

Why things don't work

—

On the extended Smale's 9th and 18th problems (the limits of AI) and methodological barriers

Anders C. Hansen (Cambridge, UiO)

Joint work with:

A. Bastounis (Edinburgh) V. Vlacic (ETH)

Lisbon, January 2022

The impact of deep learning is unprecedented

The New Yorker quotes Geoffrey Hinton (April 2017):

”They should stop training radiologists now.”

How do we determine the foundations of DL?

Google's Ali Rahimi, winner of the Test-of-Time award 2017 (NIPS), "Machine learning has become alchemy. ... I would like to live in a society whose systems are built on top of verifiable, rigorous, thorough knowledge, and not on alchemy."



Yann LeCun

December 6 at 8:57am · 🌐



My take on [Ali Rahimi's](#) "Test of Time" award talk at NIPS.

Ali gave an entertaining and well-delivered talk. But I fundamentally disagree with the message.

The main message was, in essence, that the current practice in machine learning is akin to "alchemy" (his word).

It's insulting, yes. But never mind that: It's wrong!

*The Achilles heel of modern AI/DL: it is
universally unstable*

Adversarial attacks on medical machine learning

Samuel G. Finlayson¹, John D. Bowers², Joichi Ito³, Jonathan L. Zittrain², Andrew L. Beam⁴, Isaac S. Kohane¹

+ See all authors and affiliations

Science 22 Mar 2019:
Vol. 363, Issue 6433, pp. 1287-1289
DOI: 10.1126/science.aaw4399

Article

Figures & Data

Info & Metrics

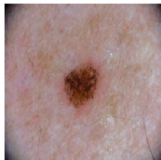
eLetters

 PDF

With public and academic attention increasingly focused on the new role of machine learning in the health information economy, an unusual and no-longer-esoteric category of vulnerabilities in machine-learning systems could prove important. These vulnerabilities allow a small, carefully designed change in how inputs are presented to a system to completely alter its output, causing it to confidently arrive at manifestly wrong conclusions. These advanced techniques to subvert otherwise-reliable machine-learning systems—so-called adversarial attacks—have, to date, been of interest primarily to computer science researchers (1). However, the landscape of often-competing interests within health care, and billions of dollars at stake in systems' outputs, implies considerable problems. We outline motivations that various players in the health care system may have to use adversarial attacks and begin a discussion of what to do about them. Far from discouraging continued innovation with medical machine learning, we call for active engagement of medical, technical, legal, and ethical experts in pursuit of efficient, broadly available, and effective health care that machine learning will enable.

Instabilities in classification/decision problems

Original image



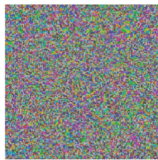
Dermoscopic image of a benign melanocytic nevus, along with the diagnostic probability computed by a deep neural network.



Diagnosis: Benign

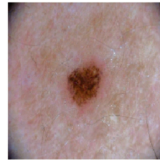


Adversarial noise



Perturbation computed by a common adversarial attack technique. See (7) for details.

Adversarial example



Combined image of nevus and attack perturbation and the diagnostic probabilities from the same deep neural network.



Diagnosis: Malignant

Adversarial rotation (8)



Adversarial text substitution (9)

The patient has a history of **back pain** and chronic **alcohol abuse** and more recently has been seen in several...

Opioid abuse risk: High

277.7 Metabolic syndrome
429.9 Heart disease, unspecified
278.00 Obesity, unspecified

Reimbursement: Denied

The patient has a history of **lumbago** and chronic **alcohol dependence** and more recently has been seen in several...

Opioid abuse risk: Low

401.0 Benign essential hypertension
272.0 Hypercholesterolemia
272.2 Hyperglyceridemia
429.9 Heart disease, unspecified
278.00 Obesity, unspecified

Reimbursement: Approved

Adversarial coding (13)

FDA NEWS RELEASE

FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems

 Share

 Tweet

 LinkedIn

 Email

 Print

For Immediate Release: April 11, 2018

[Español](#)

The U.S. Food and Drug Administration today permitted marketing of the first medical device to use artificial intelligence to detect greater than a mild level of the eye disease diabetic retinopathy in adults who have diabetes.

Diabetic retinopathy occurs when high levels of blood sugar lead to damage in the blood vessels of the retina, the light-sensitive tissue in the back of the eye. Diabetic retinopathy is the most common cause of vision loss among the more than 30 million Americans living with diabetes and the leading cause of vision impairment and blindness among working-age adults.

Concern expressed in Science Magazine

companies are beginning to require other data types such as imaging and text to prove that claims are valid. As they do so, other styles of adversarial attacks may be used as well to try to continue to dodge detection.

For example, if an insurance company requires that an image from a mole be run through a melanoma classifier before approving reimbursement for an excision, fraudsters may at first be inclined to submit moles from different patients to achieve approval. If insurance companies then begin utilizing human audits or technical tests to try to ensure that the images are coming from the correct patient, the next round would be to move to full adversarial attacks with imperceptible alterations, such as in the top figure. Simpler techniques such as the rotation in the bottom figure could constitute an ethical gray zone—given that a dermatologist could, in theory, hold the camera at any angle.

Potential applications of adversarial attacks in the

A PATH FORWARD

An essential question remains: when and how to intervene. Here, the early history of the internet offers a lesson. The approach to network architecture introduced at the advent of the internet was centered around the deferral of problems. In their essential 1984 paper, Saltzer *et al.* describe a design ethos whereby problems are to be solved at the end points of a network by users rather than preemptively within the architecture of the network itself (14). There is frequently an advantage (in terms of simplicity, flexibility, and scalability) to leaving future problems unsolved until their time has come. Another description for this is the “procrastination principle” (15).

The procrastination principle frames a difficult question: Should the adversarial-examples problem in health care

systems be addressed now—in the early, uncertain days of medical AI algorithms—or later, when algorithms and the protocols governing their use have been firmed up? At best, acting now could avoid

original hash to that of the data fed through a targeted algorithm would allow investigators to determine if that data had been tampered with or changed after acquisition. Such an intervention would rely on a health IT infrastructure capable of supporting the capture and secure storage of these hashes. But as a strictly regulated field with a focus on accountability and standards of procedure, health care may be very well suited to such adaptations.

The coalescence of strong motives to manipulate algorithms and the rapid proliferation of algorithms vulnerable to manipulation makes health care a plausible ground zero for the emergence of adversarial examples into real-world practice. As adversarial examples emerge across a range of domains, we will have to make choices again and again about whether and how to intervene early at the risk of stifling development, and how to balance the promises of ubiquitous machine learning against liabilities imposed by these emerging vulnerabilities. And the stakes will remain high—autonomous vehicles and AI-driven weapons systems will be just as

“An essential question remains: when and how to intervene.”

Deep Fool was established at EPFL in order to study the stability of neural networks.



Alhussein Fawzi, Seyed-Mohsen Moosavi-Dezfooli,
and Pascal Frossard

The Robustness of Deep Networks

A geometrical perspective

Deep Fool: Universal perturbations

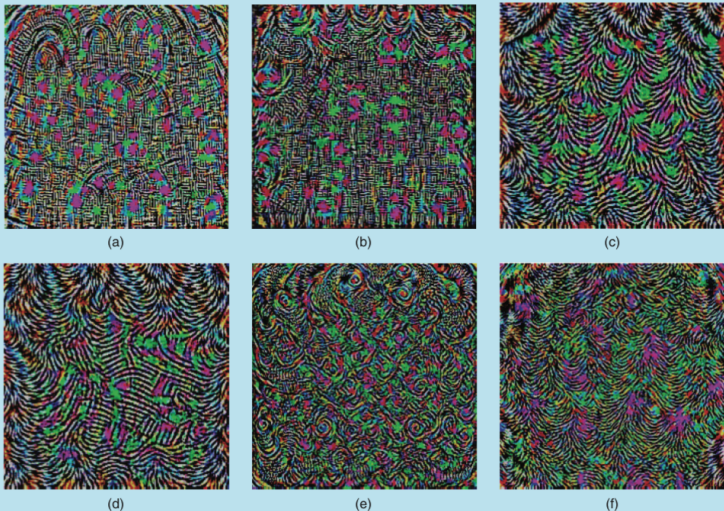



FIGURE 3. Universal perturbations computed for different deep neural network architectures. The pixel values are scaled for visibility. (a) CaffeNet, (b) VGG-F, (c) VGG-16, (d) VGG-19, (e) GoogLeNet, and (f) ResNet-152.

Deep Fool: Examples




FIGURE 4. Examples of natural images perturbed with the universal perturbation and their corresponding estimated labels with GoogLeNet. (a)–(h) Images belonging to the ILSVRC 2012 validation set. (i)–(l) Personal images captured by a mobile phone camera. (Figure used courtesy of [22].)

DL is unstable in inverse problems

Submit About Contact Journal Club Subscribe Log in 

PNAS Proceedings of the National Academy of Sciences of the United States of America

Keyword, Author, or DOI  Advanced Search

Home **Articles** Front Matter News Podcasts Authors

NEW RESEARCH IN


PHYSICAL SCIENCES

On instabilities of deep learning in image reconstruction and the potential costs of AI

Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C. Hansen

PNAS first published May 11, 2020 <https://doi.org/10.1073/pnas.1907377117>

Edited by David L. Donoho, Stanford University, Stanford, CA, and approved March 12, 2020 (received for review June 4, 2019)

 **Sign up for Article Alerts**

The press reports on instabilities

ARTIFICIAL INTELLIGENCE

Don't trust deep-learning algos to touch up medical scans: Boffins warn 'highly unstable' tech leads to

14 May 2020

Research finds shortcomings of AI techniques in medical imaging

By GC Staff Writer

Medical images reconstructed using artificial intelligence (AI) techniques are unreliable, according to recent research by an international team of mathematicians. The team found that deep learning tools that create

Images (including small circular perforations in) and are normally associated with repair networks (DSD, DSDs) are often used without such methods. (Source: Proceedings of the National Academy of Sciences 10.1073/pnas.200777117)

Medical images reconstructed using artificial intelligence (AI) techniques are unreliable, according to recent research by an international team of mathematicians. The team found that deep learning tools that create

myScience

AI techniques in medical imaging may lead to incorrect diagnoses

May 13, 2020
Whitney J. Pal

AI-Based Image Reconstruction Techniques Could Lead to Misdiagnosis

May 13, 2020
Whitney J. Pal

Relevant To: Algorithms can lead to new tests.

AI techniques in medical imaging may lead to positives and false negatives

Download PDF Copy

Reviewed by Emily Henderson, B.Sc.

Machine learning and AI are highly unstable in medical image reconstruction, and may in false negatives, a new study suggests.

A team of researchers, led by the University of Cambridge and Simon Fraser University, used medical image reconstruction algorithms based on AI and deep learning, and found in myriad artifacts, or unwanted alterations in the data, among other major errors in the were typically not present in non-AI based imaging techniques.

The phenomenon was widespread across different types of artificial neural networks, as well as different types of AI-based image reconstruction algorithms. The researchers caution that relying on AI-based image reconstruction could determine treatment and ultimately do harm to patients. Their re

HealthCareBusiness

New research provides a sobering reality check to AI optimists

By John S. Eicher, Senior Reporter | May 14, 2020

For all the revolutionary promise that artificial intelligence holds

RADIOLOGY BUSINESS

FOR LEADERS NAVIGATING VALUE-BASED CARE

Artificial intelligence-based image reconstruction may lead to incorrect diagnoses

Why has the instability problem in DL not been solved?

Hilbert's program on the foundations of mathematics

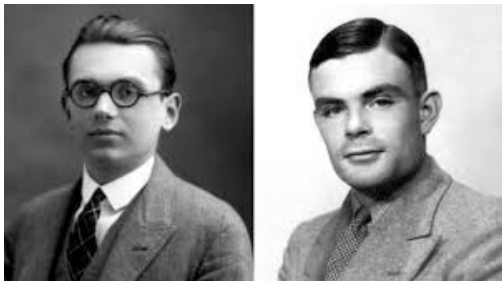


Hilbert's 10th problem:

Find an algorithm to determine whether a given polynomial Diophantine equation with integer coefficients has an integer solution.

(Source: Wikipedia)

Hilbert's program on the foundations of mathematics



Gödel and Turing turned Hilbert's optimism upside down by showing how there are true statements in mathematics that cannot be proven and that there are problems that cannot be computed by an algorithm.

Hilbert's 10th problem (Solution): No algorithm exists!

Smale's 18th problem:

What are the limits of artificial intelligence?

A program determining the foundations/limits of DL and AI is needed.

We need to understand the methodological boundaries.

Neural networks

Let $\mathcal{NN}_{N,L,d}$, with $N = (N_L, N_{L-1}, \dots, N_1, N_0 = d)$ denote the set of all L -layer neural networks. That is, all mappings $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^{N_L}$ of the form

$$\phi(x) = W_L(\rho(W_{L-1}(\rho(\dots\rho(W_1(x)))))), \quad x \in \mathbb{R}^d.$$

$$W_j y = A_j y - b_j, \quad A_j \in \mathbb{R}^{N_j \times N_{j-1}}, \quad b_j \in \mathbb{R}^{N_j}$$

$$\rho : \mathbb{R} \rightarrow \mathbb{R}$$

is some non-linear function that acts pointwise on a vector.

Classification Problems

Consider

$$f : [0, 1]^d \rightarrow \{0, 1\}. \quad (1)$$

We assume that the cost function \mathcal{R} is an element of

$$\mathcal{CF}_r = \{\mathcal{R} : \mathbb{R}^r \times \mathbb{R}^r \rightarrow \mathbb{R}_+ \cup \{\infty\} \mid \mathcal{R}(v, w) = 0 \text{ iff } v = w\}. \quad (2)$$

As we will discuss the stability of neural networks, we introduce the idea of *well-separated and stable sets*. Specifically, we define the family of well-separated and stable sets \mathcal{S}_δ^f as follows:

$$\mathcal{S}_\delta^f = \left\{ \{x^1, \dots, x^m\} \subset [0, 1]^d \mid m \in \mathbb{N}, \right. \\ \left. \min_{x^i \neq x^j} \|x^i - x^j\|_\infty \geq 2\delta, f(x^j + y) = f(x^j) \text{ for } \|y\|_\infty < \delta \right\}.$$

Why DL is unstable in classification

Theorem 1 (Bastounis, H, Vlacic)

There is an uncountable collection \mathcal{C}_1 of classification functions $f : [0, 1]^d \rightarrow \{0, 1\}$, – with fixed $d \geq 2$ – and a constant $C > 0$ such that the following holds. For every $f \in \mathcal{C}_1$, any norm $\|\cdot\|$ and every $\epsilon > 0$, there is an uncountable family \mathcal{C}_2 of probability distributions on $[0, 1]^d$ so that for any $\mathcal{D} \in \mathcal{C}_2$, any neural network dimensions $N = (N_L = 1, N_{L-1}, \dots, N_1, N_0 = d)$ with $L \geq 2$, any $p \in (0, 1)$, any positive integers q, r, s with

$$r + s \geq C \max \{ p^{-3}, q^{3/2} [(N_1 + 1) \cdots (N_{L-1} + 1)]^{3/2} \}, \quad (3)$$

any training data $\mathcal{T} = \{x^1, \dots, x^r\}$ and validation data $\mathcal{V} = \{y^1, \dots, y^s\}$, where the x^j and y^j are drawn independently at random from \mathcal{D} , the following happens with probability exceeding $1 - p$.

- (i) **(Success – great generalisability)**. We have $\mathcal{T}, \mathcal{V} \in \mathcal{S}_{\epsilon((r \vee s)/p)}^f$, where $\epsilon(n) = (Cn)^{-4}$, and, for every $\mathcal{R} \in \mathcal{CF}_r$, there exists a ϕ such that

$$\phi \in \operatorname{argmin}_{\varphi \in \mathcal{NN}_{N,L}} \mathcal{R}(\{\varphi(x^j)\}_{j=1}^r, \{f(x^j)\}_{j=1}^r) \quad (4)$$

and

$$\phi(x) = f(x) \quad \forall x \in \mathcal{T} \cup \mathcal{V}. \quad (5)$$

- (ii) **(Any successful NN in $\mathcal{NN}_{N,L}$ – regardless of architecture – becomes universally unstable)**. Yet, for any $\hat{\phi} \in \mathcal{NN}_{N,L}$ (and thus, in particular, for $\hat{\phi} = \phi$) and any monotonic $g : \mathbb{R} \rightarrow \mathbb{R}$, there is a subset $\tilde{\mathcal{T}} \subset \mathcal{T} \cup \mathcal{V}$ of the combined training and validation set of size $|\tilde{\mathcal{T}}| \geq q$, such that there exist uncountably many universal adversarial perturbations $\eta \in \mathbb{R}^d$ so that for each $x \in \tilde{\mathcal{T}}$ we have

$$|g \circ \hat{\phi}(x + \eta) - f(x + \eta)| \geq 1/2, \quad \|\eta\| < \epsilon, \quad |\operatorname{supp}(\eta)| \leq 2. \quad (6)$$

- (iii) **(Other stable and accurate NNs exist)**. However, there exists a stable and accurate neural network ψ that satisfies $\psi(x) = f(x)$ for all $x \in \mathcal{B}_\epsilon^\infty(\mathcal{T} \cup \mathcal{V})$, when $\epsilon \leq \epsilon((r \vee s)/p)$.

Interpreting Theorem 1

- (i) No training model where the dimensions of the NNs are fixed can cure instability.
- (ii) Variable dimensions are necessary for stability and accuracy.
- (iii) There are accurate and stable NNs, but DL methods do not find them.
- (iv) Why instability? – Unstable correlating features are picked up by the trained NN.

[Submitted on 13 Sep 2021]

The mathematics of adversarial attacks in AI -- Why deep learning is unstable despite the existence of stable neural networks

Alexander Bastounis, Anders C Hansen, Verner Vlačić



Newsjournal of the Society for Industrial and Applied Mathematics

snews.siam.org

Volume 54/ Issue 8
October 2021

Deep Learning: What Could Go Wrong?

By Alexander Bastounis, Anders C. Hansen, Desmond J. Higham, Ivan Y. Tyukin, and Verner Vlačić

In a field of research where algorithms can misinterpret stop signs as speed limit signs with the addition of minimal graffiti [3], many commentators are wondering whether current artificial intelligence (AI) solutions are sufficiently robust, resilient, and trustworthy. How can the research community quantify and address such issues?

Many empirical approaches investigate the generation of *adversarial attacks*: small, deliberate perturbations to an input that cause dramatic changes in a system's output. Changes that are essentially imperceptible to the human eye may alter predictions in the field of image classification, which has implications in many high-stakes and safety-critical settings. The rise of algorithms that construct attacks—and heuristic techniques that identify or guard against them—has led to a version of conflict escalation wherein attack and defense strategies become increasingly ingenious [10].

These issues concern the *conditioning* of the underlying problem and *stability* of the algorithms in use. Recent research has utilized mathematical tools—totally born from numerical analysis, applied prob-

ability, and high-dimensional geometry—to shed light on this field. However, many open problems remain.

Inevitability of Attacks

A simple but powerful example helps illustrate the way in which adversarial attacks may arise [4]. Imagine that we have data points $x \in \mathbb{R}^n$, which may be pixels in an image that are stacked into a vector. Suppose that the images come from two categories: cats and dogs. Given some fixed vector $w \in \mathbb{R}^n$ and scalar α , a linear classifier will classify a new point x as a cat or dog depending upon whether $w^T x$ is less than or greater than α . Here, w would be constructed according to some sort of best-fit procedure on a training set of labeled images.

If we perturb x to $x + \Delta x$, the output from the linear classifier changes by $w^T \Delta x$. Suppose that we are able to perturb each pixel in the input image by at most ϵ ; that is, $\|\Delta x\|_1 \leq \epsilon$. If we know the vector w , we can then increase the classifier's output as much as possible by selecting a perturbation with every component $\Delta x_i = \epsilon \text{sign}(w_i)$; similarly, the maximum decrease occurs with $\Delta x_i = -\epsilon \text{sign}(w_i)$. In this way, we can alter the output by $\|\epsilon\|_1$. If m is the average size of components in w , then a *per pixel* change of ϵ can

lead to a change of $n\epsilon m$ in the classifier output. The classifier is vulnerable to this type of attack when the dimension n —the number of pixels—is large.

This simple illustration highlights a number of issues. First, any smooth map can be well described locally by a first-order (linear) Taylor series approximation, meaning that

this type of attack is relevant whenever the attacker has access to gradient information. In the *black box* setting where attackers can only choose inputs and observe the corresponding outputs, they could use finite difference approximations to build up the necessary gradient information for the

See Deep Learning on page 7

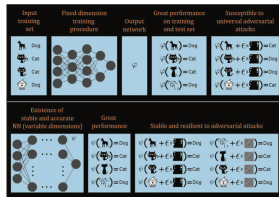
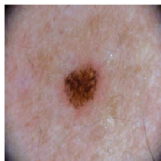


Figure 1. A paradox of instabilities in deep learning, as formalized in [2]. Trained neural networks (NNs) of fixed dimension are unstable, but stable and accurate NNs of variable dimension exist. Figure courtesy of the authors.

Can we make AI trustworthy?

Instabilities in classification/decision problems

Original image



Dermoscopic image of a benign melanocytic nevus, along with the diagnostic probability computed by a deep neural network.



Diagnosis: Benign



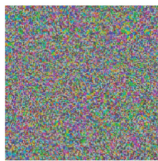
The patient has a history of **back pain** and chronic **alcohol abuse** and more recently has been seen in several...

Opioid abuse risk: High

277.7 Metabolic syndrome
429.9 Heart disease, unspecified
278.00 Obesity, unspecified

Reimbursement: Denied

Adversarial noise



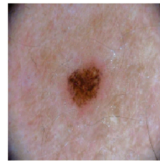
Perturbation computed by a common adversarial attack technique. See (7) for details.

Adversarial rotation (8)

Adversarial text substitution (9)

Adversarial coding (13)

Adversarial example



Combined image of nevus and attack perturbation and the diagnostic probabilities from the same deep neural network.



Diagnosis: Malignant



The patient has a history of **lumbago** and chronic **alcohol dependence** and more recently has been seen in several...

Opioid abuse risk: Low

401.0 Benign essential hypertension
272.0 Hypercholesterolemia
272.2 Hyperglyceridemia
429.9 Heart disease, unspecified
278.00 Obesity, unspecified

Reimbursement: Approved

European Commission's outline for a legal framework for AI:

"In the light of the recent advances in artificial intelligence (AI), the serious negative consequences of its use for EU citizens and organisations have led to multiple initiatives from the European Commission to set up the principles of a trustworthy and secure AI. Among the identified requirements, the concepts of robustness and explainability of AI systems have emerged as key elements for a future regulation of this technology."

– Europ. Comm. JCR Tech. Rep. (Jan 2020).

"On AI, trust is a must, not a nice to have. [...] The new AI regulation will make sure that Europeans can trust what AI has to offer. [...]"

High-risk AI systems will be subject to strict obligations before they can be put on the market: [requiring] High level of **robustness, security and accuracy.**"

— Europ. Comm. outline for legal AI (April 2021).

EUROPEAN COMMISSION

Robustness required in high-risk AI

High-risk: AI systems identified as high-risk include AI technology used in:

- **Critical infrastructures** (e.g. transport), that could put the life and health of citizens at risk;
- **Educational or vocational training**, that may determine the access to education and professional course of someone's life (e.g. scoring of exams);
- **Safety components of products** (e.g. AI application in robot-assisted surgery);
- **Employment, workers management and access to self-employment** (e.g. CV-sorting software for recruitment procedures);
- **Essential private and public services** (e.g. credit scoring denying citizens opportunity to obtain a loan);
- **Law enforcement** that may interfere with people's fundamental rights (e.g. evaluation of the reliability of evidence);
- **Migration, asylum and border control management** (e.g. verification of authenticity of travel documents);
- **Administration of justice and democratic processes** (e.g. applying the law to a concrete set of facts).

Do algorithms fail?

—

*... and can we determine when they are
wrong?*

The Problems:

We consider two concrete examples: the linear program

$$\min_{x \in \mathbb{R}^2} x_1 + x_2 \quad \text{subject to} \quad x_1 + (1 - \delta)x_2 = 1, \quad x_1, x_2 \geq 0, \quad (7)$$

where $\delta > 0$ is a parameter.

Testing MATLAB's linprog

+3	The solution is feasible with respect to the relative <code>ConstraintTolerance</code> tolerance, but is not feasible with respect to the absolute tolerance.
+1	Function converged to a solution <code>x</code> .
0	Number of iterations exceeded <code>options.MaxIterations</code> or solution time in seconds exceeded <code>options.MaxTime</code> .
-2	No feasible point was found.
-3	Problem is unbounded.
-4	NaN value was encountered during execution of the algorithm.
-5	Both primal and dual problems are infeasible.
-7	Search direction became too small. No further progress could be made.
-9	Solver lost feasibility.

Table: The `EXITFLAG` is used to verify the correctness of the solution. Possible values for the `EXITFLAG` output of `linprog` as well as their corresponding interpretations are displayed in this table. Note that a value of 1 indicates the correctness of the solution, whereas other values indicate various types of failure.

Testing MATLAB's linprog

δ	'dual-simplex'		'interior-point'		'interior-point-legacy'	
	Error	EXITFLAG	Error	EXITFLAG	Error	EXITFLAG
2^{-1}	0	1	0	1	$6.0 \cdot 10^{-12}$	1
2^{-15}	0	1	0	1	$3.0 \cdot 10^{-5}$	1
2^{-20}	0	1	0	1	$7.0 \cdot 10^{-7}$	1
2^{-24}	0	1	0	1	$7.1 \cdot 10^{-8}$	1
2^{-26}	1.4	1	1.4	1	$1.2 \cdot 10^{-1}$	1
2^{-28}	1.4	1	1.4	1	$4.6 \cdot 10^{-1}$	1
2^{-30}	1.4	1	1.4	1	$7.1 \cdot 10^{-1}$	1

Table: Testing the output of `linprog` applied to the problem in (10) for the three algorithms 'dual-simplex', 'interior-point' and 'interior-point-legacy'. The table shows the error $\|\hat{x} - \tilde{x}\|_{\ell^2}$ and the value of EXITFLAG, where \hat{x} is the true minimiser of (10) and \tilde{x} is the computed approximate minimiser. Note that machine epsilon is $\epsilon_{\text{mach}} = 2^{-52}$.

Linear Programming

Let

$$z \in \underset{x}{\operatorname{argmin}} \langle x, c \rangle \text{ such that } Ax = y, \quad x \geq 0,$$

where $A \in \mathbb{R}^{m \times N}$, $y \in \mathbb{R}^m$, $c \in \mathbb{R}^N$.

Input: A, y and c .

Problem: Find an algorithm that computes a minimiser z .

In mathematics of information one wants minimisers, not the objective function.

Smale's 9th Problem

"Is there a polynomial time algorithm over the real numbers which decides the feasibility of the linear system of inequalities $Ax \geq y$, and if so, outputs such an x ?"

— S. Smale (*Problem 9 from the list of mathematical problems for the 21st century*)

"But real number computations and algorithms which work only in exact arithmetic can offer only limited understanding. Models which process approximate inputs and which permit round-off computations are called for."

— *S. Smale (from the list of mathematical problems for the 21st century)*

Discrete vs continuous

We must be able to handle inaccurate input as $\sqrt{2}$, $\cos(3)$ or $e^{2\pi i/5}$ will never be represented exactly.

Also, when running floating point arithmetic even $1/3$ is approximated by a base-2 number.

LP in P (NY Times 1979)



LP in P proved by L. Khachiyan – based on work by N. Shor, D. Yudin, A. Nemirovski.

Karmarkar's algorithm

From Wikipedia, the free encyclopedia

Karmarkar's algorithm is an [algorithm](#) introduced by [Narendra Karmarkar](#) in 1984 for solving [linear programming](#) problems. It was the first reasonably efficient algorithm that solves these problems in [polynomial time](#). The [ellipsoid method](#) is also polynomial time but proved to be inefficient in practice.

Denoting n as the number of variables and L as the number of bits of input to the algorithm, Karmarkar's algorithm requires $O(n^{3.5}L)$ operations on $O(L)$ digit numbers, as compared to $O(n^6L)$ such operations for the ellipsoid algorithm. The runtime of Karmarkar's algorithm is thus

$$O(n^{3.5}L^2 \cdot \log L \cdot \log \log L)$$

using [FFT-based multiplication](#) (see [Big O notation](#)).

Karmarkar's algorithm falls within the class of [interior point methods](#): the current guess for the solution does not follow the boundary of the [feasible set](#) as in the [simplex method](#), but it moves through the interior of the feasible region, improving the approximation of the optimal solution by a definite fraction with every iteration, and converging to an optimal solution with rational data.^[1]

The Extended Model

Given a domain $\Omega \subset \mathbb{R}^n$ of inputs, the algorithm cannot access $\iota \in \Omega$, but rather, for any $k \in \mathbb{N}$, it can call the oracle \mathcal{O} to obtain $\tilde{\iota} = \mathcal{O}(\iota, k) \in \mathbb{R}^n$ satisfying

$$\|\mathcal{O}(\iota, k) - \iota\|_{\infty} \leq 2^{-k}, \quad \forall \iota \in \Omega, \forall k \in \mathbb{N}, \quad (8)$$

and the time cost of accessing $\mathcal{O}(\iota, k)$ is polynomial in k .

The extended model with inexact input is considered in many areas of mathematics including in the work of E. Bishop; M. Braverman & S. Cook; F. Cucker & S. Smale; C. Fefferman & B. Klartag; K. Ko and L. Lovász.

Key Problems in Mathematics of Information (38)

Linear Programming

$$z \in \underset{x}{\operatorname{argmin}} \langle x, c \rangle \text{ such that } Ax = y, \quad x \geq 0,$$

Semidefinite Programming

$$Z \in \underset{X \in \mathbb{S}^n}{\operatorname{argmin}} \langle C, X \rangle_{\mathbb{S}^n} \text{ such that } \langle A_k, X \rangle_{\mathbb{S}^n} = b_k, \quad X \succeq 0, \quad k \leq m$$

Basis Pursuit

$$z \in \underset{x}{\operatorname{argmin}} \mathcal{J}(x) \text{ such that } \|Ax - y\| \leq \delta, \quad \delta \geq 0,$$

Unconstrained Lasso

$$z \in \underset{x}{\operatorname{argmin}} \|Ax - y\|_2^2 + \lambda \mathcal{J}(x), \quad \lambda > 0,$$

Constrained Lasso

$$z \in \underset{x}{\operatorname{argmin}} \|Ax - y\|_2 \text{ such that } \mathcal{J}(x) \leq \tau, \quad \tau > 0$$

where $A \in \mathbb{C}^{m \times N}$, $y \in \mathbb{C}^m$ and $\mathcal{J}(x) = \|x\|_1$ or $\mathcal{J}(x) = \|x\|_{\text{TV}}$.

Key Papers in Mathematics of Information

L. I. Rudin, S. Osher, and E. Fatemi. 'Nonlinear total variation based noise removal algorithms',
Physica D: Nonlinear Phenomena (1992).

R. Tibshirani. 'Regression shrinkage and selection via the lasso',
Journal of the Royal Statistical Society, Series B (1996).

E. J. Candès, J. Romberg, and T. Tao. 'Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information',
IEEE Trans. Inform. Theory (2006).

D. L. Donoho. 'Compressed sensing',
IEEE Trans. Inform. Theory (2006).

These papers are cited all together about 100,000 times.

The Extended Smale's 9th problem

The Extended Smale's 9th Problem

Problem 2 (**The extended Smale's 9th problem**)

Given any of the problems in (38), represented by the solution map Ξ mapping a class of inputs Ω into a metric space $(\mathcal{M}, d_{\mathcal{M}})$, is there an algorithm which decides the feasibility of the problem, and if so, produces an output that is correct up to K digits (where the error is measured via $\text{dist}_{\mathcal{M}}$) and whose computational cost is bounded by a polynomial in K and the number of variables n ?

The extended Smale's 9th – computing solutions

Theorem 3 (The extended Smale's 9th – computing solutions)

Let Ξ denote the solution map to any of the problems (38) with the regularisation parameters satisfying $\delta \in [0, 1]$, $\lambda \in (0, 1/3]$, and $\tau \in [1/2, 2]$ (and additionally being rational in the Turing case) and consider the $\|\cdot\|_p$ -norm for measuring the error, for an arbitrary $p \in [1, \infty]$. Let $K > 2$ be an integer. There exists a class Ω of feasible inputs so that we have the following.

- (i) No algorithm can produce K correct digits on each input in Ω . Moreover, for any $p > \frac{1}{2}$, no randomised algorithm can produce K correct digits with probability greater than or equal to p on each input in Ω .
- (ii) If we allow randomised algorithms with a non-zero probability of not halting (not producing an output), then, for any $p > \frac{2}{3}$, no such algorithm can produce K correct digits with probability greater than or equal to p on each input in Ω . However, there does exist such an algorithm that can produce K correct digits on each input in Ω with probability $2/3$.
- (iii) There does exist an algorithm (a Turing or a BSS machine) that produces $K - 1$ correct digits for all inputs in Ω . However, any such algorithm will need an arbitrarily long time to achieve this. In particular, for any fixed dimensions m, N , any $T > 0$, and any algorithm Γ , there exists an input $\iota \in \Omega_{m,N}$ such that either Γ on input ι does not produce $K - 1$ correct digits for $\Xi(\iota)$ or the runtime of Γ on ι exceeds T . Moreover, for any randomised algorithm Γ^{ran} and $p < 1/2$ there exists an input $\iota \in \Omega_{m,N}$ such that

$\mathbb{P}(\Gamma^{\text{ran}}(\iota) \text{ does not produce } K - 1 \text{ correct digits for } \Xi(\iota) \text{ or the runtime of } \Gamma \text{ on } \iota \text{ exceeds } T) > p$.

- (iv) There exists a polynomial $\text{pol} : \mathbb{R} \rightarrow \mathbb{R}$, as well as a Turing machine and a BSS machine that both produce $K - 2$ correct digits for all inputs in Ω , so that the number of arithmetic operations for both machines is bounded by $\text{pol}(n)$, where $n = m + mN$ is the number of variables, and the number of digits required from the oracle (8) is bounded by $\text{pol}(\log(n))$. Moreover, the space complexity of the Turing machine is bounded by $\text{pol}(n)$.

The theorem is valid with bounded condition numbers

Condition of a matrix: $\text{Cond}(A) = \|A\| \|A^{-1}\|$.

Condition of the mapping $\Xi : \Omega \subset \mathbb{C}^n \rightarrow \mathbb{C}^m$, linear or non-linear, is often given by

$$\text{Cond}(\Xi) = \sup_{x \in \Omega} \lim_{\epsilon \rightarrow 0^+} \sup_{\substack{x+z \in \Omega \\ 0 < \|z\| \leq \epsilon}} \frac{\text{dist}(\Xi(x+z), \Xi(x))}{\|z\|}.$$

Feasibility condition number. Define

$$\rho(A, y) = \sup\{\delta \mid \|\tilde{A}\|, \|\tilde{y}\| \leq \delta \Rightarrow (A + \tilde{A}, y + \tilde{y}) \in \Omega \text{ are feasible}\},$$

and this yields the Feasibility Primal (FP) condition number

$$C_{\text{FP}}(A, y) := \frac{\max(\|A\|, \|y\|)}{\rho(A, y)}.$$

The results in the theorem are valid with uniform bounds on the condition numbers and input.

Can the 'exit flag' be computed?

Can the 'exit flag' be computed?

Problem 4 (Can the 'exit flag' be computed?)

Consider an algorithm designed to compute any of the problems (38). Suppose that the algorithm should produce K correct digits. Can we compute the 'exit flag' for this algorithm, i.e., the function taking on the value 1 if the algorithm succeeds in producing K correct digits, and 0 otherwise?

Given $\alpha > 0$ and make the following assumption on the algorithm:

$$\text{dist}_{\mathcal{M}}(\Gamma(\iota), \Xi(\Omega)) < \alpha \text{ for all } \iota \in \Omega. \quad (9)$$

Impossibility of computing the 'exit flag'

Theorem 5 (Impossibility of computing the 'exit flag')

Let Ξ denote the solution map to any of the problems (38) with the regularisation parameters satisfying $\delta \in [0, 1]$, $\lambda \in (0, 1/3]$, and $\tau \in [1/2, 2]$ (and additionally being rational in the Turing case) and consider the $\|\cdot\|_p$ -norm for measuring the error, for an arbitrary $p \in [1, \infty]$. Let $K \in \mathbb{N}$ and fix real α and ω so that $0 < \alpha \leq \omega < 10^{-K}$. Then, for any fixed dimensions $N > m \geq 4$, there exists a class of inputs Ω for Ξ such that, if Γ is an algorithm satisfying (9) with parameter α for the computational problem of approximating Ξ with K correct digits, then we have the following.

- (i) No algorithm, even randomised with access to an exact solution oracle of precision ω , can compute the exit flag of Γ (with probability exceeding $p > 1/2$ in the randomised case).
- (ii) If we allow randomised algorithms with non-zero probability of not halting (producing an output), then no such algorithm, even with access to an exact solution oracle of precision ω , can compute the exit flag of Γ with probability exceeding $p > 1/2$.
- (iii) The problem of computing the exit flag of Γ is strictly harder than computing K correct digits of Ξ in the following sense: if one is given the exit flag as an oracle then it is possible to construct an algorithm that computes K correct digits of Ξ . However, if one is instead given an oracle providing a K -digit approximation to Ξ , then it is still not possible to compute the exit flag of Γ .
- (iv) For linear programming and basis pursuit, however, there exists a class of inputs $\Omega^\# \neq \Omega$ such that no algorithm, even randomised with non-zero probability of not halting, can compute the exit flag of Γ (with probability exceeding $p > 1/2$ in the randomised case), yet one can compute the exit flag with a deterministic algorithm with access to an exact solution oracle of precision ω .

A. Bastounis, A. C. Hansen, and V. Vlacic. 'The extended Smale's 9th problem – On computational barriers and paradoxes in estimation, regularisation, computer-assisted proofs and learning', *Preprint* (2021).

A. C. Hansen. 'On the Solvability Complexity Index, the n -pseudospectrum and Approximations of Spectra of Operators', *J. Amer. Math. Soc.* (2011).

Compressive Imaging: Structure, Sampling, Learning

(Cambridge University Press)



Ben Adcock & Anders C. Hansen

The Problems:

We consider two concrete examples: the linear program

$$\min_{x \in \mathbb{R}^2} x_1 + x_2 \quad \text{subject to} \quad x_1 + (1 - \delta)x_2 = 1, \quad x_1, x_2 \geq 0, \quad (10)$$

where $\delta > 0$ is a parameter, and the centred and standardised (so that the columns of the design matrix are normalised) Lasso problem

$$\min_{x \in \mathbb{R}^N} \frac{1}{m} \|A_\delta D_\delta x - y\|_2^2 + \lambda \|x\|_1, \quad (11)$$

where $m = 3$, $N = 2$, $\lambda \in (0, 1/\sqrt{3}]$,

$$A_\delta = \begin{pmatrix} \frac{1}{\sqrt{2}} - \delta & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} - \delta & -\frac{1}{\sqrt{2}} \\ 2\delta & 0 \end{pmatrix} \in \mathbb{R}^{3 \times 2}, \quad y = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{pmatrix}^T \in \mathbb{R}^3, \quad (12)$$

and D_δ is the unique diagonal matrix such that each column of $A_\delta D_\delta$ has norm \sqrt{m}

Random matrices – Non-computability is not rare

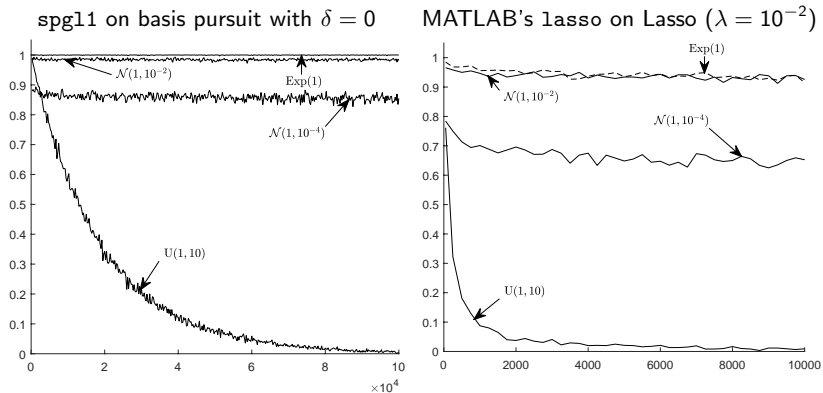


Figure: The vertical axis represents the success rate $\frac{\# \text{ of successes}}{\# \text{ of trials}}$. Success \Leftrightarrow computed solution is accurate to at least $K = 2$ digits ($\|\cdot\|_\infty$ norm). The horizontal axis shows the dimension N . In all cases, $A \in \mathbb{R}^{1 \times N}$ is iid – according to the distributions $U(a, b)$, $\text{Exp}(\nu)$ and $\mathcal{N}(\mu, \sigma^2)$ – in particular, the uniform distribution on $[a, b]$, the exponential distribution with parameter ν and the normal distribution with mean μ and variance σ .

Testing MATLAB's lasso

δ	Default settings			'RelTol' = ϵ_{mach}			'RelTol' = ϵ_{mach} 'MaxIter' = $\epsilon_{\text{mach}}^{-1}$		
	Error	Runtime	Warn	Error	Runtime	Warn	Error	Runtime	Warn
2^{-1}	$1 \cdot 10^{-16}$	< 0.01s	0	$1 \cdot 10^{-16}$	< 0.01s	0	$1 \cdot 10^{-16}$	< 0.01s	0
2^{-7}	0.68	< 0.01s	0	$2 \cdot 10^{-16}$	0.02s	0	$2 \cdot 10^{-16}$	0.02s	0
2^{-15}	1.17	< 0.01s	0	1.17	0.33s	1	$1 \cdot 10^{-11}$	1381.5s	0
2^{-20}	1.17	< 0.01s	0	1.17	0.33s	1	no output	> 12h	0
2^{-24}	1.17	< 0.01s	0	1.17	0.34s	1	no output	> 12h	0
2^{-26}	1.17	< 0.01s	0	1.17	0.34s	1	no output	> 12h	0
2^{-28}	1.17	< 0.01s	0	1.17	< 0.01s	0	1.17	< 0.01s	0
2^{-30}	1.17	< 0.01s	0	1.17	< 0.01s	0	1.17	< 0.01s	0

Table: The output of lasso applied to (11) with inputs as in (12) and $\lambda = 0.1$. The table shows the error $\|\hat{x} - \tilde{x}\|_{\ell^2}$ (where \hat{x} is the true minimiser and \tilde{x} is the computed minimiser), the CPU runtime, and a boolean value indicating whether a Warning was issued.