Problem Solving Strategies: Sampling and Heuristics

Kevin H. Knuth
Department of Physics
University at Albany
Albany NY, USA

Outline

Methodological Differences

Inverses vs. Inferences

Problem Transformation

From Inference to Searching

Rewarding Insights

Why Source Separation is Difficult

Solution

Heuristics and Sampling

The Curse of Uncertainty

A system can be in one state of a set of possible states

S1 •

S2 •

S3 •

S4 •

S5 •

The Curse of Uncertainty

A system can be in one state of a set of possible states
We can make measurements of the system

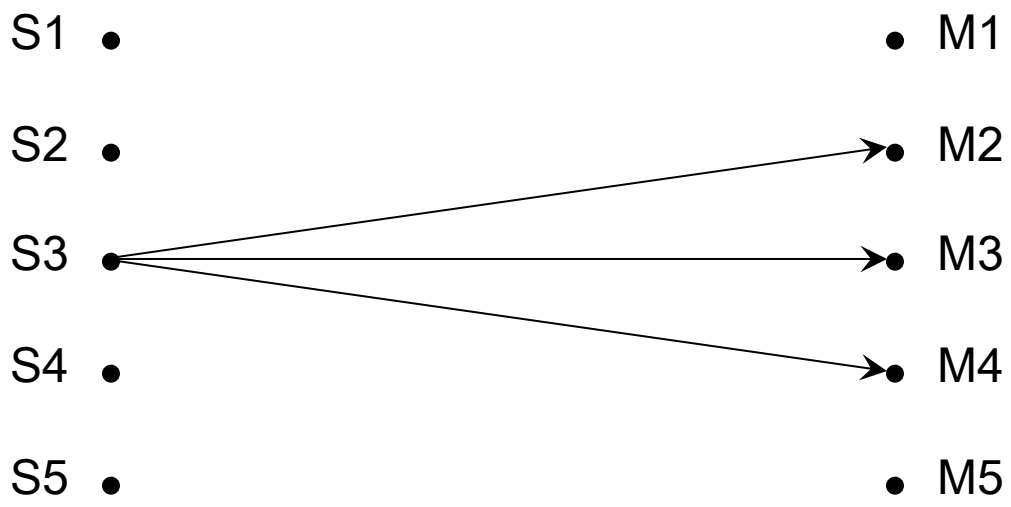
The Curse of Uncertainty

A system can be in one state of a set of possible states
We can make measurements of the system

Ideally, each state maps to one and only one measurement

The Curse of Uncertainty

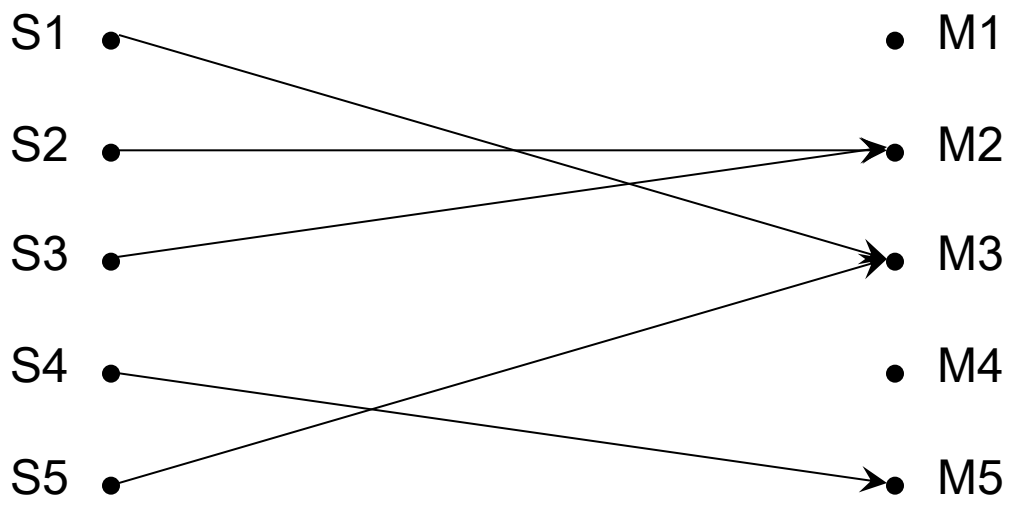
A system can be in one state of a set of possible states
We can make measurements of the system



Ideally, each state maps to one and only one measurement
However, noise can corrupt the measurement

The Inverse is Most Likely Not Defined

This uncertainty means that we are not assured of having a one-to-one mapping from the states to the measurements



Thus, the inverse mapping is not assured to exist!

Source Separation as Inference

Since there is not enough information to obtain a unique solution, we are left to infer the most probable solution.

Likelihood

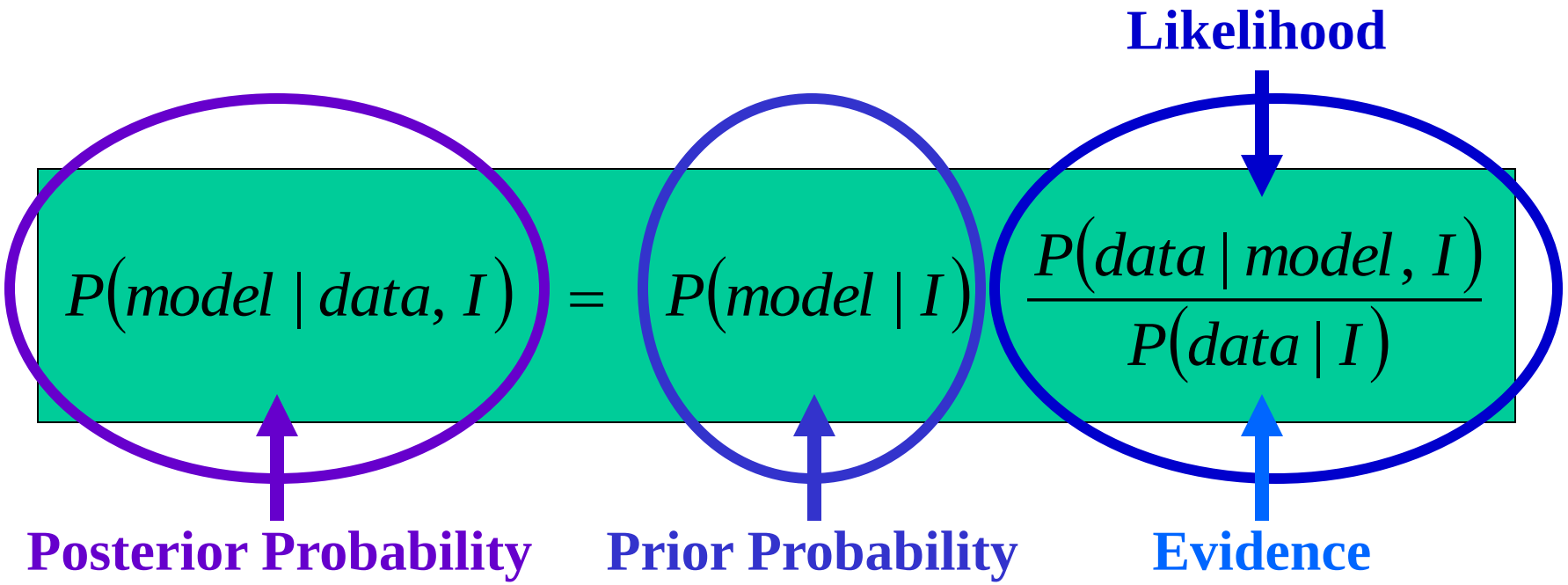
$$P(model \mid data, I) = P(model \mid I) \frac{P(data \mid model, I)}{P(data \mid I)}$$

Posterior Probability **Prior Probability** **Evidence**

Bayes Theorem describes how our prior knowledge about a model, based on our prior information I, is modified by the acquisition of new information or data:

Machine Learning

Bayes Theorem is a learning rule.
The Prior Probability describes what you first knew. Multiply this by a term that describes the effect of your new information, and the result is what you know after you have taken into account your new data.



Outline

Methodological Differences

Inverses vs. Inferences

Problem Transformation

From Inference to Searching

Rewarding Insights

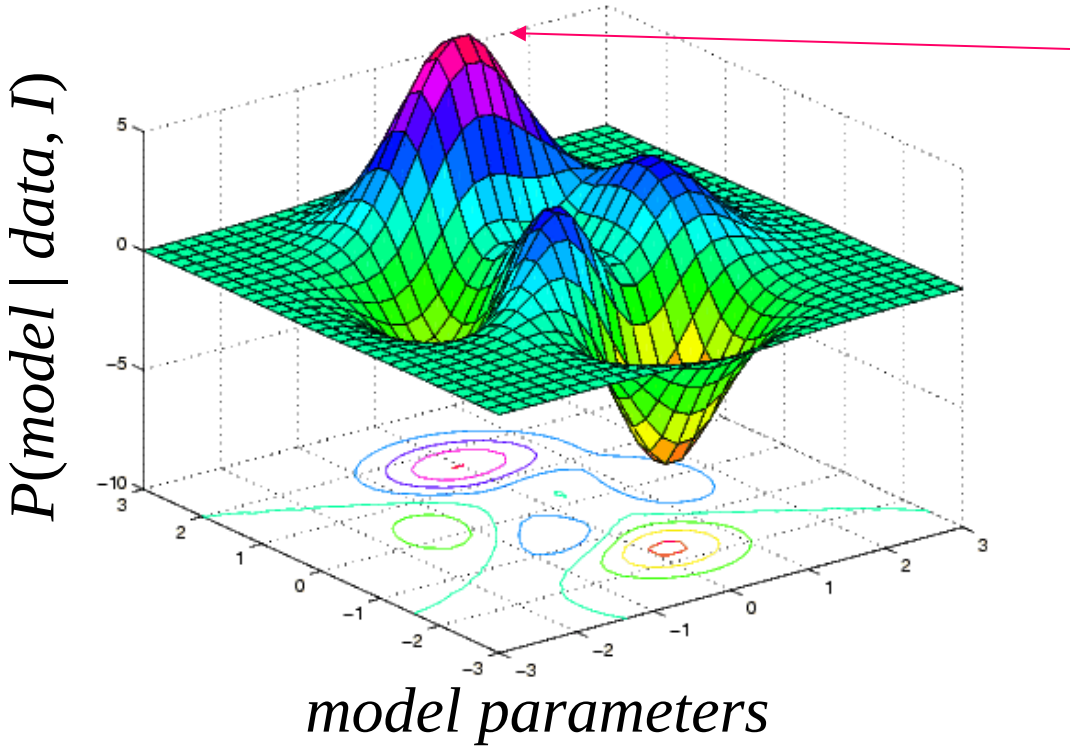
Why Source Separation is Difficult

Solution

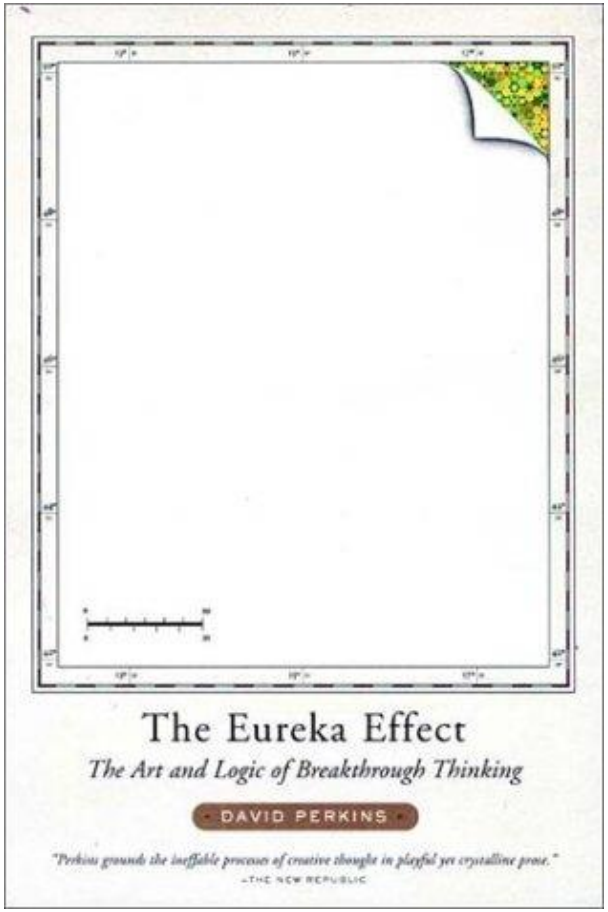
Heuristics and Sampling

From Inference to Searching

The Posterior Probability can be viewed as an optimization criterion that is to be maximized. This *recasts the inference problem into a search problem* where we search the space of model parameters to find the high probability regions



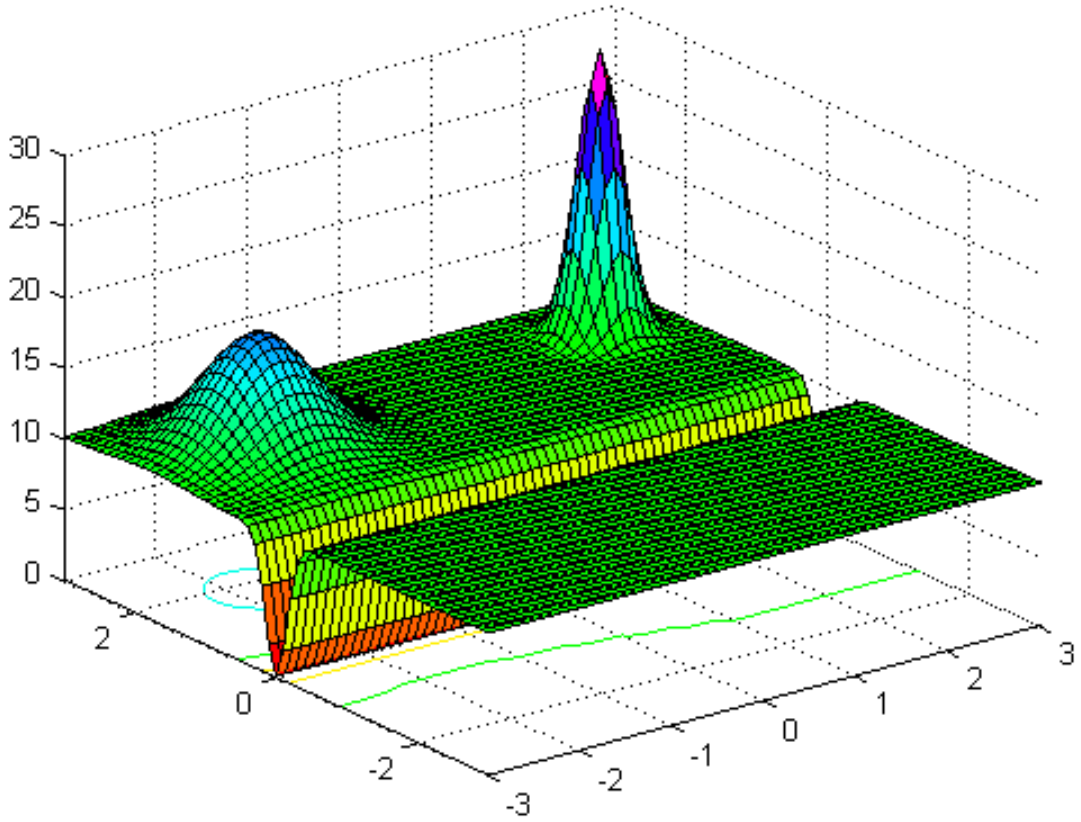
The Art and Logic of Breakthrough Thinking



David Perkins (Harvard)

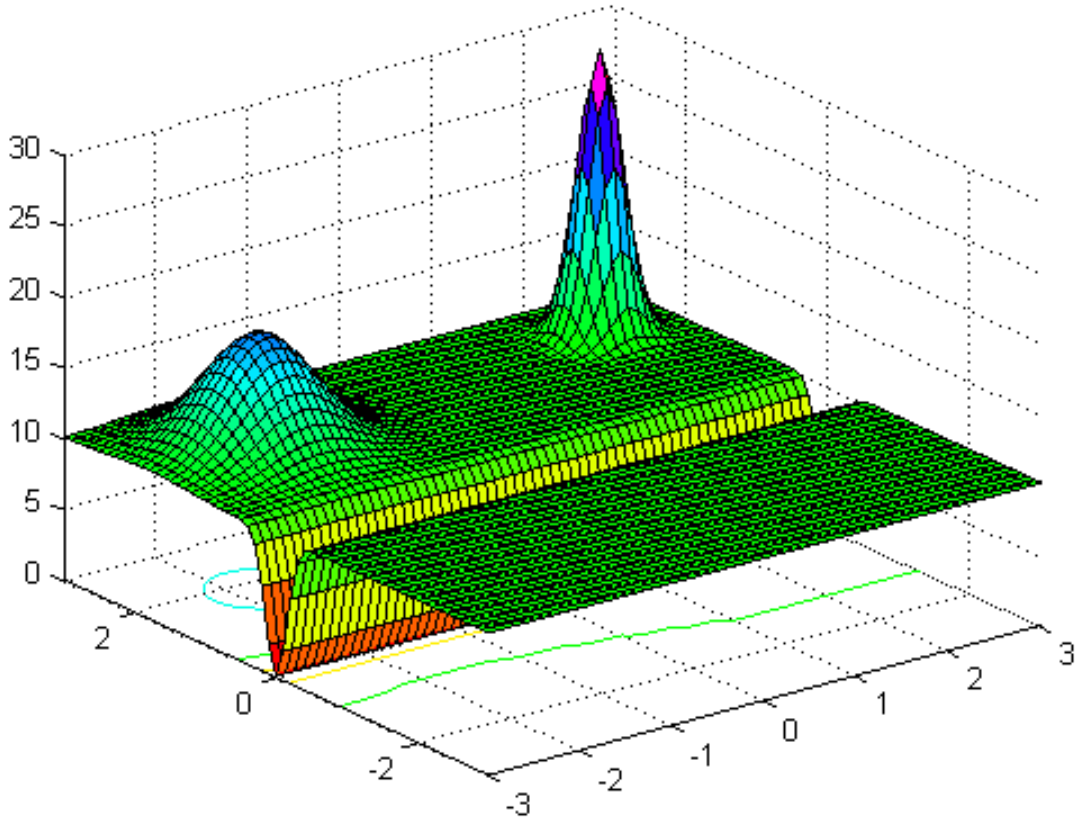
Klondike Space

David Perkins describes these spaces as Klondike spaces.
There is gold in the hills, but the terrain is treacherous!



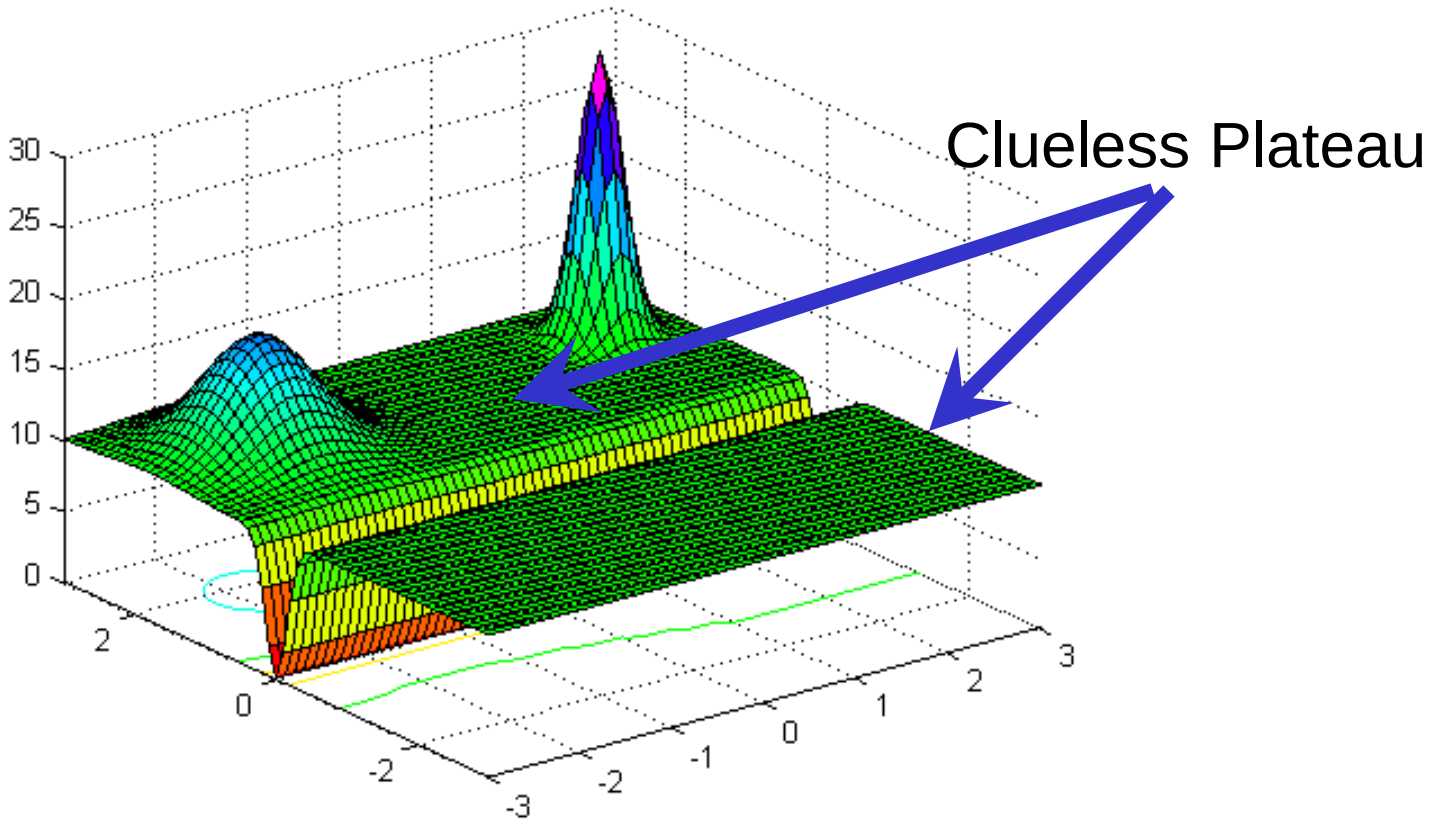
Wilderness Trap

The **Wilderness Trap** is characterized by the space having a large volume to search over. This is increasingly problematic in higher-dimensions.



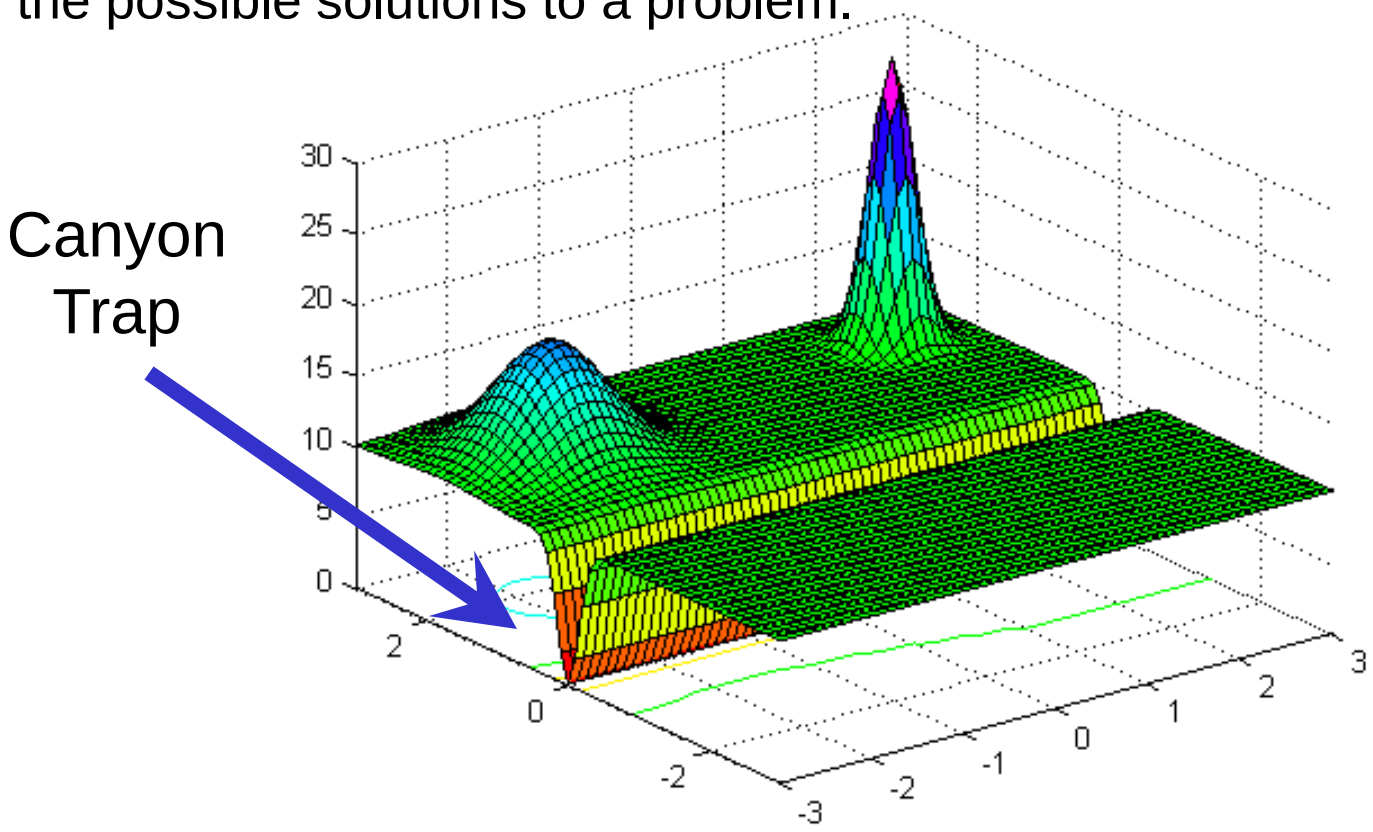
Clueless Plateau

The **Clueless Plateau** is characterized by a flat region of the space that provides no hints as to where the peak might be.



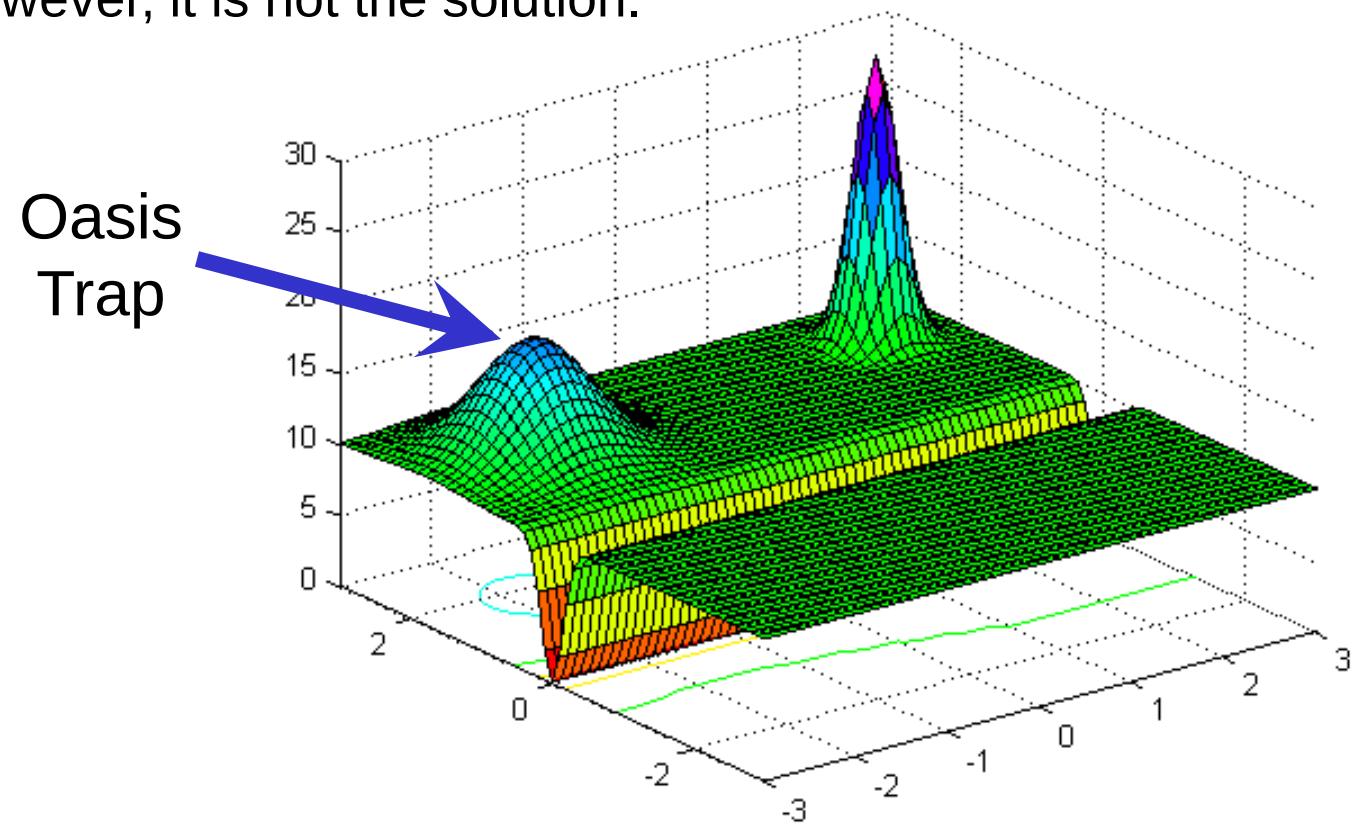
Narrow Canyon of Exploration

The **Canyon Trap** is characterized by artificially constrained exploration within an extensive low probability region. This arises from not considering all the possible solutions to a problem.



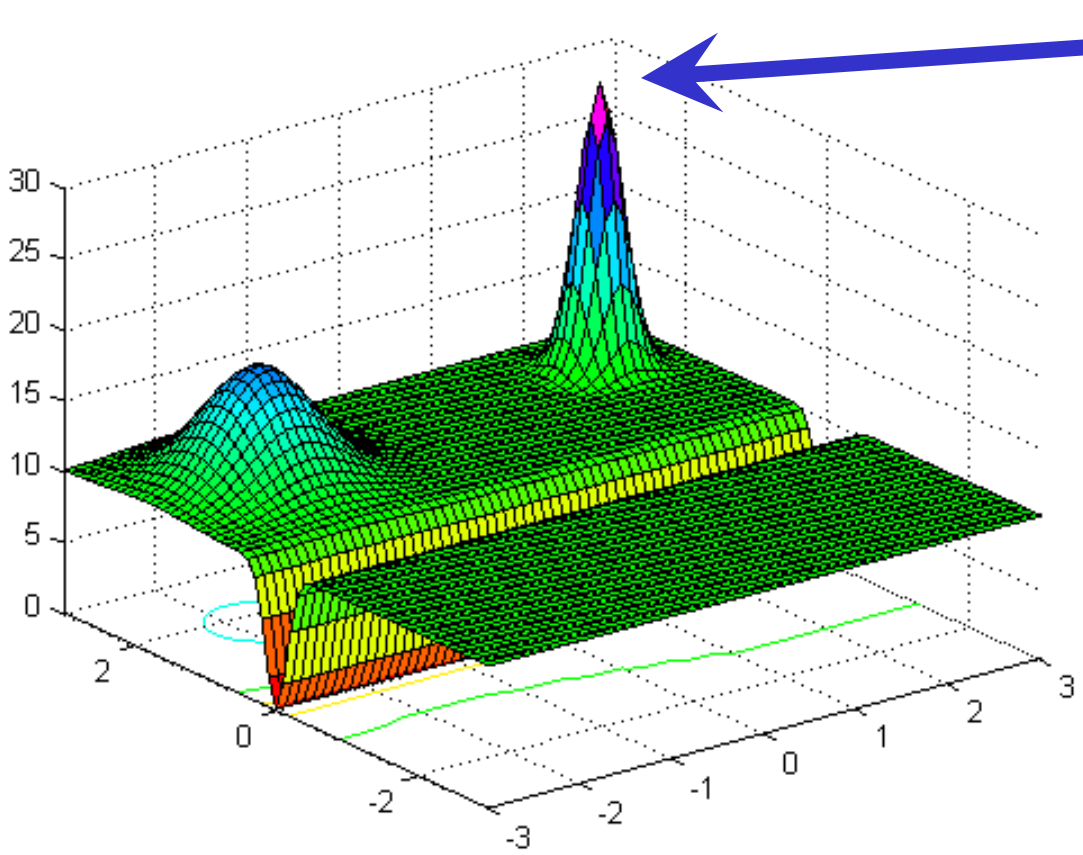
Oasis of False Promise

The **Oasis Trap** is characterized by a local maximum that is so tempting that it is almost impossible to move away from.
However, it is not the solution.

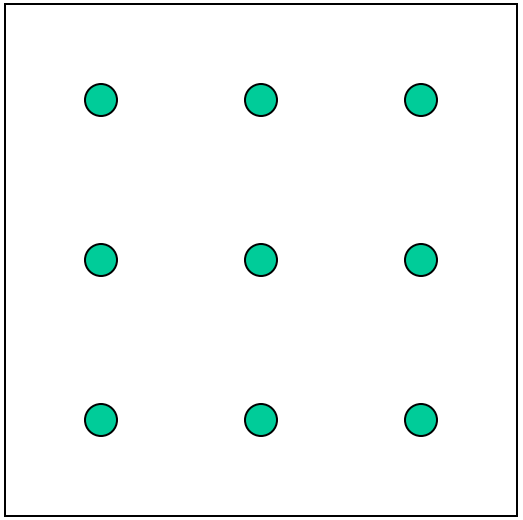


True Solution

The **Solution** may occupy a relatively small volume in the space making it extremely difficult to find.

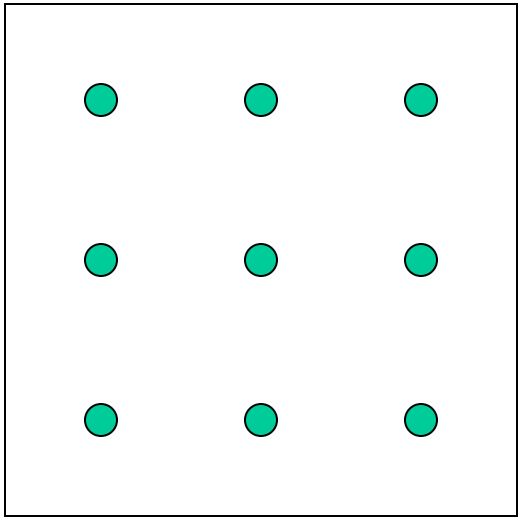


Puzzles as Challenging Problems



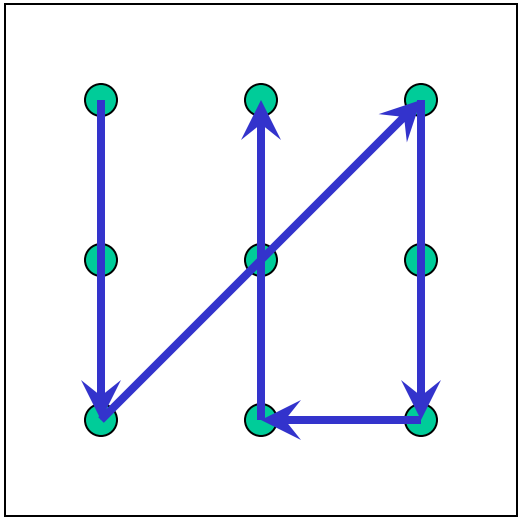
Without lifting your pencil off the page,
connect the dots using only 4 straight lines

Why is this Problem Difficult?



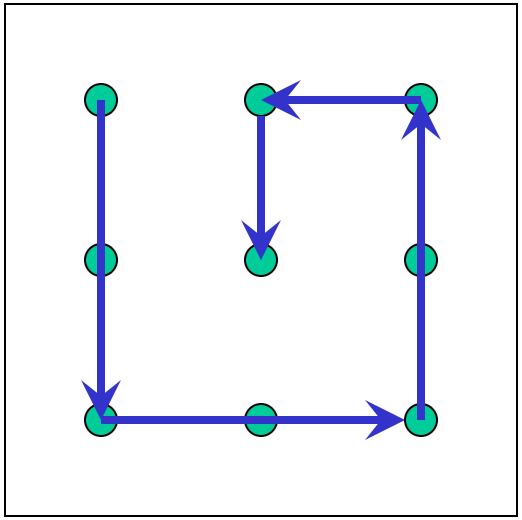
Wilderness Trap: Very many ways of drawing 4 straight lines

Why is this Problem Difficult?



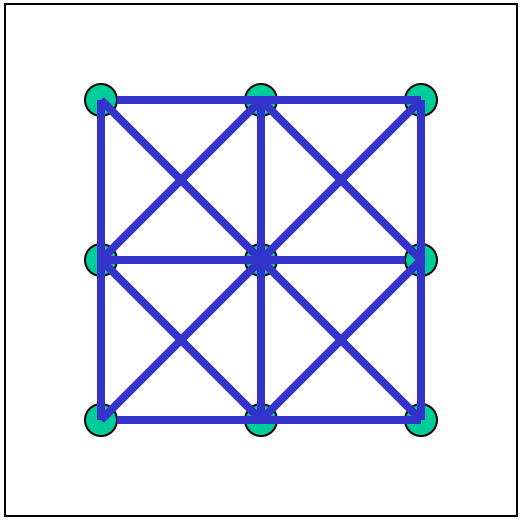
Wilderness Trap: Very many ways of drawing 4 straight lines
Oasis Trap: Many close solutions using 5 lines

Why is this Problem Difficult?



Wilderness Trap: Very many ways of drawing 4 straight lines
Oasis Trap: Many close solutions using 5 lines
Clueless Plateau: Close solutions do not provide insight

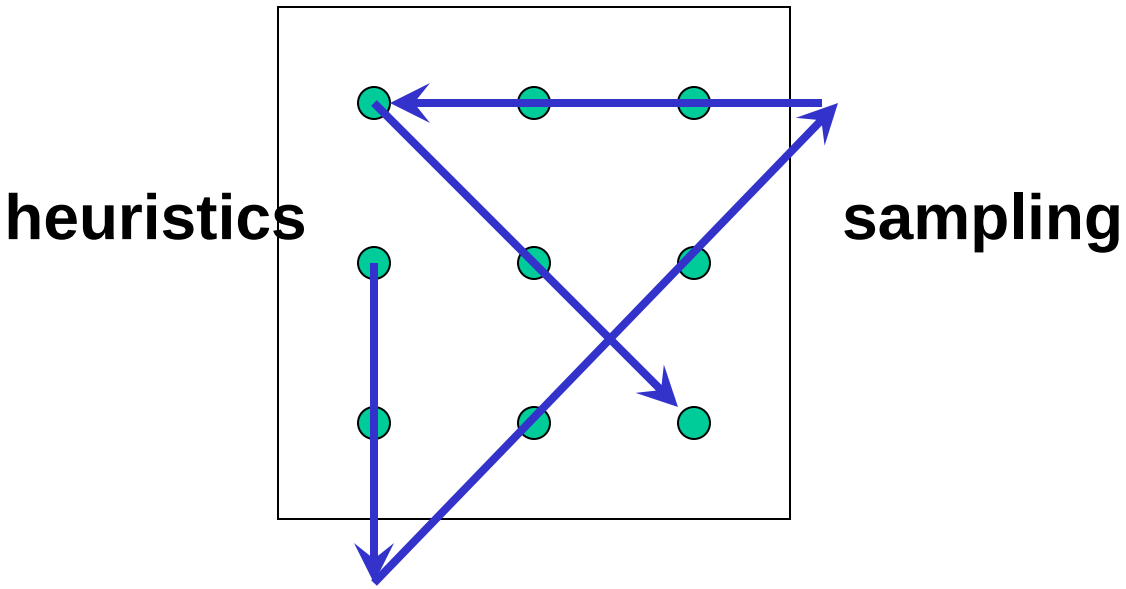
Why is this Problem Difficult?



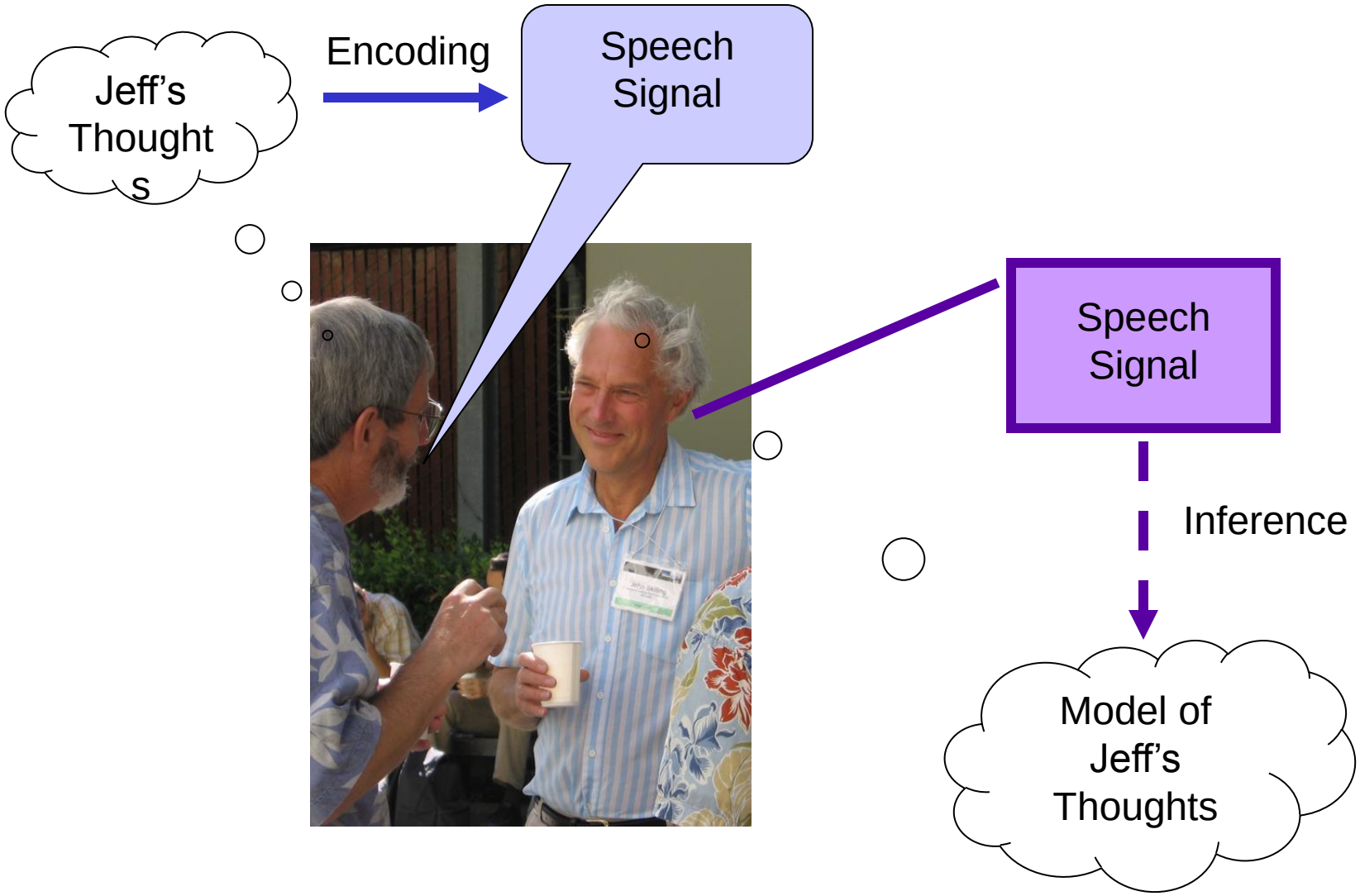
Wilderness Trap: Very many ways of drawing 4 straight lines
Oasis Trap: Many close solutions using 5 lines
Clueless Plateau: Close solutions do not provide insight
Canyon Trap: Focused on subspace of possible solutions

Search Strategies

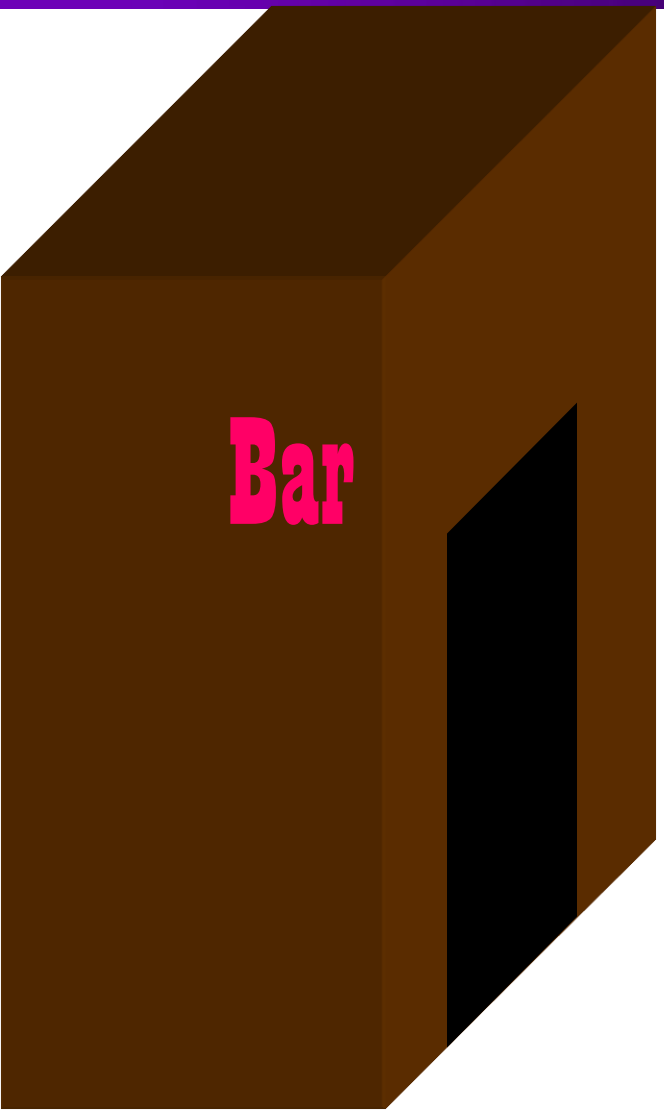
To overcome the topological difficulties of the search space, one must **balance educated guesses with exploration**



Communication as a Problem



Jokes Rely on Difficult Topography



A String Walks into a Bar

He Sits Down and Orders a Beer

Jokes as Challenging Problems

The Bartender says,
“We don’t serve strings here!”

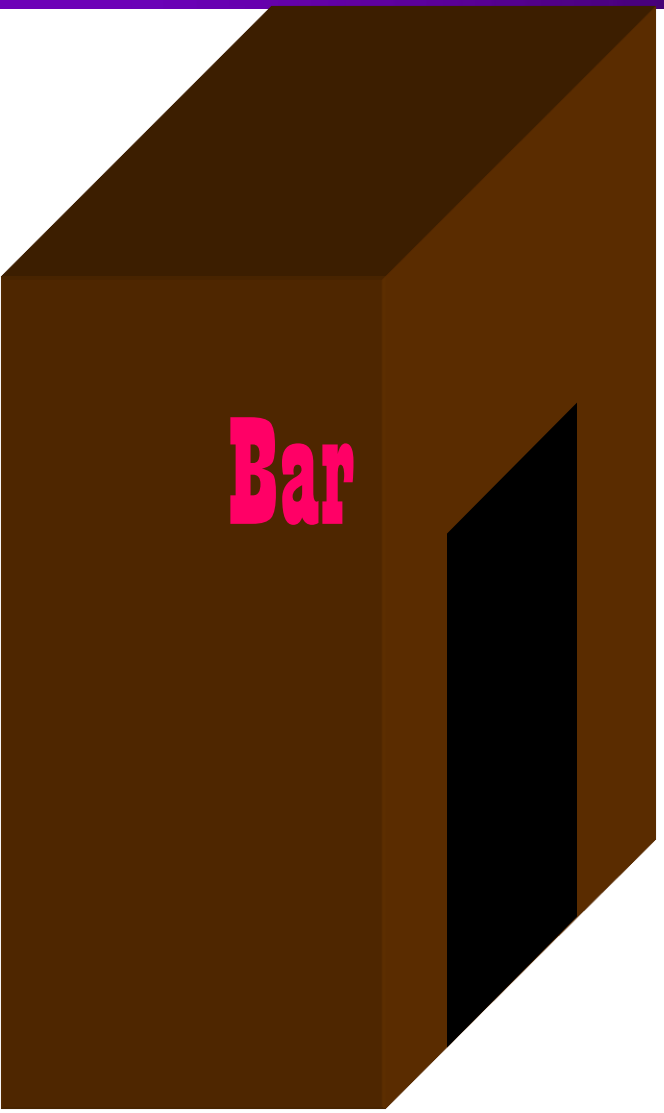
“Go on! Get outta here!!”

Jokes as Challenging Problems

The String leaves Frustrated.

When he gets outside, he messes up his hair and tangles himself until he is virtually unrecognizable.

Jokes as Challenging Problems



The String goes back into the Bar.

He sits down and orders a Beer.

Jokes as Challenging Problems

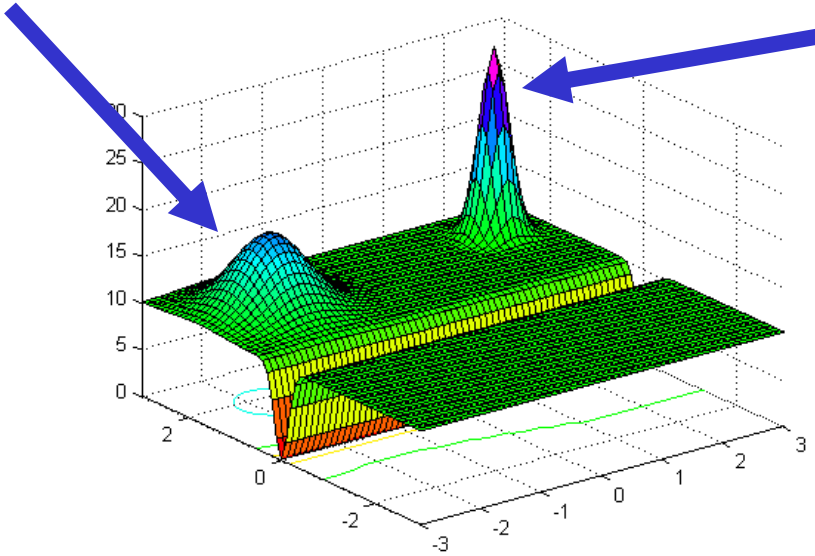
The Bartender looks at him funny and says, “Aren’t you that stupid String who was just in here?”

The String replies “No …”

Why Jokes are Funny

I'm afraid not!

I'm a frayed knot!



Several classes of jokes are funny because they pose problems that are difficult to solve. The problem is that of inferring the my meaning (model) from my speech (data).

This joke relies on an Oasis Trap!

Physics Joke

What is the difference between:

A person who drives and steers their way to work every day
and
The set of operations {driving and steering}?

Physics Joke

What is the difference between:

A person who drives and steers their way to work every day
and
The set of operations {driving and steering}?

One commutes and the other doesn't!

Problem of Limited Solution Space

This joke is again funny because of the Oasis Trap.

However, this joke fails, because the solution lies in a part of the space not accessible by many members of the audience. There is a Canyon Trap imposed by the nature of the joke itself.

In Comedy, this is a failure on the part of the comedian. However, in algorithm design, this is a failure on the part of the problem solver (audience).

Another Difficult Problem

Why is one side of the V in a flock of Geese usually longer than the other?

Related Problems

Why is one side of the V in a flock of Geese usually longer than the other?

Because there are more Geese in it!

This joke sets up a Canyon Trap that leads you to believe that the answer is profound; whereas it is instead simple.

Outline

Methodological Differences

Inverses vs. Inferences

Problem Transformation

From Inference to Searching

Rewarding Insights

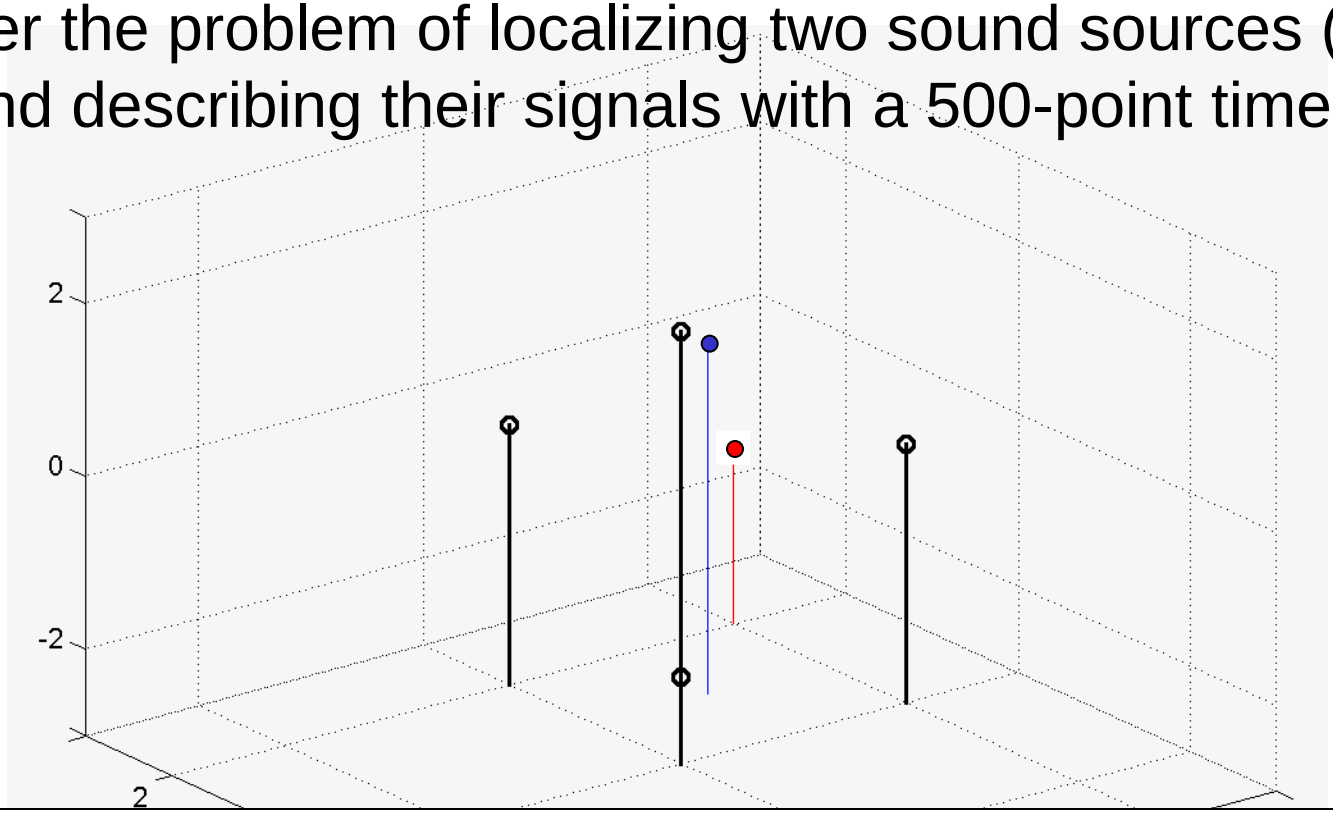
Why Source Separation is Difficult

Solution

Heuristics and Sampling

The Problem of Source Separation

Consider the problem of localizing two sound sources (red and blue) and describing their signals with a 500-point time-series.



This is a 2*3 + 2*500 = 1006 dimensional problem.

Expected Difficulties

With a 1006 dimensional problem we must deal with:

Wilderness Trap: Enormous volume to search. Even if the solution occupies 1/10 the volume along each parameter, the relative volume of the solution space is 10^{-1006}!

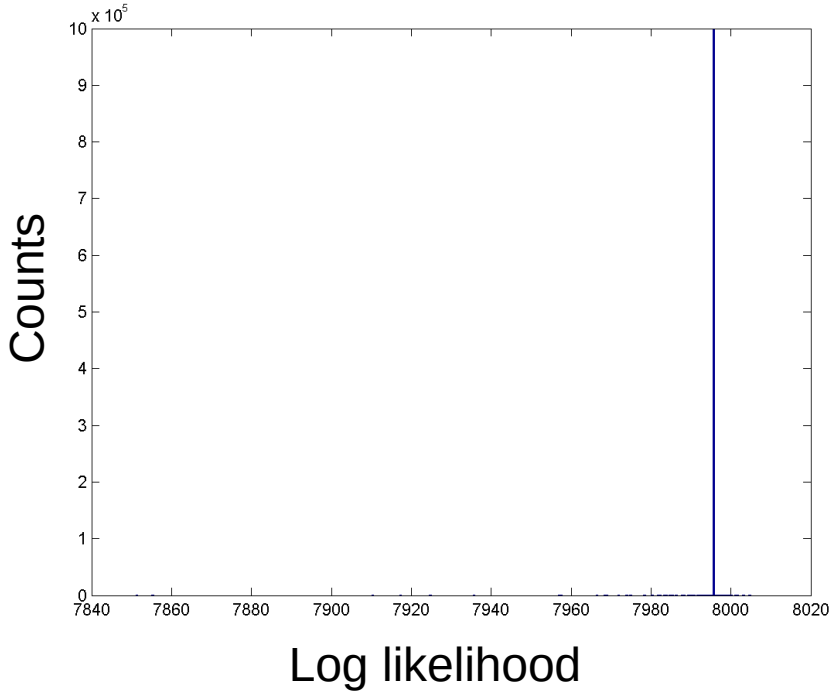
Oasis Traps: We expect local solutions. These are expected to increase dramatically with dimension.

But this may be surprising…

Clueless Plateau

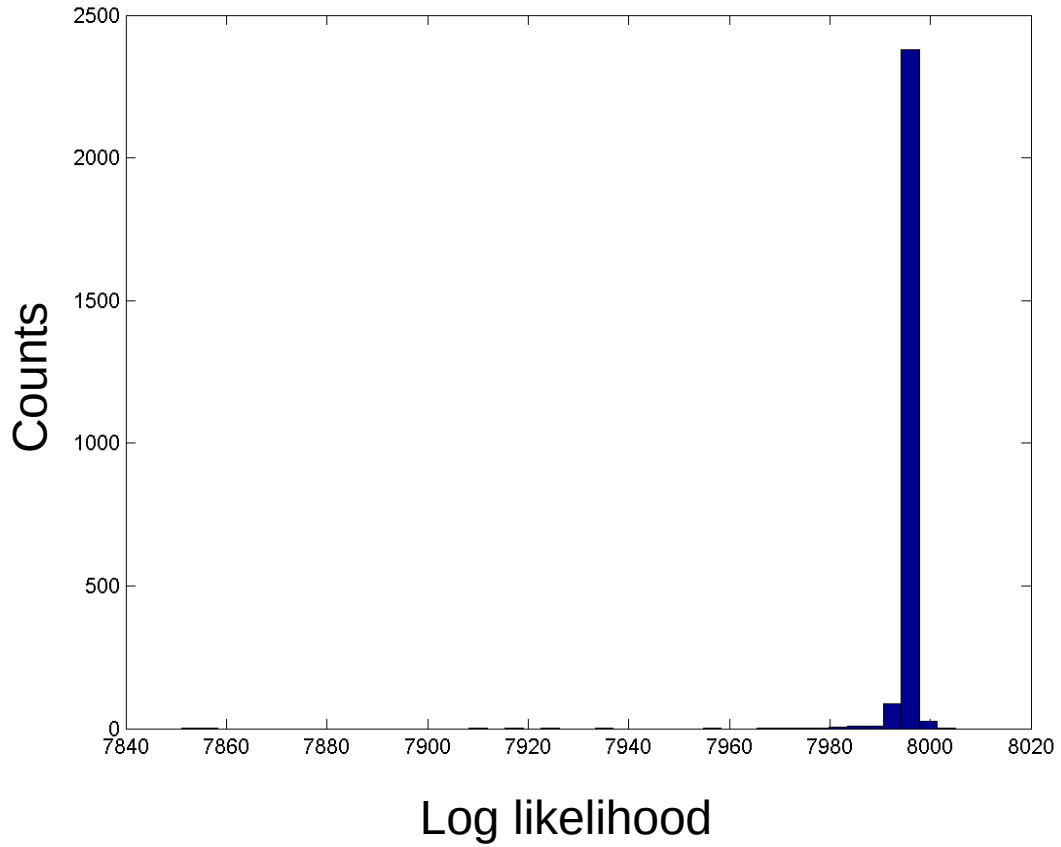
This is a histogram of the log likelihood of 1 MILLION samples uniformly drawn from the 1006 dimensional space.

The space is VERY flat!



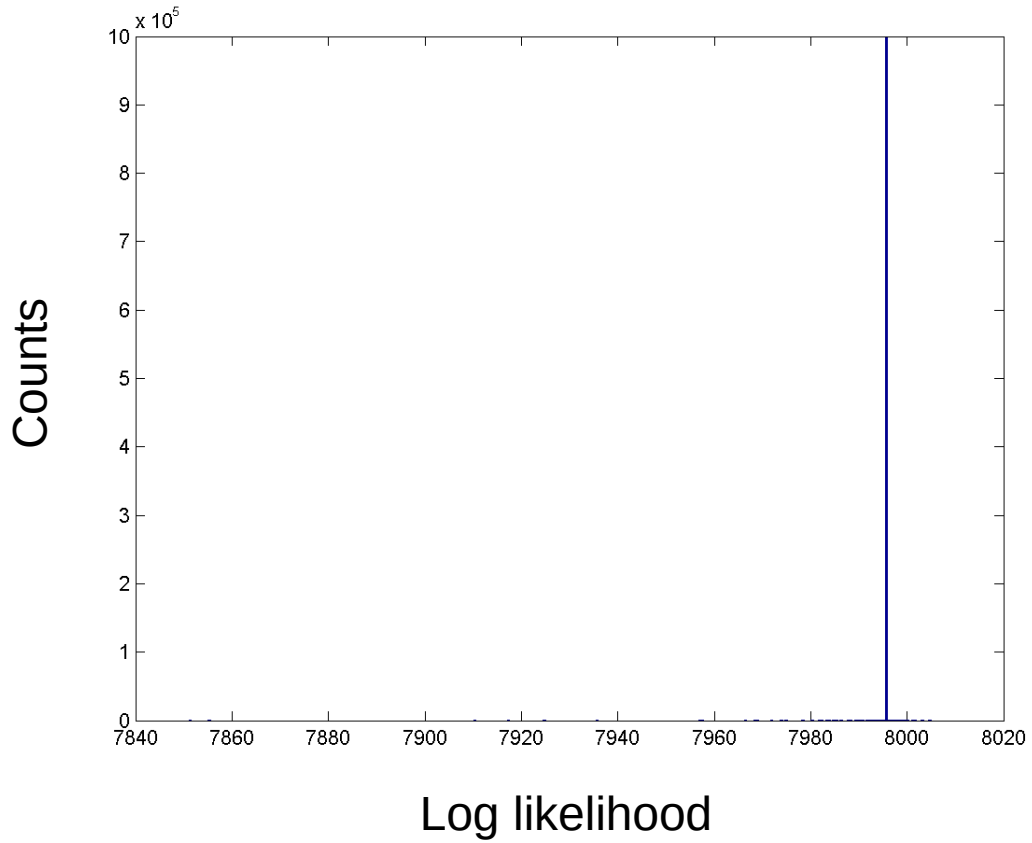
The Space is EXTREMELY FLAT!

Out of 1 MILLION samples, only 2527 samples were greater than 7996.3 or less than 7995.3



Nowhere Near a Solution

The peak Log Likelihood is at 17700, with ONE MILLION samples, we are still exp(10000) away from the solution!



Why Source Separation is Difficult

Wilderness Trap: Enormous volume to search. Even if the solution occupies 1/10 the volume along each parameter, the relative volume of the solution space is 10^{-1006}!

Oasis Traps: We expect local solutions. These are expected to increase dramatically with dimension.

Clueless Plateau: Extensive regions of the space are non-informative

Small Solution Volume: The solution occupies an extremely small volume in the entire solution space. Moreover, there are often gates, which are peaks on top of peaks.

Outline

Methodological Differences

Inverses vs. Inferences

Problem Transformation

From Inference to Searching

Rewarding Insights

Why Source Separation is Difficult

Solution

Heuristics and Sampling

Heuristics

Heuristic Solutions constitute Educated Guesses

These solutions rely on a set of various ideas:

- **General Observations** about the problem that constrain possible parameter values
- **Approximations** that constrain possible parameter values
- **Idealizations** or Simplifications that focus on a sub-problem and thus solve the problem within a subspace of the original problem
- **Marginalization** over a subspace focused the inference on a sub-problem

Examples of Heuristics

Many of the clever solutions seen in the literature are relying on heuristics. These are often computationally fast, but used alone, **heuristics are not consistently successful**

Examples in source separation

- Examine frequency bands unique to a source to pin down its contribution to the mixture

- In conversation, use fact that sources are not emitting at the same time to find temporal regions where the recordings are from a single source

- Decompose mixtures into a sparse library

Exploration via Sampling

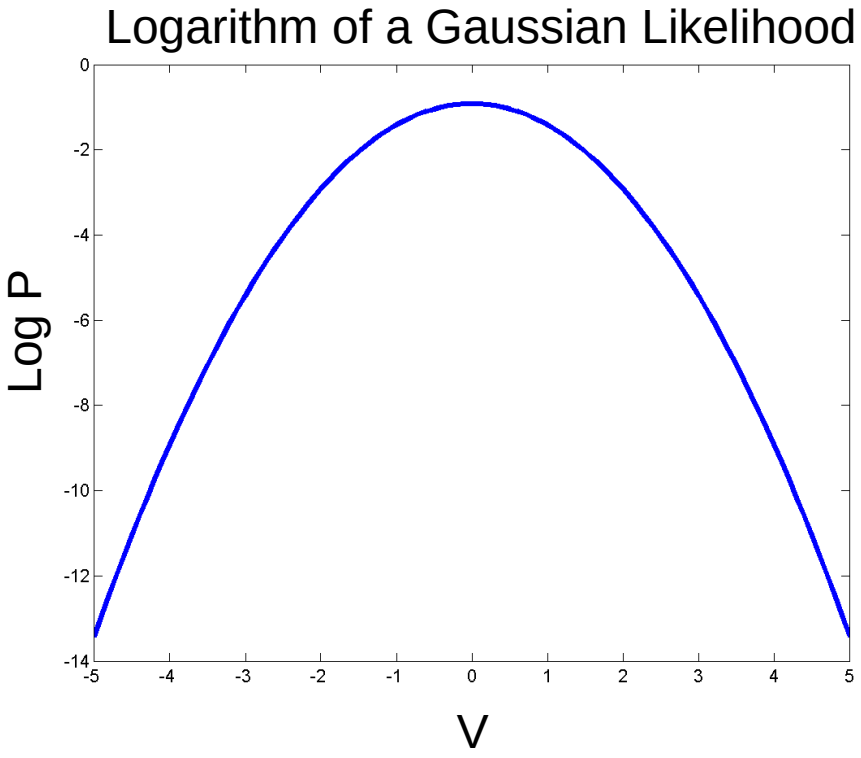
Sampling Techniques such as those collectively labeled Monte Carlo methods constitute Exploration.

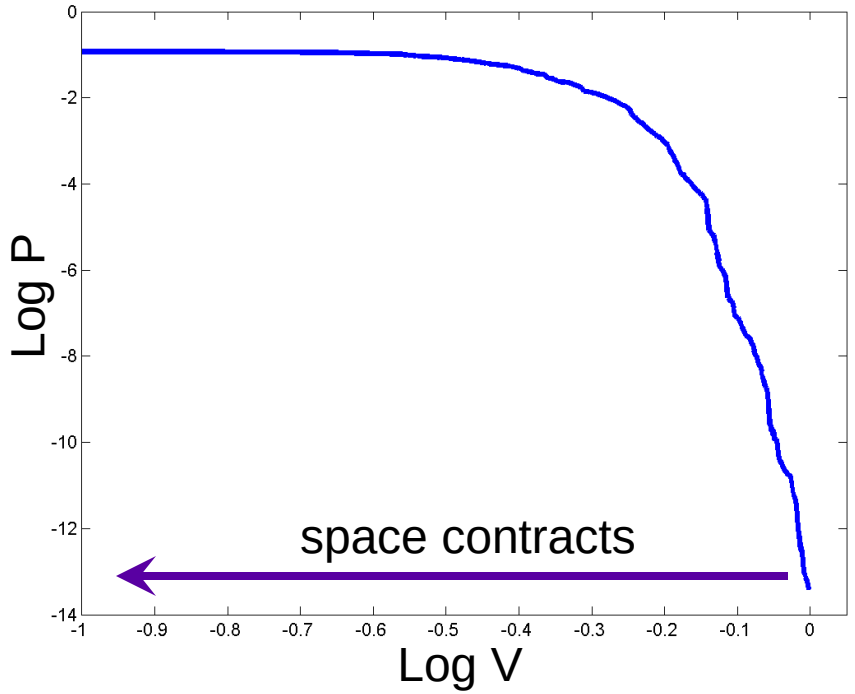
- **Metropolis-Hastings**
- **Gibbs Sampling**
- **Particle Filters**
- **Importance Sampling**
- **Nested Sampling (Skilling 2005)**

The advantages are that they are robust. The disadvantages are that in difficult problems, they can be slow and may not converge to the right solution.

Sorted Log Likelihoods reveal Structure

Skilling's Nested Sampling relies on sorting the log Likelihood values of samples drawn from the Prior Probability



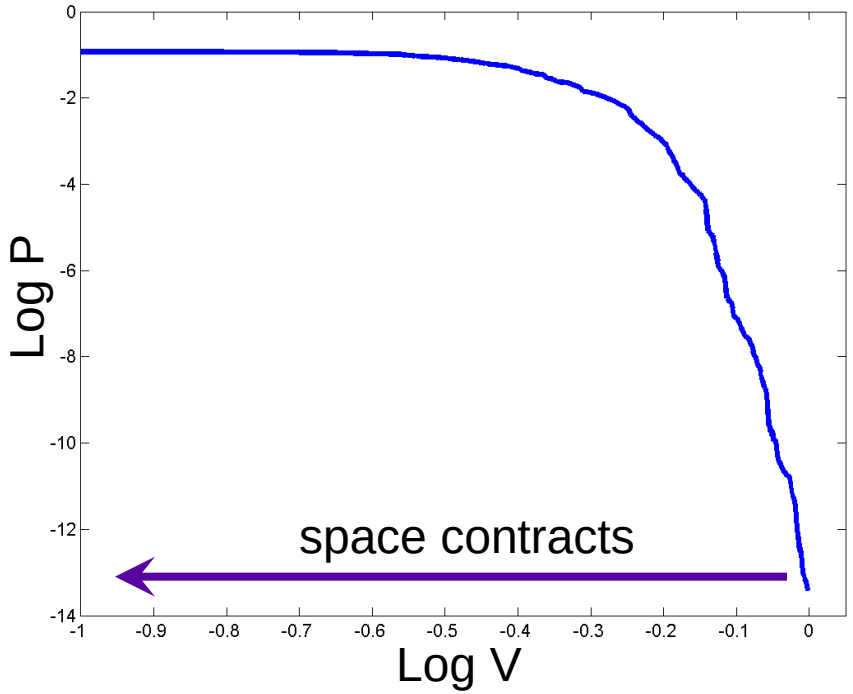


This turns all problems into 1-D problems

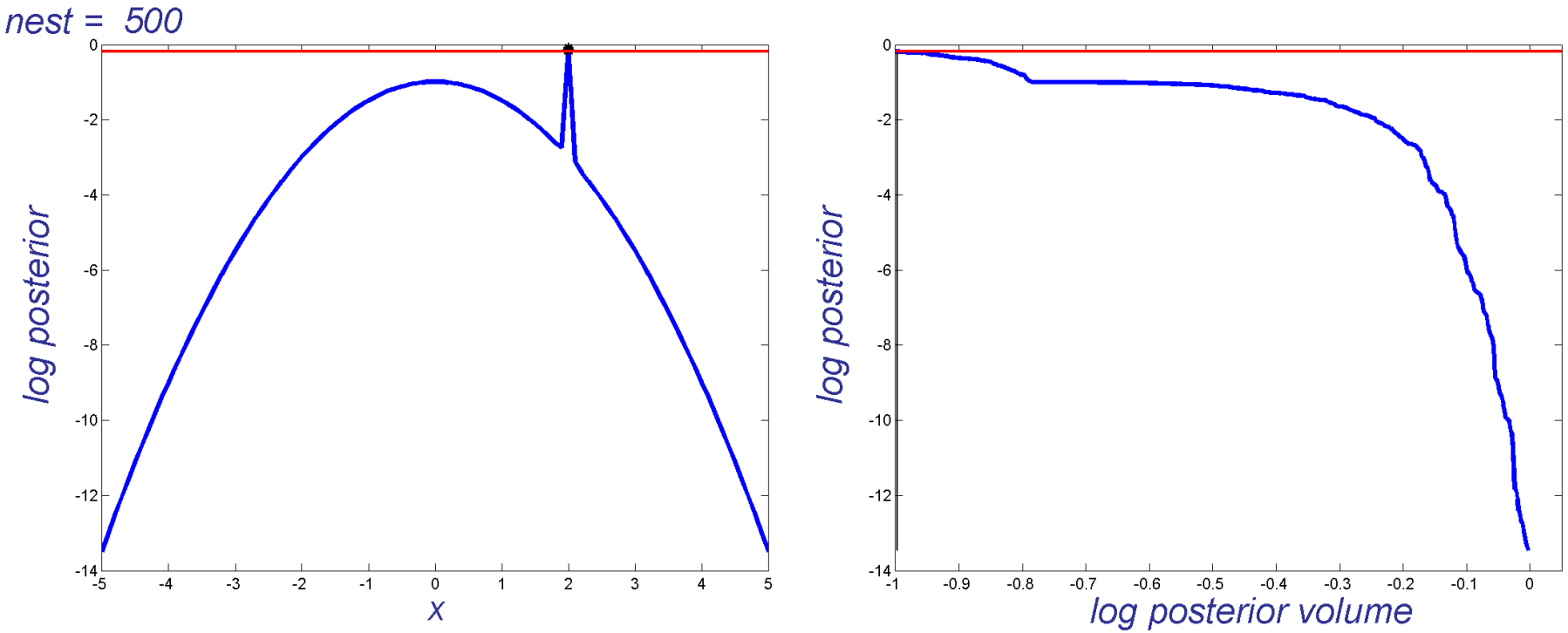
Typical Behavior

In a well-behaved problem the log likelihood increases monotonically with a monotonically decreasing volume.

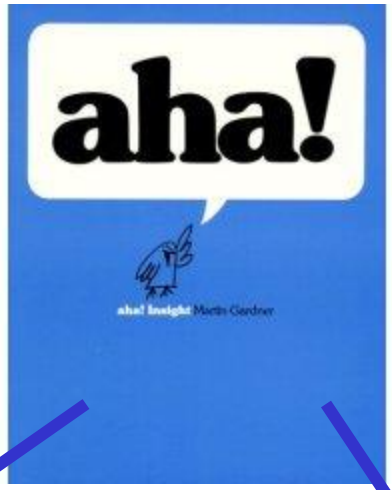
This can be viewed as a density of states.

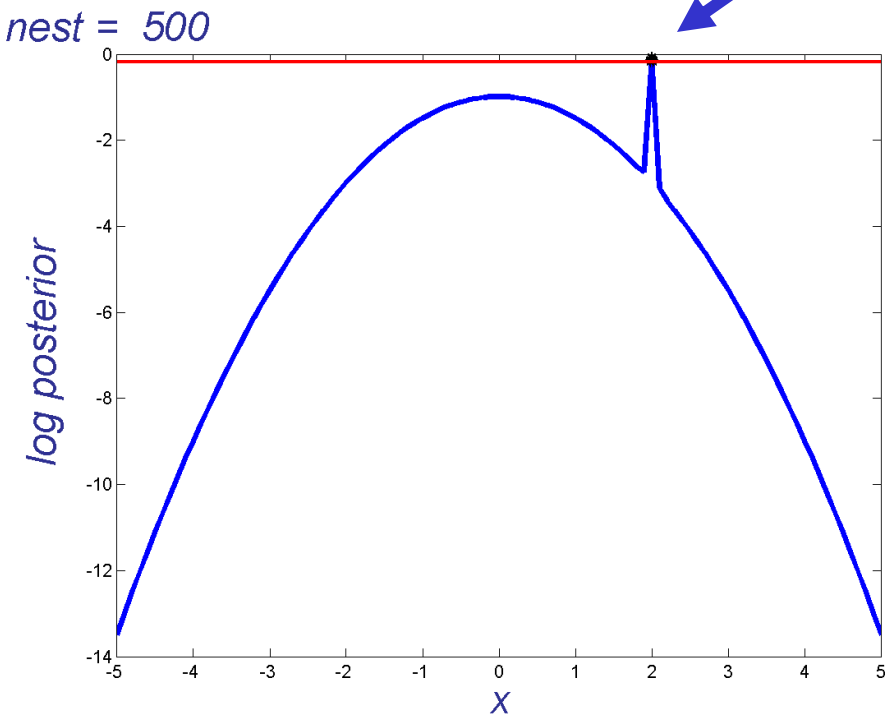


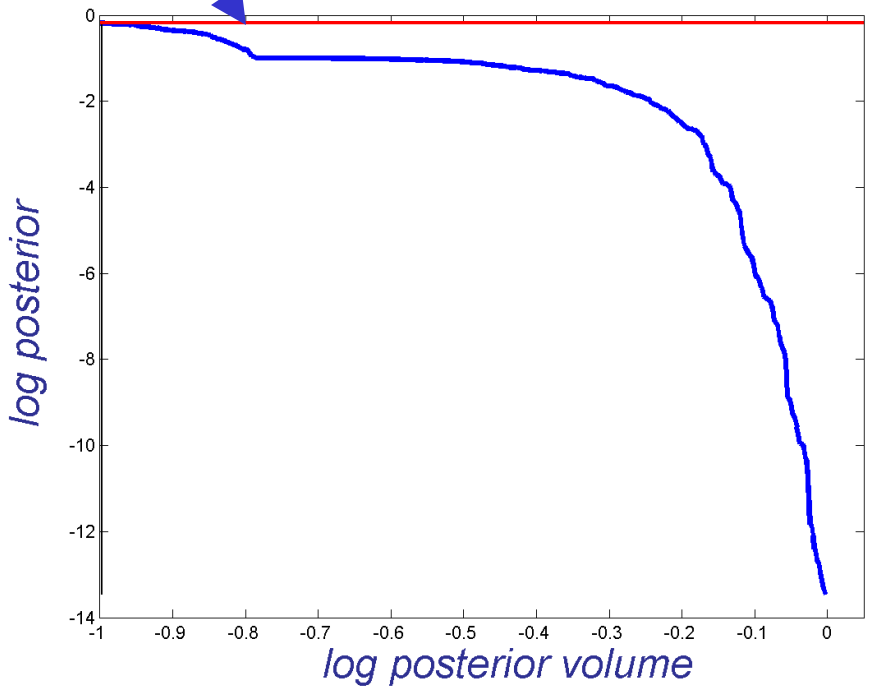
Typical Behavior



Problems with peaks on peaks are mildly difficult.
In statistical mechanics, one interprets these inflections as phase transitions.


aha!
aha! Insight Martin Gardner


nest = 500
log posterior
x
0
-2
-4
-6
-8
-10
-12
-14
-5
-4
-3
-2
-1
0
1
2
3
4
5


log posterior
log posterior volume
0
-2
-4
-6
-8
-10
-12
-14
-1
-0.9
-0.8
-0.7
-0.6
-0.5
-0.4
-0.3
-0.2
-0.1
0

Conclusions

Algorithm Design is Problem-Solving!

Effective problem-solving relies both on exploration and educated guesses.

Choosing samples randomly encourages exploration.

Including samples recommended by heuristics incorporates educated guesses.

Such a union is still fully-Bayesian!

Thank You for Your Kind Attention

This work is supported in part by:
NASA Applied Information Systems Research Program (AISPR)
and by
NASA ESTO Applied Information Systems Technology (AIST)