Based on the estimation of the minimum and the maximum eigenvalues, we propose using the Chebyshev method, which converges much faster than Gauss-Seidel, and involves only simple, local and parallel operations. We can choose $8\theta sin^2(\pi h/2)$ and $8\eta cos^2(\pi h/2) + \lambda h^2C^2$ as lower and upper bounds of λ_{min} and λ_{max}, respectively. For small h, the bounds are tight, and therefore, the Chebyshev method is numerically stable. Chebyshev method is also optimal in terms of computational complexity, see [40]. For the Chebyshev methods, see [8] ch.4 - 6 and Appendix A, which includes FORTRAN subroutines.

5.1.4 Existence and Uniqueness of the Solution

One checks that the block matrix

$$\begin{pmatrix} (\partial^2\kappa/\partial u_{i,j}\partial u_{k,l})_{i,j;k,l} & (\partial^2\kappa/\partial u_{i,j}\partial v_{k,l})_{i,j;k,l} & (\partial^2\kappa/\partial u_{i,j}\partial w_{k,l})_{i,j;k,l} \\ (\partial^2\kappa/\partial v_{i,j}\partial u_{k,l})_{i,j;k,l} & (\partial^2\kappa/\partial v_{i,j}\partial v_{k,l})_{i,j;k,l} & (\partial^2\kappa/\partial v_{i,j}\partial w_{k,l})_{i,j;k,l} \\ (\partial^2\kappa/\partial w_{i,j}\partial u_{k,l})_{i,j;k,l} & (\partial^2\kappa/\partial w_{i,j}\partial v_{k,l})_{i,j;k,l} & (\partial^2\kappa/\partial w_{i,j}\partial w_{k,l})_{i,j;k,l} \end{pmatrix}$$

equals $2\lambda^{-1}h^{-2}M^*$, which is positive definite. Therefore, the unique solution of TT4.6 obtains the minimal value of (46). We have

Proposition 5.3 There exists a unique solution for minimizing (46), which is the unique solution of the system of linear equations (52). []

5.2 Interpolating Splines

The interpolating spline approach is to find u, v and w, which satisfy the information constraint (44) exactly and minimize

$$\int_D \{ [(u_x)^2 + (u_y)^2 + (v_x)^2 + (v_y)^2] + \mu^{-1}[(w_x)^2 + (w_y)^2] \}dx\, dy, \qquad (60)$$

where μ is a chosen parameter.

Similar to the spline-smoothing approach, we discretize (60) on the unit square region D with mesh size h. Then the corresponding *discrete interpolating spline* satisfies

$$p_{i,j}u_{i,j} + q_{i,j}v_{i,j} + w_{i,j} + r_{i,j} = 0, \quad i,j = 1, \cdots, K, \quad K + 1 = h^{-1}, \tag{61}$$

and minimizes

$$\kappa = \sum_{i,j} [\, a_{i,j} + \mu^{-1} b_{i,j} \,], \tag{62}$$

where

$$a_{i,j} = [(u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2 + (v_{i+1,j} - v_{i,j})^2 + (v_{i,j+1} - v_{i,j})^2],$$

and

$$b_{i,j} = [(w_{i+1,j} - w_{i,j})^2 + (w_{i,j+1} - w_{i,j})^2].$$

(There is a short discussion of this approach in [14] for w≡0).

To solve this constrained minimization problem, we apply the method of Lagrange multipliers, and we have

$$- M \begin{pmatrix} U \\ V \\ W \\ \Lambda \end{pmatrix} = - \begin{pmatrix} 0 \\ 0 \\ 0 \\ R \end{pmatrix}, \tag{63}$$

where $\Lambda = [\lambda_{1,1}, \cdots, \lambda_{1,K}, \cdots, \lambda_{K,K}]^T$, the coefficient matrix

$$M = \begin{pmatrix} A & 0 & 0 & P \\ 0 & A & 0 & Q \\ 0 & 0 & A & \mu I \\ P & Q & I & 0 \end{pmatrix}, \tag{64}$$

and U, V, W, P, Q and R are defined similarly as in (49). The coefficient matrix M is neither symmetric nor positive definite. It is easy to symmetrize M by dividing the third row by μ. From (63) and (64), we have

$$AU + P\Lambda = 0, \quad AV + Q\Lambda = 0, \quad AW + \mu\Lambda = 0, \quad \text{and}$$

$$PU + QV + W = -R.$$

So

$$U = -A^{-1}P\Lambda, \quad V = -A^{-1}Q\Lambda, \quad W = -\mu A^{-1}\Lambda, \tag{65}$$

$$P(-A^{-1}P\Lambda) + Q(-A^{-1}Q\Lambda) + (-\mu A^{-1}\Lambda) = -R, \text{ i.e.,}$$

$$(PA^{-1}P + QA^{-1}Q + \mu A^{-1})\Lambda = R.$$

Therefore, the problem is reduced to solving a system of linear equations

$$M^* \Lambda = R, \tag{66}$$

where

$$M^* = PA^{-1}P + QA^{-1}Q + \mu A^{-1}.$$

After solving (66), U, V and W can be obtained directly from (65). We analyze the matrix M^*, and then discuss iterative methods for solving the system of linear equations (66).

5.2.1 Matrix M^*

We estimate the minimum and the maximum eigenvalues of the matrix M^* in (66), since this is crucial for applying the Chebyshev iterative method. We have

$$\langle M^* x , x\rangle =$$

$$\langle (PA^{-1}P)x , x\rangle + \langle (QA^{-1}Q)x , x\rangle + \langle \mu A^{-1}x , x\rangle,$$

$$\inf \{ \langle \mu A^{-1}x , x\rangle \} = \mu[8cos^2(\pi h/2)]^{-1} \text{ and}$$

$$\sup \{ \langle \mu A^{-1}x , x\rangle \} = \mu[8sin^2(\pi h/2)]^{-1}.$$

$$\langle (PA^{-1}P)x , x\rangle + \langle (QA^{-1}Q)x , x\rangle =$$

$$||P||_2^2\langle A^{-1}(Px)/||P||_2,(Px)/||P||_2\rangle + ||Q||_2^2\langle A^{-1}(Qx)/||Q||_2,(Qx)/||Q||_2\rangle.$$

$$\sup \{ <(PA^{-1}P)x , x> + <(QA^{-1}Q)x , x>: \|x\|_2 = 1 \}$$

$$\leq \|P\|_2^2 \lambda_{max}(A^{-1}) + \|Q\|_2^2 \lambda_{max}(A^{-1})$$

$$= C[8sin^2(\pi h/2)]^{-1}, \quad \text{where } C = \max \{ p_{i,j}^2 \} + \max \{ q_{i,j}^2 \}.$$

$$\lambda_{min}(M^*) \geq \inf \{ <\mu A^{-1}x , x>: \|x\|_2 = 1 \} = \mu[8cos^2(\pi h/2)]^{-1}.$$

$$\lambda_{max}(M^*)$$

$$\leq \sup\{<(PA^{-1}P)x,x> + <(QA^{-1}Q)x , x>\} + \sup\{<\mu A^{-1}x , x>\}$$

$$\leq C[8sin^2(\pi h/2)]^{-1} + \mu[8sin^2(\pi h/2)]^{-1} = (C + \mu)[8sin^2(\pi h/2)]^{-1}.$$

$$\lambda_{max}(M^*) \geq \sup \{ <\mu A^{-1}x , x> \}$$

$$= \mu[8sin^2(\pi h/2)]^{-1}.$$

Thus we have

Proposition 5.4 The minimum and the maximum eigenvalues of the positive definite and symmetric matrix M^*

$$\lambda_{min}(M^*) \geq \mu sec^2(\pi h/2)/8,$$

and (67)

$$\mu csc^2(\pi h/2)/8 \leq \lambda_{max}(M^*) \leq (C + \mu)csc^2(\pi h/2)/8, \quad (68)$$

where $C = \max \{ p_{i,j}^2 \} + \max \{ q_{i,j}^2 \}$.

The condition number of matrix M^* is

$$\lambda_{max}(M^*)/\lambda_{min}(M^*) \leq \{(C + \mu)/\mu\}cot^2(\pi h/2). \quad (69)$$

□

For small h, the condition number of M^* is of order h^{-2}.

Since M^* is non-degenerate, we have

Corollary 5.2 The systems of linear equations (66) and (63) have one and only one solution. []

5.2.2 Iterative Methods and their Implementation

Since the coefficient matrix M^* of (66) is symmetric and positive definite, the conjugate gradient iterative method can be used.

As explained in the spline-smoothing approach, based on the estimation of the minimum and the maximum eigenvalues of M^*, we propose using the Chebyshev method.

The coefficient matrix M^* is dense, and each iterative step requires multiplying this matrix by a vector, which costs $O(N^2)$, using the conventional matrix multiplication. However, we can decompose M^* into a sum of $PA^{-1}P$, $QA^{-1}Q$ and μA^{-1}. Since A^{-1} has a special structure, as indicated in Subsection 4.2.1, we can use the FFT, to reduce the cost of each matrix multiplication to O(N (*log* N)), and therefore, each matrix multiplication by M^* costs O(N (*log* N)). After solving (66), we can compute U, V and W from (65), with cost O(N (*log* N)).

3. Chapter 3

Comparator with Completion Signal

Abstract

We provide a design of a binary comparator with completion signal, for the purpose of optimizing the average processing time. The average propagation delay is a constant, independent of n, the number of inputs, while the logic complexity is a linear function of n.

1 Introduction

Binary comparators can be implemented in tree form. A tree circuit of a binary comparator is shown in Fig.1-1. The inputs a_i and b_i to the primary module P_i represent the i-th bits of binary numbers A and B, which are to be compared. (The most significant bits are a_1 and b_1). The output of P_i specifies, in suitably coded form, whether a_i is equal to, smaller than or greater than b_i, referred to as E, S or G, respectively. Each I-module receives signals derived from two other modules (I or P) . The I-module processes these data and emits a signal that indicates the relative size of the corresponding parts of A and B, confined by the leaves of the subtree with the I-module as the root. The output from the root I-module of the whole tree provides the final result of the comparison of A and B. The functions performed by the P- and I-modules are shown in Fig.1-2 and 1-3, respectively, and the coding will be discussed in the next section. It is obvious that the gate complexity is a linear function of n, the number of inputs, and the propagation delay is proportional to the logarithm of n. For tree-like comparators, see [28] and [43], and for the implementation, see [26], [30], and [36].

2 A Binary Comparator with Completion Signal

From the I-module function it is obvious that once the left input is S (or G) the output of this module will be 'trapped' into S (or G). Therefore, if the left input to the leftmost I-modules on a level is S (or G), then the outputs of this module and all the leftmost I-modules down to the root I-module will be all the same. Thus, if the left input to a leftmost I-module is S (or G), then the final output will be the same. If we could identify the first leftmost I-module, which emits an S (or G) output, 'extract' it as the final answer, and signal the completion, then we might reduce the propagation delay on the average, since we do not have to wait until the signal propagates down to the root I-module.

We design a binary comparator with a tree structure, which generates a completion signal at the earliest level of the tree circuit and delivers the final correct result, with no hazard. The added cost relative to the circuit of Fig.1-1 is small. We first design the P- and I-modules and then complete the design of the comparator. The delay analysis is given in the next section. The worst case delay is, of course, still $\log n$, but the average case delay is a constant, independent of the number of inputs. For the basic concepts and a general discussion of combinational circuits with completion signal, see [42] and [41].

The functions performed by the P- and I-modules are the same as that in Fig.1-2 and 1-3, respectively. The coding and logic expressions of the P- and I-modules are given in Fig.2-1, 2-2 and 2-3, respectively. We choose the 1-hot code

to avoid a hazard; this will be made clear later when we discuss the completion signal.

Fig.2-4 depicts a design of a binary comparator with a completion signal. The S- and G-outputs of all the leftmost modules are merged into the OR gates S and G, respectively. The outputs of gates S and G along with the E-output of the root I-module provide the final result in coded form, as specified in Fig.2-1, when the processing has been completed. On the other hand, these outputs are merged into another OR gate C, which was 0 before processing and signifies the completion of processing by emitting a 1. Thus as soon as the output of gate C becomes 1, information of the relative size of A and B is ready. From the design it is clear now that the 1-hot code of the P- and I-modules avoids the delay hazard, which might cause false completion signal and a premature erroneous result. The logic complexity is a linear function of n, i.e., 2n - 1 P- and I-modules and three added OR gates. The fan-in to and fan-out from the P- and I-module are not a problem, and the fan-in to the OR gates S (and G) and C are (log n) + 1 and 3, respectively, which is not a problem either, except for very large n.

The key issue which facilitates the design is that the flow table of the binary comparator has 'trap' rows, i.e. , once the system enters into the state S or G, corresponding to a 'trap' row, it will be 'trapped' in that state, independent of the inputs. The design is applicable in general to other logic functions with trap rows in their flow tables.

3 Delay analysis

In this section we analyze the average delay of the binary comparator designed in the previous section. Assume that the delay through either a P- or an I- module is d and that n, the number of bit positions of A and B, is a power of 2, i.e. , $n = 2^k$. Label the levels of the tree circuit from the P-modules (leaves) to the root I-module as 1 to k+1.

For each bit position there are 4 possible bit configurations and so there are 4^n configurations of inputs. Assign a uniform probability distribution to the configurations, i.e., each configuration has a probability measure $1/4^n$.

We first compute S_i, the number of configurations which generate completion signals exactly at level i, i = 1,...,k+1. The leftmost 2^{i-1} bits of A and B are involved in generating completion signals at level i, and the remaining $(n - 2^{i-1})$ bits are arbitrary, which correspond to $4^{n-2^{i-1}}$ configurations. For i=1, only one bit position of A and B is involved; the other n-1 bits are arbitrary. Hence $S_1 = 2 \cdot 4^{n-1}$. For i>1, the leftmost 2^{i-2} bits must generate ESG = 100 (equal), and the other 2^{i-2} bits of the 2^{i-1} involved bits must generate ESG = 010 or 001 (smaller than or greater than), since otherwise the completion signal would be generated either before or after the i-th level. The 2^{i-2} bits generating ESG = 100 correspond to $2^{2^{i-2}}$ configurations. The other 2^{i-2} bits correspond to $(4^{2^{i-2}} - 2^{2^{i-2}})$ configurations, since the total number of configurations is $4^{2^{i-2}}$, among which $2^{2^{i-2}}$ correspond to ESG = 100. Thus for i>1:

$$S_i = 2^{2^{i-2}} (4^{2^{i-2}} - 2^{2^{i-2}}) 4^{n-2^{i-1}} = 4^n (2^{-2^{i-2}} - 2^{-2^{i-1}}).$$

The average delay through P- and I-modules (in unit d)

$$D = \{ S_1 + \sum_{i=2}^{k+1} i\, S_i \}/4^n$$

$$= \{ 2*4^{n-1} + \sum_{i=2}^{k+1} i\, 4^n\, (2^{-2^{i-2}} - 2^{-2^{i-1}}) \}/4^n$$

$$= 2^{-1} + \sum_{i=2}^{k+1} i\, (2^{-2^{i-2}} - 2^{-2^{i-1}})$$

$$< 2^{-1} + \sum_{i=2}^{\infty} i\, (2^{-2^{i-2}} - 2^{-2^{i-1}})$$

$$= 2^{-1} + \sum_{i=1}^{\infty} (i+1)\, 2^{-2^{i-1}} - \sum_{i=2}^{\infty} i\, 2^{-2^{i-1}}$$

$$= 2^{-1} + 1 + \sum_{i=2}^{\infty} 2^{-2^{i-1}}$$

$$= 1 + 2^{-1} + 2^{-2} + 2^{-4} + 2^{-8} + \sum_{i=4}^{\infty} 2^{-2^i}.$$

Since

$$\sum_{i=4}^{\infty} 2^{-2^i} < \int_3^{\infty} 2^{-2^x}\, dx = \int_8^{\infty} 2^{-y}\, \frac{dy}{y} < 2^{-3} \int_8^{\infty} 2^{-y} dy = \frac{2^{-11}}{\ln 2},$$

$$D < 1 + 2^{-1} + 2^{-2} + 2^{-4} + 2^{-8} + \frac{2^{-11}}{\ln 2}$$

< 1.8187.

If double rail inputs of A and B are available, then from the P- and I-module logic expressions, it is obvious that d is 2 AND or OR gate delays, whenever we use two level logic. Therefore, the average delay through the P- and I-modules is 3.6374 gate delay. Including the delay of the OR gates, which generate the completion signal and the final result, the time duration of processing ranges from 4 to $2(log\ n) + 4$ gate delay, with the average less than 5.6374 gate delay. If only single rail inputs of A and B are available, then we have to add in one more inverter delay.

We have assumed that the input configurations have a uniform distribution. However, in different implementation environments, the probability distributions might be different. Thus in certain cases, the average computing time might be larger than the value derived here. This is particularly true when the equality of A and B, the worst case for the computation delay, occurs with relatively large frequency.

The number of inputs to the OR gates S (and G) is $(log\ n) + 1$. Let r be the gate fan-in. Then for $(log\ n) + 1 \leq r$, i.e., $n \leq 2^{r-1}$, this will not cause any problem. For very large n, such that $(log\ n) + 1 > r$, i.e., $n > 2^{r-1}$, the completion circuit could be implemented by a tree type multi-level gate network. This would introduce an additional delay approximately $log_r(log\ n)$. On the other hand, to implement this tree type gate network, approximately $2\{(log\ n) - 1\}/(r - 1)$ OR gates are required.

References

1. Babenko, K.I. (editor). *Theoretical Foundations and Construction of Computational Algorithms for the Problems of Mathematical Physics.* , Moscow, 1979(in Russian).

2. Bruss, A. R. "Some Properties of Discontinuities in the Image Irradiance Equation." *AI Memo, AI Lab. MIT, 517* (1979).

3. Cornelius, N., and Kanade, T. Adapting Optical-flow to Measure Object Motion in Reflectance and X-ray Image Sequences. Proceeding of Image Understanding Workshop, DARPA,, July, 1983.

4. Diaconis, P. and Freedman, D. "Frequency Properties of Bayes Rules." *Scientific Interference, Data Analysis, and Robustness* (1983), 105-115.

5. Duchon, J. "Interpolation de Fonctions de Deux Variables Suivant le Principe de la Flexion des Plaques Minces." *Revue Francaise d'Automatique, Informatique et Recherche Operationnelle* (Dec. 1976), 5-12.

6. Gaffney, P.W. and Powell, M.J.D. "Optimal Interpolation." *Numerical Analysis (G.A. Watson, ed.), Lecture Notes in Math. 506,* (Springer Verlag, Berlin and New York, 1976), 90-100.

7. Grimson, W. E. L.. *From Images to Surfaces.* MIT Press, 1981.

8. Hageman, L.A., and Young, D.M.. *Applied Iterative Methods.* Academic Press, 1981.

9. Horn, B.K.P. "Shape from Shading: a Method for Obtaining the Shape of a Smooth Opaque Object from One View." *TechnicalReportMAC-TR-79* , (Project MAC, MIT 1970) .

10. Horn, B.K.P. Determining Shape from Shading. In *The Psychology of Computer Vision*, Winston, P.H., Ed.,McGraw-Hill, New York, 1975, ch. 4.

11. Horn, B.K.P. "Understanding Image Intensities." *Artificial Intelligence 8,* 2 (April 1977), 201-231.

12. Horn, B.K.P., Woodham, R.J. and Silver, W.M. "Determing Shape and Reflectance Using Multiple Images." *AIMemo* , AI Lab. MIT 490 (1978) .

13. Horn, B.K.P., and Sjoberg, R.W. "Calculating the Reflectance Map." *Applied Optics 18* (June 1979), 1770-1779.

14. Horn, B.K.P., and Schunck, B.G. "Determining Optical Flow." *Artificial Intelligence 17* (1981), 185-203.

15. Ikeuchi, K. and Horn, B. K .P. "Numerical Shape from Shading and Occluding Boundaries." *Artificial Intelligence 17* (1981), 141-184.

16. Kadane, J. B., Wasilkowski, G. W. and Wozniakowski, H. "Can Adaption Help on the Average for Stochastic Information?" *Dept. of Comp. Sci. Rep., Columbia University* (1984) .

17. Kender, J. R.. *AI Research Notes. Volume : Shape from Texture.* Pitman Publishing Ltd. London, Accepted for publication, 1984.

18. Kender, J. R., and Lee, D. "The Information-Centered Approach to Algorithms Applied to the 2-1/2 D Sketch: Splines are Optimal Interpolators." *Dept. of Comp. Sci. Rep., Columbia University* (1984).

19. Kimeldorf, G. S. and Wahba, G. "A Correspondence between Bayesian Estimation on Stochastic Process and Smoothing by Splines." *Ann. Math. Statisti. 41,2* (1970), 495-502.

20. Kimeldorf, G. S. and Wahba, G. "Spline Functions and Stochastic Processes." *Sankhya, Ser. A 32* (1970), 173-180.

21. Kuo, H. *Gaussian Measure in Banach Spaces.* Lecture Notes in Math. 463, Springer-Verlag, Berlin, 1975.

22. Larkin, F.M. "Gaussian Measure in Hilbert Space and Application in Numerical Analysis." *Rocky Mt. J. Math. 2* (1972), 378-421.

23. Laurent, P. J. *Approximation et optimisation.* Hermann, Paris, 1972.

24. Meinguet, J. "Multivariate Interpolation at Arbitrary Points Made Simple." *J. of Applied Mathematics and Physics (ZAMP)* , 30 (1979), 292-304.

25. Meinguet, J. "Surface Spline Interpolation: Basic Theory and Computational Aspects." *Institut de Mathematique Pure dt Appliquee, Universite Catholique de Louvain* , 35 (1983).

26. Molesworth, W. H. "Simpler Carry Look-ahead Digital Comparison." *Electron. Eng. 50, no. 610* (Aug. 1978), 20.

27. Nicodemus, F. E., Richmond, J. C., Hsia, J. J., Ginsberg, I. W., and Limperis, T. "Geometrical Considerations and Nomenclature for Reflectance." *NBS monograph 160 US Department of Commerce* (1977), National Bureau of Standards.

28. Papachristou, C. A. "Parallel Implementation of Binary Comparison Circuits." *Int. J. Electron. 47* (1979), no.2, 187-92.

29. Parthasarathy, K.R. *Probability Measures on Metric Spaces.* Academic Press, New York, 1967.

30. Rao, M. V. Subba, and Mittal, S. C. "SN 7485 Gives Carry Look Ahead Digital Comparison." *Electron. Eng. 50,* 603 (1978), 21.

31. Sacks, J. and Ylvisaker, D. "Statistical Designs and Integral Approximation." *Proc. of the 12th Biennial Seminar of the Canadian Math. Congress* (1970), 115-136.

32. Silver, W. "Determining Shape and Reflectance Using Multiple Images." *Maser Th. Dept. of E. E. and CS* (MIT 1980).

33. Smith, G. D.. *Numerical Solution of Partial Differential Equations: Finite Difference Methods.* Oxford University Press, 1978.

34. Suldin, A. V. "Wiener Measure and its Applications to Approximation Methods." *I, II, Izv. Vyss. Ucebn. Zaved. Mat. 6,5,* 13,18 (1959, 1960, Math. Rev.28 #722, 29 #1482), 145-158, 165-179.

35. Terzopoulos, D. *Multiresolution Computation of Visible-Surface Representations.* Ph.D. Th., MIT Artificial Intelligence Lab., 1984.

36. (Ed.). *The TTL Data Book for Design Engineers.* Texas Instruments inc., 2nd ed.

37. Traub, J. F., and Woźniakowski, H. *A General Theory of Optimal Algorithms.* Academic Press, 1980.

38. Traub, J.F., Wasilkowski, G.W., and Woźniakowski, H. *Information, Uncertainty, Complexity.* Addison-Wesley,Reading,MA, 1983.

39. Traub, J. F. and Woźniakowski, H. Information and Computation. Advances in Computers, Vol.23, 1984, pp. 35-92.

40. Traub, J. F. and Woźniakowski, H. "On the Optimal Solution of Large Linear Systems." *J. of ACM 31,* 3 (1984), 545-559.

41. Unger, S. H.. *Asynchronous Sequential Switching Circuits.* New York: Wiley, 1969.

42. Unger, S. H. "The Generation of Completion Signals in Iterative Combinational Circuits." *IEEE Trans: Compt., C-26,* no. 1 (Jan. 1977).

43. Unger, S. H. "Tree Realization of Iterative Circuits." *IEEE Trans: Compt., C-26,* no. 4 (April 1977).

44. Vakhania, N. N. *Probability Distributions on Linear Spaces.* North Holland, New York, 1981.

45. Wahba, G. "Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression." *J. R. Statist. Soc. B 40,* 3 (1978), 364-372.

46. Wasilkowski, G. W. "Optimal Algorithms for Linear Problems with Gaussian Measures." *To appear in Rocky Mt. J. Math.*

47. Wasilkowski, G. W. "Information of Varying Cardinality." (in preparation).

48. Wasilkowski, G. W. "Local Average Errors." , (Dept. of Comp. Sci. Rep., Columbia University, 1983) .

49. Wasilkowski, G. W., and Woźniakowski, H. "Can Adaption Help on the Average?" *Num. Math. 44,* (1984), 169-190.

50. Wasilkowski, G. W., and Woźniakowski, H. "On Optimal Algorithms in an Asymptotic Model with Gaussian Measure." , (Dept. of Comp. Sci. Rep., Columbia University, 1984) .

51. Wiener, N. "Differential Space." *J. Math. Phys. 2,* (1923), 131-174.

52. Wiener, N. "The Average Value of a Functional." *Proc. London Math. Soc. 22,* (1924), 454-467.

53. Woodham, R.J. A Cooperative Algorithm for Determining Surface Orientations from a Single View. Proceedings of the Fifth International Joint Conference on Art ificial Intelligence, Aug., 1977, pp. 635-641.

54. Woodham, R.J. *Reflectance Map Techniques for Analyzing Defects in Metal Castings.* Ph.D. Th., MIT Artificial Intelligence Lab., June 1978. Available as AI-TR-457.

55. Woodham, R.J. "Photometric Method for Determining Surface Orientation from Multiple Images." *OpticalEngineering 19(1)* (1980), 139-144.

56. Wozniakowski, H. "Private communication."

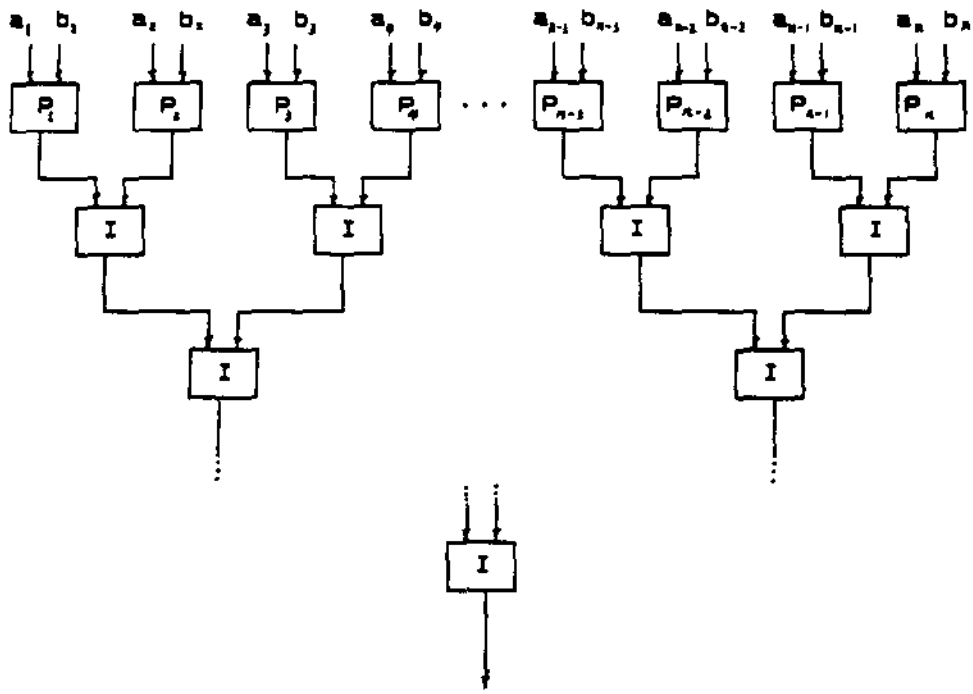


Fig. 1-1 Tree Circuit of Binary Comparator

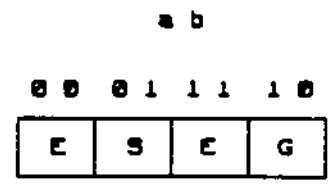


Fig. 1-2 P-module Function

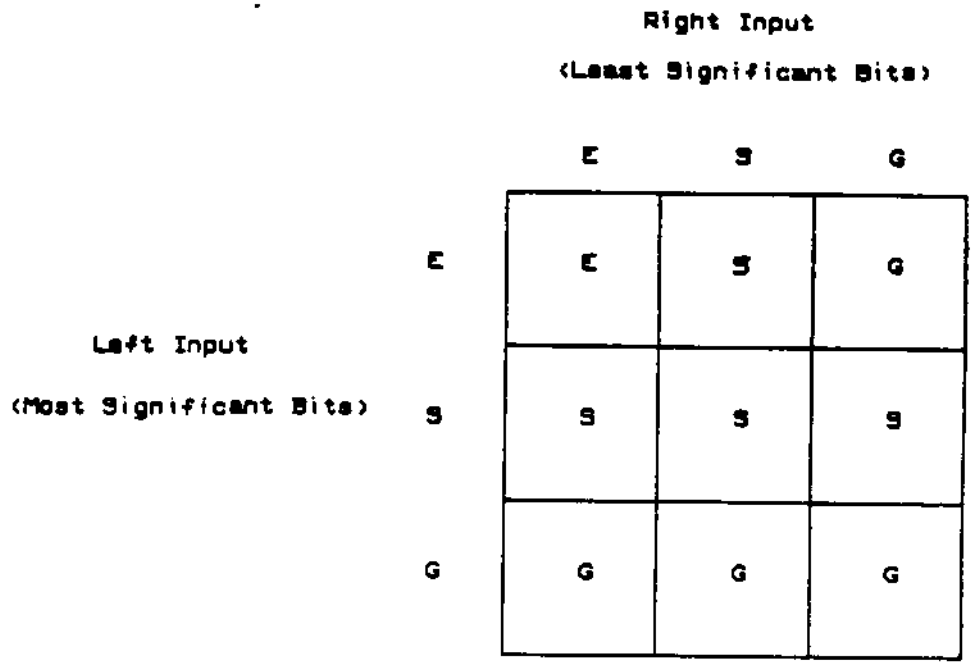
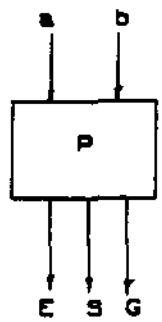


Fig. 1-3 I-module Function

Conditions	E	S	G
A = B	1	0	0
A < B	0	1	0
A > B	0	0	1

Fig. 2-1 Coding



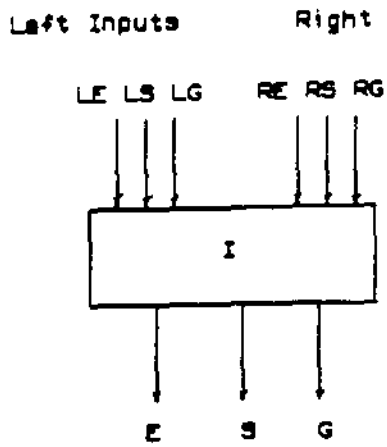
a	b	E	S	G
0	0	1	0	0
1	1			
0	1	0	1	0
1	0	0	0	1

$E = ab + \bar{a}\bar{b}$

$S = \bar{a}b$

$G = a\bar{b}$

Fig. 2-2 P-module, its Truth Table and Logic Expressions



Left Inputs			Right Inputs			Outputs		
LE	LS	LG	RE	RS	RG	E	S	G
1	0	0				RE	RS	RG
0	1	0	-	-	-	0	1	0
0	0	1	-	-	-	0	0	1

$$E = LE * RE$$

$$S = LE * RS + LS$$

$$G = LE * RG + LG$$

Fig. 2-3 I-module, its Truth Table and Logic Expressions

Fig. 2-4 Binary Comparator with Completion Signal C and Outputs E, S and G