in Proceedings of the Eighth Inverse Problems in Engineering Seminar, A. Maniatty, ed., Rensselaer Polytechnic Institute, 1997

Forward Modeling for an Inverse Linear Elasticity Problem Using Bayesian Statistics

Christopher J.R. Clutz and Antoinette M. Maniatty

Abstract

The statistical analysis of the recovery of boundary traction information in linear elasticity is considered. Specifically, the statistical analysis carried out in [8] is extended to include modelization error. The general approach used here is based on Bayesian statistics and was developed in [1]. Specifically, a Gaussian probability density function is used to describe the forward problem of obtaining internal strains or displacements from a given boundary traction distribution. A method using finite element analysis is proposed for defining the central point and the "spread" (covariance operator) of this Gaussian function. Two numerical examples are presented.

1 Introduction

The problem of determining unknown traction boundary distributions on a linear elastic body when some information about the displacement field is known is an inverse problem. Such problems, in general, are ill–posed in the sense that the constraints of existence, uniqueness and continuous dependence on the data which apply to solutions in linear elasticity [13], do not hold. Ill–posed problems in general are examined in [6, 10, 4]. Calculating boundary information using internal data in elasticity was considered in [3] using a least squares minimization method, which proved to be unstable in many cases. Recovery of boundary information in thermoelasticity was analyzed in [7] and recovery of boundary tractions from internal data in elasticity was addressed in [9]. There, least squares minimization was used with a penalty term to stabilize the solution with respect to changes (or uncertainties) in the data. This approach was improved in [12] and was compared with a statistical approach based on Bayesian statistics in [8]. It is the statistical analysis performed in [8], based on methods developed in [1], which is the basis for this paper.

2 A Bayesian Statistical Approach

The Bayesian statistical approach to general inverse problems, developed in [1], is based on describing information about a physical system in terms of probability density functions on some parameter space. This space is usually a subset of $\mathbb{R}^n$ and represents the possible values of a finite set of parameters which characterize the physical system. For instance, if the system is an ideal gas it can be parameterized by its volume, temperature, and pressure. The parameters of a system may be classified as "model" or "data" parameters. Suppose there are N_M model parameters and N_D data parameters. The former represent the non-observable properties of the system, while the latter characterize those properties which can be measured experimentally. The set of all possible N_D-tuples of data parameter

values defines the data space, D, and the set of all possible N_M-tuples of model parameter values defines the model space, M. Thus the total parameter space for the physical system is $D \times M$.

Here, the physical system is a linear elastic body subject to unknown x- and y- traction distributions on part of its boundary, say $\partial B_{\tilde{\mathbf{t}}}$. Also, displacement vectors are known at a set of points on the interior $\mathbf{x}_i$, $i = 1 \ldots N_D$. Let us define the model parameters to be the x- and y- traction distribution values at a finite set of points on $\partial B_{\tilde{\mathbf{t}}}$, say $\mathbf{p}_j$, $j = 1 \ldots N_M$. The data parameters will be the x- and y- displacements at $\mathbf{x}_i$, $i = 1 \ldots N_D$. We write a point in model space as "$\tilde{\mathbf{t}} \in M$" and a point in data space as "$\mathbf{u} \in D$", where $\tilde{\mathbf{t}}$ is a $2 * N_M \times 1$ column vector, and $\mathbf{u}$ is $2 * N_D \times 1$.

The general approach of [1] is founded on the notion that information about the solution to the inverse problem can be obtained by combining information about the theory — "forward" information on how a $\mathbf{u} \in D$ and a $\tilde{\mathbf{t}} \in M$ are correlated according to the laws of elastostatics — with prior information such as continuity constraints on the unknown traction distribution and observed displacements $\mathbf{u}_{obs}$ at the interior points $\mathbf{x}_i$, $i = 1 \ldots N_D$. Tarantola proposed [1] that the most general way to describe information about a system parametrized by a finite set of (scalar) variables is to use probability density finctions defined on the associated parameter space. In light of this, the idea of combining states of information can be cast in more mathematical language: the probability density for the solution, written $\sigma(\tilde{\mathbf{t}})$, is some combination of the probability densities for the forward modeling information, denoted $\theta(\mathbf{u}, \tilde{\mathbf{t}})$, and for the prior information, written $\rho(\mathbf{u}, \tilde{\mathbf{t}})$. As indicated, θ and ρ are defined on $D \times M$.

The theory probability density can be decomposed as the product of two independent probability density functions [1]:

$$\theta(\mathbf{u}, \tilde{\mathbf{t}}) = \theta_{D|M}(\mathbf{u}|\tilde{\mathbf{t}})\theta_M(\tilde{\mathbf{t}})$$

where $\theta_{D|M}(\mathbf{u}|\tilde{\mathbf{t}})$ is the probability density at $\mathbf{u} \in D$ under the condition that $\tilde{\mathbf{t}} \in M$ is specified. This is called the "conditional" probability density function for $\mathbf{u}$. The function $\theta_M(\tilde{\mathbf{t}})$ is the "marginal" probability density function for $\tilde{\mathbf{t}}$. It is simply the integral of $\theta(\mathbf{u}, \tilde{\mathbf{t}})$ over all $\mathbf{u} \in M$. Thus, $\theta_M(\tilde{\mathbf{t}})$ represents the *total* probability density for any $\tilde{\mathbf{t}}$.

Intuitively, if only the theory information is considered, there is no reason to think that any $\tilde{\mathbf{t}}$ should have greater total probability density than any other. This is because the theory information is (usually) concerned only with *correlating* each $\mathbf{u} \in D$ with a given $\tilde{\mathbf{t}} \in M$. Mathematically, the equal total probability density of each $\tilde{\mathbf{t}}$ is expressed by making $\theta_M(\tilde{\mathbf{t}})$ constant. Thus,

$$\theta(\mathbf{u}, \tilde{\mathbf{t}}) \propto \theta_{D|M}(\mathbf{u}|\tilde{\mathbf{t}}).$$

The function $\theta_{D|M}(\mathbf{u}|\tilde{\mathbf{t}})$ will be modeled as a Gaussian on D with central point $\mathbf{u_0}$ and with some "spread" over $\mathbf{u}$. Under this Gaussian assumption, the central point $\mathbf{u_0}$ is the maximum value of the probability density function $\theta_{D|M}(\mathbf{u}|\tilde{\mathbf{t}})$ and thus is the most likely data point $\mathbf{u_0} \in D$ for a given $\tilde{\mathbf{t}} \in M$. The mapping $\mathbf{g} : \tilde{\mathbf{t}} \in M \longrightarrow \mathbf{u_0} \in D$ is simply the governing physical law of the system. The spread in the Gaussian probability density function is a multi-dimensional variance. It is defined using a linear operator on D called the covariance operator, $[\mathbf{C}_T]$. We can construct $[\mathbf{C}_T]$ by examining the definition of its components.

Let f be some probability density function on a parameter space X. We have (see [1]):

$$C_T^{ij} \stackrel{def}{=} \int_X (x^i - x_m^i)(x^j - x_m^j) f(\mathbf{x}) d\mathbf{x}, \tag{1}$$

where $\mathbf{x}_m$ is the mean, or central value, of $f(\mathbf{x})$. The diagonal components C_T^{ii} are the variances $(\sigma^i)^2$ of the parameter x^i under f (see [1]):

$$C_T^{ii} = \int_X (x^i - x_m^i)^2 f(\mathbf{x}) d\mathbf{x}. \tag{2}$$

However, the meaning of the off–diagonal components is not so apparent.

It can be shown that the integral in equation 1 is a generalization of the linear correlation coefficient for a collection of discrete values (x_l, y_l), $l = 1 \ldots N$ [5]. This coefficient is denoted r and is defined to be

$$r \stackrel{def}{=} \frac{1}{C} \sum_{l=1}^{N} (x_l - x_m)(y_l - y_m) \tag{3}$$

where x_m and y_m are the mean values of x and y in the data set (x_l, y_l), $l = 1 \ldots N$ and C is a normalization constant chosen so r takes on values between -1 and $+1$. The value $r = 1$ represents "complete positive correlation" of the x– and y–values in the data set (x_l, y_l), $l = 1 \ldots N$ while $r = -1$ represents "complete negative correlation". When $r = 0$, the x– and y–values are uncorrelated in the data set.

Passing from a discrete set of points to a continuum of ordered pairs (x, y), let us consider the quantity $f(x_0, y_0)dx_0dy_0$, where dx_0dy_0 is some infinitessimal neighborhood centered on the point (x_0, y_0). As f is a probability density function, this quantity is the probability that the parameters (x, y) will be "close" to the values (x_0, y_0) (i.e. in the neighborhood dx_0dy_0). To recover the discrete case we enforce $f(x_0, y_0)dx_0dy_0 \to 1$ if one of the data points (x_l, y_l), $l = 1 \ldots N$ is in dx_0dy_0 and $f(x_0, y_0)dx_0dy_0 \to 0$ if none of the data points are in dx_0dy_0. From this, the correlation coefficient defined by the sum in equation 3 can be expressed as an integral:

$$r = \frac{1}{C} \int (x - x_m)(y - y_m) f(x, y) dx dy, \tag{4}$$

with

$$\begin{aligned} f(x, y) &= \delta(x - x_1)\delta(y - y_1) + \ldots + \delta(x - x_N)\delta(y - y_N) \\ &= \sum_{i=1}^{N} \delta(x - x_i)\delta(y - y_i). \end{aligned} \tag{5}$$

As the linear correlation coefficient r describes how closely x and y vary together in the data set (x_l, y_l), $l = 1 \ldots N$, it might just as well be called the linear *co-variance* coefficient. Thus, by extension, the integral expression in equation 1 for C_T^{ij} is (proportional to) the linear covariance coefficient for the parameters x^i and x^j.

It should be noted that the mathematical definitions in equations 1 and 2 require full knowledge of the probability density function. When we are *constructing* the probability density function, however, we must rely on the intuitive meaning of equations 1 and 2 to define $[\mathbf{C}_T]$. That is, we must use the knowledge that the diagonal components are the variances and the off–diagonal components are linear correlation coefficients for continuous valued parameters. With these intuitive definitions, we can construct the components of $[\mathbf{C}_T]$ from our physical understanding of the system.

Once the central point $\mathbf{u}_0$ and the covariance operator $[\mathbf{C}_T]$ have been constructed, the Gaussian probability density function $\theta_{D|M}(\mathbf{u}|\tilde{\mathbf{t}})$ can be defined as follows:

$$\theta_{D|M}(\mathbf{u}|\tilde{\mathbf{t}}) = (const.) \exp\left[-\frac{1}{2}(\mathbf{u} - \mathbf{u}_0)[\mathbf{C}_T]^{-1}(\mathbf{u} - \mathbf{u}_0)\right],$$

where the exponential argument has the properties of the square of a norm, $\|\mathbf{u} - \mathbf{u}_0\|^2$.

3 A Method for Determining the Central Point and Covariance Operator

Assuming the body is accurately modeled as linear elastic, there are two sources of uncertainty in determining the displacement values $\mathbf{u}$ associated with a set of traction values $\tilde{\mathbf{t}}$. One source of uncertainty arises from the definition of M, the other from the definition of D. A point in model space $\tilde{\mathbf{t}} \in M$ is not just *any* collection of N_M x- and y-traction values. Each $\tilde{\mathbf{t}} \in M$ is a finite sampling of a traction distribution $\tilde{\mathbf{t}}(s)$ on $\partial B_{\tilde{\mathbf{t}}}$ *which is a possible solution to the actual inverse problem.* That is, a possible solution to the question: What is the traction distribution on $\partial B_{\tilde{\mathbf{t}}}$ when the displacements $\mathbf{u}_{obs}$ at $\mathbf{x}_i$, $i = 1 \ldots N_D$ are known a-priori? Furthermore, the space of possible actual solutions $\tilde{\mathbf{t}}(s) \in \mathfrak{S}$ is restricted by prior information (continuity, boundedness, etc.).

Because the model space used to describe information about the actual solution $\tilde{\mathbf{t}}(s)$ is finite dimensional, $\tilde{\mathbf{t}}(s)$ cannot be exactly described by the $2*N_M$-dimensional point $\tilde{\mathbf{t}} \in M$, unless $\mathfrak{S}$ is an appropriate finite dimensional space. The other source of uncertainty is the actual displacement values $\mathbf{u} \in D$ caused by applying a full traction distribution $\tilde{\mathbf{t}}(s) \in \mathfrak{S}$ to $\partial B_{\tilde{\mathbf{t}}}$. In general, these displacements cannot be determined analytically and must be numerically approximated. The nature of the uncertainty in $\mathbf{u}$ depends on the numerical approximation method used.

Here, finite element analyses on uniform meshes are used to approximate $\mathbf{u}$. For linear, elliptic problems like this one, it is known that as element size parameter h goes to zero, the finite element displacements $\mathbf{u}_h$ at the points $\mathbf{x}_i$, $i = 1 \ldots N_D$ will converge to the solution $\mathbf{u}$ [11]. In theory, a finite element analysis can be performed on a uniform mesh with any h value. In practice, however, there is a lower limit to the size of h, say h_{min}. Beyond this limit, roughly, the processing time for the analysis becomes prohibitively (or inconveniently) large. Naturally, h_{min} depends on the computer, the type of element being used, and other details of the finite element analysis.

Summarizing, there are two sources of random error in this approach:

1. Uncertainty due to the finite dimensional approximation of the traction distribution $\tilde{\mathbf{t}}(s)$.

2. Uncertainty due to the practical limits of refinement in the finite element analyses used to compute the most likely point $\mathbf{u}_0 \in D$ from $\tilde{\mathbf{t}} \in M$.

In this analysis, the "best" F.E.–based estimate for $\mathbf{u}$ is used as the most likely set of displacements, $\mathbf{u}_0$. The best F.E. estimate is obtained using a Richardson extrapolation as follows. A point in model space $\tilde{\mathbf{t}} \in M$ consisting of a sampling of x- and y- traction values at a fixed set of points $\mathbf{p}_j$, $j = 1 \ldots N_M$ is used to define two polynomial distributions of order $N_M - 1$: $P_x(s)$ and $P_y(s)$. These tractions are applied as boundary conditions in F.E. analyses on a sequence of uniform meshes $\{M_{h_i}\}_{i=1}^{N_{msh}}$ where $h_{min} \leq h_{N_{msh}} \leq \ldots \leq h_1 \leq h_{max}$, where h_{max} is assumed to be small enough to yield a reasonable approximation to the actual displacements $\mathbf{u} \in D$. For each component of $\mathbf{u}$, say u^j, there will be a sequence of solutions $\{u^j_{h_i}\}_{i=1}^{N_{msh}}$ corresponding to the mesh sequence $\{M_{h_i}\}_{i=1}^{N_{msh}}$. These solutions are fit to a polynomial in h, called the Richardson polynomial for u^j. This function is then evaluated at $h = 0$ and that value is taken to be the j^{th} component u^j_0 of the central point $\mathbf{u}_0$ (the most likely set of displacements). The order of the Richardson

polynomial was determined by assuming $O(h^2)$ global convergence for linear elements [11]. It was further assumed that the points $\mathbf{x}_i$, $i = 1 \ldots N_D$ corresponding to $\mathbf{u} \in D$ are nodes in each mesh of the sequence $\{M_{h_i}\}_{i=1}^{N_{msh}}$ and that each mesh is uniform and consists entirely of linear elements.

From the above description, it is easily seen that determining $\mathbf{u}_0$ from $\mathbf{t}$ (i.e. the mapping $\mathbf{g} : \mathbf{t} \in M \longmapsto \mathbf{u}_0 \in D$) is a linear process with three steps:

1. Define the two polynomials P_x, P_y of order $N_M - 1$ which are uniquely determined by the N_M x- and y- traction values of the vector $\mathbf{t} \in M$.

2. For each mesh h_i in the sequence $\{M_{h_i}\}_{i=1}^{N_{msh}}$, an F.E. analysis with traction boundary conditions $P_x(s)$ and $P_y(s)$ is performed and the approximations $u_{h_i}^j$, $j = 1 \ldots N_D$ to u^j are recorded.

3. For each component j, the sequence of solution values $\{u_{h_i}^j\}_{i=1}^{N_{msh}}$ is fitted to a quadratic Richardson polynomial, $P_R^j(h)$. The j^{th} component of the central value $\mathbf{u}_0$, is taken to be $u_0^j = \lim_{h \to 0} P_R^j(h)$

Since each step is linear, considerable computational savings can be obtained by determining a-priori the displacements $\mathbf{u}_0$ for each monomial of the x- and y- traction distributions $P_x(s)$ and $P_y(s)$. Writing $\mathbf{u}_0^{t_x=k}$ for $\mathbf{u}_0$ when the traction polynomials are $P_x = k$ and $P_y = 0$, and writing $\mathbf{u}_{0_x}^{t_x=ks}$ for $\mathbf{u}_0$ when $P_x = ks$ and $P_y = 0$ etc., we are easily able to construct $\mathbf{u}_0$ when $P_x(s)$ and $P_y(s)$ are arbitrary $O(s^{N_M-1})$ polynomials. Writing

$$P_x(s) = \alpha_0 + \alpha_1 s + \ldots + \alpha_{N_M-1} s^{N_M-1} \tag{6}$$

$$P_y(s) = \beta_0 + \beta_1 s + \ldots + \beta_{N_M-1} s^{N_M-1}, \tag{7}$$

it follows that

$$\begin{aligned} \mathbf{u}_{0_x} = & \frac{\alpha_0}{k}\mathbf{u}_{0_x}^{t_x=k} + \frac{\alpha_1}{k}\mathbf{u}_{0_x}^{t_x=ks} + \ldots + \frac{\alpha_{N_M-1}}{k}\mathbf{u}_{0_x}^{t_x=ks^{N_M-1}} + \\ & \frac{\beta_0}{k}\mathbf{u}_{0_x}^{t_y=k} + \frac{\beta_1}{k}\mathbf{u}_{0_x}^{t_y=ks} + \ldots + \frac{\beta_{N_M-1}}{k}\mathbf{u}_{0_x}^{t_y=ks^{N_M-1}} \end{aligned} \tag{8}$$

and similarly for $\mathbf{u}_{0_y}$, where k is a scaling factor. Although this approach requires substantial pre-processing work, once the "fundamental" solutions $\mathbf{u}_{0_x}^{t_x=k}$, $\mathbf{u}_{0_y}^{t_x=k}$, etc. are determined, $\mathbf{g}(\mathbf{t})$ for any $\mathbf{t}$ is computed directly from the coefficients of $P_x(s)$ and $P_y(s)$.

Recall that the Gaussian statistics of the random variable $\mathbf{u}$, for a given $\mathbf{t} \in M$, are defined by two quantities: the central value $\mathbf{u}_0 = \mathbf{g}(\mathbf{t})$ and the covariance operator $[\mathbf{C}_T]$. The former has just been defined and the latter describes the error in each component u_0^j as well as the correlations between these errors. The variance of the j^{th} component of the solution is given by the diagonal components C_T^{jj} and the correlations are given by the off–diagonal components C_T^{ij}. The diagonal components may be defined as follows. For simplicity, we consider only the x–components, so $\mathbf{u}_0 = \mathbf{u}_{0_x}$. Each diagonal component $C_T^{jj} = \sigma_j^2$ is the error in the j^th component of $\mathbf{u}_{0_x}$. As described above, there are two sources of error contributing to σ_j^2 — the F.E. error and the uncertainty introduced by representing general traction distributions $\tilde{\mathbf{t}}(s) \in \mathfrak{S}$ with finite dimensional functions $P_x(s)$ and $P_y(s)$, defined by $\tilde{\mathbf{t}} \in M$. The F.E. error is the sum of the errors in each of the terms of the polynomial (8). So,

$$\sigma_j^2\Big|_{fea} = \frac{\alpha_0}{k}(\sigma_j^2)^{t_x=k} + \frac{\alpha_1}{k}(\sigma_j^2)^{t_x=ks} + \cdots, \tag{9}$$

where $(\sigma_j^2)^{t_x=k}$ and $(\sigma_j^2)^{t_x=ks}$ are the errors in the j^{th} component of $\mathbf{u}_{0_x}^{t_x=k}$ and $\mathbf{u}_{0_x}^{t_x=ks}$, respectively. Each of these uncertainties is due to the error in the Richardson extrapolation from the sequence of F.E. solutions. This extrapolation error may be taken as the average of the square of the difference between the F.E. solution and the Richardson fit at each of the N_{msh} data points. Thus, for instance, the error in the j^{th} component of $\mathbf{u}_{0_x}^{t_x=k}$ is

$$(\sigma_j^2)^{t_x=k} = \frac{1}{N_{msh}} \sum_{i=1}^{N_{msh}} \left| P_R^{j,t_x=k}(h=h_i) - u_{h_i}^{j,t_x=k} \right|^2 . \tag{10}$$

Here, $P_R^{j,t_x=k}$ is the Richardson polynomial and $u_{h_i}^{j,t_x=k}$ is the i^{th} F.E. solution value for $\mathbf{u}_{0_x}^{t_x=k}$. Then the total error will be

$$C_T^{jj} = \sigma_j^2 = \sigma_j^2\Big|_{fea} + \sigma_j^2\Big|_{\mathfrak{S}-M}, \tag{11}$$

where the last term on the right is the error in approximating the actual traction distribution $\tilde{\mathbf{t}}(s) \in \mathfrak{S}$ with the polynomials $P_x(s)$ and $P_y(s)$ defined by $\tilde{\mathbf{t}} \in M$.

The off–diagonal components of the covariance operator were taken to be

$$C_T^{ij} = \frac{1}{2}(\sigma_i^2 + \sigma_j^2) \exp\left(-\frac{1}{2} \frac{|\mathbf{x}_{s(i)} - \mathbf{x}_{s(j)}|^2}{\Delta L^2} \right), \tag{12}$$

where $s(j)$ is the interior point at which the j^{th} component of $\mathbf{u} \in D$ is sampled. The "correlation length" is ΔL. This is the distance over which the actual displacement *field* $\mathbf{u}(\mathbf{x})$ changes significantly. That is, the approximate distance beyond which two points $\mathbf{x}_1$ and $\mathbf{x}_2$ will in general have unrelated displacements. It is a measure of how rapidly $\mathbf{u}(\mathbf{x})$ changes locally. Of course, this will vary from region to region within the domain, but for simplicity a single value may be chosen. This value is determined by the traction distributions $P_x(s)$ and $P_y(s)$ defined by $\mathbf{t} \in M$ as well as by the positions of the interior points $\mathbf{x}_i$, $i = 1 \ldots N_D$ used to define $\mathbf{u} \in M$. If $P_x(s)$ and $P_y(s)$ are rapidly varying traction distributions on $\partial B_{\tilde{\mathbf{t}}}$, rapidly varying displacements will be induced in the elastic body near the boundary $\partial B_{\tilde{\mathbf{t}}}$. It follows that the characteristic length of the actual displacement field will increase with increasing distance from $\partial B_{\tilde{\mathbf{t}}}$ (St. Venant's principle) and with increasing characteristic lengths of $P_x(s)$ and $P_y(s)$. The value of ΔL may be chosen, then, using the distance of the points $\mathbf{x}_i$, $i = 1 \ldots N_D$ from $\partial B_{\tilde{\mathbf{t}}}$ and using the characteristic lengths of $P_x(s)$ and $P_y(s)$.

4 Numerical Examples

4.1 Data Space is Displacement at Sensor Nodes

A sequence of five uniformly refined meshes of linear quadrilaterals were used to obtain z–displacements at four nodes, with fixed coordinates. These coordinates correspond to the points $\mathbf{x}_i$, $i = 1 \ldots N_D$ used to define the data space D, as discussed in the previous section. A point $\mathbf{u} \in D$, in this example, is the 4–tuple of z–displacement values at each of the four fixed node points. Each mesh was subject to three different loadings and the z–displacements were recorded at the four nodal points for each loading. The analyses were axisymmetric with coordinates r and z. The five uniform meshes had the following average mesh sizes: $h_1 = 1.293061$, $h_2 = 1.091147$, $h_3 = 0.855974$, $h_4 = 0.573205$, $h_5 = 0.388952$ (the characteristic size of the model was roughly 40). Traction was applied only in the

z–direction according to the distributions $t_z(r) = 100,\ 100r,$ and $100r^2$. The results are summarized in Tables 1–5.

For each combination of z–displacement solution at $\mathbf{x}_i$, $i = 1 \ldots N_D$ and applied traction distribution, there is a sequence of five finite element solutions associated with the meshes $h_1,\ h_2,\ h_3,\ h_4,\ h_5$. The solutions were used to define quadratic polynomials in mesh size parameter h. These are the Richardson polynomials described in the previous section. The polynomials have no linear term, so that the extreme value occurs at $h = 0$, corresponding to the estimate of the exact solution. Each Richardson polynomial was obtained from a series of data points $\{(x_i, y_i)\}_{i=1}^{5}$ by finding the second order polynomial $f = \alpha_0 + \alpha_2 x^2$ which minimizes the error

$$E = \sum_{i=1}^{N} |y_i - f(x_i)|^2. \tag{13}$$

Writing f as

$$f(x_i) = \alpha_0 + \alpha_2 x_i^2 = \begin{bmatrix} \alpha_0 & \alpha_2 \end{bmatrix} \begin{bmatrix} 1 \\ x_i^2 \end{bmatrix} = \boldsymbol{\alpha}^T \mathbf{x}_i, \tag{14}$$

we have

$$E = \sum_{i=1}^{N} \left| y_i - \boldsymbol{\alpha}^T \mathbf{x}_i \right|^2. \tag{15}$$

Differentiating E with respect to $\boldsymbol{\alpha}$ yields

$$\frac{dE}{d\alpha_j} = \sum_{i=1}^{N} \frac{d}{d\alpha_j} |y_i - \alpha_k x_{ki}|^2 = \sum_{i=1}^{N} 2 |y_i - \alpha_k x_{ki}| (-x_{ki}\delta_{jk}). \tag{16}$$

The condition of minimizing E can be written in direct notation as

$$\frac{dE}{d\boldsymbol{\alpha}} = 0 \ \Rightarrow \ \sum_{i=1}^{N} (\mathbf{x}_i \otimes \mathbf{x}_i)\boldsymbol{\alpha} - \mathbf{x}_i y_i = 0, \tag{17}$$

which can be written

$$\mathbf{A}\boldsymbol{\alpha} = \mathbf{b} \text{ where } \mathbf{A} = \sum_{i=1}^{N} (\mathbf{x}_i \otimes \mathbf{x}_i) \text{ and } \mathbf{b} = \sum_{i=1}^{N} \mathbf{x}_i y_i. \tag{18}$$

Solving equation 18 yields the Richardson polynomial coefficients, $\boldsymbol{\alpha}$.

The Richardson polynomial for each combination of applied traction distribution ($t_z(r) = 100,\ 100r,\ 100r^2$) and z–displacement at one of the four fixed nodes is denoted $R_1^{100}(h)$, $R_2^{100r}(h)$, etc. The subscript refers to the fixed–position node and the superscript refers to the load distribution. One example of these polynomials, and the corresponding data points, is shown in Figure 1. Evaluating the Richardson polynomials at $h = 0$ gives the "fundamental" solutions, described in the previous section. These solutions are used to construct the central value $\mathbf{u}_0$ of the Gaussian probability density for the forward problem, $\theta(\mathbf{u}|\tilde{\mathbf{t}})$. They are summarized in Tables 6 – 8. Thus, for any quadratic traction distribution $t_z(r) = a_0 + a_1 r + a_2 r^2$, the components of the central value $\mathbf{u}_0$ of $\theta(\mathbf{u}|\tilde{\mathbf{t}})$ are

$$u_{0_j} = \frac{a_0}{100} R_j^{100}(h=0) + \frac{a_1}{100} R_j^{100r}(h=0) + \frac{a_2}{100} R_j^{100r^2}(h=0),\ j = 1 \ldots 4. \tag{19}$$

4.2 Data Space is Strain at Sensor Nodes

Usually the data in an inverse problem is obtained directly from experimental measurements. For the problem of recovering unknown boundary traction distributions, it is easier to measure internal strains than displacements. In this case, we simply redefine the data space to be the strain components ϵ_{kl} at $\mathbf{x}_i$, $i = 1 \ldots N_D$. This was done for the meshes and the loading conditions described above. The strain component ϵ_{zz} was obtained at the sensor nodes located at $\mathbf{x}_i$, $i = 1 \ldots N_D$ using two methods: extrapolation from integration points followed by nodal averaging, and a patch–wise recovery method proposed in [2]. The finite element analyses and the nodal averaging were done using ABAQUS, while the patch–wise recovery was implemented separately, taking as input the ϵ_{zz} values at element integration points obtained from ABAQUS. Figure 2 shows the typical convergence behavior observed for nodal strain values obtained using nodal averaging and patch–wise recovery. The nodal averaging solutions exhibit a piecewise convergence which is at first quadratic and then flat, while there is a distinct $O(h)$ convergence for the nodal values obtained by the method of [2].

With the data space defined to be strain values at the four fixed nodes, the identical analysis carried out in the previous section (except with a lower order Richardson polynomial) can be performed to define a new probability density for the forward problem — $\theta(\boldsymbol{\epsilon}_0|\tilde{\mathbf{t}})$. This is a Gaussian function defined by its central vector, $\boldsymbol{\epsilon}_0$, and its covariance operator. Just as in the previous example, the central vector is obtained from Richardson extrapolations of finite element solutions. The covariance operator contains information about the uncertainty in the extrapolations and in the approximation of physical traction distributions by finite dimensional functions, as discussed in Section 3.

5 Conclusion

A method for defining the central point and covariance operator for the Gaussian probability density describing the forward problem of linear elasticity has been proposed. The technique for determining the central point is simple and is based on well established principles (linearity, existence, uniqueness). Defining the "spread", or covariance operator, for the Gaussian probability density is a more subjective matter. The diagonal components of the covariance operator reflect modelization uncertainties introduced by two assumptions: the restriction of boundary traction distributions to a finite dimensional space, and the approximation of the forward solution by extrapolation from a sequence of finite element analyses. These assumptions are reasonable and are consistent with the subjective nature of Bayesian statistical analysis.

Acknowledgement

This work has been supported by the NSF under grant DMI–9358123.

	$t_z = 100$	$t_z = 100r$	$t_z = 100r^2$
u_z at node 1	1.0666×10^{-6}	7.7790×10^{-6}	7.5094×10^{-5}
u_z at node 2	1.0207×10^{-6}	8.2922×10^{-6}	8.1208×10^{-5}
u_z at node 3	9.4386×10^{-7}	8.5967×10^{-6}	8.9513×10^{-5}
u_z at node 4	7.9949×10^{-7}	7.8334×10^{-6}	8.7009×10^{-5}

Table 1: Finite element solutions for mesh $h_1 = 1.293061$

	$t_z = 100$	$t_z = 100r$	$t_z = 100r^2$
u_z at node 1	1.0667×10^{-6}	7.7793×10^{-6}	7.5112×10^{-5}
u_z at node 2	1.0210×10^{-6}	8.2931×10^{-6}	8.1197×10^{-5}
u_z at node 3	9.4388×10^{-7}	8.5949×10^{-6}	8.9490×10^{-5}
u_z at node 4	7.9981×10^{-7}	7.8408×10^{-6}	8.7147×10^{-5}

Table 2: Finite element solutions for mesh $h_2 = 1.091147$

	$t_z = 100$	$t_z = 100r$	$t_z = 100r^2$
u_z at node 1	1.0670×10^{-6}	7.7798×10^{-6}	7.5118×10^{-5}
u_z at node 2	1.0210×10^{-6}	8.2898×10^{-6}	8.1172×10^{-5}
u_z at node 3	9.4395×10^{-7}	8.5939×10^{-6}	8.9459×10^{-5}
u_z at node 4	8.0112×10^{-7}	7.8593×10^{-6}	8.7416×10^{-5}

Table 3: Finite element solutions for mesh $h_3 = 0.855794$

	$t_z = 100$	$t_z = 100r$	$t_z = 100r^2$
u_z at node 1	1.0669×10^{-6}	7.7820×10^{-6}	7.5154×10^{-5}
u_z at node 2	1.0211×10^{-6}	8.2906×10^{-6}	8.1193×10^{-5}
u_z at node 3	9.4405×10^{-7}	8.5936×10^{-6}	8.9463×10^{-5}
u_z at node 4	8.0175×10^{-7}	7.8653×10^{-6}	8.7479×10^{-5}

Table 4: Finite element solutions for mesh $h_4 = 0.573205$

	$t_z = 100$	$t_z = 100r$	$t_z = 100r^2$
u_z at node 1	1.0670×10^{-6}	7.7831×10^{-6}	7.5175×10^{-5}
u_z at node 2	1.0212×10^{-6}	8.2914×10^{-6}	8.1200×10^{-5}
u_z at node 3	9.4415×10^{-7}	8.5941×10^{-6}	8.9466×10^{-5}
u_z at node 4	8.0232×10^{-7}	7.8727×10^{-6}	8.7579×10^{-5}

Table 5: Finite element solutions for mesh $h_5 = 0.388952$

	$t_z = 100$
u_z at node 1	$R_1^{100}(h=0) = 1.0671 \times 10^{-6}$
u_z at node 2	$R_2^{100}(h=0) = 1.0212 \times 10^{-6}$
u_z at node 3	$R_3^{100}(h=0) = 9.4413 \times 10^{-7}$
u_z at node 4	$R_4^{100}(h=0) = 8.0246 \times 10^{-7}$

Table 6: Extrapolation to $h = 0$ of the Richardson polynomials for loading $t_z(r) = 100$. These are the "fundamental solutions" used to construct the central value $\mathbf{u}_0$ of the probability density $\theta(\mathbf{u}|\tilde{\mathbf{t}})$.

	$t_z = 100r$
u_z at node 1	$R_1^{100r}(h=0) = 7.7828 \times 10^{-6}$
u_z at node 2	$R_2^{100r}(h=0) = 8.2905 \times 10^{-6}$
u_z at node 3	$R_3^{100r}(h=0) = 8.5686 \times 10^{-6}$
u_z at node 4	$R_4^{100r}(h=0) = 7.8757 \times 10^{-6}$

Table 7: Extrapolation to $h = 0$ of the Richardson polynomials for loading $t_z(r) = 100r$.

	$t_z = 100r^2$
u_z at node 1	$R_1^{100r^2}(h=0) = 7.5171 \times 10^{-5}$
u_z at node 2	$R_2^{100r^2}(h=0) = 8.1188 \times 10^{-5}$
u_z at node 3	$R_3^{100r^2}(h=0) = 8.9451 \times 10^{-5}$
u_z at node 4	$R_4^{100r^2}(h=0) = 8.7634 \times 10^{-5}$

Table 8: Extrapolation to $h = 0$ of the Richardson polynomials for loading $t_z(r) = 100r^s$.

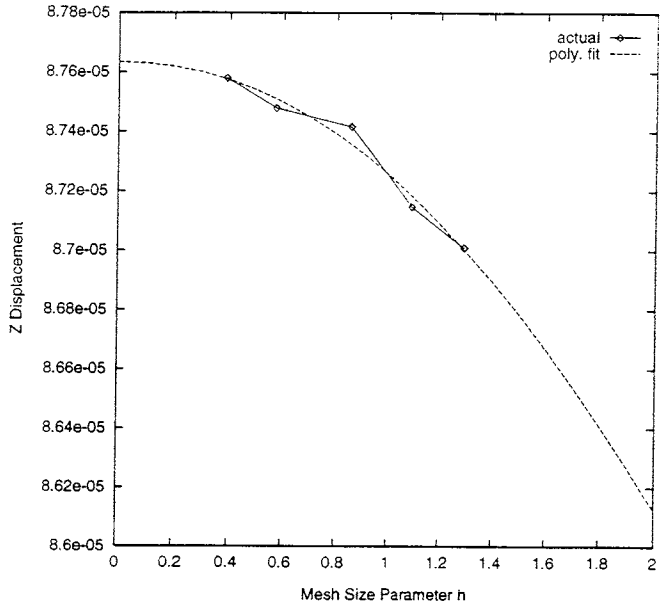

Figure 1: The Richardson polynomial $R_4^{100r^2}$ and the sequence of finite element solutions to which it was fit.

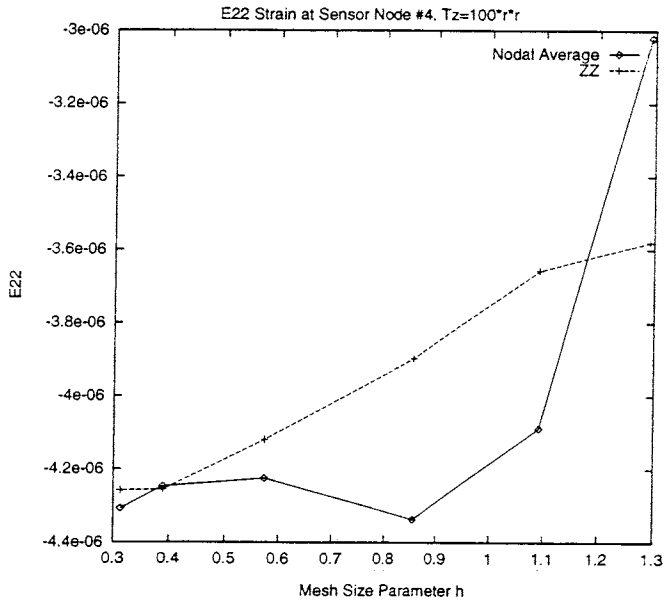

Figure 2: The strain value ϵ_{22} at the fourth sensor node obtained using both the patch–wise recovery method of [2] and nodal averaging (solid lines) for the loading $\tilde{t}_z(r) = 100r^2$.

References

[1] Tarantola A. *Inverse problem theory.* Elsevier, United States, 1987.

[2] Zienkiewicz O C and Zhu J Z. The superconvergent path recovery and a posteriori error estimates. part i: the recovery technique. *International Journal for Numerical Methods in Engineering*, 33:1331–1364, 1992.

[3] Sutton M A et al. A study of the use of measured displacement data with integral equations of elasticity. *Engrg. Anal. J.*, 2:395–403, 1986.

[4] Romanov V G. *Inverse problems of mathematical physics.* VNU Science Press, Utrecht, The Netherlands, 1987.

[5] Press W H, Flannery B P, Teukolsky S A, and Vetterling W T. *Numerical Recipes in C.* Cambridge University Press, Cambridge, U.K., 1988.

[6] Hadamard J. *Lectures on Cauchy's problems in linear partial differential equations.* Yale University Press, New Haven, Conn., 1923.

[7] Grysa K, Cialkowski M J, and Kaminski H. An inverse temperature field problem of the theory of thermal stresses. *Nucl. Engrg. Design*, 64:169–184, 1981.

[8] Maniatty A M and Zabaras N J. Investigation of regularization parameters and error estimating in inverse elasticity problems. *International Journal for Numerical Methods in Engineering*, 37:1039–1052, 1994.

[9] Maniatty A M, Zabaras N J, and Stelson K. Finite element analysis of some inverse elasticity problems. *Journal of Engineering Mechanics*, 115:1303–1317, 1989.

[10] Tikhonov A N and Arsenin V Y. *Solution of ill-posed problems.* B.H. Winston and Sons, Washingtion, D.C., 1977.

[11] Hughes T J R. *The finite element method: linear static and dynamic finite element analysis.* Prentice–Hall, Inc., Englewood Cliffs, New Jersey, 1987.

[12] Schnur D S and Zabaras N. Finite element solution of two–dimensional elasticity problems using spatial smoothing. *International Journal for Numerical Methods in Engineering*, 30:57–75, 1990.

[13] Sokolnikoff I S. *Mathematical theory of elasticity.* McGraw–Hill, New York, N.Y., 1956.