Towards Analytical Techniques for Optimizing Knowledge Acquisition, Processing, Propagation, and Use in Cyberinfrastructure

L. Octavio Lerma
Computational Science Program
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
lolerma@episd.org

1. Introduction

- Knowledge-related processes are important: we rely on them when we drive, communicate, etc.
- Surprisingly, the very process of acquiring and propagating information is the least automated.
- At present, to decide on the best way to place sensors or propagate data, we mostly use numerical models.
- These models are very resource-consuming, rely on supercomputers, not ready for everyday applications.
- We therefore need analytical models – which would allow easier optimization and application.
- Developing such models is our main objective.

2. Outline of the Dissertation

- We describe analytical models for all stages of knowledge processing.
- We start with knowledge acquisition: optimal sensor placement for stationary and mobile sensors.
- We then deal with data and knowledge processing: how to best organize computing power and research teams.
- We deal with knowledge propagation and resulting knowledge enhancement; we analyze:
  - how early stages of idea propagation occur;
  - how to assess the initial knowledge level;
  - how to present the material and how to provide feedback.
- Finally, we analyze how knowledge is used.

3. Outline of the Presentation

- In this presentation, we will focus mainly on the new results, obtained after the Master's thesis.
- Our three main new results are:
  - an analysis of knowledge propagation, on the example of the Out of Eden Walk;
  - an explanation of why increased climate variability is more visible than global warming; and
  - an analysis of the software migration and modernization process.
- After that, we will briefly overview other results from this dissertation. Most of these other results:
  - either have already been largely presented in the thesis,
  - or are incremental improvements over the thesis' results.

4. Analytical Techniques for Knowledge Propagation, on the Example of the Out of Eden Walk

- To improve teaching and learning, it is important to understand how knowledge propagates.
- Traditional knowledge propagation models are based on diff. equations – similar to epidemics propagation.
- In these models, for large times t, the number of new learners decreases as $r(t) \approx A \cdot \exp(-\alpha \cdot t)$.
- Some empirical data suggests that this decrease follows the power law: $r(t) \approx A \cdot t^{-\alpha}$.
- Power laws are ubiquitous in real life.
- These laws underlie *fractal* techniques pioneered by B. Mandelbrot.
- In this part of the talk, we check which model is better.

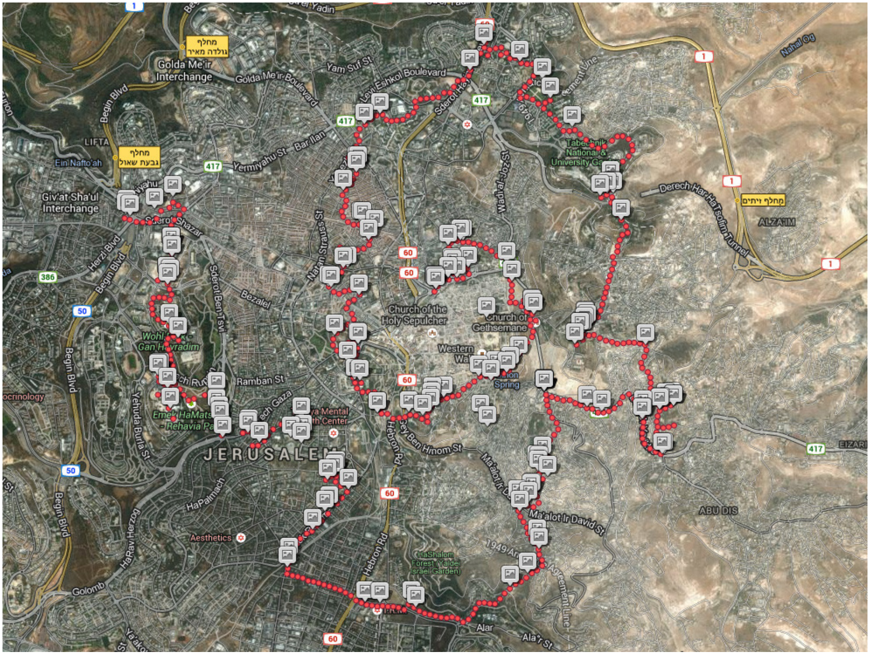
5. Out of Eden Walk Project: A Description

- Commenced on January 10th, 2013 in Ethiopia.
- The Out of Eden Walk is a 7-year, 21,000 mile long, storytelling journey created by Paul Salopek.
- Paul Salopek is a two-time Pulitzer Prize winning journalist.
- This project is sponsored by the National Geographic Society.
- Reports from this journey regularly appear:
  - in the National Geographic magazine;
  - in leading newspapers: NY Times, Washington Post, Chicago Tribune, Los Angeles Times, etc.;
  - on the US National Public Radio (NPR).

6. The Journey Starts in Ethiopia

7. The Journey Starts in Ethiopia (cont-d)

8. Walking Through Jerusalem



9. It Is Not Only About Beauty of the Faraway Lands: Refugees

10. This Project Has Important Educational and Knowledge Propagation Goals

- Main objective: to enhance education and knowledge propagation as main features of journalism.
- Main idea: *slow journalism* – revealing human stories and world events from the ground, at a walking pace.
- The project has largely succeeded in this goal:
  - the website has thousands of followers worldwide,
  - there are also many Facebook and Twitter followers;
  - over 200 schools worldwide regularly use Salopek's reports to teach about world's cultures.

11. Out of Eden Walk Project: Technical Details

- After visiting an area, Paul Salopek publishes a *dispatch* describing his impressions and thoughts.
- As of now, there are more than 100 dispatches.
- Followers are welcome to add comments after each dispatch.
- After two weeks, each dispatch gathers from 15 to more than 250 comments.
- These comments are part of the knowledge propagation process.
- We trace how the number of comments made by the readers changes with time.
- This number reflects how the knowledge contained in a dispatch propagates with time.

12. Power Law Model vs. Traditional Approach

- In the power law model, the number of comments $r(t)$ decreases with t as $r(t) = A \cdot t^{-\alpha}$.
- This model has two parameters: A and $\alpha > 0$.
- Traditional models use differential equations:

$$\frac{dr}{dt} = -f(r).$$

- When $r = 0$, we have $f(r) = 0$.
- The simplest function $f(r)$ with $f(0) = 0$ is linear: $f(r) = \alpha \cdot r$.
- For this $f(r)$, we already get a 2-parametric family of solutions $r(t) = A \cdot \exp(-\alpha \cdot t)$.
- So, we compare power law with this exponential model.

13. How We Compare: Technical Details

- How the number of comments $r(t)$ depends on time t?
  - exponential model: $r(t) \approx r_0(t) = A \cdot \exp(-\alpha \cdot t)$;
  - power law model: $r(t) \approx r_0(t) = A \cdot t^{-\alpha}$.
- To check which model is more adequate, we use the chi-square criterion

$$\chi^2 \stackrel{\text{def}}{=} \sum_t \frac{(r(t) - r_0(t))^2}{r_0(t)}.$$

- To estimate A and α, we use both Least Squares $\sum_i e_i^2 \to \min$ and robust (ℓ^1) estimation $\sum_i |e_i| \to \min$.
- Result: the power law is more adequate:
  - for exponential model H_0, $p \ll 0.05$, so H_0 is rejected;
  - for power law model H_0, $p \gg 0.05$, so H_0 is not rejected.

14. Comparison Results

Dispatch Title	N_c	χ^2_p	$\chi^2_{p,1}$	χ^2_e	p_p	$p_{p,1}$	p_e
Let's Walk	271	30.6	30.0	31,360	0.33	0.37	0.00
Sole Brothers	61	22.1	22.8	83	0.76	0.74	0.00
The Glorious Boneyard	59	16.3	18.6	262	0.96	0.91	0.00
The Self-Love Boat	67	63.1	60.0	124	0.00	0.00	0.00
Go Slowly–Work Slowly	91	33.0	31.5	821	0.24	0.29	0.00
The Camel and the Gyrocopter	52	28.4	24.6	72	0.45	0.65	0.00
Lines in Sand	69	21.4	18.3	89	0.81	0.92	0.00

15. Out of Eden Walk: Conclusions

- To improve teaching and learning, it is important to know how knowledge propagates.
- Traditional models of knowledge propagation are similar to differential-equations-based models in physics.
- Recently, an alternative fractal-motivated power-law model of knowledge propagation was proposed.
- We compare this model with the traditional model on the example of the Out of Eden Walk project.
- It turns out that for the related data, the power law is indeed a more adequate description.
- This shows that the fractal-motivated power law is a more adequate description of knowledge propagation.

16. Analytical Techniques for Knowledge Use, on the Example of Climate Variability

- Global warming is a statistically confirmed long-term phenomenon.
- Somewhat surprisingly, its most visible consequence is:
  - not the warming itself but
  - the increased climate variability.
- In this talk, we explain why increased climate variability is more visible than the global warming itself.
- In this explanation, use general system theory ideas.

17. Formulation of the Problem

- *Global warming* usually means statistically significant long-term increase in the average temperature.
- Researchers have analyzed the expected future consequences of global warming:
  - increase in temperature,
  - melting of glaciers,
  - raising sea level, etc.
- A natural hypothesis was that at present, we would see the same effects, but at a smaller magnitude.
- This turned out not to be the case.
- Some places do have the warmest summers and the warmest winters in record.
- However, other places have the coldest summers and the coldest winters on record.

18. Formulation of the Problem (cont-d)

- What we actually observe is unusually high deviations from the average.
- This phenomenon is called *increased climate variability*.
- A natural question is: why is increased climate variability more visible than global warming?
- A usual answer is that the increased climate variability is what computer models predict.
- However, the existing models of climate change are still very crude.
- None of these models explains why temperature increase has slowed down in the last two decades.
- It is therefore desirable to provide more reliable explanations.

19. A Simplified System-Theory Model

- Let us consider the simplest model, in which the state of the Earth is described by a single parameter x.
- In our case, x can be an average Earth temperature or the temperature at a certain location.
- We want to describe how x changes with time.
- In the first approximation, $\dfrac{dx}{dt} = u(t)$, where $u(t)$ are external forces.
- We know that, on average, these forces lead to a global warming, i.e., to the increase of $x(t)$.
- Thus, the average value u_0 of $u(t)$ is positive.
- We assume that the random deviations $r(t) \stackrel{\text{def}}{=} u(t) - u_0$ are i.i.d., with some standard deviation σ_0.

20. Conclusions

- By solving this equation, we get an analytical model, in which:
  - the systematic part $x_s(t)$ corresponds to global warming, while
  - the random part $x_r(t)$ corresponds to climate variability.
- It turns out that on the initial stages of this process:
  - climate variability effects are indeed much larger
  - than the effects of global warming.
- This is exactly what we currently observe.
- Similar conclusions can be made if we consider more complex multi-parametric models.

21. Analytical Techniques for Knowledge Use, on the Example of Software Migration

- In many aspects of our daily life, we rely on computer systems:
  - computer systems record and maintain the student grades,
  - computer systems handle our salaries,
  - computer systems record and maintain our medical records,
  - computer systems take care of records about the city streets,
  - computer systems regulate where the planes fly, etc.
- Most of these systems have been successfully used for years and decades.
- Every user wants to have a computer system that, once implemented, can effectively run for a long time.

22. Need for Software Migration/Modernization

- Computer systems operate in a certain environment; they are designed:
  - for a certain computer hardware – e.g., with support for words of certain length,
  - for a certain operating system, programming language, interface, etc.
- Eventually, the computer hardware is replaced by a new one.
- While all the efforts are made to make the new hardware compatible with the old code, there are limits.
- As a result, after some time, not all the features of the old system are supported.
- In such situations, it is necessary to adjust the legacy software so that it will work on a new system.

23. Software Migration and Modernization Is Difficult

- At first glance, software migration and modernization sounds like a reasonably simple task:
  - the main intellectual challenge of software design is usually when we have to invent new techniques;
  - in software migration and modernization, these techniques have already been invented.
- Migration would be easy if every single operation from the legacy code was clearly explained and justified.
- The actual software is far from this ideal.
- In search for efficiency, many "tricks" are added by programmers that take into account specific hardware.
- When the hardware changes, these tricks can slow the system down instead of making it run more efficiently.

24. How Migration Is Usually Done

- When a user runs a legacy code on a new system, the compiler produces thousands of error messages.
- Usually, a software developer corrects these errors one by one.
- This is a very slow and very expensive process:
  - correcting each error can take hours, and
  - the resulting salary expenses can run to millions of dollars.
- There exist tools that try to automate this process by speeding up the correction of each individual error.
- These tools speed up the required time by a factor of even ten.
- However, still thousands of errors have to be handled individually.

25. Resulting Problem: Need to Speed up Migration and Modernization

- Migration and modernization of legacy software is a ubiquitous problem.
- It is thus desirable to come up with ways to speed up this process.
- In this dissertation:
  - we propose such an idea, and
  - we show how expert knowledge can help in implementing this idea.

26. Our Main Idea

- Modern compilers do not simply indicate an error.
- They usually provide a reasonably understandable description of the type of an error; for example:
  - it may be that a program is dividing by zero,
  - it may be that an array index is out of bound.
- Some of these types of error appear in numerous places in the software.
- Our experience shows that in many such places, these errors are caused by the same problem in the code.
- So, instead of trying to "rack our brains" over each individual error, a better idea is
  - to look at all the errors of the given type, and
  - come up with a solution that would automatically eliminate the vast majority of these errors.

27. Need for Analytical Models

- This idea saves time only if we have enough errors of a given type.
- We thus need to predict how many errors of different type we will encounter.
- There are currently no well-justified software models that can predict these numbers.
- What we do have is many system developers who have an experience in migrating and modernizing software.
- It is therefore desirable to utilize their experience.
- So, we need to build an analytical model based on expert knowledge.

28. Expert Knowledge about Software Migration

- A reasonable idea is to start with n_1 errors of the most frequent type.
- Then, we should concentrate on n_2 errors of the second most frequent type, etc.
- So, we want to know the numbers $n_1, n_2, \ldots,$ for which

$$n_1 \geq n_2 \geq \ldots \geq n_{k-1} \geq n_k \geq n_{k+1} \geq \ldots$$

- We know that for every k, n_{k+1} is somewhat smaller than n_k.
- Similarly, n_{k+2} is more noticeably smaller than n_k, etc.
- After formalizing the $n_k < n_{k+1}$ rule, we get $n_{k+1} = f(n_k)$.
- Which function $f(n)$ should we choose?

29. Which Function $f(n)$ Should We Choose?

- A migrated software package usually consists of two (or more) parts.
- We can estimate n_{k+1} in two different ways:
  - We can use $n_k = n_k^{(1)} + n_k^{(2)}$ to predict
  $$n_{k+1} \approx f(n_k) = f(n_k^{(1)} + n_k^{(2)}).$$
  - Or, we can use $n_k^{(1)}$ to predict $n_{k+1}^{(1)}$, $n_k^{(2)}$ to predict $n_{k+1}^{(2)}$, and add them: $n_{k+1} \approx f(n_k^{(1)}) + f(n_k^{(2)})$.
- It is reasonable to require that these estimates coincide:
$$f(n_k^{(1)} + n_k^{(2)}) = f(n_k^{(1)}) + f(n_k^{(2)}).$$
- So, $f(a+b) = f(a) + f(b)$ for all a and b, thus $f(a) = f(1) + \ldots + f(1)$ (a times), and $f(a) = f(1) \cdot a$.
- Thus, $n_{k+1} = c \cdot n_k$, i.e., $n_{k+1}/n_k = \text{const}$.

30. Empirical Data: Values n_k for Migrating a Health-Related C Package from 32 to 64 Bits

Here, n_{ab} is stored in the a-th column (marked ax) and b-th row (marked xb).

	0x	1x	2x	3x	4x	5x	6x	7x
x0	–	308	95	47	13	5	2	1
x1	7682	301	91	38	13	4	2	1
x2	4757	266	85	34	12	4	2	1
x3	3574	261	81	34	12	4	2	1
x4	2473	241	76	30	11	3	2	1
x5	2157	240	69	24	9	3	2	1
x6	956	236	58	21	8	3	2	1
x7	769	171	57	19	8	3	1	1
x8	565	156	50	17	8	2	1	1
x9	436	98	47	17	6	2	1	–

31. Empirical Data: Values n_k for Migrating a Health-Related C Package from 32 to 64 Bits

Here, n_{ab} is stored in the a-th column (marked ax) and b-th row (marked xb); e.g., $n_{23} = 81$.

	0x	1x	**2x**	3x	4x	5x	6x	7x
x0	–	308	95	47	13	5	2	1
x1	7682	301	91	38	13	4	2	1
x2	4757	266	85	34	12	4	2	1
x3	3574	261	**81**	34	12	4	2	1
x4	2473	241	76	30	11	3	2	1
x5	2157	240	69	24	9	3	2	1
x6	956	236	58	21	8	3	2	1
x7	769	171	57	19	8	3	1	1
x8	565	156	50	17	8	2	1	1
x9	436	98	47	17	6	2	1	–

32. How Accurate is This Estimate?

- One can easily see that for $k \le 9$, we indeed have $n_{k+1} \approx c \cdot n_k$, with $c \approx 0.65$-0.75.
- Thus, the above simple rule described the most frequent errors reasonably accurately.
- However, starting with $k = 10$, the ratio n_{k+1}/n_k becomes much closer to 1.
- Thus, the one-rule estimate is no longer a good estimate.
- A natural idea is this to use two rules:
  - in addition to the rule that n_{k+1} is somewhat smaller than n_k,
  - let us also use the rule that n_{k+2} is more noticeably smaller than n_k.

33. Two-Rules Approach: Results

- Similar arguments lead to the following analytical model

$$n_k = A_1 \cdot \exp(-b_1 \cdot k) + A_2 \cdot \exp(-b_2 \cdot k).$$

- This double-exponential model indeed describes the above data reasonably accurately:
  - for $k \le 9$, the data is a good fit with an an exponential model for which $\rho = n_{k+1}/n_k \approx 0.65\text{-}0.75$;
  - for $k \ge 10$, the data is a good fit with another exponential model, for which $\rho^{10} \approx 2\text{-}3$.

34. Practical Consequences

- For small k, the dependence n_k rapidly decreases with k.
- So, the values n_k corresponding to small k constitute the vast majority of all the errors.
- In the above example, 85 percent of errors are of the first 10 types; thus:
  - once we learn to repair errors of these types,
  - the remaining number of un-corrected errors decreases by a factor of seven.
- This observation has indeed led to a significant speed-up of software migration and modernization.

35. Software Migration: Conclusion

- In many practical situations, we need to migrate legacy software to a new hardware and system environment.
- If we run the software package in the new environment, we get thousands of difficult-to-correct errors.
- As a result, software migration is very time-consuming.
- A reasonable way to speed up this process is to take into account that:
  - errors can be naturally classified into categories,
  - often all the errors of the same category can be corrected by a single correction.
- Coming up with such a joint correction is also somewhat time-consuming.
- The corresponding additional time pays off only if we have sufficiently many errors of this category.

36. Software Migration: Conclusion (cont-d)

- Coming up with a joint correction is time-consuming.
- This additional time pays off only if we have sufficiently many errors of this category.
- So, it is desirable to be able to estimate the number of errors n_k of different categories k.
- We show that expert knowledge leads to a double-exponential model in good accordance w/observations.

37. Other Results from the Dissertation: A Brief Overview

- *Problem:* optimal sensor placement under uncertainty.
- *First situation:* placing bio-weapon detectors.
- *Objective:* minimize the expected risk.
- *Second situation:* patrolling the border with UAVs.
- *Objective:* minimize the probability of undetected smuggling.
- *Third situation:* placing meteorological sensors.
- *Objective:* maximize accuracy with which we know all relevant quantities.
- *Solution:* we describe analytical expressions for optimal sensor placement for all three situations.

38. Other Results from the Dissertation (cont-d)

- *Problem:* optimal placement tests.
- *Situation:* inability to solve problems causes discomfort which hinders ability to solve further problems.
- *Objective:* minimize this discomfort.
- *Solution:* we come up with an analytical expression for the optimal placement testing.
- *Problem:* optimal selection of class-related feedback for students.
- *Empirical fact:* immediate feedback makes learning, on average, twice faster.
- *Solution:* an analytical expression derived from first principles explains the empirical fact.

39. Acknowledgments

- My deep gratitude to my committee: Drs. V. Kreinovich, D. Pennington, C. Tweedie, S. Starks, & O. Kosheleva.
- I also want to thank all the faculty, staff, and students of the Computational Science program.
- My special thanks to Drs. A. Gates and M.-Y. Leung, to J. Maxey and L. Valera, and to C. Aguilar-Davis.
- Last but not the least, my thanks and my love to my family:
  - to my sons Sam and Joseph,
  - to my parents Hortensia and Antonio,
  - to my sister Martha, and
  - to my brothers Antonio, Victor, and Javier.

Appendix 1: Global Warming vs. Climate Variability

40. Towards the Second Approximation

- Most natural systems are resistant to change: otherwise, they would not have survived.
- So, when $y \stackrel{\text{def}}{=} x - x_0 \neq 0$, a force brings y back to 0: $\dfrac{dy}{dt} = f(y)$; $f(y) < 0$ for $y > 0$, $f(y) > 0$ for $y < 0$.
- Since the system is stable, y is small, so we keep only linear terms in the Taylor expansion of $f(y)$:

$$f(y) = -k \cdot y, \text{ so } \frac{dy}{dt} = -k \cdot y + u_0 + r(t).$$

- Since this equation is linear, its solution can be represented as $y(t) = y_s(t) + y_r(t)$, where

$$\frac{dy_s}{dt} = -k \cdot y_s + u_0; \quad \frac{dy_r}{dt} = -k \cdot y_r + r(t).$$

- Here, $y_s(t)$ is the *systematic* change (global warming).
- $y_r(t)$ is the *random* change (climate variability).

41. An Empirical Fact That Needs to Be Explained

- At present, the climate variability becomes more visible than the global warming itself.
- In other words, the ratio $y_r(t)/y_s(t)$ is much higher than it will be in the future.
- The change in y is determined by two factors:
  - the external force $u(t)$ and
  - the parameter k that describes how resistant is our system to this force.
- Some part of global warming may be caused by the variations in Solar radiation.
- Climatologists agree that global warming is mostly caused by greenhouse effect etc., which lowers resistance k.
- What causes numerous debates is which proportion of the global warming is caused by human activities.

42. An Empirical Fact to Be Explained (cont-d)

- Since decrease in k is the main effect, in the 1st approximation, we consider only this effect.
- In this case, we need to explain why the ratio $y_r(t)/y_s(t)$ is higher now when k is higher.
- To gauge how far the random variable $y_r(t)$ deviates from 0, we can use its standard deviation $\sigma(t)$.
- So, we fix values u_0 and σ_0, st. dev. of $r(t)$.
- For each k, we form the solutions $y_s(t)$ and $y_r(t)$ corresponding to $y_s(0) = 0$ and $y_r(0) = 0$.
- We then estimate the standard deviation $\sigma(t)$ of $y_r(t)$.
- We want to prove that, when k decreases, the ratio $\sigma(t)/y_s(t)$ also decreases.

43. Estimating the Systematic Deviation $y_s(t)$

- We need to solve the equation $\dfrac{dy_s}{dt} = -k \cdot y_s + u_0$.
- If we move all the terms containing $y_s(t)$ to the left-hand side, we get $\dfrac{dy_s(t)}{dt} + k \cdot y_s(t) = u_0$.
- For an auxiliary variable $z(t) \stackrel{\text{def}}{=} y_s(t) \cdot \exp(k \cdot t)$, we get

$$\frac{dz(t)}{dt} = \frac{dy_s(t)}{dt} \cdot \exp(k \cdot t) + y_s(t) \cdot \exp(k \cdot t) \cdot k =$$

$$\exp(k \cdot t) \cdot \left(\frac{dy_s(t)}{dt} + k \cdot y_s(t) \right).$$

- Thus, $\dfrac{dz(t)}{dt} = u_0 \cdot \exp(k \cdot t)$, so $z(t) = u_0 \cdot \dfrac{\exp(k \cdot t) - 1}{k}$, and

$$y_s(t) = \exp(-k \cdot t) \cdot z(t) = u_0 \cdot \frac{1 - \exp(-k \cdot t)}{k}.$$

44. Estimating the Random Component $y_r(t)$

- For the random component, we similarly get

$$y_r(t) = \exp(-k \cdot t) \cdot \int_0^t r(s) \cdot \exp(k \cdot s)\, ds, \text{ so}$$

$$y_r(t)^2 = \exp(-2k\cdot t)\cdot \int_0^t ds \int_0^t dv\, r(s)\cdot r(v)\cdot \exp(k\cdot s)\cdot \exp(k\cdot v),$$

$$\text{and } \sigma^2(t) = E[y_r(t)^2] =$$

$$\exp(-2k \cdot t) \cdot \int_0^t ds \int_0^t dv\, E[r(s) \cdot r(v)] \cdot \exp(k \cdot s) \cdot \exp(k \cdot v).$$

- Here, $E[r(s)\cdot r(v)] = E[r(s)]\cdot E[r(v)] = 0$ and $E[r^2(s)] = \sigma_0^2$, so

$$\sigma^2(t) = E[y_r(t)^2] = \exp(-2k\cdot t)\cdot \int_0^t ds\, \sigma_0^2 \cdot \exp(k\cdot s)\cdot \exp(k\cdot s).$$

- Thus, $\sigma^2(t) = \sigma_0^2 \cdot \dfrac{1 - \exp(-2k \cdot t)}{2k}$.

45. Analyzing the Ratio $\sigma(t)/y_s(t)$

- $\sigma^2(t) = \sigma_0^2 \cdot \dfrac{1 - \exp(-2k \cdot t)}{2k}$, $y_s(t) = u_0 \cdot \dfrac{1 - \exp(-k \cdot t)}{k}$.
- Thus, $S(t) \stackrel{\text{def}}{=} \dfrac{\sigma^2(t)}{y_s^2(t)} = \dfrac{\sigma_0^2}{u_0^2} \cdot \dfrac{(1 - \exp(-2k \cdot t)) \cdot k^2}{2k \cdot (1 - \exp(-k \cdot t))^2}$.
- Here, $1 - \exp(-2k \cdot t) = (1 - \exp(-k \cdot t)) \cdot (1 + \exp(-k \cdot t))$, so $S(t) = \dfrac{\sigma_0^2}{u_0^2} \cdot \dfrac{(1 + \exp(-k \cdot t)) \cdot k}{2 \cdot (1 - \exp(-k \cdot t))}$.
- When the k is large, $\exp(-k \cdot t) \approx 0$, and $S(t) \approx \dfrac{\sigma_0^2}{u_0^2} \cdot \dfrac{k}{2}$.
- This ratio clearly decreases when k decreases.
- So, when the Earth's resistance k will decrease, the ratio $\sigma(t)/y_s(t)$ will also decrease.
- Thus, we will start observing mainly the direct effects of global warming – unless we do something about it.

46. Discussion

- We made a simplifying assumption that the climate system is determined by a single parameter x (or y).
- A more realistic model is when the climate system is determined by several parameters $y_1, \ldots, y_n$.
- In this case, in the linear approximation, the dynamics is described by a system of linear ODEs

$$\frac{dy_i}{dt} = -\sum_{j=1}^{n} a_{ij} \cdot y_j(t) + u_i(t).$$

- In the generic case, all eigenvalues λ_k of the matrix a_{ij} are different.
- In this case, a_{ij} can be diagonalized by using the linear combinations $z_k(t)$ corresponding to eigenvectors:

$$\frac{dz_k}{dt} = -\lambda_k \cdot z_k(t) + \widetilde{u}_k(t).$$

47. Discussion (cont-d)

- Reminder: we have a system of equations
$$\frac{dz_k}{dt} = -\lambda_k \cdot z_k(t) + \widetilde{u}_k(t).$$
- For each of these equations, we can arrive at the same conclusion:
  - the current ratio of the random to systematic effects is much higher
  - than it will be in the future.
- So, our explanations holds in this more realistic model as well.

Appendix 2. Two-Rules Approach to Software Migration

48. Two-Rules Approach

- Once we know n_k and n_{k+1}, we need to estimate $n_{k+2} = f(n_k, n_{k+1})$.
- When the software package consists of two parts, we can estimate n_{k+2} in two different ways:
  - We can use the overall numbers $n_k = n_k^{(1)} + n_k^{(2)}$ and $n_{k+1} = n_{k+1}^{(1)} + n_{k+1}^{(2)}$ and predict

$$n_{k+2} \approx f(n_k, n_{k+1}) = f(n_k^{(1)} + n_k^{(2)}, n_{k+1}^{(1)} + n_{k+1}^{(2)}).$$

  - Alternatively, we can predict the values $n_{k+2}^{(1)}$ and $n_{k+2}^{(2)}$, and add up these predictions:

$$n_{k+2} \approx f(n_k^{(1)}, n_{k+1}^{(1)}) + f(n_k^{(2)}, n_{k+1}^{(2)}).$$

- It is reasonable to require that these two approaches lead to the same estimate, i.e., that we have

$$f(n_k^{(1)} + n_k^{(2)}, n_{k+1}^{(1)} + n_{k+1}^{(2)}) = f(n_k^{(1)}, n_{k+1}^{(1)}) + f(n_k^{(2)}, n_{k+1}^{(2)}).$$

49. Two-Rules Approach (cont-d)

- Reminder: for all $a \geq a'$ and $b \geq b'$, we have
$$f(a+b, a'+b') = f(a, a') + f(b, b').$$
- One can show that this leads to $n_{k+2} = c \cdot n_k + c' \cdot n_{k+1}$ for some c and c', and thus, to
$$n_k = A_1 \cdot \exp(-b_1 \cdot k) + A_2 \cdot \exp(b_2 \cdot k).$$
- In general, b_i are complex numbers – leading to oscillating sinusoidal terms.
- We want $n_k \geq n_{k+1}$, so there are no oscillations, both b_i are real.
- Without losing generality, we can assume that $b_1 < b_2$.
- If $A_1 > A_2$, then the first term always dominates.
- But we already know that an exponential function is not a good description of n_k.

50. Two-Rules Model Fits the Data

- Thus, to fit the empirical data, we must use models with $A_1 < A_2$. In this case:
  - for small k, the second – faster-decreasing – term dominates: $n_k \approx A_2 \cdot \exp(-b_2 \cdot k)$;
  - for larger k, the first – slower-decreasing – term dominates: $n_k \approx A_1 \cdot \exp(-b_1 \cdot k)$.
- This double-exponential model indeed describes the above data reasonably accurately:
  - for $k \le 9$, the data is a good fit with an an exponential model for which $\rho = n_{k+1}/n_k \approx 0.65$-$0.75$;
  - for $k \ge 10$, the data is a good fit with another exponential model, for which $\rho^{10} \approx 2$-3.

Appendix 3. First Case of Sensor Placement: Placing Bio-Weapon Detectors

51. Sensor Placement: Case Study

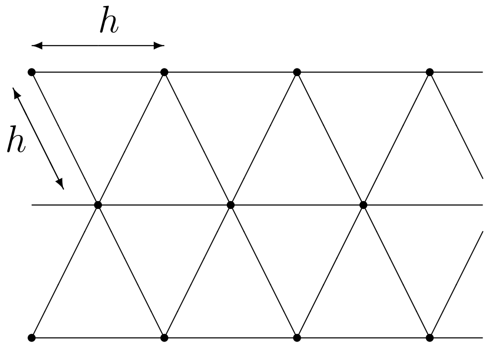
- Biological weapons are difficult and expensive to detect.
- Within a limited budget, we can afford a limited number of bio-weapon detector stations.
- It is therefore important to find the optimal locations for such stations.
- A natural idea is to place more detectors in the areas with more population.
- However, such a commonsense analysis does not tell us how many detectors to place where.
- To decide on the exact detector placement, we must formulate the problem in precise terms.

52. Towards Precise Formulation of the Problem

- The adversary's objective is to kill as many people as possible.
- Let $\rho(x)$ be a population density in the vicinity of the location x.
- Let N be the number of detectors that we can afford to place in the given territory.
- Let d_0 be the distance at which a station can detect an outbreak of a disease.
- Often, $d_0 = 0$ – we can only detect a disease when the sources of this disease reach the detecting station.
- We want to find $\rho_d(x)$ – the density of detector placement.
- We know that $\int \rho_d(x)\,dx = N$.

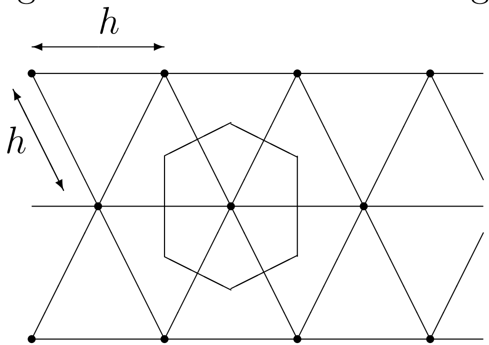
53. Optimal Placement of Sensors

- We want to place the sensors in an area in such a way that
  - the largest distance D to a sensor
  - is as small as possible.
- It is known that the smallest such number is provided by an equilateral triangle grid:

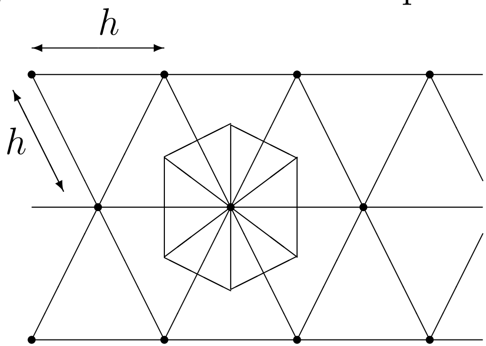


Introduction
Case 1: Out of Eden Walk
Case 2: Climate . . .
Case 3: Software . . .
Placing Bio-Weapon . . .
Meteorological Sensors
UAVs Patrolling the . . .
Optimal Placement Tests
Feedback for Students
Home Page
Title Page
Page 57 of 106
Go Back
Full Screen
Close
Quit

For the equilateral triangle placement, points which are closest to a given detector forms a hexagonal area:



This hexagonal area consists of 6 equilateral triangles:



54. Optimal Placement of Sensors (cont-d)

- In each $\triangle$, the height $h/2$ is related to the side s by the formula $\dfrac{h}{2} = s \cdot \cos(60^\circ) = s \cdot \dfrac{\sqrt{3}}{2}$, hence $s = h \cdot \dfrac{\sqrt{3}}{3}$.
- Thus, the area A_t of each triangle is equal to

$$A_t = \frac{1}{2} \cdot s \cdot \frac{h}{2} = \frac{1}{2} \cdot \frac{\sqrt{3}}{3} \cdot \frac{1}{2} \cdot h^2 = \frac{\sqrt{3}}{12} \cdot h^2.$$

- So, the area A_s of the whole set is equal to 6 times the triangle area: $A_s = 6 \cdot A_t = \dfrac{\sqrt{3}}{2} \cdot h^2$.
- In a region of area A, there are $A \cdot \rho_d(x)$ sensors, they cover area $(A \cdot \rho_d(x)) \cdot A_s$.
- The condition $A = (A \cdot \rho_d(x)) \cdot A_s = (A \cdot \rho_d(x)) \cdot \dfrac{\sqrt{3}}{2} \cdot h^2$ implies that $h = \dfrac{c_0}{\sqrt{\rho_d(x)}}$, with $c_0 \stackrel{\text{def}}{=} \sqrt{\dfrac{2}{\sqrt{3}}}$.

55. Estimating the Effect of Sensor Placement

- The adversary places the bio-weapon at a location which is the farthest away from the detectors.
- This way, it will take the longest time to be detected.
- For the grid placement, this location is at one of the vertices of the hexagonal zone.
- At these vertices, the distance from each neighboring detector is equal to $s = h \cdot \dfrac{\sqrt{3}}{3}$.
- By know that $h = \dfrac{c_0}{\sqrt{\rho_d(x)}}$, so $s = \dfrac{c_1}{\sqrt{\rho_d(x)}}$, with

$$c_1 = \frac{\sqrt{3}}{3} \cdot c_0 = \frac{\sqrt[4]{3} \cdot \sqrt{2}}{3}.$$

- Once the bio-weapon is placed, it starts spreading until it reaches the distance d_0 from the detector.

56. Effect of Sensor Placement (cont-d)

- The bio-weapon is placed at a distance $s = \dfrac{c_1}{\sqrt{\rho_d(x)}}$ from the nearest sensor.
- Once the bio-weapon is placed, it starts spreading until it reaches the distance d_0 from the detector.
- In other words, it spreads for the distance $s - d_0$.
- During this spread, the disease covers the circle of radius $s - d_0$ and area $\pi \cdot (s - d_0)^2$.
- The number of affected people $n(x)$ is equal to:

$$n(x) = \pi \cdot (s - d_0)^2 \cdot \rho(x) = \pi \cdot \left(\frac{c_1}{\sqrt{\rho_d(x)}} - d_0 \right)^2 \cdot \rho(x).$$

57. Precise Formulation of the Problem

- For each location x, the number of affected people $n(x)$ is equal to:

$$n(x) = \pi \cdot \left(\frac{c_1}{\sqrt{\rho_d(x)}} - d_0 \right)^2 \cdot \rho(x).$$

- The adversary will select a location x for which this number $n(x)$ is the largest possible:

$$n = \max_x \left(\pi \cdot \left(\frac{c_1}{\sqrt{\rho_d(x)}} - d_0 \right)^2 \cdot \rho(x) \right).$$

- Resulting problem:
  - given population density $\rho(x)$, detection distance d_0, and number of sensors N,
  - find a function $\rho_d(x)$ that minimizes the above expression n under the constraint $\int \rho_d(x)\,dx = N$.

58. Main Lemma

- *Reminder:* we want to minimize the worst-case damage $n = \max\limits_x n(x)$.
- *Lemma:* for the optimal sensor selection, $n(x) = \text{const}$.
- *Proof by contradiction:* let $n(x) < n$ for some x; then:
  - we can slightly increase the detector density at the locations where $n(x) = n$,
  - at the expense of slightly decreasing the location density at locations where $n(x) < n$;
  - as a result, the overall maximum $n = \max\limits_x n(x)$ will decrease;
  - but we assumed that n is the smallest possible.
- *Thus:* $n(x) = \text{const}$; let us denote this constant by n_0.

59. Towards the Solution of the Problem

- We have proved that $n(x) = \text{const} = n_0$, i.e., that

$$n_0 = \pi \cdot \left(\frac{c_1}{\sqrt{\rho_d(x)}} - d_0\right)^2 \cdot \rho(x).$$

- Straightforward algebraic transformations lead to:

$$\rho_d(x) = \frac{2 \cdot \sqrt{3}}{9} \cdot \frac{1}{\left(d_0 + \frac{c_2}{\sqrt{\rho(x)}}\right)^2}.$$

- The value c_2 must be determined from the equation

$$\int \rho_d(x)\, dx = N.$$

- Thus, we arrive at the following solution.

60. Solution

- *General case:* the optimal detector location is characterized by the detector density
$$\rho_d(x)=\frac{2\cdot\sqrt{3}}{9}\cdot\frac{1}{\left(d_0+\dfrac{c_2}{\sqrt{\rho(x)}}\right)^2}.$$
- Here the parameter c_2 must be determined from the equation $\displaystyle\int\frac{2\cdot\sqrt{3}}{9}\cdot\frac{1}{\left(d_0+\dfrac{c_2}{\sqrt{\rho(x)}}\right)^2}\,dx=N.$
- *Case of $d_0=0$:* in this case, the formula for $\rho_d(x)$ takes a simplified form $\rho_d(x)=C\cdot\rho(x)$ for some constant C.
- In this case, from the constraint, we get:
$$\rho_d(x)=\frac{N}{N_p}\cdot\rho(x),\ \text{ where } N_p \text{ is the total population.}$$

61. Towards More Relevant Objective Functions

- We assumed that the adversary wants to maximize the number $\int \rho(x)\,dx$ of people affected by the bio-weapon.
- The actual adversary's objective function may differ from this simplified objective function.
- For example, the adversary may take into account that different locations have different publicity potential.
- In this case, the adversary maximizes the weighted value $\int_A \widetilde{\rho}(x)\,dx$, where $\widetilde{\rho}(x) \stackrel{\text{def}}{=} w(x) \cdot \rho(x)$.
- Here, $w(x)$ is the importance of the location x.
- From the math. viewpoint, the problem is the same – w/ "effective population density" $\widetilde{\rho}(x)$ instead of $\rho(x)$.
- Thus, if we know $w(x)$, we can find the optimal detector density $\rho_d(x)$ from the above formulas.

Appendix 4. Second Case of Optimal Sensor Placement: Placing Meteorological Sensors

62. How Temperatures etc. Change from One Spatial Location to Another: A Model

- Each environmental characteristic q changes from one spatial location to another.
- A large part of this change is unpredictable (i.e., random).
- A reasonable value to describe the random component of the difference $q(x) - q(x')$ is the variance

$$V(x, x') \stackrel{\text{def}}{=} E[((q(x) - E[q(x)]) - (q(x') - E[q(x')]))^2].$$

- *Comment:* we assume that averages are equal.
- Locally, processes should not change much with shift $x \to x + s$: $V(x + s, x' + s) = V(x, x')$.
- For $s = -x'$, we get $V(x, x') = C(x - x')$ for

$$C(x) \stackrel{\text{def}}{=} V(x, 0).$$

63. A Model (cont-d)

- In general, the further away the points x and x', the larger the difference $C(x - x')$.
- In the isotropic case, $C(x - x')$ depends only on the distance $D = |x - x'|^2 = (x_1 - x'_1)^2 + (x_2 - x'_2)^2$.
- It is reasonable to consider a scale-invariant dependence $C(x) = A \cdot D^\alpha$.
- In practice, we may have more changes in one direction and less change in another direction.
- E.g., 1 km in x is approximately the same change as 2 km in y.
- The change can also be mostly in some other direction, not just x- and y-directions.
- Thus, in general, in appropriate coordinates (u, v), we have $C = A \cdot D^\alpha$ for $D = (u - u')^2 + (v - v')^2$.

64. Model: Final Formulas

- In general, $C = A \cdot D^\alpha$, for $D = (u - u')^2 + (v - v')^2$ in appropriate coordinates (u, v).
- In the original coordinates x_1 and x_2, we get:

$$C(x - x') = A \cdot D^\alpha, \text{ where}$$

$$D = \sum_{i=1}^{2} \sum_{j=1}^{2} g_{ij} \cdot (x_i - x'_i) \cdot (x_j - x'_j) =$$

$$g_{11} \cdot (x_1 - x'_1)^2 + 2g_{12} \cdot (x_1 - x'_1) \cdot (x_2 - x'_2) + g_{22} \cdot (x_2 - x'_2)^2.$$

- From the computational viewpoint, we can include A into g_{ij} if we replace g_{ij} with $A^{1/\alpha} \cdot g_{ij}$, then

$$C(x - x') =$$

$$\left(g_{11} \cdot (x_1 - x'_1)^2 + 2g_{12} \cdot (x_1 - x'_1) \cdot (x_2 - x'_2) + g_{22} \cdot (x_2 - x'_2)^2\right)^\alpha$$

- We can use these formulas to find the optimal sensor locations.

Appendix 5. Third Case of Optimal Sensor Placement: Patrolling the Border with UAVs

65. Optimal Use of Mobile Sensors: Case Study

- Remote areas of international borders are used by the adversaries: to smuggle drugs, to bring in weapons.
- It is therefore desirable to patrol the border, to minimize such actions.
- It is not possible to effectively man every single segment of the border.
- It is therefore necessary to rely on other types of surveillance.
- Unmanned Aerial Vehicles (UAVs):
  - from every location along the border, they provide an overview of a large area, and
  - they can move fast, w/o being slowed down by clogged roads or rough terrain.
- Question: what is the optimal trajectory for these UAVs?

66. How to Describe Possible UAV Patrolling Strategies

- Let us assume that the time between two consequent overflies is smaller the time needed to cross the border.
- Ideally, such a UAV can detect all adversaries.
- In reality, a fast flying UAV can miss the adversary.
- We need to minimize the effect of this miss.
- The faster the UAV goes, the less time it looks, the more probable that it will miss the adversary.
- Thus, the velocity $v(x)$ is very important.
- By a patrolling strategy, we will mean a f-n $v(x)$ describing how fast the UAV flies at different locations x.

67. Constraints on Possible Patrolling Strategies

1) The time between two consequent overflies should be smaller the time T needed to cross the border:
   - the time during which a UAV passes from the location x to the location $x+\Delta x$ is equal to $\Delta t = \dfrac{\Delta x}{v(x)}$;
   - thus, the overall flight time is equal to the sum of these times:
   $$T = \int \frac{dx}{v(x)}.$$
2) UAV has the largest possible velocity V, so we must have $v(x) \le V$ for all x.

It is convenient to use the value $s(x) \stackrel{\text{def}}{=} \dfrac{1}{v(x)}$ called *slowness*, so

$$T = \int s(x)\,dx; \quad s(x) \ge S \left(\stackrel{\text{def}}{=} \frac{1}{V} \right).$$

68. Simplification of the Constraints

- Since $s(x) \geq S$, the value $s(x)$ can be represented as $S + \Delta s(x)$, where $\Delta s(x) \stackrel{\text{def}}{=} s(x) - S$.
- The new unknown function satisfies the simpler constraint $\Delta s(x) \geq 0$.
- In terms of $\Delta s(x)$, the requirement that the overall time be equal to T has a form $T = S \cdot L + \int \Delta s(x)\,dx$.
- This is equivalent to:

$$T_0 = \int \Delta s(x)\,dx, \text{ where:}$$

  - L is the total length of the piece of the border that we are defending, and
  - $T_0 \stackrel{\text{def}}{=} T - S \cdot L$.

69. Detection at Crossing Point x

- Let h be the width of the border zone from which an adversary (A) is visible.
- Then, the UAV can potentially detect A during the time $h/v(x) = h \cdot s(x)$.
- So, the UAV takes $(h \cdot s(x))/\Delta t$ photos, where Δt is the time per photo.
- Let p_1 be the probability that one photo misses A.
- It is reasonable to assume that different detection errors are independent.
- Then, the probability $p(x)$ that A is not detected is $p_1^{(h \cdot s(x))/\Delta t}$, i.e., $p(x) = \exp(-k \cdot s(x))$, where:

$$k \stackrel{\text{def}}{=} \frac{2h}{\Delta t} \cdot |\ln(p_1)|.$$

70. Strategy Selected by the Adversary

- Let $w(x)$ denote the utility of the adversary succeeding in crossing the border at location x.
- Let us first assume that we know $w(x)$ for every x.
- According to decision theory, the adversary will select a location x with the largest expected utility

$$u(x) = p(x) \cdot w(x) = \exp(-k \cdot s(x)) \cdot w(x).$$

- Thus, for each slowness function $s(x)$, the adversary's gain $G(s)$ is equal to

$$G(s) = \max_x u(x) = \max_x \left[\exp(-k \cdot s(x)) \cdot w(x)\right].$$

- We need to select a strategy $s(x)$ for which the gain $G(s)$ is the smallest possible.

$$G(s) = \max_x u(x) = \max_x \left[\exp(-k \cdot s(x)) \cdot w(x)\right] \to \min_{s(x)}.$$

71. Towards an Optimal Strategy for Patrolling the Border

- Let x_m be the location at which the utility $u(x) = \exp(-k \cdot s(x)) \cdot w(x)$ attains its largest possible value.
- If we have a point x_0 s.t. $u(x_0) < u(x_m)$ and $s(x_0) > S$:
  - we can slightly decrease the slowness $s(x_0)$ at the vicinity of x_0 (i.e., go faster in this vicinity) and
  - use the resulting time to slow down (i.e., to go slower) at all locations x at which $u(x) = u(x_m)$.
- As a result, we slightly decrease the value

$$u(x_m) = \exp(-k \cdot s(x_m)) \cdot w(x_m).$$

- At x_0, we still have $u(x_0) < u(x_m)$.
- So, the overall gain $G(s)$ decreases.
- Thus, when the adversary's gain is minimized, we get

$$u(x) = u_0 = \text{const whenever } s(x) > S.$$

72. Towards an Optimal Strategy (cont-d)

- *Reminder:* for the optimal strategy,

  $u(x) = w(x) \cdot \exp(-k \cdot s(x)) = u_0$ whenever $s(x) > S$.

- So, $\exp(-k \cdot s(x)) = \dfrac{u_0}{w(x)}$, hence

  $s(x) = \dfrac{1}{k} \cdot (\ln(w(x)) - \ln(u_0))$ and $\Delta s(x) = \dfrac{1}{k} \cdot \ln(w(x)) - \Delta_0$.

- Here, $\Delta_0 \stackrel{\text{def}}{=} \dfrac{1}{k} \cdot \ln(u_0) - S$.

- When $s(x)$ gets to $s(x) = S$ and $\Delta s(x) = 0$, we get $\Delta s(x) = 0$.

- Thus, we conclude that

$$\Delta s(x) = \max\left(\frac{1}{k} \cdot \ln(w(x)) - \Delta_0, 0\right).$$

73. An Optimal Strategy: Algorithm

- *Reminder:* for some Δ_0, the optimal strategy has the form
$$\Delta s(x) = \max\left(\frac{1}{k}\cdot \ln(w(x)) - \Delta_0, 0\right).$$
- *How to find Δ_0:* from the condition that
$$\int \Delta s(x)\,dx =$$
$$\int \max\left(\frac{1}{k}\cdot \ln(w(x)) - \Delta_0, 0\right)\,dx = T_0.$$
- *Easy to check:* the above integral monotonically decreases with Δ_0.
- *Conclusion:* we can use bisection to find the appropriate value Δ_0.

Appendix 6. Designing Optimal Placement Tests

74. Assessing the Initial Knowledge Level

- Computers enable us to provide individualized learning, at a pace tailored to each student.
- In order to start the learning process, it is important to find out the current level of the student's knowledge.
- Usually, such placement tests use a sequence of N problems of increasing complexity.
- If a student is able to solve a problem, the system generates a more complex one.
- If a student cannot solve a problem, the system generates an easier one, etc.
- Once we find the exact level of student's knowledge, the actual learning starts.
- It is desirable to get to actual leaning as soon as possible, i.e., to minimize the # of placement problems.

75. Bisection – Optimal Search Procedure

- At each stage, we have:
  - the largest level i at which a student can solve, &
  - the smallest level j at which s/he cannot.
- Initially, $i = 0$ (trivial), $j = N + 1$ (very tough).
- If $j = i + 1$, we found the student's level of knowledge.
- If $j > i + 1$, give a problem on level $m \stackrel{\text{def}}{=} (i + j)/2$:
  - if the student solved it, increase i to m;
  - else decrease j to m.
- In both cases, the interval $[i, j]$ is decreased by half.
- In s steps, we decrease the interval $[0, N + 1]$ to width $(N + 1) \cdot 2^{-s}$.
- In $s = \lceil \log_2(N+1) \rceil$ steps, we get the interval of width ≤ 1, so the problem is solved.

76. Need to Account for Discouragement

- Every time a student is unable to solve a problem, he/she gets discouraged.
- In bisection, a student whose level is 0 will get $\approx \log_2(N+1)$ negative feedbacks.
- For positive answers, the student simply gets tired.
- For negative answers, the student also gets stressed and frustrated.
- If we count an effect of a positive answer as one, then the effect of a negative answer is $w > 1$.
- The value w can be individually determined.
- We need a testing scheme that minimizes the worst-case overall effect.

77. Analysis of the Problem

- We have $x = N + 1$ possible levels of knowledge.
- Let $e(x)$ denote the smallest possible effect needed to find out the student's knowledge level.
- We ask a student to solve a problem of some level n.
- If s/he solved it (effect $= 1$), we have $x - n$ possible levels $n, \ldots, N$.
- The effect of finding this level is $e(x - n)$, so overall effect is $1 + e(x - n)$.
- If s/he didn't (effect w), his/her level is between 0 and n, so we need effect $e(n)$, with overall effect $w + e(n)$.
- Overall worst-case effect is $\max(1+e(x-n), w+e(n))$.
- In the optimal test, we select n for which this effect is the smallest, so $e(x) = \min\limits_{1\le n<x} \max(1+e(x-n), w+e(n))$.

78. Resulting Algorithm

- For $x = 1$, i.e., for $N = 0$, we have $e(1) = 0$.
- We know that $e(x) = \min\limits_{1 \le n < x} \max(1 + e(x - n), w + e(n))$.
- We can use this formula to sequentially compute the values $e(2)$, $e(3)$, ..., $e(N + 1)$.
- We also compute the corresponding minimizing values $n(2)$, $n(3)$, ..., $n(N + 1)$.
- Initially, $i = 0$ and $j = N + 1$.
- At each iteration, we ask to solve a problem at level $m = i + n(j - i)$:
  - if the student succeeds, we replace i with m;
  - else we replace j with m.
- We stop when $j = i + 1$; this means that the student's level is i.

79. Example 1: $N = 3$, $w = 3$

- Here, $e(1) = 0$.
- When $x = 2$, the only possible value for n is $n = 1$, so

$$e(2) = \min_{1 \le n < 2} \{\max\{1 + e(2-n), 3 + e(n)\}\} =$$

$$\max\{1 + e(1), 3 + e(1)\} = \max\{1, 3\} = 3.$$

- Here, $e(2) = 3$, and $n(2) = 1$.
- To find $e(3)$, we must compare two different values $n = 1$ and $n = 2$:

$$e(3) = \min_{1 \le n < 3} \{\max\{1 + e(3-n)), 3 + e(n)\}\} =$$

$$\min\{\max\{1+e(2), 3+e(1)\}, \max\{1+e(1), 3+e(2)\}\} =$$

$$\min\{\max\{4, 3\}, \max\{1, 6\}\} = \min\{4, 6\} = 4.$$

- Here, min is attained when $n = 1$, so $n(3) = 1$.

80. Example 1: $N = 3$, $w = 3$ (cont-d)

- To find $e(4)$, we must consider three possible values $n = 1$, $n = 2$, and $n = 3$, so

$$e(4) = \min_{1 \le n < 4} \{\max\{1 + e(4 - n), 3 + e(n))\}\} =$$

$$\min\{\max\{1 + e(3), 3 + e(1)\}, \max\{1 + e(2), 3 + e(2)\},$$

$$\max\{1 + e(1), 3 + e(3)\}\} =$$

$$\min\{\max\{5, 3\}, \max\{4, 6\}, \max\{1, 7\}\} =$$

$$\min\{5, 6, 7\} = 5.$$

- Here, min is attained when $n = 1$, so $n(4) = 1$.

81. Example 1: Resulting Procedure

- First, $i = 0$ and $j = 4$, so we ask a student to solve a problem at level $i + n(j - i) = 0 + n(4) = 1$.
- If the student fails level 1, his/her level is 0.
- If s/he succeeds at level 1, we set $i = 1$, and we assign a problem of level $1 + n(3) = 2$.
- If the student fails level 2, his/her level is 1.
- If s/he succeeds at level 2, we set $i = 2$, and we assign a problem of level $2 + n(3) = 3$.
- If the student fails level 3, his/her level is 2.
- If s/he succeeds at level 3, his/her level is 3.
- We can see that this is the most cautious scheme, when each student has at most one negative experience.

82. Example 2: $N = 3$ and $w = 1.5$

- We take $e(1) = 0$.
- When $x = 2$, then

$$e(2) = \min_{1 \le n < 2} \{\max\{1 + e(2-n), 3 + e(n)\}\} =$$

$$\max\{1 + e(1), 1.5 + e(1)\} = \max\{1, 1.5\} = 1.5.$$

- Here, $e(2) = 1.5$, and $n(2) = 1$.
- To find $e(3)$, we must compare two different values $n = 1$ and $n = 2$:

$$e(3) = \min_{1 \le n < 3} \{\max\{1 + e(3-n)), 1.5 + e(n)\}\} =$$

$$\min\{\max\{1+e(2), 1.5+e(1)\}, \max\{1+e(1), 1.5+e(2)\}\} =$$

$$\min\{\max\{2.5, 1.5\}, \max\{1, 3\}\} = \min\{2.5, 3\} = 2.5.$$

- Here, min is attained when $n = 1$, so $n(3) = 1$.

83. Example 2: $N = 3$ and $w = 1.5$ (cont-d)

- To find $e(4)$, we must consider three possible values $n = 1$, $n = 2$, and $n = 3$, so

$$e(4) = \min_{1 \le n < 4} \{\max\{1 + e(4 - n), 1.5 + e(n))\}\} =$$

$$\min\{\max\{1+e(3), 1.5+e(1)\}, \max\{1+e(2), 1.5+e(2)\},$$

$$\max\{1 + e(1), 1.5 + e(3)\}\} =$$

$$\min\{\max\{3.5, 1.5\}, \max\{2.5, 3\}, \max\{1, 4\}\} =$$

$$\min\{3.5, 3, 4\} = 3.$$

- Here, min is attained when $n = 2$, so $n(4) = 2$.

84. Example 2: Resulting Procedure

- First, $i = 0$ and $j = 4$, so we ask a student to solve a problem at level $i + n(j - i) = 0 + n(4) = 2$.
- If the student fails level 2, we set $j = 2$, and we assign a problem of level $0 + n(2) = 1$:
  - if the student fails level 1, his/her level is 0;
  - if s/he succeeds at level 1, his/her level is 1.
- If s/he succeeds at level 2, we set $i = 2$, and we assign a problem at level $2 + n(2) = 3$:
  - if the student fails level 3, his/her level is 2;
  - if s/he succeeds at level 3, his/her level is 3.
- We can see that in this case, the optimal testing scheme is bisection.

85. A Faster Algorithm May Be Needed

- For each n from 1 to N, we need to compare n different values.
- So, the total number of computational steps is proportional to $1 + 2 + \ldots + N = O(N^2)$.
- When N is large, N^2 may be too large.
- In some applications, the computation of the optimal testing scheme may take too long.
- For this case, we have developed a faster algorithm for producing a testing scheme.
- The disadvantage of this algorithm is that it is only asymptotically optimal.

86. A Faster Algorithm for Generating an Asymptotically Optimal Testing Scheme

- First, we find the real number $\alpha \in [0, 1]$ for which $\alpha + \alpha^w = 1$.
- This value α can be obtained, e.g., by applying bisection to the equation $\alpha + \alpha^w = 1$.
- At each iteration, once we know bounds i and j, we ask the student to solve a problem at the level
$$m = \lfloor \alpha \cdot i + (1 - \alpha) \cdot j \rfloor.$$
- This algorithm is similar to bisection, expect that bisection corresponds to $\alpha = 0.5$.
- This makes sense, since for $w = 1$, the equation for α takes the form $2\alpha = 1$, hence $\alpha = 0.5$.
- For $w = 2$, the solution to the equation $\alpha + \alpha^2 = 1$ is the well-known golden ratio $\alpha = \dfrac{\sqrt{5} - 1}{2} \approx 0.618$.

Appendix 7. Towards Optimal Feedback for Students

87. Towards Optimal Teaching

- One of the main objectives of a course – calculus, physics, etc. – is to help students understand its main concepts.
- Of course, it is also desirable that the students learn the corresponding methods and algorithms.
- However, understanding is the primary goal.
- If a student does not remember a formula by heart, she can look it up.
- However:
  - if a student does not have a good understanding of what, for example, is a derivative,
  - then even if this student remembers some formulas, he will not be able to decide which formula to apply.

88. How to Gauge Student Understanding

- To properly gauge student's understanding, several disciplines have developed *concept inventories*.
- These are sets of important basic concepts and questions testing the students' understanding.
- The first such Force Concept Inventory (FCI) was developed to gauge the students' understanding of forces.
- A student's degree of understanding is measured by the percentage of the questions that are answered correctly.
- The class's degree of understanding is measured by averaging the students' degrees.
- An ideal situation is when everyone has a perfect 100% understanding; in this case, the average score is 100%.
- In practice, the average score is smaller than 100%.

89. How to Compare Different Teaching Techniques

- We can measure the average score μ_0 before the class and the average score μ_f after the class.
- Ideally, the whole difference $100 - \mu_0$ disappears, i.e., the students' score goes from μ_0 to $\mu_f = 100$.
- In practice, of course, the students' gain $\mu_f - \mu_0$ is somewhat smaller than the ideal gain $100 - \mu_0$.
- It is reasonable to measure the success of a teaching method by which portion of the ideal gain is covered:

$$g \stackrel{\text{def}}{=} \frac{\mu_f - \mu_0}{100 - \mu_0}.$$

90. Empirical Results

- It turns out that the gain g does not depend on the initial level μ_0, on the textbook used, or on the teacher.
- Only one factor determines the value g: the absence or presence of immediate feedback.
- In traditionally taught classes,
  - where the students get their major feedback only after their first midterm exam,
  - the average gain is $g \approx 0.23$.
- For the classes with an immediate feedback, the average gain is twice larger: $g \approx 0.48$.
- In this talk, we provide a possible geometric explanation for this doubling of the learning rate.

91. Why Geometry

- Learning means changing the state of a student.
- At each moment of time, the state can be described by the scores $x_1, \ldots, x_n$ on different tests.
- Each such state can be naturally represented as a point $(x_1, \ldots, x_n)$ in the n-dimensional space.
- In the starting state S, the student does not know the material.
- The desired state D describes the situation when a student has the desired knowledge.
- When a student learns, the student's state of knowledge changes continuously.
- It forms a (continuous) trajectory γ which starts at the starting state S and ends up at the desired state D.

92. First Simplifying Assumption: All Students Learn at the Same Rate

- Some students learn faster, others learn slower.
- The above empirical fact, however, is not about their *individual* learning rates.
- It is about the *average* rates of student learning, averaged over all kinds of students.
- From this viewpoint, it makes sense to assume that all the students have the same average learning rate.
- In geometric terms, this means that the leaning time is proportional to the length of the corresponding curve γ.
- We thus need to show that learning trajectories corr. to immediate feedback are, on average, twice shorter.

93. Second Simplifying Assumption: the Shape of the Learning Trajectories

- At first, a student has misconceptions about physics or calculus, which lead him in a wrong direction.
- We can thus assume that at first, a student moves in a random direction.
- After the feedback, the student corrects his/her trajectory.
- In the case of immediate feedback, this correction comes right away, so the students goes in the right direction.
- In the traditional learning, with a midterm correction:
  - a student first follows a straight line of length $d/2$ which goes in a random direction,
  - and then takes a straight line to the midpoint M.
- Then, a student goes from M to the destination D.

94. 3rd Simplifying Assumption: 1-D State Space

- We can think of different numerical characteristics describing different aspects of student knowledge.
- In practice, to characterize the student's knowledge, we use a single number – the overall grade for the course.
- It is therefore reasonable to assume that the state of a student is characterized by only one parameter x_1.
- In case of immediate feedback, the learning trajectory has length d.
- To make a comparison, we must estimate the length of a trajectory corresponding to the traditional learning.
- This trajectory consists of two similar parts: connecting S and M and connecting M and D.
- To estimate the total average length, we can thus estimate the average length from S to M and double it.

95. Analysis: Case of Traditional Leaning

- A student initially goes either in the correct direction or in the opposite (wrong) direction.
- Randomly means that both directions occur with equal probability $1/2$.
- If the student moves in the right direction, she gets exactly into the desired midpoint M.
- In this case, the length of the S-to-M part of the trajectory is exactly $d/2$.
- If the student starts in the wrong direction, he ends up at a point at distance $d/2$ – on the wrong side of S.
- Getting back to M then means first going back to S and then going from S to M.
- The overall length of this trajectory is thus $3d/2$.

96. Resulting Geometric Explanation

- Here:
  - with probability $1/2$, the length is $d/2$;
  - with probability $1/2$, the length is $3d/2$.
- So, the average length of the S-to-M part of the learning trajectory is equal to

$$\frac{1}{2} \cdot \frac{d}{2} + \frac{1}{2} \cdot \frac{3d}{2} = d.$$

- The average length of the whole trajectory is double that, i.e., $2d$.
- This average length is twice larger than the length d corresponding to immediate feedback.
- This explains why immediate feedback makes learning, on average, twice faster.

97. How to Use the Resulting Knowledge: Case Study

- How can we use the acquired knowledge?
- In many practical situations, we have a well-defined problem, with a clear well-formulated objective.
- Such problems are typical in engineering:
  - we want a bridge which can withstand a given load,
  - we want a car with a given fuel efficiency, etc.
- However, in many practical situations, it is important to also take into account subjective user preferences.
- This subjective aspect of decision making is known as *Kansei engineering*.