

Lecture Notes on Info-Gap Learning

Yakov Ben-Haim

Yitzhak Moda'i Chair in Technology and Economics

Faculty of Mechanical Engineering

Technion — Israel Institute of Technology

Haifa 32000 Israel

<http://info-gap.com> <http://www.technion.ac.il/yakov>

yakov@technion.ac.il

Source material: Yakov Ben-Haim, 2006, *Info-Gap Decision Theory: Decisions Under Severe Uncertainty*, 2nd edition, Academic Press. Chapter 8.

A Note to the Student: These lecture notes are not a substitute for the thorough study of books. These notes are no more than an aid in following the lectures.

Contents

1	Learning and Deciding	2
2	Info-Gap Supervision of a Classifier	8
2.1	Robustness of a Classifier	8
2.2	Asymptotic Robustness	10
2.3	Robust Optimal Classifier	13
2.4	Robust Severe Tests of Truth	16
2.5	Up-Dating Info-Gap Models	19
2.6	Plantar Pressures in Metatarsal Pathology	25
3	Acoustic Noise	30
3.1	Empirical Robustness	31
3.2	Up-Dating the Acoustic Uncertainty Model	32

1 Learning and Deciding

§ The outline of this section is summarized in a transparency, see p. 7.

¶ When you study a decision problem and reach a decision to do such-and-such, you can be said to have learned something.

¶ The process of reaching the decisions:

- “This bird is an albatross.”
- “That snark was a boojum.”

can each be seen as a process of learning something.

Decision: sorting out a vague and unclear situation.

A decision is the content of what we have learned.

¶ We will consider the questions:

- What is learning?
- How do the general aspects of learning pertain to decision processes?

¶ Learning can be viewed from at least 3 different perspectives:

- **Content:** what has been learned?
- **Function:** what effect does the learning
have on subsequent behavior?
- **Structure:** what are the processes by which learning occurs?

¶ Piaget's theory of intelligence:

- Focuses on **structure and processes** of thoughtful behavior.
- De-emphasizes **content** of intelligent thought.
- E.g.:
Study logical structure of a child's discovery process.
Rather than
Study subjects which children are interested in discovering.

¶ In decision theory we focus much more on the **structure of the decision process** than on the **content** or **function** of specific decisions.

¶ Another taxonomy of learning (also due to Piaget) distinguishes between:

- Immediate acquisition and processing of information.
- Modification of cognitive structures governing the learning process.

Examples:

- Classification: immediate acquisition of information.
 - E.g. learning to identify albatrosses or snarks.
 - E.g. learning that not all 4-legged fuzzy animals are dogs.
- Intellectual maturation: modification of cognitive structures.
 - E.g. Learning that albatrosses evolved from proto-albatrosses, rather than by being created from nothing.
 - E.g. learning that taxonomic classification is ambiguous.

¶ Rousseau: key to intellectual maturation is **error**.

Applying this concept of learning to decision processes,
the key questions are:

- How to decide a given decision algorithm is **defective**?
- What constitutes **evidence against** a given decision model?
- How is such evidence used to improve the decision model?

A modern expression of the same idea is: **falsification** (Popper):

Improvement by exclusion of inadequate elements.

This is an **evolutionary methodology**.

¶ Evolutionary theories are of two sorts:

- Historicist, deterministic, strongly predictive.
 - Based on “laws” (natural, social, logical, etc.)
 - Examples:
 - All classical physics.
 - Hegel: Historical dialectic: thesis, antithesis, synthesis.
 - Marx: 5 irrevocable stages of history.
 - primitive, feudal, capitalist, socialist, communist.
 - Darwin: biological evolution by survival of fittest.
 - Freud: universal principals of human psychology:
 - Sex drive, dream symbols, super-ego, etc.
- Adaptive, interactive, undeterministic and unpredictable.
 - Driven by non-causal innovation: “random”, un-lawlike.
 - Examples:
 - Darwin: biological evolution by random modification. (Gould)
 - Some modern physics: quantum mechanics, chaotic systems.
 - Popper: Piecemeal social engineering.
 - Unpredictability of intellectual discovery.
 - Non-computability: E.g. Penrose’s toy universe:
 1. Next state determined by, but not computable from, history.
 2. Tiling problem: given a set of polygons, determine whether or not it will tile the plane. There is no algorithm solution that solves this in finite time for all sets of polygons.
 3. The “toy universe” progresses with this rule:
 - The state at time step n given by S_n , a set of polygons.
 - The universe moves to state (set) S_{n+1} at $t + 1$ if S_n will tile the plane.
 - The universe moves to state (set) S_{n+2} at $t + 1$ if S_n will not tile the plane.
 - This universe is deterministic but not computable.

We will want to keep track of both types of evolutionary methods.

1

¹Penrose, Roger, *Shadows of the Mind: A Search for the Missing Science of Consciousness*. Oxford University Press, 1994. See pp.29–33.

¶ We will study a two tiered decision situation:

- Lower level:

A decision algorithm is making operative decisions.

- Upper level:

IGDT supervises, modifies and improves the operative algorithm.

¶ In section 2 we will consider

info-gap supervision of a classifier.

Lower level: operative decision algorithm

assimilates and classifies data.

Upper level: Classifier is modified.

Limitation of the classifier: inadequacy of the info-gap model.

¶ Immediate learning:

Results of the classification.

¶ Structural Learning:

- Modifying the info-gap model, (structural learning)

- Improving the classifier, (behavioral learning).

§ Learning and deciding:

- **Content.**
- **Function.**
- **Structure.**

§ Piaget: structure, not content of children's learning.**§ Piaget:**

- **Immediate acquisition of knowledge.**
- **Intellectual maturation.**

§ Rousseau: Learning from error.**§ Popper:**

- **Falsification.**
- **Evolutionary process.**

§ Evolutionary theories:

- **Historicist, deterministic, predictive.**
- **Adaptive, interactive, uncertain, non-predictive.**

§ 2-tiered decision process:

- **Make operative decision.**
- **Modify decision algorithm.**

2 Info-Gap Supervision of a Classifier

2.1 Robustness of a Classifier

¶ Classification: use measured vector u
to select from among a number of classes.

¶ The problem:

- No class produces a unique u -vector.
- The u -vectors of distinct classes may be identical: classes overlap.

¶ Severe uncertainty: u -vectors for n th class is an info-gap model:

$$\mathcal{U}_n(h, \tilde{u}_n), \quad h \geq 0 \quad (1)$$

¶ The class index can be taxonomically informative. E.g.

$$n = (730, 02, 1) \quad (2)$$

represents the orthopedic pathology:

Acute osteomyelitis of the 2nd right metatarsal.

¶ Distance measure between classes, $\| \cdot \|$:
Degree of qualitative difference between the classes.
For instance:

$$\|n - m\| = \text{'large'} \quad (3)$$

means that these classes are quite different.

¶ Classification algorithm:

$$C(u) = n \quad (4)$$

Measurement u interpreted as arising from class n .

¶ Robustness function for algorithm C :

$$\hat{h}(C, r_c) = \max \{h : \|C(u) - n\| \leq r_c, \text{ for all } u \in \mathcal{U}_n(h, \tilde{u}_n) \text{ and for all } n\} \quad (5)$$

¶ Note usual trade-off between \hat{h} and r_c , fig. 1.

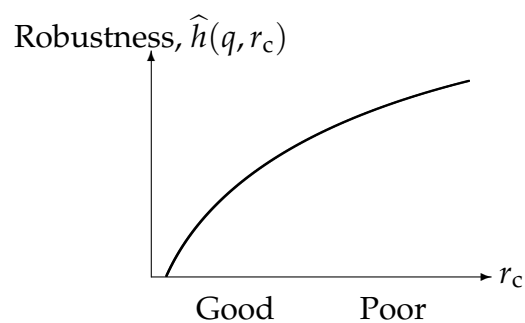


Figure 1: Trade-off of robustness against performance.

¶ Two-tiered decision process:

- $\hat{h}(C, r_c)$ is “supervising” the classification alg. C .
- $C(u)$ decides on the provenance of measurement u .
- In $\hat{h}(C, r_c)$ we consider C as a decision function:
 - \hat{h} depends on structure of classifier.
 - \hat{h} can be improved by modifying C .

¶ Calibrate $\hat{h}(C, r_c)$ in terms of asymptotic robustness, which we now explain.

2.2 Asymptotic Robustness

¶ Definition:

$\mathcal{U}(h, \tilde{u})$ is an **unbounded info-gap model** if:

For any vector u ,

There exists an h such that

$u \in \mathcal{U}(h, \tilde{u})$.

Most of the common info-gap models are unbounded. E.g.:

$$\mathcal{U}(h, \tilde{u}) = \left\{ u : (u - \tilde{u})^T W (u - \tilde{u}) \leq h^2 \right\}, \quad h \geq 0 \quad (6)$$

¶ If the info-gap models representing class uncertainty are unbounded,

Then any u could come from any info-gap model.

This is the cause of **classification ambiguity**.

¶ **Asymptotic robustness:**

$$\hat{h}_\infty = \max \{ h : \mathcal{U}_n(h, \tilde{u}_n) \cap \mathcal{U}_m(h, \tilde{u}_m) = \emptyset \text{ for all } m \neq n \} \quad (7)$$

¶ **Meaning of \hat{h}_∞ :**

- \hat{h}_∞ independent of classification algorithm.
- \hat{h}_∞ depends only on the info-gap models.

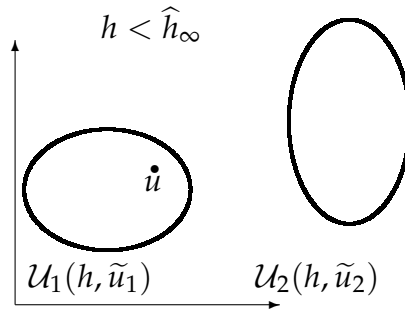


Figure 2: Non-intersecting info-gap models, illustrating unambiguous classification for $h < \hat{h}_\infty$.

¶ If the info-gap models accurately represent the uncertainty.

And if $h < \hat{h}_\infty$, (fig. 2)

then any measured u is consistent with exactly one info-gap model.

- In this case an exhaustive-search algorithm will always choose the correct class.
- Some other algorithm may also always choose the correct class.
- A different classifier may sometimes err.

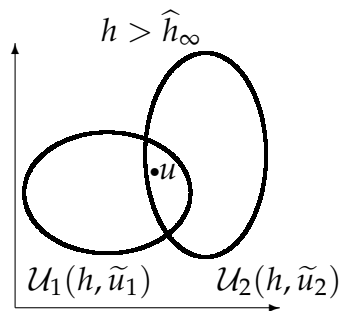


Figure 3: Intersecting info-gap models, illustrating ambiguous classification for $h > \hat{h}_\infty$.

¶ If the info-gap models accurately represent the uncertainty.

And if $h > \hat{h}_\infty$, (fig. 3)

then some measured u 's are consistent with several info-gap models.

- In this case **no** algorithm will always choose the correct class.
- Any u in the intersection can "fool" any algorithm.

¶ Meaning of \hat{h}_∞ :

- \hat{h}_∞ : limiting level of info-gap beyond which classification ambiguity occurs.
- \hat{h}_∞ : limiting error-free robustness of any realizable algorithm $C(u)$.
- For any classifier $C(u)$:
$$\hat{h}(C, 0) \leq \hat{h}_\infty$$
- If $\hat{h}(C, 0) \ll \hat{h}_\infty$ then C could be improved substantially.
- If $\hat{h}(C, 0)$ only slightly $< \hat{h}_\infty$
then C is about as good as can be demanded **with these info-gap models**.
- If the info-gap models can be improved, **then** \hat{h}_∞ can be improved.

¶ Value judgment:

Calibration of $\hat{h}(C, 0)$ in terms of \hat{h}_∞ is an **analogical inference** as discussed in lecture on value judgments.

2.3 Robust Optimal Classifier

¶ Definition. $C(u)$ is a **robust optimal classifier** if:

$$\widehat{h}(C, 0) = \widehat{h}_\infty \quad (8)$$

No algorithm can have greater robustness at $r_c = 0$.

¶ Why are we interested in robust optimal classifiers?

- Their performance is “optimal”.
- We will see their pivotal importance in up-dating the info-gap models in the structural learning of a classification task.

¶ Definition. **Gap function:**

$$\Gamma(u) = \min \{h : u \in \mathcal{U}(h, \tilde{u})\} \quad (9)$$

- $\Gamma(u)$ is the lowest h at which u is consistent with $\mathcal{U}(h, \tilde{u})$.
- If $\Gamma(u)$ is small then u is highly consistent with $\mathcal{U}(h, \tilde{u})$.
- If $\Gamma(u)$ is large then u arises from $\mathcal{U}(h, \tilde{u})$ only under extraordinary circumstances.

Given a collection of info-gap models, $\mathcal{U}_n(h, \tilde{u}_n)$

We denote their gap functions by $\Gamma_n(u)$, respectively.

¶ Consider the following classification algorithm:

$$C(u) = n \text{ if } \Gamma_n(u) \leq \Gamma_m(u) \text{ for all } m \neq n \quad (10)$$

(With some tie-breaking rule.)

This is a nearest-neighbor decision rule.

Probabilistic analog: maximum likelihood decision algorithm.

¶ This robust optimal classifier depends upon the structure of the info-gap models.

¶ We will see that eq.(10) is a **robust optimal classifier**.

We must show that $\hat{h}(C, 0) = \hat{h}_\infty$.

In other words, we must show that:

If, for some h :

$$\mathcal{U}_n(h, \tilde{u}_n) \cap \mathcal{U}_m(h, \tilde{u}_m) = \emptyset \text{ for all } m \neq n \quad (11)$$

then, for the same h :

$$C(u) = n \text{ for all } u \in \mathcal{U}_n(h, \tilde{u}_n) \quad (12)$$

Brief proof:

1. If eq.(11) holds,

then an observed u belongs to one and only one info-gap model.
at this h .

2. Suppose $u \in \mathcal{U}_n(h, \tilde{u}_n)$.

Thus $\Gamma_n(u) \leq h$.

3. But $u \notin \mathcal{U}_m(h, \tilde{u}_m)$ for all $m \neq n$.

Thus $\Gamma_m(u) > h$.

4. Hence $\Gamma_n(u) < \Gamma_m(u)$

Thus $C(u) = n$.

2.4 Robust Severe Tests of Truth

¶ Current status:

- We have found a robust optimal classifier.
- **If** the underlying info-gap models are correct,
- **then** no other algorithm has greater error-free robustness.
- The task: **“Improve the classifier.”**
- thus becomes the new task:
- **“Improve the info-gap models.”**

¶ But perhaps the info-gap models are accurate.

We don't want to tamper with them arbitrarily.

The questions thus become:

- What constitutes **evidence against** a set of info-gap models?
- How to use that evidence to **up-date** the info-gap models.?

¶ Methodology:

- **Falsification** of an info-gap model is based on a **criterion of truth**.
- **Search for truer models** is based on a **criterion of error**.

¶ In this subsection we study **severe test of truth** as a tool for falsifying info-gap models.

¶ In the next subsection we study **up-dating info-gap models** based on a criterion of error.

¶ Truth and falsity are **complements**.

¶ Learning data:

Measurements u_1, \dots, u_K of known provenance.

u_k came from the n_k th info-gap model, $\mathcal{U}_{n_k}(h, \tilde{u}_{n_k})$.

¶ “Reasonable data” assumption:

None of the measurements is “extraordinary” or “pathological”.

¶ Correct classification by the robust optimal classifier of eq.(10)

requires:

$$\Gamma_{n_k}(u_k) < \Gamma_{n_i}(u_k) \quad (13)$$

for all $k = 1, \dots, K$ and for all $i \neq k$.

¶ So, still under “reasonable data” assumption, suppose:

$$\Gamma_{n_i}(u_k) \leq \Gamma_{n_k}(u_k) \quad (14)$$

for some k and for some $i \neq k$.

This means that u_k is mis-classified.

This is **evidence against** one of both of the info-gap models:

$$\mathcal{U}_{n_i}(h, \tilde{u}_{n_i}) \quad \text{or} \quad \mathcal{U}_{n_k}(h, \tilde{u}_{n_k}) \quad (15)$$

¶ If all the learning data are “reasonable”, **then** modify the info-gap models so the learning data are correctly classified.

¶ That is, a **severe test** of the info-gap models is:

A set of info-gap models is accepted as **provisionally true** if only extraordinary data are mis-classified by a robust-optimal classifier.

¶ Stated as an **hypothesis test**:

○ The info-gap models have passed a “**severe test of truth**” if they correctly classify all but extraordinary data.

○ Conversely:

We reject a collection of info-gap models if they do not pass this test.

2.5 Up-Dating Info-Gap Models

¶ Given:

- Learning data: u_k from class n_k , $k = 1, \dots, K$.
- Uncertainty in the measurements of class n is represented (to the best of our knowledge) by $\mathcal{U}_n(h, \tilde{u}_n)$.
- Gap function for the n th info-gap model, eq.(9) on p. 13, is $\Gamma_n(u)$.

¶ Now suppose:

Robust optimal classifier, eq.(10), errs on one or more elements in the learning set.

That is, the classifier fails the severe test.

What is to be done?

¶ Three types of corrective action are possible:

- Basic event-classes are wrong and must be revised.
- Structure of the info-gap models is wrong and must be revised.
- Parameters of the info-gap models are wrong and must be revised.

¶ We will consider the last option:

often suitable for severe lack of information.

¶ q = vector of parameters of info-gap models.

= decision vector.

¶ We will formulate an **empirical robustness function**

which depends on q and with which we up-date the info-gap models.

¶ **Empirical robustness function** $\hat{h}_e(q, r_c)$ for classification algorithm $C(u)$:

Estimate of the greatest h for which $C(u)$ errs no more than r_c .

¶ $\hat{h}_e(q, r_c)$ is evaluated from

the performance of $C(u)$ in sorting the learning data $(u_1, n_1), \dots, (u_K, n_K)$.

¶ Measurement u_k comes from class n_k .

$\Gamma_{n_k}(u_k)$ is the gap function for the corresponding info-gap model $\mathcal{U}_{n_k}(h, \tilde{u}_{n_k})$.

$\hat{h}_e(q, r_c)$ is estimated from the gap functions evaluated on the learning data:

$$\hat{h}_e(q, r_c) = \max \left\{ \Gamma_{n_i}(u_i) : \begin{array}{l} \|C(u_k) - n_k\| \leq r_c \text{ if } \Gamma_{n_k}(u_k) \leq \Gamma_{n_i}(u_i), \\ k = 1, \dots, K, \quad i = 1, \dots, K \end{array} \right\} \quad (16)$$

($\hat{h}_e(q, r_c) = 0$ if the set defined in (16) is empty.)

¶ Rough explanation:

$\hat{h}_e(q, r_c)$ is the greatest “uncertainty horizon” within which the learning data are all classified with error no greater than r_c .

¶ More precisely,

$\hat{h}_e(q, r_c)$ is the value of the greatest measured gap function, $\Gamma_{n_i}(u_i)$,

for which all measured vectors u_k (including u_i)

whose corresponding gap functions are no greater than $\Gamma_{n_i}(u_i)$,

are classified with error no larger than r_c .

¶ Further explanation. The condition:

$$\Gamma_{n_k}(u_k) \leq \Gamma_{n_i}(u_i) \quad (17)$$

is equivalent to:

$$u_k \in \mathcal{U}_{n_k}(\Gamma_{n_i}(u_i), \tilde{u}_{n_k}) \quad (18)$$

Thus all measurements in the learning set which satisfy:

$$u_k \in \mathcal{U}_{n_k}(\hat{h}_e(q, r_c), \tilde{u}_{n_k}) \quad (19)$$

are classified by $C(u)$ with error no greater than r_c .

There may be some other data points that are classified with greater error.

¶ The error-free ($r_c = 0$) empirical robustness function is:

$$\begin{aligned} \hat{h}_e(q, 0) = \max \{ \Gamma_{n_i}(u_i) : & \text{if } \Gamma_{n_k}(u_k) \leq \Gamma_{n_i}(u_i), \\ & \text{then } C(u_k) = n_k \text{ for all} \\ & k = 1, \dots, K, i = 1, \dots, K \} \end{aligned} \quad (20)$$

¶ In particular, for the robust-optimal classifier in eq.(10) on p. 14,
the error-free empirical robustness function is:

$$\begin{aligned} \hat{h}_e(q, 0) = \max \{ \Gamma_{n_i}(u_i) : & \text{if } \Gamma_{n_k}(u_k) \leq \Gamma_{n_i}(u_i), \\ & \text{then } \Gamma_{n_k}(u_k) < \Gamma_{n_j}(u_k) \text{ for all} \\ & j \neq k, k = 1, \dots, K, i = 1, \dots, K \} \end{aligned} \quad (21)$$

¶ An empirical robustness function can take any non-negative value.

¶ $\hat{h}_e(q, 0)$ can be **small**:

- If the data are very unusual,
or if the info-gap models are very unrealistic.
Then few or none of the learning data will be correctly classified.

¶ $\hat{h}_e(q, 0)$ can be **small** for an utterly opposite reason:

- If the data are highly compatible with the info-gap models
then a low h will be sufficient to “catch” the data.

¶ $\hat{h}_e(q, 0)$ can be **large**:

- If the info-gap models poorly reflect ambient uncertainty
then they must be greatly distended
in order to “catch” the learning data.

¶ $\hat{h}_e(q, 0)$ can be **large** for an utterly opposite reason:

- If the info-gap models well reflect ambient uncertainty
then $C(u)$ will “catch” even extreme data at large h .

- ¶ In our earlier study of the “regular” robustness function $\hat{h}(q, r_c)$ we learned that “bigger is better”.
- ¶ However, with the empirical robustness function $\hat{h}_e(q, r_c)$ it is not necessarily true that “bigger is better” when $\hat{h}_e(q, r_c)$ is used to up-date info-gap models. We must know **why** $\hat{h}_e(q, r_c)$ is large.
- ¶ A large value of $\hat{h}_e(q, r_c)$ may be either good or bad. We will still vary the decision vector q to modify $\hat{h}_e(q, r_c)$. However, a combined increase-and-decrease strategy must be employed.
- ¶ Special interpretation of robustness function is needed when up-dating an info-gap model.
Why?
Why is this so different from all other decision problems?

¶ Explanation:

- All decisions with uncertainty (unlike deductions) involve judgments of pragmatic truth based on severe tests.
- A severe test depends upon an assessment of uncertainty. That is, a severe test employs an uncertainty model.
- Why? (re-iterate previous discussion of severe tests):
 - A proposition is (pragmatically) true to the extent that only highly uncertain evidence contradicts the proposition.
 - Thus one must be able to assess degrees of uncertainty.
 - For this one uses \hat{h} based on an info-gap model or a probability function in statistical decisions.
- When the proposition being tested is:
“This uncertainty model is false.”
The severe test collapses: one has no uncertainty model to use.

2.6 Plantar Pressures in Metatarsal Pathology

¶ Plantar pressures vary :

- During gait.
- Across the sole of the foot.
- Between individuals.
- Between activities.
- Due to pathology.

¶ We will up-date an info-gap model for pathological plantar pressure distribution.

¶ Common orthopedic pathology: hallux valgus.

- Enlarged angle between 1st metatarsus and 1st phalanx.
- Can produce excessive pressure on 1st metatarsal-phalangeal joint.
- Surgical correction needed in extreme cases.

¶ We will consider two info-gap models:

- Normal plantar pressures: $\mathcal{U}_1(h, \tilde{u}_1), h \geq 0$.
- Pathological plantar pressures: $\mathcal{U}_2(h, \tilde{u}_2), h \geq 0$.

¶ We will consider only the pressures under 1st and 5th metatarsal-phalangeal joint.

¶ In the normal condition 1st and 5th plantar pressures are:

- roughly equal.
- display wide fluctuations between:
 - people.
 - activities.

¶ In the pathological condition 1st and 5th plantar pressures are:

- quite different.
- display wide fluctuations between:
 - people.
 - activities.

¶ $u = (u_1, u_2)^T$ = vector of uncertain plantar pressures.

¶ Info-gap models:

$$\mathcal{U}_1(h, \tilde{u}_1) = \left\{ u = \tilde{u}_1 + \phi : \phi^T \phi \leq h^2 \right\}, \quad h \geq 0 \quad (22)$$

$$\mathcal{U}_2(h, \tilde{u}_2) = \left\{ u = \tilde{u}_2 + \phi : \phi^T W \phi \leq h^2 \right\}, \quad h \geq 0 \quad (23)$$

W is a positive definite diagonal matrix: $W = \text{diag}(w_1, w_2)$.

¶ Nominal pressures in the two states, relative to average body weight:

$$\tilde{u}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \tilde{u}_2 = \begin{pmatrix} 1.3 \\ 0.8 \end{pmatrix} \quad (24)$$

¶ Gap function for measurement u , defined in eq.(9) on p. 13:

Least h for which the info-gap model
is consistent with the measurement.

Gap function $\Gamma_1(u)$ for healthy-state:

$$\Gamma_1(u) = \sqrt{(u - \tilde{u}_1)^T (u - \tilde{u}_1)} \quad (25)$$

Gap function $\Gamma_2(u)$ for the pathological state:

$$\Gamma_2(u) = \sqrt{(u - \tilde{u}_2)^T W (u - \tilde{u}_2)} \quad (26)$$

¶ Given learning data, we use these gap functions to:

- Evaluate the empirical robustness of the robust-optimal classifier.
- Up-date the pathological info-gap model.

¶ Learning data:

- K pathological pressure vectors, u_1, \dots, u_K .
- $K = 6$: Open circles in fig. 4.
- Bullets are nominal vectors.

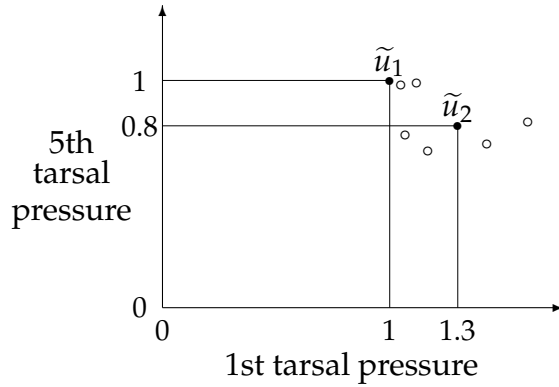


Figure 4: Nominal measurements (•) \tilde{u}_1 and \tilde{u}_2 , and learning data (○) u_1, \dots, u_K .

¶ Empirical error-free empirical robustness eq.(21) on p. 21, of the robust-optimal classifier:

Empirical estimate of the maximum h up to which pathology is not mis-classified:

$$\hat{h}_e(q, 0) = \max \{ \Gamma_2(u_k) : \Gamma_2(u_k) < \Gamma_1(u_k), \quad k = 1, \dots, K \} \quad (27)$$

If this set is empty then $\hat{h}_e(q, 0) = 0$.

¶ Parameters to up-date: diagonal shape matrix, so $q = (w_1, w_2)$.

Goal: Up-date $\mathcal{U}_2(h, \tilde{u}_2)$ by varying **shape** of ellipse.

Area of ellipse: $\pi / \sqrt{w_1 w_2}$.

Thus: vary w_1 and adjust w_2 according to $w_2 = 1/w_1$.

¶ Calculated \hat{h}_e vs. shape parameter w_1 : fig. 5.

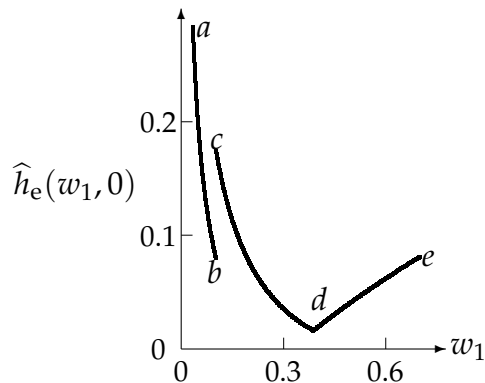


Figure 5: Empirical robustness function $\hat{h}_e(w_1, 0)$ vs. shape parameter w_1 .

¶ Discussion of fig. 5:

- Only Γ_2 (pathological) varies with w_1 .
- From a to b :
 - Only 3 points correctly classified.
 - \hat{h}_e decreasing: \mathcal{U}_2 improving: becoming better adapted to data.
- Discontinuous jump from b to c :
 - Additional datum correctly classified.
 - Gap function better, even though \hat{h}_e is now larger.
- From c to d :
 - Info-gap model improving:
 - Info-gap model adjusting to the 4 captured points.
- Discontinuous slope at d :
 - Different captured point is now critical.
- From d to e :
 - Info-gap model deteriorating.

¶ Optimal shape matrix:

- Local minimum of \hat{h}_e .
- Global minimum of \hat{h}_e **not** optimum.
- Optimum: $w_1 = 0.38$ and $w_2 = 1/w_1 = 2.61$.
- Semi-axis ratio: $\sqrt{w_2/w_1} = 2.62$.
- Horizontal axis is long axis of the ellipse.

¶ Summary. Info-gap model optimized by:

- Locally minimizing \hat{h}_e .
- Identifying **reason** for minimum.
- Different from “bigger is better” for “regular” \hat{h} .

3 Acoustic Noise

¶ Not all decisions are classifications, and not all learning is supervision of a classifier.

¶ Problem: Noise suppression in acoustic cavity.

- Variable and uncertain noise sources.
- Measure sound pressure at M points.
- Impose acoustic signal from actuators.
- Good performance if total residual sound energy is small.

¶ Info-gap model of acoustic source.

$f(x)$ = unknown actual acoustic pressure at position x in cavity.

$\tilde{f}(x)$ = known estimated acoustic pressure at position x in cavity.

$\mathcal{U}(h, \tilde{f})$ = info-gap model for uncertainty in $f(x)$.

$\mathcal{U}(h, \tilde{f})$ underlies choice of imposed acoustic signal.

¶ Learning:

- Given measured acoustic pressures u_1, \dots, u_M at positions x_1, \dots, x_M .
- Do these measurements provide evidence against the info-gap model $\mathcal{U}(h, \tilde{f})$?
- If the info-gap model is indicted, how do we improve it?
- q = vector of parameters or properties of the info-gap model which can be modified to improve $\mathcal{U}(h, \tilde{f})$, if necessary.

3.1 Empirical Robustness

¶ Gap function: compatibility of datum u_m with info-gap model $\mathcal{U}(h, \tilde{f})$:

$$\Gamma(u_m) = \min \left\{ h : f(x_m) = u_m \text{ for some } f(x) \in \mathcal{U}(h, \tilde{f}) \right\} \quad (28)$$

¶ Empirical robustness function.

- “Reasonable” measurement: not extraordinary or unusual.
- $\mathcal{U}(h, \tilde{f})$ compatible with data if gap functions of reasonable measurements are not large.
- Empirical robustness function:

$$\hat{h}_e(q, r_c) = \max \left\{ \Gamma(u_m) : \begin{array}{l} |u_n| \leq r_c \text{ if } \Gamma(u_n) \leq \Gamma(u_m), \\ m = 1, \dots, M, \quad n = 1, \dots, M \end{array} \right\} \quad (29)$$

- Explanation:
 - $|u_n| \leq r_c$ means u_n is not extraordinary measurement.
 - $\Gamma(u_n) \leq \Gamma(u_m)$ means u_n at least as compatible with $\mathcal{U}(h, \tilde{f})$ as u_m .
 - Thus $\hat{h}_e(q, r_c)$ is the greatest horizon of uncertainty at which measurements are not extraordinary.
- $\hat{h}_e(q, r_c)$ assesses fidelity of $\mathcal{U}(h, \tilde{f})$ to the measurements, and is the basis for updating the info-gap model.

¶ $\widehat{h}_e(q, r_c)$ can be large for either of two opposite reasons:

- $\mathcal{U}(h, \widetilde{f})$ is a true and accurate characterization of the acoustic uncertainty, and is highly compatible with all the data, even extreme measurements.
- $\mathcal{U}(h, \widetilde{f})$ is so erroneous that extremely large gap-function values are needed to capture even a few data points.

¶ Up-dating $\mathcal{U}(h, \widetilde{f})$:

- $\widehat{h}_e(q, r_c)$ depends on info-gap model parameters q .
- We must understand *why* $\widehat{h}_e(q, r_c)$ is changing as we modify q .

3.2 Up-Dating the Acoustic Uncertainty Model

¶ 1-dimensional acoustic cavity: $0 \leq x \leq 1$.

¶ Acoustic pressure field, $f(x)$.

- $f(x)$ is complex, and changing with circumstances.
- Roughly constant, or biased predominantly to left or right.
- Envelope-bound info-gap model:

$$\mathcal{U}(h, 0) = \{f(x) : |f(x)| \leq h\psi(x)\}, \quad h \geq 0 \quad (30)$$

- Linear envelope functions:

$$\psi(x) = q \left(x - \frac{1}{2} \right) + 1 \quad (31)$$

- All these functions cross $(0.5, 1)$.
- Slope determined by q which must be in $[-2, 2]$ to assure non-negativity.

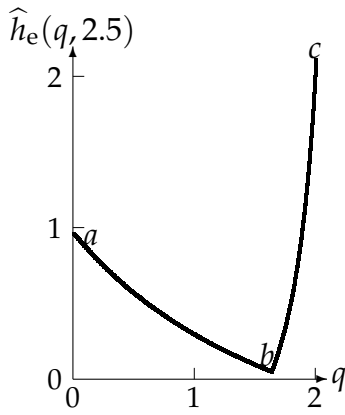


Figure 6: Empirical robustness function $\hat{h}_e(q, 2.5)$ vs. slope parameter q .

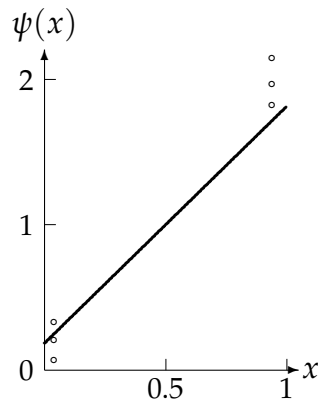


Figure 7: Six pressure measurements (\circ) and the updated envelope function $\psi(x)$ with $q = 1.63$.

¶ Measured pressures at $x = 0.05$ and $x = 0.95$.

$x = 0.05$: 0.07, 0.21, 0.33.

$x = 0.95$: 1.83, 1.97, 2.15.

¶ Empirical robustness function in fig. 6. $r_c = 2.5$. All data captured.

- Left branch, a to b :
 - robustness decreasing as q rises from 0 to 1.63.
 - Info-gap model improving as it becomes more consistent with measurements.
- Right branch, b to c :
 - robustness increasing as q rises above 1.63.
 - Info-gap model degenerating as it becomes less consistent with measurements.
 - Saw-tooth for larger q values as data points are “lost”.
- $q = 1.63$ is robust-satisficing value of envelope slope.
- Essentially same results with any r_c large enough to capture at least one value on both sides of cavity.

¶ Optimal slope, fig. 7.

- Positive slope due to predominantly right-side pressure field.
- Clearly not least-squares best-fit.
- No presumption of linear pressure field.