ON C.A. FEINSTEIN'S ELEGANT ARGUMENT THAT P NEQ NP

J.H. MARTEL

Introduction

This paper aims to provide the mathematical foundations for C.A. Feinstein's elegant argument that $\mathbf{P} \neq \mathbf{NP}$ [2, 1]. The primary claim in C.A. Feinstein's argument is that every deterministic algorithm which decides zero subset sums is necessarily random search "guess and check" over exponentially many subsets. This claim implies every deterministic algorithm deciding zero subset sums has exponential computational complexity. The difficulty in accepting this simple claim as sufficient proof that $\mathbf{P} \neq \mathbf{NP}$ is the idea that there are *yet possibly more undiscovered* algorithms. But we claim that the "space of algorithms" has an extremely simple structure, essentially consisting of random search. A key hypothesis of this paper is that random search is the universal deterministic algorithm. In simple terms, every deterministic algorithm reduces to a sequence of "guess and checks" conditioned on various subsets "look randomly here, and not there". This idea is made precise by a form of Tao-Szemeredi's Regularity lemma (see 1, 2 below). This hypothesis implies that random search on unstructured sets is algorithmically irreducible, and therefore has computational complexity equal to the Shannon entropy $\log \# X$. To illustrate with a simple example, if Alice asks Bob "What colour am I [Alice] thinking of?" then Bob has no deterministic strategy except "guess and check" over the colour space. Our point is that any more elaborate strategy, based on the conditioning of deterministic random variables, is a waste of effort. Bob has only one strategy to decide Alice's colour: random search.

1. Deciding Supports is the Universal Decision Problem

The next hypothesis is that deciding the supports of Borel-Radon measures is the universal decision problem in mathematics. This is the weak

Date: 2024-05-23.

star limit of the set membership decision problem $isElement(X, a)$ deciding "$a \in X$". This hypothesis implies every decision problem is equivalent to deciding whether a point x belongs to the support of a Borel measure μ. Formally $x \in spt(\mu)$ if and only if $\mu(U) > 0$ for every every open neighborhood U containing x. As motivated below, we are specifically interested in correlation measures between finite measure spaces. Correlations are measures π on $X \times Y$ with marginals satisfying $\text{proj}_X \pi = \sigma$ and $\text{proj}_Y \pi = \tau$. If $F : (X, \sigma) \to (Y, \tau)$ is a morphism, i.e. a measurable map $F : X \to Y$ pushing forward σ to $\tau = F\#\sigma$, then the graph of F is a correlation $(Id_X \times F)\#\sigma$ on $X \times Y$. Deciding the support of a graph is equivalent to deciding $y = F(x)$ almost everywhere. With respect to subset sum "knapsack" problems, the relevant measures are the images in $\mathbf{Z}$ of the subset sums $\epsilon(f)$ for all $f \in 2^X$.

2. Algorithms are Explicit Definitions

The computational complexity of a decision problem is defined variationally as the minimum complexity over all algorithms which deterministically decide the decision problem. But what is the space of all algorithms? We propose that the space of all deterministic algorithms is representable as spaces of correlation measures between large finite measure spaces. Algorithms are precisely and simply characterized as completely explicit definitions. Algorithms consist of explicit functions depending on no implicit expressions or symbols. The purpose of algorithms is to "bridge the gap" between implicit abstract definitions and explicit computation. Thus we define algorithms as explicit definitions. This hypothesis allows us to represent the space of all algorithms more concretely as consisting of correlations between explicit numerical inputs and outputs, and all decision problems consist in deciding the supports of these correlations.

3. Guess and Check is the Universal Deterministic Algorithm

A basic element of C.A. Feinstein's elegant argument is the claim that: every deterministic algorithm is a composition of random search "guess and check", and deterministic function evaluations which reduce the support of the random search, i.e. "Look randomly here, not there". Thus we consider "guess and check" the universal algorithm, even the *unique* algorithm. We motivate this hypothesis with some examples.

Deciding the Graph $y = f(x)$. Consider a function $f : X \to Y$ and the problem of deciding whether (x, y) belongs to the support of $\pi = \text{graph}(f)$ on $X \times Y$. There are two random variables by which we

can condition the event $(x, y) \in \text{spt}\pi$, these are the source and target variables x, y. If we first condition on $x \in X$, then we are looking for y such that $y = f(x)$. To find such events we simply need compute $f(x)$ explicitly and find $y \in Y$ such that $y = f(x)$. Alternatively if we condition on $y \in Y$, then we need find x such that $f(x) = y$, in otherwords we are searching $x \in X$ for elements in the fibre subset $f^{-1}(y) \subset X$. Therefore algorithms for deciding the support of the graph of f either consists of explicit function evaluations or fibre search "guess and check".

Euclid's Algorithm. Euclid's algorithm is a sequence of operations which returns the target $gcd(p, q)$ in a finite number of steps. The algorithm replaces the evaluation of $gcd(p, q)$ with the evaluation of gcd on smaller reduced inputs p', q'. This is based on the identity

$$gcd(p, q) = gcd(qs + r, q) = gcd(r, q)$$

which holds for any integers s, r satisfying $p = qs + r$. Euclid's algorithm replaces the gcd with smaller arguments, and terminating with a gcd of the form $gcd(d, f) := gcd(d, ds)$ which is evidently equal to d. This final step requires knowing that $f = d.s$. The efficiency of Euclid's algorithm is based on it's linear structure, which means gcd is correlated with many different inputs. Euclid's algorithm succeeds by using the identities implied by these linear structures.

Gauss Elimination. Gauss Elimination is the classic algorithm to solve systems of linear equations $Tx = y$. It consists of simple rules to reduce the support of linear systems to their simplest form, where the solution is immediately "readable" and explicit. This is analogous to Euclidean algorithm which reduces the arguments of $gcd(,)$ until we find a "trivial" argument and gcd is evaluated explicitly.

A review of further algorithms is consistent with the hypothesis that every algorithm is a composition of conditional random searches. This is formalized in Tao-Szemeredi's Regularity lemma discussed below.

4. Complexity and Entropy

Our next hypothesis is that C.E. Shannon's information entropy and computational complexity are the same quantity. Informally the complexity of a measure is understood as the "size" of it's support, i.e. the logarithm of the number of distinct memory states necessary for the computation. Shannon derived the analytic formula $\sum p \log p$ for entropy from the following axioms [3]:

(a) If X is finite homogeneous probability space, then

$$H(X) = \log \#X.$$

(b) (Composition Formula) If $Z = Y \circ X$ is a composition of finite probability spaces, then

$$H(Z) = H(Y \circ X) = H(X) + \int H(Y|X(x))\, dX(x).$$

Thus the total uncertainty of $Z = Y \circ X$ is equal to the uncertainty in the first choice $H(X)$ plus the X-average uncertainty of the second choice Y as conditioned by the first outcome $H(Y|X(x))$. These hypotheses imply the usual analytic Planck formula for entropy, i.e. if π is a correlation measure on $X \times Y$, then the total entropy of π is

$$H(\pi) = \int d\pi(x, y) \log d\pi(x, y).$$

The entropy is interpreted as the uncertainty of the pair (X, Y) as correlated by π.

Complexity is subadditive with respect to composition. This means that decomposing π into a very long composition $\pi = \cdots \circ \pi_i \circ \pi_{i-1} \circ \cdots$ does not necessarily reduce the complexity, but contrariwise tends to increase the total uncertainty.

Now we introduce some auxiliary definitions. The entropy of a measure X conditional on Y is defined by

$$H(X|Y) := H(X, Y) - H(Y)$$

where $H(X, Y)$ is the joint distribution of X, Y. The conditional entropy is interpreted as the uncertainty on average about X given Y. We have $H(X|Y) = H(X, Y) - H(Y) \geq 0$ with equality if and only $X = f(Y)$ deterministically. The mutual information content of X, Y is defined by

$$I(X : Y) := H(X) + H(Y) - H(X, Y),$$

and is interpreted as a measure of the codependance of X, Y. Indeed we have

$$H(X) + H(Y) \geq H(X, Y)$$

with equality if and only if X, Y are independant. The mutual information of X, Y conditioned on a deterministic variable Z is defined by

$$I(X : Y|Z) := H(X|Z) + H(Y|Z) - H(X, Y|Z) = H(X|Z) - H(X|Y, Z).$$

Likewise we have $I(X:Y|Z) \geq 0$ with equality if and only if X, Y are independant when conditioned on Z. The positivity of $I(X:Y|Z) \geq 0$ is equivalent to the strong subadditivity property

$$H(X,Y,Z) + H(Z) \leq H(X,Z) + H(Y,Z)$$

with equality if and only if $X \mapsto Z \mapsto Y$ defines a Markov chain.

5. Factorization and Tao-Szemeredi Regularity Lemma

A review of various algorithms suggests the idea that every deterministic algorithm is a composition of random searches and function evaluations. We thank M. Sunohara for referring us to T. Tao's proof of Szemeredi Regularity Lemma [4] which presents the dichotomy between coarse structures (where only a small number of bits are known about events) and fine structures (where a large number of bits are known about events). Tao's basic idea is that every correlation π is a hybrid of these two extremes. Tao's presentation of Szemeredi is based on constructing small number of bits $Z_1.Z_2$ of X, Y which correlate with π, and such that the distribution π is approximately independant of conditioning on further bits of X, Y. This independance implies that π conditioned on the coarse bits Z_1, Z_2 is only decided by random search.

Moreover Tao-Szemeredi constructs a fine approximation Z_1', Z_2' by iterating the following minimization program. Let π, σ, τ be described as above. Let $F : \mathbf{R}_+ \to \mathbf{R}_+$ be a monotone function, e.g. $F(x) = 1 + x$.

If $X \mapsto Z_1$, $Y \mapsto Z_2$ are deterministic mappings, then consider the following minimization program:

$$\min_{Z_1', Z_2'} [H(\pi|Z_1', Z_2') + \frac{H(Z_1', Z_2')}{F(H(Z_1, Z_2))}] \tag{1}$$

where the minimum is taken over all random variables Z_1', Z_2' satisfying the determinism constraints $\sigma \mapsto Z_1' \mapsto Z_1$ and $\tau \mapsto Z_2' \mapsto Z_2$. The program (1) is initialized by setting $Z_1 = 0$ and $Z_2 = 0$.

The basic step in the Tao-Szemeredi lemma is updating $Z_1 := Z_1'$ and $Z_2 := Z_2'$ whenever $H(\pi|Z) - H(\pi|Z') \geq \epsilon$ and restarting the minimimization program 1. The goal of the minimization program (1) is to find deterministic reductions such that both $H(\pi|Z')$ is small (i.e. π is approximately determined by the deterministic marginals $X \mapsto Z_1'$, $Y \mapsto Z_2'$) and such that $H(Z')/F(H(Z))$ is small (i.e. the joint distribution $Z' = (Z_1', Z_2')$ has bounded uncertainty relative to the uncertainty

of $H(Z)$. So given $Z = (Z_1, Z_2)$ we want a finer deterministic resolution $Z' = (Z_1', Z_2')$ such that π is almost determined by Z' and Z' is a coarse approximation relative to Z.

The existence of minimizers in 1 is key element of Szemeredi's Lemma, and we quote directly from T. Tao's information theoretic presentation [4, Lemma 4.3]:

Lemma 1 (Tao-Szemeredi Regularity Lemma)**.** *Let π, σ, τ be defined as above. Suppose $H(\pi) = m$ is finite. For every $\epsilon > 0$ there exists random variables Z_1', Z_2' (the "fine approximation") and Z_1, Z_2 (the "coarse approximation") with the following properties:*

(Determinism) We have the determinism relations

$$\sigma \mapsto Z_1' \mapsto Z_1, \quad \tau \mapsto Z_2' \mapsto Z_2.$$

(Coarse Approximation has Bounded Entropy) We have

$$H(Z_1, Z_2) \leq H(Z_1', Z_2') \leq O_{F,\epsilon,m}(1).$$

(Coarse and Fine Approximations are ϵ-close) We have

$$I(\pi : Z_1', Z_2' | Z_1, Z_2) = H(\pi | Z_1, Z_2) - H(\pi | Z_1', Z_2') \leq \epsilon$$

given the determinism relations.

(Fine Approximation is Nearly Optimal) For any random variables with $\sigma \mapsto W_1$, $\tau \mapsto W_2$ we have

$$I(\pi : W_1, W_2 | Z_1', Z_2') \leq \frac{H(W_1, W_2)}{F(H(Z_1, Z_2))}.$$

6. Zero Subset Sums and $\mathbf{P} \neq \mathbf{NP}$

Given a finite measure f on $\mathbf{Z}$, let $DP(f)$ be the decision problem of deciding whether f contains a zero subset sum. We recall that $DP(f)$ is the universal **NP** problem (called "**NP** complete" in the literature). C.A. Feinstein's argument that $\mathbf{P} \neq \mathbf{NP}$ is that every explicit algorithm which decides $DP(f)$ is necessarily "guess and check" over exponentially many subsets, specifically $2^{(\#f)/2}$ per the Meet in the Middle (MITM) algorithm. MITM depends on the observation that the support of any nontrivial solution to $\epsilon(f) = 0$ has cardinality ≥ 2. This implies $\epsilon(f) = 0$ is equivalent to an identity $\epsilon(f) = \epsilon(g)$ where f, g are both nonzero. This implies the well known lower bound of $O(2^{\#(f)/2})$ complexity, but otherwise the supports of $\epsilon(f)$ and $\epsilon(g)$ are unstructured.

We apply Tao-Szemeredi's lemma by studying the graph of the augmentation map $\epsilon : 2^f \to \mathbf{Z}$ and the general problem is deciding whether given pairs (f', m) where f' is a subset of f and m is the target sum $\epsilon(f') = m$. The initial zero subset problem is to decide whether $(f', 0)$ belongs to the graph of ϵ. The necessity of exponential guess and check in deciding zero subset sum is obtained via Tao-Szemeredi lemma 1 by demonstrating that the entropy $H(\epsilon_f)$ is not reduced by conditioning on proper coarse bits of f. Tao-Szemeredi implies that $DP(f)$ is independant of all proper sub-bits of f. Having partial information on the bits of f is generally insufficient to decide $DP(f)$ over $\mathbf{Z}$. That is $DP(f)$ generally can only be decided by complete information on *all* the bits of f.

Tao-Szemeredi's Regularity Lemma is based on finding minimizers to an entropy program which looks for coarse bits which maximally determine π almost everywhere.

Proposition 2. *Let f be a finite measure on $\mathbf{Z}$. Let ϵ_f be the pushforward of the graph of $\epsilon : 2^f \to \mathbf{Z}$.*

1. The minimization program 1 is minimized by the trivial reductions $Z_1 = 1_{2^f}$ and $Z_2 = 0$. In otherwords the Tao-Szemeredi coarse approximation of ϵ_f is trivial.

2. Moreover $H(\epsilon_f | Z_1, Z_2) \geq H(\epsilon_f | 1_{2^f}, 0)$ for all deterministic $f \mapsto Z_1$ and $Y \mapsto Z_2 \mapsto 0$ with $Z_1 \neq f$.

Item 2. implies that ϵ_f is approximately independant of all the proper sub-bits of the inputs 2^f. Therefore ϵ_f cannot be correlated to any structure relating to the proper sub-bits of f, but is only generally decided by precise information about the complete bits of f. We claim this establishes the nonidentity $\mathbf{P} \neq \mathbf{NP}$. In other words, the only coarse approximation to ϵ_f is the uniform measure itself.

References

[1] Stephen Cook. "The P versus NP problem". In: *Clay Mathematics Institute* 2 (2000), p. 6.

[2] Craig Alan Feinstein. "An Elegant Argument that $P \neq NP$". In: *Progress in Physics* 2 (2011), pp. 30–31.

[3] Claude Elwood Shannon. "A Mathematical Theory of Communication". In: *The Bell system technical journal* 27.3 (1948), pp. 379–423.

[4] Terence Tao. “Szemeredi’s Regularity Lemma Revisited”. In: *arXiv preprint math/0504472* (2005).