*Journal of Strain Analysis for Engineering Design, 47(3): 153–162 (2012)*

# Modeling and Design of a Hertzian Contact: An Info-Gap Approach

Yakov Ben-Haim
Yitzhak Moda'i Chair in Technology and Economics
Faculty of Mechanical Engineering
Technion—Israel Institute of Technology
Haifa 32000 Israel
yakov@technion.ac.il

## Contents



**Keywords:** Uncertainty, info-gaps, robustness, modeling, design, Hertzian contact

### Abstract

We illustrate the generic properties of info-gap robustness analysis by studying the design and data-based modeling of a Hertzian contact. We begin with an example of choosing the load and other operational parameters, to achieve a specified mechanical response, given uncertainty in the material properties of the system. We then study the updating of a mathematical model, based on data, given uncertain similarity of the measured system to the future operational system. We illustrate the universal trade off between outcome-quality and robustness to uncertainty, and we discuss the implications of this trade off for modeling and design decisions.

## 1 Overview of Info-Gap Concepts

Models are used by engineers to support design and planning decisions. We validate models in order to augment the confidence in the subsequent decisions that are made. Engineers view their empirical or science-based models differently from how models (or theories) are viewed by natural

\Papers\Shanghai2011\Paper\shang04.tex Based on \people\cogan\IMAC2012\ppr02imac2012.tex 17.1.2012 and physical scientists. The engineer has a utilitarian attitude: truth is useful, but usefulness does not require truth. This epistemological distinction between engineering and scientific modeling is particularly important in model validation under severe uncertainty. 'Severe uncertainty' refers to situations where the underlying processes are imperfectly understood, or where final-use shapes, loads and environmental conditions may differ substantially from those known at the time of design and testing. Under severe uncertainty, models must be validated while taking account of the disparities between what is known by the analyst, and what the analyst would need to know in order to make a fully confident design. Robustness to ignorance, to error and to surprise is cardinal in validating engineering models under severe uncertainty.

In this paper we will use info-gap decision theory for validating engineering models under severe uncertainty. The following main points will emerge through discussing an example of modeling the Hertzian contact between solid spheres under compressive loading.

**Info-gaps.** Info-gap theory is a methodology for supporting model-based decisions under severe uncertainty (Ben-Haim, 2006). An info-gap is a disparity between what *is known,* and what *needs to be known* in order to make a comprehensive and reliable decision. The future may differ from the past, so our models may err in ways we cannot know. Our data may lack evidence about surprises: catastrophes or windfalls. Our scientific and technical understanding may be incomplete. These are info-gaps: incomplete understanding of the system being managed.

Info-gaps arise in all areas of activity, and they are resolved only when a surprise occurs, or a new fact is uncovered, or when our knowledge and understanding improve. We know very little about the substance of an info-gap. For instance, we rarely know what unusual event will delay the completion of a task. Even more strongly, we cannot know what is not yet discovered, such as tomorrow's news, or future scientific theories or technological inventions. The ignorance of these things are info-gaps.

An info-gap is a Knightian uncertainty (Knight, 1921) since it is not characterized by a probability distribution. A probability distribution may exist, but it is unknown, or imperfectly known, to the analyst. Non-parametric statistical tools are available for handling some such situations (DeGroot, 1986). In some situations one can use robust statistical methods (Huber, 1981). However, in some situations the assumptions needed to justify these statistical tools—such as large samples or independence of measurements—are not valid.

An info-gap model of uncertainty is axiomatically distinct from a probability distribution (Ben-Haim, 1999). For instance, the absence of a probability density function (pdf) is not equivalent to assuming a rectangular (equal weight) pdf, as discussed in (Ben-Haim, 2006, section 2.2).

Nonetheless, info-gap and probabilistic methods can be combined in various ways. For instance, uncertainty in probability distributions can be quantified using non-probabilistic info-gap models of uncertainty (Ben-Haim, 2006; Burgman *et al.,* 2010). Info-gap is particularly useful when modelling uncertain parameters or uncertain functional relations in the absence of probabilistic information, as will be illustrated in this paper.

**Outcome-quality and robustness against uncertainty.** Modeling and design decisions have two attributes: the quality of the outcomes, and the robustness of that quality to deviation between reality and the design-base models. The former attribute is sometimes over-emphasized. In particular, a common design strategy is to use the best available model to find the design with the best predicted outcome. This best-model optimization strategy may ignore severe uncertainty, even if the models are themselves probabilistic.

**Trade off between outcome-quality and robustness.** Higher quality is better than lower quality, when all else is the same. However, a basic theorem in info-gap theory asserts that higher quality is achieved only by accepting lower robustness against uncertainty. That is, outcome-quality trades off against robustness to severe uncertainty.

This trade off also holds for model fidelity. An important attribute of validated models is their fidelity to experimental data. Higher fidelity is better than lower fidelity, when all else is the same. However, a model constructed to have high fidelity to data will have low robustness to error in the model form or in the data structure. That is, fidelity trades off against robustness to severe uncertainty.

**Zeroing.** Models are used to predict outcomes. However, model errors may be transmitted to the predictions. A basic info-gap theorem asserts that predicted outcomes have zero robustness to error in models or data upon which the predictions are based.

**Preference reversal.** As noted above, designs have two attributes: quality of the outcome, and robustness against uncertainty. We also noted that these attributes trade off, one against the other. These two attributes of a design can be expressed differently. A design's nominal performance is the outcome-quality obtained if the design-base assumptions, models and data are correct. A design's cost of robustness expresses how much the outcome-quality must be reduced in order to augment the robustness against uncertainty. A situation of particular importance is where the nominal performance is great, but the cost of robustness is high. In this case, the designer may prefer a nominally sub-optimal design whose cost of robustness is low.

This paper will illustrate these properties of the info-gap robust-satisficing strategy for modeling and design through a series of examples in sections 2–4. Section 5 discusses some generic issues.

# 2 Hertzian Contact of Spheres

Given two spheres in contact under axial load $f$, with radii $r_1$ and $r_2$, Young's moduli $E_1$ and $E_2$, and Poisson ratio $\nu$. Define:

$$E = \frac{2E_1E_2}{E_1+E_2} \tag{1}$$

$$\frac{1}{r} = \frac{1}{r_1}+\frac{1}{r_2} \tag{2}$$

The magnitude of the maximum normal compressive stress and the depth of penetration are (Beitz and Küttner, 1994):

$$\sigma = \frac{1}{\pi}\left(\frac{1.5fE^2}{r^2(1-\nu^2)^2}\right)^{1/3} \tag{3}$$

$$g = \left(\frac{2.25(1-\nu^2)^2f^2}{E^2r}\right)^{1/3} \tag{4}$$

$g$ is the sum of compression of both spheres in the region of contact.

Eqs.(3) and (4) depend on several assumptions: (a) The geometry and material properties are known. (b) The material is homogeneous and isotropic. (c) Hooke's law holds. (d) Only normal forces act at the surface. (e) Deformations are small in relation to the radii of the spheres.

# 3 Uncertainty and Its Representation

Let us consider various uncertain or unknown violations of the assumptions underlying the Hertzian model described in section 2.

## 3.1 Excess Stiffness

Suppose that the materials have undergone phase change, perhaps due to thermal cycling, making them stiffer than expected to an unknown degree. Then we expect the combined Young's modulus to be larger than anticipated to an unknown degree. The functional form of eqs.(3) and (4) remain the same, but the coefficients change. In the present case the change is in the combined Young's modulus $E$.

$\widetilde{E}$ is the estimated combined modulus before stiffening. A fractional-error info-gap model for uncertainty in $E$ after stiffening, which reflects the asymmetric uncertainty about the true value of

$E$, is:

$$\mathcal{U}(h) = \left\{ E:\ 0 \le E - \widetilde{E} \le \widetilde{E}h \right\}, \quad h \ge 0 \tag{5}$$

Like all info-gap models of uncertainty, eq.(5) is an unbounded family of nested sets of possible realizations (values of $E$ in this example). No worst case is known, and the uncertainty is represented without probability distributions. This info-gap model illustrates the properties of nesting and contraction. The sets are nested and become more inclusive as $h$ grows. This gives $h$ its meaning as an horizon of uncertainty. In the absence of uncertainty, when $h = 0$, the sets contract to a singleton set containing the estimate, $\widetilde{E}$.

Let's consider a slightly different example. Suppose we have an estimate, $s$, of the magnitude of error of $\widetilde{E}$, though we do not regard this as an actual upper limit of the error. The info-gap model of eq.(5) would be modified as:

$$\mathcal{U}(h) = \left\{ E:\ 0 \le E - \widetilde{E} \le sh \right\}, \quad h \ge 0 \tag{6}$$

In this info-gap model the horizon of uncertainty is calibrated in units of the estimated error, unlike eq.(5) where $h$ is a fractional error with respect to the estimated value itself.

Finally, suppose we have an estimate, $\widetilde{p}(E)$, of the pdf of the Young's modulus, but the true pdf, $p(E)$, is unknown. A fractional-error info-gap model for this uncertainty is:

$$\mathcal{U}(h) = \left\{ p(E):\ p(E) \ge 0,\ \int_0^\infty p(E)\,\mathrm{d}E = 1,\ |p(E) - \widetilde{p}(E)| \le \widetilde{p}(E)h \right\}, \quad h \ge 0 \tag{7}$$

The uncertain entity is a pdf, $p(E)$, but the uncertainty expressed by the info-gap model is non-probabilistic: no measure function describes the disparity between what we know (the estimate, $\widetilde{p}(E)$) and what we would like to know (the true pdf, $p(E)$).

There are many other info-gap models of uncertainty (Ben-Haim, 2006). The choice of an info-gap model is a crucial judgment made by the analyst, based on the extent of, as well as perceived limitations to, existing information and knowledge.

## 3.2 Heterogeneity

Now let us consider uncertain heterogeneity of one or both spheres. In particular, suppose that the materials in the immediate region of contact have the anticipated Young's moduli, but adjacent regions are stiffer by an unknown amount.

A plot of force, $f$, vs. compression, $g$, would follow eq.(4) for values of compression which do not yet significantly reach the stiffer domain. However, larger compression will require greater force. Thus the true shape of the entire curve, $f(g)$, will lie on or above the theoretical relation, $\widetilde{f}(g)$, obtained by inverting eq.(4), and may display a kink when the compressed region reaches the stiffer region.

An info-gap model for uncertain shape of the force-compression relation, obtained in analogy to eq.(5), is:

$$\mathcal{U}(h) = \left\{ f(g):\ 0 \le f(g) - \widetilde{f}(g) \le \widetilde{f}(g)h \right\}, \quad h \ge 0 \tag{8}$$

Suppose we have an estimate, $s(g)$, of the magnitude of error of $\widetilde{f}(g)$, though we do not regard this as an actual upper limit of the error. For instance, we may know that the stiffer inclusion will be reached at a compression no less than a value $g_o$. One might then choose $s(g) = 0$ for $g < g_o$ and $s(g) = 1$ for $g \ge g_o$. That is, the deviation between $f(g)$ and $\widetilde{f}(g)$ is "turned on" only for $g \ge g_o$. The info-gap model of eq.(8) would be modified, in analogy to eq.(6), as:

$$\mathcal{U}(h) = \left\{ f(g):\ 0 \le f(g) - \widetilde{f}(g) \le s(g)h \right\}, \quad h \ge 0 \tag{9}$$

There are many other info-gap models which can be formulated. For instance, one can impose a monotonicity constraint on the functions $f(g)$.

# 4 Robustness

We now consider two examples which illustrate the info-gap robustness analysis. In section 4.1 we choose the load and other operational parameters, to achieve a specified critical compression, given uncertainty in the stiffness of the Hertzian contact. In section 4.2 we update the Hertzian model based on data, given uncertain fidelity of the test system with the future operational system.

## 4.1 Choosing the Load Given Uncertain Young's Modulus

### 4.1.1 Formulation

Let $g(f, E)$ denote the compression, eq.(4), with load $f$ and Young's modulus $E$. We wish to choose the load, $f$, so that the total compression, $g$, exceeds a critical value $g_{\rm c}$. That is, the outcome requirement is:

$$g(f, E) \geq g_{\rm c} \tag{10}$$

Consider uncertainty in the Young's modulus as represented by the info-gap model of eq.(6). By choosing the materials from which the spheres are composed we can choose the estimate, $\widetilde{E}$, of this info-gap model. Thus $\widetilde{E}$ can be a design variable.

We may also be able to choose how well the Young's modulus is estimated, meaning that we can choose the value of the error estimate $s$.

The robustness of design $(f, \widetilde{E}, s)$ is the greatest horizon of uncertainty, $h$, up to which all realizations of the true modulus, $E$, result in compression no less than $g_{\rm c}$:

$$\widehat{h}_{\rm g}(g_{\rm c}) = \max \left\{ h : \left( \min_{E \in \mathcal{U}(h)} g(f, E) \right) \geq g_{\rm c} \right\} \tag{11}$$

We will sometimes denote the design variables explicitly by writing $\widehat{h}_{\rm g}(g_{\rm c}, f, \widetilde{E}, s)$.

Let $\mu_{\rm g}(h)$ denote the inner minimum in eq.(11). $\mu_{\rm g}(h)$ decreases as $h$ increases because the sets of the info-gap model become more inclusive as $h$ increases. This means that the robustness is the greatest value of $h$ at which $\mu_{\rm g}(h) = g_{\rm c}$. In other words, $\mu_{\rm g}(h)$ is the inverse of the robustness function: a plot of $h$ vs $\mu_{\rm g}(h)$ is the same as a plot of $\widehat{h}_{\rm g}(g_{\rm c})$ vs $g_{\rm c}$. Stating this explicitly:

$$\mu_{\rm g}(h) = g_{\rm c} \quad \text{if and only if} \quad \widehat{h}_{\rm g}(g_{\rm c}) = h \tag{12}$$

From examination of eq.(4) we see that $\mu_{\rm g}(h)$ occurs when $E$ takes its maximal value at horizon of uncertainty $h$: $E = \widetilde{E} + sh$. Thus:

$$\mu_{\rm g}(h) = c \left( \frac{f}{\widetilde{E} + sh} \right)^{2/3} \tag{13}$$

where:

$$c = \left( \frac{2.25(1 - \nu^2)^2}{r} \right)^{1/3} \tag{14}$$

Equating $\mu_{\rm g}(h)$ to $g_{\rm c}$ and solving for $h$ yields the robustness:

$$\widehat{h}_{\rm g}(g_{\rm c}) = \frac{1}{s} \left( \left( \frac{c}{g_{\rm c}} \right)^{3/2} f - \widetilde{E} \right) \tag{15}$$

or zero if this is negative.

### 4.1.2 Properties of the Robustness Function

Eq.(15) illustrates the basic properties of all info-gap robustness functions.

**Zeroing.** The estimated compression is:

$$\widetilde{g}(f) = c\left(\frac{f}{\widetilde{E}}\right)^{2/3} \tag{16}$$

This is the best model-based estimate of the compression, so one might be inclined to use this relation to choose $f$ for achieving specified compression. However, from eq.(15) we see that the robustness equals zero if the demanded compression, $g_c$, equals the estimated compression, $\widetilde{g}$:

$$\widehat{h}_g(\widetilde{g}) = 0 \tag{17}$$

Achieving the predicted outcome, $\widetilde{g}(f)$, has zero robustness to error in the model. This illustrates the general info-gap theorem that predicted outcomes have zero robustness to error in models or data upon which the predictions are based. Eq.(16) cannot be used directly for selecting the load to achieve predicted compression. Rather, we must consider the robustness to model uncertainty.

**Trade off.** Robustness decreases as greater compression is demanded. That is, $\widehat{h}_g(g_c)$ decreases as $g_c$ increases. This illustrates the info-gap theorem that higher quality is achieved only by accepting lower robustness against uncertainty. In other words, outcome-quality trades off against robustness to uncertainty.

**Cost of robustness.** The robustness curve—$\widehat{h}_g$ vs $g_c$—decreases monotonically, expressing the trade off between robustness and outcome. A steep slope implies that the robustness can be increased substantially in exchange for small decrease in the demanded compression. A shallow slope implies that large decrease in $g_c$ is required in order to obtain a substantial increase in robustness. A *steep slope* implies a *low cost of robustness.* Conversely, a *shallow slope* implies a *high cost of robustness.*

**Preference reversal.** Consider the choice between two options, $(f_1, s_1)$ and $(f_2, s_2)$ where:

$$f_1 > f_2 \quad \text{and} \quad \frac{s_1}{f_1} > \frac{s_2}{f_2} \tag{18}$$

In the first design, $(f_1, s_1)$, greater force is applied and the modulus is less accurately estimated. Which design is preferred, assuming that both designs satisfy the outcome requirement if the estimated modulus is correct?

The estimated outcomes are ranked as:

$$\widetilde{g}(f_1) > \widetilde{g}(f_2) > g_c \tag{19}$$

The first design is estimated—based on $\widetilde{E}$—to satisfy the outcome requirement by a larger margin than the second design. One might therefore tend to prefer the first design.

However, we know from the zeroing property that the robustness of estimated outcomes is zero, suggesting that eq.(19) is not a reliable basis for prioritizing the designs. From the trade off property we know that only less demanding outcomes—lower values of $g_c$—have positive robustness. From eq.(15) we see that the cost of robustness is determined by the ratio $f/s$. A large value of $f/s$ implies a low cost of robustness. Thus, from the specification in eq.(18), we find that the robustness curve for the second design is steeper than for the first design. Consequently the cost of robustness is lower for the second design. These considerations show that the robustness curves cross one another at $g_\times$:

$$g_\times = c\left(\frac{s_2 f_1 - s_1 f_2}{(s_1 - s_2)\widetilde{E}}\right)^{2/3} \tag{20}$$

Design 1 is more robust than design 2 for $g_c > g_\times$, and design 2 is more robust than design 1 otherwise. Thus the robust preference between the designs depends on our outcome requirement, $g_c$.

Crossing of robustness curves, and reversal of preference between the designs, can occur with fixed error estimate, $s$, and different forces and combined moduli. Specifically, crossing occurs if:

$$f_1 > f_2 \quad \text{and} \quad \frac{f_1}{\widetilde{E}_1} < \frac{f_2}{\widetilde{E}_2} \tag{21}$$

Crossing of the robustness curves occurs if the force is greater in the first design and the combined modulus of the first design is dis-proportionately greater.

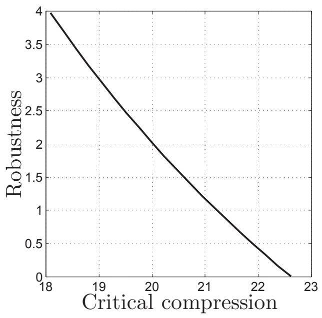


Figure 1: Robustness vs critical compression [microns].

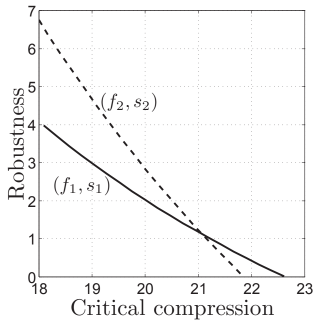


Figure 2: Robustness vs critical compression [microns] for two different designs.

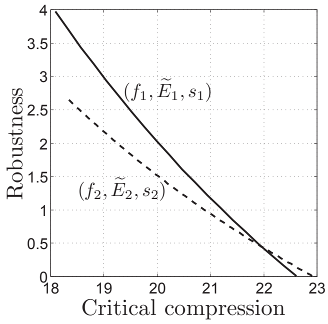


Figure 3: Robustness vs critical compression [microns] for two different designs.

### 4.1.3 Example

We illustrate the properties discussed in section 4.1.2 using the robustness function of eq.(15) from section 4.1.1.

Consider Hertzian contact between a steel and a bronze sphere, with estimated Young's moduli of 200GPa and 110GPa respectively, and radii of 0.01m and 0.04m respectively. The combined estimated modulus is $\widetilde{E} = 141.9$GPa. The Poisson ratio is 0.3.

Fig. 1 shows a robustness curve for force $f = 1000$N with error estimate $s = 0.1\widetilde{E} = 14.19$GPa. The trade off and zeroing properties are illustrated in this figure. The horizontal intercept equals the estimated compression, which from eq.(16) is $\widetilde{g} = 22.6$ microns.

Fig. 2 shows robustness curves for two choices of $f$ and $s$ with the same moduli as in fig. 1. $(f_1, s_1) = (1000, 0.1\widetilde{E})$ which is the case portrayed in fig. 1. $(f_2, s_2) = (950, 0.05\widetilde{E})$, so the conditions for preference reversal in eq.(18) are satisfied. The robustness curves cross one another, implying reversal of preference between these two designs, depending on the critical compression.

Fig. 3 shows robustness curves for two choices of $f$, $\widetilde{E}$ and $s$, so now we are considering different materials. The first design is the same steel and bronze spheres as in fig. 1, so $f = 1000$N, $\widetilde{E} = 141.9$GPa, and $s = 0.1\widetilde{E} = 14.19$GPa. The second design uses wrought iron and aluminium spheres, with estimated Young's moduli of 110GPa and 69GPa respectively, and radii of 0.01m and 0.04m respectively. The combined estimated modulus is $\widetilde{E} = 84.8$GPa. The moduli are more variable and less accurately estimated, so $s = 0.15\widetilde{E} = 12.72$GPa. The Poisson ratio is 0.3. The robustness curves cross one another as expected from eq.(21).

## 4.2 Choosing the Load Given Data for Estimating the Uncertain Young's Modulus

### 4.2.1 Formulation

$N$ observations have been made of the relation between load, $f$, and total compression, $g$. Denote these data $(\widetilde{\phi}_i, \widetilde{\gamma}_i)$, $i = 1, \ldots, N$, where the $\widetilde{\phi}_i$'s are the observed loads and the $\widetilde{\gamma}_i$'s are the corresponding observed compressions. The vectors of observed forces and compressions are $\widetilde{\phi}$ and $\widetilde{\gamma}$. Arbitrary vectors of loads and compressions are denoted $\phi$ and $\gamma$.

We want to use these data to estimate the combined Young's modulus, $E$, which we could do using eq.(4) and the method of least squares estimation. We want to use the estimate of $E$ to choose a load, $f$, so that future applications will result in compression no less than $g_c$. That is, our outcome requirement for the loading is eq.(10).

However, we are uncertain that the modulus for these test data is the same as the modulus that will be encountered in practice. In fact, as in section 4.1, we believe that the actual future modulus could be greater—by an unknown amount—than the value which holds for the test data. We will use the info-gap model of eq.(6) to represent this uncertainty in the future modulus.

Likewise, we expect that greater forces would be required—on future real specimens—to produce the compressions that we have observed in the tests. In other words, the available test data are biased—in comparison to future real applications—by having forces that are too low for the corresponding compressions. Consider an observed pair, $(\widetilde{\phi}_i, \widetilde{\gamma}_i)$. We expect that, for any future compression equal to $\widetilde{\gamma}_i$, the corresponding force could exceed $\widetilde{\phi}_i$ by an unknown amount. So let us define the following info-gap model for uncertain bias of the $N$ observed forces, corresponding to the $N$ observed compressions, analogous to eq.(8):

$$\mathcal{F}(h) = \left\{ \phi : \ 0 \leq \phi_i - \widetilde{\phi}_i \leq \widetilde{\phi}_i h, \ , i = 1, \ldots, N \right\}, \quad h \geq 0 \tag{22}$$

For any choice of the modulus, $E$, the squared error of data $(\phi, \gamma)$ is:

$$S(E; \phi, \gamma) = \sum_{i=1}^{N} [\gamma_i - g(\phi_i, E)]^2 \tag{23}$$

where $g(\phi_i, E)$ is calculated from eq.(4).

If we believed that the observations $(\widetilde{\phi}, \widetilde{\gamma})$ were representative of future applications, then we would choose $E$ to minimize $S(E; \widetilde{\phi}, \widetilde{\gamma})$. However, we believe that the data are biased to an unknown amount as expressed by the info-gap model of eq.(22).

We will therefore choose $E$ to make the squared error acceptably small—though perhaps not minimal—for a wide range of realizations of the forces corresponding to the observed compressions. We denote the chosen modulus by $E_{\text{est}}$.

Let $S_c$ be an acceptably small value of the squared error. This might be chosen as a value not too much larger than the nominal value of the least squared error. We want to choose $E_{\text{est}}$ so that $S(E_{\text{est}}; \phi, \widetilde{\gamma})$ is less than $S_c$ for a wide range of realizations of $\phi$ in the info-gap model $\mathcal{U}(h)$. Our outcome requirement for the estimation is:

$$S(E_{\text{est}}; \phi, \widetilde{\gamma}) \leq S_c \tag{24}$$

The robustness—to force uncertainty—of estimated modulus $E_{\text{est}}$, is the greatest horizon of uncertainty in force vectors, $\phi$, up to which the squared error satisfies the requirement in eq.(24):

$$\widehat{h}_e(E_{\text{est}}, S_c) = \max \left\{ h : \ \left( \max_{\phi \in \mathcal{F}(h)} S(E_{\text{est}}; \phi, \widetilde{\gamma}) \right) \leq S_c \right\} \tag{25}$$

Note that eq.(25) uses the observed compressions, $\widetilde{\gamma}$. The forces, $\phi$, are uncertain. The observed forces, $\widetilde{\phi}$, appear only as the end points of the info-gap model $\mathcal{F}(h)$ in eq.(22). A large value of $\widehat{h}_e(E_{\text{est}}, S_c)$ implies that this estimate of $E$ has adequate fidelity to a wide range of data that is less biased than our actual observations. A large value of $\widehat{h}_e$ is preferable over a small value of $\widehat{h}_e$.

However, we also want to select the load, $f$, and the error-estimate $s$, to satisfy the outcome requirement in eq.(10). That is, we have to make three decisions: an estimate of $E$ based on the observations, and a choice of $f$ and $s$ based on this estimate of $E$. We want these decisions to have high fidelity to truly representative data (which is unknown) as expressed by eq.(24), and we want these decisions to satisfy the compression requirement of eq.(10).

To consider both requirements, and all three decisions, we combine the robustness functions of eqs.(11) and (25):

$$\widehat{h}_{\text{eg}}(E_{\text{est}}, f, s, S_{\text{c}}, g_{\text{c}}) = \max \left\{ h : \ \left( \max_{\phi \in \mathcal{F}(h)} S(E_{\text{est}}; \phi, \widetilde{\gamma}) \right) \leq S_{\text{c}}, \ \left( \min_{E \in \mathcal{U}(h, \widetilde{E})} g(f, E) \right) \geq g_{\text{c}} \right\} \tag{26}$$

The righthand side contains three symbols for the modulus: $E$, $\widetilde{E}$ and $E_{\text{est}}$. The latter denotes the estimated modulus that we must choose, and it appears in the squared error function, $S(E_{\text{est}}; \phi, \widetilde{\gamma})$. The endpoint in the info-gap model $\mathcal{U}(h, \widetilde{E})$ of eq.(6) uses the nominal estimate, $\widetilde{E}$, based on eq.(1). The term $E$ represents the uncertain modulus in the argument of $g(f, E)$ and the elements of the info-gap model $\mathcal{U}(h, \widetilde{E})$.

### 4.2.2 Evaluating the Robustness

The robustness $\widehat{h}_{\text{eg}}(E_{\text{est}}, f, s, S_{\text{c}}, g_{\text{c}})$ in eq.(26) is the greatest horizon of uncertainty, $h$, at which *both* of the requirements are satisfied. The first requirement—on $S_{\text{c}}$—appears in the robustness function $\widehat{h}_{\text{e}}(E_{\text{est}}, S_{\text{c}})$ in eq.(25), so $\widehat{h}_{\text{e}}(E_{\text{est}}, S_{\text{c}})$ is the greatest $h$ which satisfies the first requirement. The second requirement appears in the robustness function of eq.(11), $\widehat{h}_{\text{g}}(f, s, g_{\text{c}})$, which is the greatest $h$ that satisfies the second requirement. Hence $\widehat{h}_{\text{eg}}(E_{\text{est}}, f, s, S_{\text{c}}, g_{\text{c}})$ is the *lesser* of the two robustnesses, $\widehat{h}_{\text{e}}(E_{\text{est}}, S_{\text{c}})$ and $\widehat{h}_{\text{g}}(f, s, g_{\text{c}})$:

$$\widehat{h}_{\text{eg}}(E_{\text{est}}, f, s, S_{\text{c}}, g_{\text{c}}) = \min \left[ \widehat{h}_{\text{e}}(E_{\text{est}}, S_{\text{c}}), \ \widehat{h}_{\text{g}}(f, s, g_{\text{c}}) \right] \tag{27}$$

An explicit expression for $\widehat{h}_{\text{g}}(f, s, g_{\text{c}})$ appears in eq.(15).

**Evaluating the inverse of $\widehat{h}_{\text{e}}(E_{\text{est}}, S_{\text{c}})$.** Let $\mu_{\text{e}}(h)$ denote the inner maximum in eq.(25), which is the inverse of $\widehat{h}_{\text{e}}(E_{\text{est}}, S_{\text{c}})$. Combining eqs.(16) and (23) we obtain:

$$\mu_{\text{e}}(h) = \max_{\phi \in \mathcal{F}(h)} \sum_{i=1}^{N} \left[ \widetilde{\gamma}_i - c \left( \frac{\phi_i}{E_{\text{est}}} \right)^{2/3} \right]^2 \tag{28}$$

This maximum occurs when each term in the sum is maximal. Each term is maximal when its value of $\phi_i$ is replaced by either $\widetilde{\phi}_i$ or $(1+h)\widetilde{\phi}_i$. Each term is maximized independently. Adopting a convenient notation, we can write $\mu_{\text{e}}(h)$ as:

$$\mu_{\text{e}}(h) = \sum_{i=1}^{N} \max_{\text{`or'}} \left[ \widetilde{\gamma}_i - c \left( \frac{\{1 \text{ or } (1+h)\}\widetilde{\phi}_i}{E_{\text{est}}} \right)^{2/3} \right]^2 \tag{29}$$

**Suppose we want to plot $\widehat{h}_{\text{eg}}(E_{\text{est}}, f, s, S_{\text{c}}, g_{\text{c}})$ vs $g_{\text{c}}$** with $S_{\text{c}}$ fixed and with the decision variables $E_{\text{est}}, f, s$ fixed. We will denote this $\widehat{h}_{\text{eg}}(g_{\text{c}} | S_{\text{c}}, \widetilde{E}, f, s)$. From eq.(27) we can express this as:

$$\widehat{h}_{\text{eg}}(g_{\text{c}} | S_{\text{c}}, E_{\text{est}}, f, s) = \begin{cases} \widehat{h}_{\text{g}}(f, s, g_{\text{c}}) & \text{if } \widehat{h}_{\text{g}}(f, s, g_{\text{c}}) \leq \widehat{h}_{\text{e}}(E_{\text{est}}, S_{\text{c}}) \\ \widehat{h}_{\text{e}}(E_{\text{est}}, S_{\text{c}}) & \text{if } \widehat{h}_{\text{g}}(f, s, g_{\text{c}}) > \widehat{h}_{\text{e}}(E_{\text{est}}, S_{\text{c}}) \end{cases} \tag{30}$$

where $\widehat{h}_{\text{e}}(E_{\text{est}}, S_{\text{c}})$ is a fixed number since $E_{\text{est}}$ and $S_{\text{c}}$ are fixed. Thus a plot of $\widehat{h}_{\text{eg}}(g_{\text{c}} | S_{\text{c}}, E_{\text{est}}, f, s)$ vs $g_{\text{c}}$ will increase as $g_{\text{c}}$ decreases, following the curve $\widehat{h}_{\text{g}}(f, s, g_{\text{c}})$ vs $g_{\text{c}}$, until $\widehat{h}_{\text{g}}(f, s, g_{\text{c}})$ reaches the value $\widehat{h}_{\text{e}}(E_{\text{est}}, S_{\text{c}})$. For all smaller values of $g_{\text{c}}$, $\widehat{h}_{\text{eg}}(g_{\text{c}} | S_{\text{c}}, E_{\text{est}}, f, s)$ is held at $\widehat{h}_{\text{e}}(E_{\text{est}}, S_{\text{c}})$. The plot of $\widehat{h}_{\text{eg}}(E_{\text{est}}, f, s, S_{\text{c}}, g_{\text{c}})$ vs $g_{\text{c}}$ will be flat for small values of $g_{\text{c}}$, reach a kink and then decrease for larger values of $g_{\text{c}}$ as in figs. 4–6 which will be discussed shortly.

Eq.(30) depends on explicit knowledge of $\widehat{h}_{\text{e}}(E_{\text{est}}, S_{\text{c}})$, while we are only able to evaluate its inverse, $\mu_{\text{e}}(h)$, eq.(29). It is thus more convenient to use an expression for the inverse of $\widehat{h}_{\text{eg}}(g_{\text{c}} | S_{\text{c}}, E_{\text{est}}, f, s)$, for fixed $S_c$. $\mu_g(h)$ denotes the inverse of $\widehat{h}_g(f, s, g_c)$, for which eq.(13) is an explicit expression. The inverse of $\widehat{h}_{eg}(g_c|S_c, E_{est}, f, s)$ vs $g_c$, for fixed $S_c$, $\widetilde{E}$, $f$ and $s$ is:

$$\mu_{eg}(h|S_c, E_{est}, f, s) = \begin{cases} \mu_g(h) & \text{if } \mu_e(h) \leq S_c \\ -\infty & \text{if } \mu_e(h) > S_c \end{cases} \tag{31}$$

The upper and lower lines of eq.(31) correspond to the upper and lower lines of eq.(30).

**Suppose we want to plot** $\widehat{h}_{eg}(E_{est}, f, s, S_c, g_c)$ **vs** $S_c$ with $g_c$ fixed and with the decision variables $E_{est}, f, s$ fixed. We will denote this $\widehat{h}_{eg}(S_c|g_c, E_{est}, f, s)$. From eq.(27) we can express this as:

$$\widehat{h}_{eg}(S_c|g_c, E_{est}, f, s) = \begin{cases} \widehat{h}_e(E_{est}, S_c) & \text{if } \widehat{h}_e(E_{est}, S_c) \leq \widehat{h}_g(f, s, g_c) \\ \widehat{h}_g(f, s, g_c) & \text{if } \widehat{h}_e(E_{est}, S_c) > \widehat{h}_g(f, s, g_c) \end{cases} \tag{32}$$

where $\widehat{h}_g(f, s, g_c)$ is a fixed number since $f, s$ and $g_c$ are fixed. Thus a plot of $\widehat{h}_{eg}(S_c|g_c, E_{est}, f, s)$ vs $S_c$ will increase as $S_c$ increases, following the curve $\widehat{h}_e(E_{est}, S_c)$ vs $S_c$, until $\widehat{h}_e(E_{est}, S_c)$ reaches the value $\widehat{h}_g(f, s, g_c)$. For all larger values of $S_c$, $\widehat{h}_{eg}(S_c|g_c, E_{est}, f, s)$ is held at $\widehat{h}_g(f, s, g_c)$. The plot will thus have a kink.

### 4.2.3 Example

| run | $\widetilde{E}_{fac}$ N | $\widetilde{E}_{est}$ GPa | $S(E_{est})$ $\mu m^2$ |
|---|---|---|---|
| 7 | 1 | 141.94 | 2.8409 |
| 6 | 1.01 | 143.35 | 3.1195 |
| 9 | 1 | 141.94 | 2.6841 |
| 10 | 1.01 | 143.35 | 3.3361 |
| 12 | 1 | 141.94 | 3.2364 |
| 13 | 1.01 | 143.35 | 3.7201 |

Table 1: Data for runs. $f = 1000$N, $r_1 = 0.01$m, $r_2 = 0.04$m, $E_1 = 200$GPa, $E_2 = 110$GPa, $\widetilde{E} = 141.94$GPa, $g(\widetilde{E}) = 22.61\mu$m, $g_{frac} = 0.03$, $s_{frac} = 0.1$, $s = 14.19$GPa, $S_{cfrac} = 5$, $\nu = 0.3$.

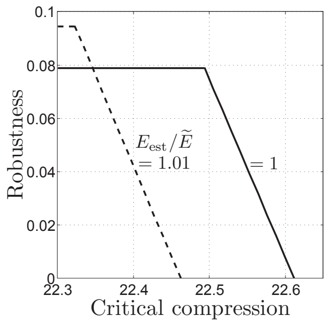


Figure 4: Robustness vs critical compression [microns]. Run 7 (—), run 6 (– –).

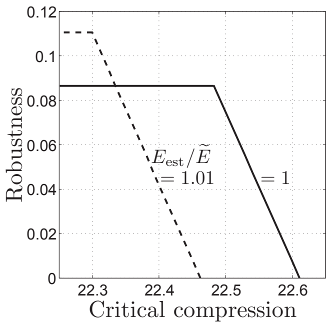


Figure 5: Robustness vs critical compression [microns]. Run 9 (—), run 10 (– –).

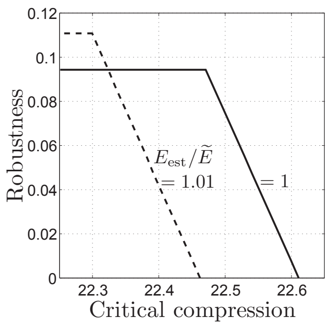


Figure 6: Robustness vs critical compression [microns]. Run 12 (—), run 13 (– –).

Figs. 4–6 show pairs of robustness curves, $\widehat{h}_{eg}(g_c|S_c, E_{est}, f, s)$ vs $g_c$ in eq.(30), for runs specified in table 1. The design load, $f$, is the same in each case. In each simulated experimental run, 8 loads, $\widetilde{\phi}_i$, are applied, equally spaced from 600N to 1200N. The 8 corresponding nominal compressions are calculated from eq.(4), and denoted $g(\widetilde{\phi}_i)$. The noisy simulated compressions, $\widetilde{\gamma}_i$, are calculated by adding zero-mean normal noise to each nominal compression with standard deviation $g_{\rm frac} g(\widetilde{\phi}_i)$:

$$\widetilde{\gamma}_i = g(\widetilde{\phi}_i) + \varepsilon_i, \quad \varepsilon_i \sim N[0, (g_{\rm frac} g(\widetilde{\phi}_i))^2] \tag{33}$$

The value of $g_{\rm frac}$, shown in table 1, is 0.03. The 2 curves in each figure are calculated with the same force-compression data. The data changes from figure to figure, which explains the differences of the robustness curves between the figures.

The critical squared error, $S_{\rm c}$, is chosen as 5 times the estimated squared error based on $\widetilde{E}_{\rm est}$:

$$S_{\rm c} = 5S(E_{\rm est}, \widetilde{\phi}, \widetilde{\gamma}) \tag{34}$$

The estimated squared errors are shown in table 1 in the column $S(E_{\rm est})$.

The solid curves in figs. 4–6 use the nominal combined modulus, $\widetilde{E}$ from eq.(1), as the estimated combined modulus, $E_{\rm est}$. The dashed curves use an estimated modulus which is 1% larger: $E_{\rm est} = 1.01\widetilde{E}$. From eq.(16) we see that the estimated compression is less for the dashed than for the solid curves.

The robustness curves in figs. 4–6 cross one another, which we now explain. The robustness at the kink of any robustness curve depends on the squared error. We can understand this as follows. $\widehat{h}_{\rm g}(g_{\rm c})$ increases as $g_{\rm c}$ decreases. According to eq.(27), the kink occurs at the value of $g_{\rm c}$ for which $\widehat{h}_{\rm g}(g_{\rm c}) = \widehat{h}(S_{\rm c})$. $S_{\rm c}$ is fixed throughout each curve, but it depends on the data and other parameters of the run through eq.(34). The dashed curves have larger squared error, $S(E_{\rm est}, \widetilde{\phi}, \widetilde{\gamma})$, because they use a sub-optimal stiffness estimate, $E_{\rm est}$. Thus the kink in the dashed curves occurs at greater robustness than in the solid curves. Since the solid curves intersect the horizontal axis further to the right, the robustness curves intersect. This intersection implies the potential for reversal of preference between the corresponding designs.

# 5 Discussion

**Robustness.** 'Robustness' has many meanings. As we have used it, the concept of robustness derives from a prior concept of non-probabilistic uncertainty. Knight (1921) distinguished between 'risk' based on known probability distributions and 'true uncertainty' for which probability distributions are not known. Similarly, Ben-Tal and Nemirovski (1999) are concerned with uncertain data within a prescribed uncertainty set, without any probabilistic information. Likewise Hites *et al.* (2006, p.323) view "robustness as an aptitude to resist to 'approximations' or 'zones of ignorance' ", an attitude adopted also by Roy (2010). We also are concerned with robustness against Knightian uncertainty. Info-gap theory has been used to model and manage uncertainty in probability distributions, but it is not based on an explicitly statistical approach to robustness as studied by Huber (1981) and many others. The concepts developed here are related to the idea of probability bounds (Ferson and Tucker, 2008), and to the concept of coherent lower previsions, as discussed elsewhere (Ben-Haim *et al.* 2009).

Many engineering researchers, beginning in the 1960s, developed estimation and control algorithms for linear dynamic systems based on sets of uncertain inputs. Schweppe (1973) for instance develops inference and decision rules based on assuming that the uncertain phenomenon can be quantified in such a way as to be bounded by an ellipsoid, with no probability function involved.

The concept of robustness in this paper is in the tradition of ideas that we have described.

**Optimizing and satisficing.** To 'satisfice' means "To decide on and pursue a course of action that will satisfy the minimum requirements necessary to achieve a particular goal." (Oxford English Dictionary). Simon was the first to use the term in this technical sense, which is an old alteration of the ordinary English word "satisfy". Simon (1956) wrote "Evidently, organisms adapt well enough to 'satisfice'; they do not, in general, 'optimize'."

In the context of engineering design, 'satisficing' is the same as 'satisfying a design specification'. Designers routinely satisfy specified bounds on performance outcomes (e.g. deflection, weight, rigidity, etc.), rather than optimize (e.g. minimize or maximize) these quantities.

The putative reason for satisfying design specs is that they specify what constitutes acceptable performance, and that is, by definition, good enough. But there is a deeper motivation for designing to spec—rather than optimizing—that we can appreciate from info-gap theory.

A satisficing outcome is not, usually, the best possible performance. It is sub-optimal in performance, meeting a specified criterion (or perhaps several criteria). But sub-optimal performance can usually be achieved by alternative designs. The designer who satisfices (rather than optimizes) thus has an additional degree of freedom in choosing the design. This freedom is exploited to maximize the robustness of the system against uncertainty, ignorance, and surprise. This is precisely the info-gap robust-satisficing strategy. The satisficer does not shirk the potential for doing better. Satisficing, when linked to robustness, is a means of maximizing the immunity to uncertainty for achieving critical performance.

**Maximizing the robustness.** A robustness function, such as in eqs.(11), (25) and (26), depends on both design and performance parameters. The robustness is monotonic in the performance parameters, as seen in figs. 1–6. This monotonicity expresses the trade off between robustness and performance discussed in section 1. The robustness function is not, in general, monotonic in the design parameters. The maximization of robustness referred to in the previous paragraph is with respect to the design parameters. This maximization is usually also subject to constraints on feasible designs.

**Info-gap robustness and min-max.** Wald (1945) studied the problem of statistical hypothesis testing based on a random sample whose probability distribution is not known, but whose distribution is known to belong to a given class of distribution functions. Wald states that "in most of the applications not even the existence of ... an a priori probability distribution [on the class of distribution functions] ... can be postulated, and in those few cases where the existence of an a priori probability distribution ... may be assumed this distribution is usually unknown." (p.267). Wald introduced a loss function expressing the "relative importance of the error committed by accepting" one hypothesized subset of distributions when a specific distribution in fact is true. (p.266). He notes that "the determination of the [loss function] is not a statistical question and is considered here as given." (p.266). Wald developed a decision procedure which "minimizes the maximum ... of the risk function." (p.267). Thus was born, in its modern form, the method of min-max.

The relation between min-max and info-gap robust-satisficing has been discussed at length elsewhere (Ben-Haim *et al.* 2009). The two methods have much apparent similarity, though also important differences. Most significantly, they depend on different prior information, and can lead to different solutions. Briefly, min-max requires knowledge of a worst case. In contrast, the horizon of uncertainty of an info-gap model is unknown and unbounded, thus deliberately avoiding the specification of a worst case. On the other hand, the info-gap robustness does require the analyst to specify the worst acceptable outcome, which in engineering design is usually a design specification.

**Opportuneness: The other side of uncertainty.** The info-gap robustness function is a tool for protecting against pernicious surprises. But some surprises are propitious. The info-gap opportuneness function enables the designer to explore the propitious side of uncertainty, and to incorporate the potential for favorable surprise into the design process (Ben-Haim 2006).

**Value of information.** Information (including knowledge and understanding) is valuable to the extent that it enhances the robustness of the designer's decisions against uncertainty and surprise. Information is not valuable in its own right, disconnected from the end use which employs the information. One implication is that separation of the modeler from the decision maker may lead to poor—unreliable—decisions. The value of information is illustrated in fig. 2 where we see robustness curves for two different levels of precision in estimating the Young's modulus. The relative error in the dashed case, $s_2/f_2$, is less than in the solid case. These two curves cross one another, so the robust preference for the more precise estimate (dash) depends on the required compression, $g_{\rm c}$. Better estimation is *not* automatically preferred. The value of the increment of information, as expressed by the increment in robustness, depends on the outcome requirement.

**Other applications of info-gap robustness.** Info-gap theory has been applied to many problems in engineering analysis and design, including fault diagnosis, reliability assessment, structural dynamics, fatigue, and so on. Info-gap theory is particularly suited for modeling and managing the uncertainty in functional shape, as well as uncertainty in the values of parameters. For instance, many applications deal with uncertainty in the shape of constitutive relations, or uncertainty in the tails of a probability distribution. Info-gap theory has also been applied widely outside of engineering, in biological conservation, economics, medicine, project management, homeland security and other areas (http://info-gap.com).

**Acknowledgements**

The author is indebted to Professors Eli Altus, David Elata and Doron Shilo for valuable insights into contact mechanics.

## 6 References

1. Beitz, W. and K.-H. Küttner, eds., 1994, *Dubbel Handbook of Mechanical Engineering,* trans. by M.J. Shields, Springer-Verlag, London, p.B38.
2. Ben-Haim, Yakov, 1999, Set-models of information-gap uncertainty: Axioms and an inference scheme, *Journal of the Franklin Institute,* 336: 1093–1117.
3. Ben-Haim, Yakov, 2006, *Info-Gap Decision Theory: Decisions Under Severe Uncertainty,* 2nd ed., Academic Press, London.
4. Ben-Haim, Y. Dacso, C.C., Carrasco J. and Rajan, N., 2009, Heterogeneous Uncertainties in Cholesterol Management, *International Journal of Approximate Reasoning,* 50: 1046–1065.
5. Ben-Tal, A., Nemirovski, A., 1999, Robust solutions of uncertain linear programs, *Oper. Res. Lett.* 25, 1–13.
6. Burgman, M.A., B.A. Wintle, C.A. Thompson, A. Moilanen, M.C. Runge, and Yakov Ben-Haim, 2010, Reconciling uncertain costs and benefits in Bayes nets for invasive species management, *Risk Analysis,* vol.30, #2, pp.277–284.
7. DeGroot, Morris H., 1986, *Probability and Statistics,* 2nd ed., Addison-Wesley, Reading, MA.
8. Ferson, S. and Tucker, W.T., 2008, Probability boxes as info-gap models, Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS) 2008, Article number 4531314.
9. Hites, R., De Smet, Y., Risse, N., Salazar-Neumann, M., Vincke, P., 2006, About the applicability of MCDA to some robustness problems, *Eur. J. Oper. Res.* 174, 322–332.
10. Huber, P.J., 1981. *Robust Statistics,* John Wiley, New York.
11. Knight, Frank H., 1921, *Risk, Uncertainty and Profit,* Hart, Schaffner and Marx. Re-issued by Harper Torchbooks, New York, 1965.
12. Oxford English Dictionary, entry 'satisfice, 2', online version, accessed 8.7.2011.
13. Roy, B., 2010, Robustness in operational research and decision aiding: A multi-faceted issue, *Eur. J. Oper. Res.* 200, 629–638.
14. Schweppe, F.C., 1973. *Uncertain Dynamic Systems,* Prentice-Hall, Englewood Cliffs.
15. Simon, Herbert, 1956, Rational choice and the structure of the environment, *Psychological Review,* 63(2), 129–138.
16. Wald, A., 1945, Statistical decision functions which minimize the maximum risk, *Annals of Mathematics,* 46(2), 265–280.

# 7 Notation

| Symbol | Meaning |
|---|---|
| $E$ | Young's modulus |
| $\widetilde{E}$ | Nominal Young's modulus |
| $E_{\mathrm{est}}$ | Estimated Young's modulus |
| $f$ | Force |
| $g$ | Compression |
| $\widetilde{g}$ | Estimated compression |
| $g_{\mathrm{c}}$ | Critical compression |
| $h$ | Horizon of uncertainty |
| $\widehat{h}$ | Robustness |
| $p$ | Probability, probability density |
| $r$ | radius |
| $s$ | Error estimate |
| $S$ | Squared error |
| $S_{\mathrm{c}}$ | Critical squared error |
| $\mathcal{U}(h)$ | Info-gap model of uncertainty |
| $\gamma$ | Vector of observed compressions |
| $\mu$ | Inverse of robustness function |
| $\phi$ | Vector of applied loads |
| $\sigma$ | Stress |

Table 2: Notation.