A NEW LINE OF ATTACK ON THE DICHOTOMY CONJECTURE

GÁBOR KUN AND MARIO SZEGEDY

Abstract. The well known dichotomy conjecture of Feder and Vardi states that for every family Γ of constraints CSP(Γ) is either polynomially solvable or NP-hard. Bulatov and Jeavons reformulated this conjecture in terms of the properties of the algebra $Pol(\Gamma)$, where the latter is the collection of those m-ary operations ($m = 1, 2, \ldots$) that keep all constraints in Γ invariant. We show that the algebraic condition boils down to whether there are arbitrarily resilient functions in $Pol(\Gamma)$. Equivalently, we can express this in the terms of the PCP theory: CSP(Γ) is NP-hard iff all long code tests created from Γ that passes with zero error admits only juntas[1]. Then, using this characterization and a result of Dinur, Friedgut and Regev, we give an entirely new and transparent proof to the Hell-Nešetřil theorem, which states that for a simple connected undirected graph H, the problem CSP(H) is NP-hard if and only if H is non-bipartite.

We also introduce another notion of resilience (we call it strong resilience), and we use it in the investigation of CSP problems that 'do not have the ability to count.' The complexity of this class is unknown. Several authors conjectured that CSP problems without the ability to count have bounded width, or equivalently, that they can be characterized by existential k-pebble games. The resolution of this conjecture would be a major step towards the resolution of the dichotomy conjecture. We show that CSP problems without the ability to count are exactly the ones with strongly resilient term operations, which might give a handier tool to attack the conjecture than the known algebraic characterizations.

.

Partially supported by NSF Grants CCF-0832797 and DMS-0835373.

[1]We shall use the term "junta" in a little weaker sense than usual: a constant number of the variables have constant influence on the outcome.

1. Introduction

Constraint satisfaction problems (CSP) are the pinnacles in NP not only because they have multiple interpretations in logic, combinatorics, and complexity theory, but also for their immense popularity in various branches of science and engineering, where they are looked at as a versatile language for phrasing search problems. This said, it is even more remarkable that some basic complexity questions about them remain unanswered.

To a finite domain D, variables $\{x_1, x_2, \ldots\}$ ranging in D, and a set Γ of finitary relations on D we can associate a problem CSP(Γ), whose instances consist of a finite set of *constraints* of the form $(x_{i_1}, \ldots, x_{i_k}) \in R_j$ for some $R_j \in \Gamma$. The size of the instance (usually denoted by n) is by definition the number of different variables involved in its constrains.

As one might expect, for the tractability of CSP(Γ) the relations in Γ matter. For instance, general Boolean CSPs are NP hard, but if all constraints are Horn clauses (i.e. disjunctions of literals, at most one of which is negative), then the problem is polynomially solvable. Other polynomially solvable cases include linear equations over finite fields and the set of all Boolean constraints that involve at most two variables.

The central question of the field is how the complexity of CSP(Γ) depends on Γ. Due to a beautiful result of Schaefer [41] we know, that if the variables are binary then CSP(Γ) is either NP-hard or polynomial time solvable for every Γ. His *Dichotomy Theorem* also gives a full description of the polynomial time solvable families.

A fundamental question raised by Feder and Vardi [21], if this theorem generalizes for arbitrary finite domain. Their *Dichotomy Conjecture* would imply the dichotomy of Monotone Monadic SNP ([21, 30], see also [31]), which is perhaps the largest natural subclass of NP, for which dichotomy can be hoped for. That the entire class NP does not have dichotomy (unless P=NP) was proved by Ladner [32].

In [21] it is established that it is sufficient to settle the dichotomy conjecture when Γ contains a single binary relation, i.e. a directed graph, H. With a slight abuse of notation we denote this problem by CSP(H). A problem instance now simply becomes a directed graph G whose vertices we want to map to the vertices of H such that edges go into edges. The problem then becomes a graph homomorphism problem. What if G is undirected? In this case dichotomy holds by a pioneering theorem due to Hell and Nešetřil (1990):

Theorem 1. *Assume that H is a simple, connected, undirected graph. Then CSP(H) is polynomial time solvable if and only if H is bipartite. Otherwise CSP(H) is NP-complete.*

We refer the reader interested in the graph homomorphic view of CSPs to an excellent survey written by the above authors, which also puts our current result into that context [25].

There is a beautiful algebraic theory due to Jeavons and his coauthors [14, 15, 27, 11, 12], that looks at maps from D^m to D ($m = 1, 2, \ldots$), which keep all relations in Γ invariant (said to be *compatible* with Γ). These maps, if we look at them as operators, form an algebra, called $Pol(\Gamma)$. The theory heavily relies on the fact that a composition $f(g_1, \ldots, g_m)$ of operators that are compatible with Γ is also compatible with Γ, hence $Pol(\Gamma)$ is closed under composition.

We can also look at these functions in an entirely different way. For fixed m the condition that $f : D^m \to D$ keeps all relations in Γ can be interpreted so that f passes the long code test associated with Γ with zero error.

This dual interpretation of $Pol(\Gamma)$ allows us to connect the algebraic theory of CSPs with Fourier analytic techniques that were successfully used in the theory of probabilistically checkable proofs.

To demonstrate the strong interaction between the theories we reprove the theorem of Hell and Nešetřil in a transparent way. We rely on theorems of Bulatov and Jeavons as well as on the Fourier analytic results of Dinur, Friedgut and Regev.

We then go farther, and give new analytic characterizations of two different classes of CSPs. The first class is known as Block Projective CSPs: This is the class that does not have "interesting" polymorphisms, provably NP-hard, and contains all known NP-hard instances. The other class is the set of CSPs to which some linear equations can be reduced. The class goes under the name "CSPs with ability to count."

With one leg our characterizations stand on the algebraic theory of CSPs, and with the other leg they rest on concepts familiar from PCP theory such as resilience to noise (random or adversarial) and the long code tests. The table below gives a little summary of our results:

Algebraic Condition		*Analytic Condition on* $Pol(\Gamma)$		*Long Code Tests*
Block Projective	$\leftrightarrow$	Lacks Asymptotically Resilient Terms	$\leftrightarrow$	Admits Only Juntas[2]
$\neg$ Block Projective	$\leftrightarrow$	Has Asymptotically Resilient Terms		
No Ability to Count	$\leftrightarrow$	Has Strongly Resilient Terms		

Our paper also contributes a little bit to the theory of higher order dynamical systems: We characterize maps from D^n to D whose high iterates are resilient to small noise. That is, for any measure on D, if k is large enough, then no matter how we change a fixed constant number of inputs *before* other input bits are set, the distribution of the function values of the k times iterated map will be decreasingly influenced as k tends to infinity. We also study a stronger notion of resilience when an adversary controling a variable cannot significantly influence the outcome of powers of a map even if she can set her input *after* seeing all other inputs. We show that the existence of these functions in $Pol(\Gamma)$ coincides with Γ being in the "Not block projective," and the "Without the ability to count" classes, respectively.

Our new characterization of the Block projective class allows us to give a new modular proof to the Hell-Nešetřil theorem (Sections 5, 6). Our characterization of the class of CSPs without the ability to count (Section 8) gives a new tool to tackle the conjecture of Feder and Vardi [21], Larose and Zádori [33], and Bulatov [8] proved all to be equivalent by Larose, Zádori and Valeriote [34], that CSP problems without the ability to count have bounded width. The resolution of this conjecture would be a major step towards the resolution of the dichotomy conjecture.

2. Long Code Tests

Fix n. The Long Code over alphabet D consists of those functions $f : D^n \to D$ that depend on a single coordinate:

$$f : (x_1, \dots x_n) \to x_i \qquad (1 \le i \le n).$$

A membership test for this code is an essential element of the PCP theory. The test must be local in the sense that it evaluates f only at a constant number of places. If the replacement values are found consistent, the test *accepts*, otherwise it *rejects*. If the test looks at k places, it is called *k-local*. The test is *random*. PCP theory is concerned with tests that with high probability accept only words that are close to some word in the Long Code. In this paper we somewhat reverse the question: given a Test, determine the set of those functions that are accepted with probability one! This sounds like an easier problem because the approximation aspect is ignored. What we have found is that even in the non-approximate setup Fourier Analytic techniques benefit us.

Some words about the test: Each known test[3] is associated with a relation R on D (or with a set of relations, in which case we run tests for each, separately). Let $R \subseteq D^k$ be a k-ary relation on D and let π be a probability distribution on k tuples $(x^{(1)}, \dots, x^{(k)}) \in (D^n)^k$ that obey the property that

$$(x_i^{(1)}, \dots, x_i^{(k)}) \in R \qquad \text{for } 1 \le i \le n. \tag{1}$$

[2]In the sense that a constant number of the variables have constant influence on the outcome.

[3]Hastad's test requires a little modification of the framework.

Then $\text{Test}_{R,\pi}$ is a procedure that takes a function $f : D^n \to D$ as its input, selects a k-tuple $(x^{(1)}, \ldots, x^{(k)}) \in (D^n)^k$ of according to π, and accepts if and only if

$$(f(x^{(1)}), \ldots, f(x^{(k)})) \in R.$$

Take Dinur's test of the long code on $D = \{0, 1\}$ for an example. She used relations: $b = \neg a$ and $a \vee b \vee c$. The first relation is automatically provided to hold everywhere by a technique known as *folding*. For the second relation Dinur used a certain non-trivial probability distribution π on triples. By Fourier analytic techniques she verified that the test checks the long code in the following strong sense: If the acceptance probability is $1 - \varepsilon$ then f must coincide with some word of the long code on $1 - O(\varepsilon)$ fraction of randomly and uniformly chosen elements of D^n.

3. The Graph Homomorphism Test

We discuss the special case of the long code test, where H is a simple, connected, undirected graph. The result of this section will be a component of our new proof of the Hell-Nešetřil theorem. We denote the vertex set of H by D (faithfully to our prior notations), and the edge set of H with E. The power set of a set is denoted by $P()$. Let

Graph	Vertices	Edges
H^n	D^n	$(\vec{v}, \vec{w}) : (v_i, w_i) \in E$ for all $1 \leq i \leq n$;
$P(H)$	$P(D)$	(S, T): $(s, t) \in E$ for all $s \in S$ and $t \in T$.

By the previous section $\text{Test}_{H,\pi}$ on $f : D^n \to D$ picks an edge $e = (v, w)$ of H^n with probability $\pi(e)$ and accepts if $(f(v), f(w))$ is an edge in H. If the support of π is the entire E^n, then f is a homomorphism from H^n to H if and only if $\text{Test}_{H,\pi}$ accepts with probability one. Our analysis requires a special measure.

The *stationary measure* on the vertices, μ_D, (edges, μ_E) of H assigns frequencies to every node (edge), with which that node (edge) is visited by an infinite random walk. It is well known that the stationary measure on the edges of a simple connected undirected graph is uniform. This implies that the stationary measure on the vertices is proportional to the degree of each node.

It is immediate that the stationary measure on the vertices (edges) of H^n is μ_D^n (μ_E^n), where

$$\mu_D^n((v_1, \ldots, v_n)) = \prod_{i=1}^{n} \mu_D(v_i); \qquad \mu_E^n((e_1, \ldots, e_n)) = \prod_{i=1}^{n} \mu_D(e_i).$$

We would like to characterize those functions $F : D^n \to D$ that are accepted by $\text{Test}_{H,\pi}$ with probability one. By our previous remark these are exactly the graph homomorphisms from H^n to H. What we show is that, independently of n, for any such f we find a constant sized set of coordinates that have non-negligible influence on the value of f. This holds when H is connected, non-bipartite. We say that a mapping $f : H^n \to G$ depends only on the subset of coordinates $L \subseteq \{1, \ldots, n\}$ if $u|_L = v|_L$ implies $f(u) = f(v)$, where $u|_L$ $(v|_L)$ denotes the restriction of $v = (v_1, \ldots, v_n)$ to the coordinates in L.

Lemma 2. *Let $H = (D, E)$ be a simple, connected, undirected graph, which is also* non-bipartite*, and let μ_D and μ_E be the stationary measure on its vertices and edges. Then for every $\varepsilon > 0$ there exists an integer $l = l(\varepsilon, H)$ such that if $f : D^n \to D$ is a homomorphism then there is a mapping $s : D^n \to P(D)$ such that:*

1. $\text{Prob}_{(v,w) \in \mu_E^n}((s(v), s(w))$ *is not an edge in* $P(H)) \leq \epsilon$*;*
2. $\text{Prob}_{v \in \mu_D^n}(f(v) \notin s(v)) \leq \epsilon$.
3. *The mapping s depends on at most l coordinates.*

Proof. Since $f : D^n \to D$ is a graph homomorphism, the inverse image, $f^{-1}(K)$, of an independent set $K \subseteq D$ is independent in H^n. We use a theorem of Dinur, Friedgut and Regev to show that $f^{-1}(K)$ has a special structure:

Theorem 3. [19] *Let $H = (D, E)$ be a simple, undirected, connected, non-bipartite graph with stationary measures μ_D and μ_E on its vertices and edges. Then for every $\delta > 0$ there exists a positive integer $j = j(\delta)$ such that to every independent set I in H^n we can associate a set of coordinates L_I and an "almost independent" set I^* that spans less than δ fraction of the edges (according to measure μ_E^n) and depends only on coordinates in L_I, such that*

$$\mu_D^n(I \setminus I^*) \leq \delta. \tag{2}$$

For an independent set $I \subseteq D^n$ let L_I and I^* as in Theorem 3. We choose δ later. Let $\mathrm{Ind}(H)$ be the system of all independent sets of H, and define the following set of coordinates:

$$L = \bigcup_{K \in \mathrm{Ind(H)}} L_{f^{-1}(K)} \qquad l = |L|.$$

We define $s : D^n \to P(D)$ via an "inverse" function $S : D \to P(D^n)$ as:

$$S(x) = \bigcap_{\substack{K \in \mathrm{Ind(H)} \\ x \in K}} f^{-1}(K)^* \tag{3}$$

$$s(v) = \{x \in D |\ v \in S(x)\}. \tag{4}$$

Lemma 4. *The mapping s depends only on its coordinates in L.*

Proof. Notice that for any $K \in \mathrm{Ind(H)}$ membership in $f^{-1}(K)^*$ depends only on the coordinates in $L_{f^{-1}(K)} \subseteq L$. Thus the same holds for $S(x)$ for any x. This makes S effectively a function from D to $P(D^l)$, hence s is a function from D^l to $P(D)$. □

We now set $\delta = \varepsilon / |\mathrm{Ind(H)}|$.

Lemma 5. *Condition (1) of Lemma 3 holds for s.*

Proof. Notice that $(s(v), s(w))$ is not an edge in $P(H)$ if and only if there are $x \in s(v)$, $y \in s(w)$ such that $\{x, y\} \in \mathrm{Ind}(H)$, which in turn by Definitions (3) and (4) implies that $v, w \in f^{-1}(\{x, y\})^*$.

Say that an edge $e = (v, w) \in E^n$ is *bad* if it is induced inside $f^{-1}(K)^*$ for some $K \in \mathrm{Ind(H)}$, and *good* otherwise. The total measure of bad edges is bounded by $\delta |\mathrm{Ind}(H)| = \varepsilon$, since for every $K \in \mathrm{Ind}(H)$ the total measure of edges induced inside $f^{-1}(K)^*$ is at most δ. An edge participates in the event we want to bound from above only if it is bad. □

Lemma 6. *Condition (2) of Lemma 3 holds for f.*

Proof. Let us call $v \in D^n$ *faulty* if for some $K \in \mathrm{Ind}(H)$ it belongs to $f^{-1}(K) \setminus f^{-1}(K)^*$. The probability that v is faulty is then at most $\delta |\mathrm{Ind}(H)|$. It is obvious from our definitions, that when v is not faulty, then $f(v) \in s(v)$. □

□

4. Dichotomy Conjecture and Algebra

From Dinur's analysis of her test, when setting ε to 0, we get that if a function $f : \{0,1\}^n \to \{0,1\}$ obeys $f(\neg x_1, \ldots, \neg x_n) = \neg f(x_1, \ldots, x_n)$ for every $\vec{x}$, and for every $\vec{x}$, $\vec{y}$ and $\vec{z}$ it obeys that $x_1 \vee y_1 \vee z_1, \ldots, x_n \vee y_n \vee z_n$ implies $f(x_1, \ldots, x_n) \vee f(y_1, \ldots, y_n) \vee f(z_1, \ldots, z_n)$, then f has to be a word of the long code.

Words of the long code, or *projections*, play a central role in the algebraic theory of CSPs developed by Jeavons and his co-authors: If for every n the set of *operations* $f : D^n \to D$ that keep all relations in Γ are only the projections, then CSP(Γ) is NP-hard. An operation f keeps an l-ary relation R iff for every $x^{(1)}, \ldots, x^{(l)} \in D^n$:

$$R(x_1^{(1)}, \ldots, x_1^{(n)}) \wedge \cdots \wedge R(x_l^{(1)}, \ldots, x_l^{(n)}) \;\to\; R(f(x_1^{(1)}, \ldots, x_1^{(l)}), \ldots, f(x_n^{(1)}, \ldots, x_n^{(l)})).$$

One can easily recognize that this is exactly saying that the long code test for relation R succeeds with probability one for any distribution π with full support on tuples that satisfy Equation (1).

That the algebraic theory of CSPs and long code tests talk about the same objects, raises a lot of questions. Why this connection has not been utilized thus far? The answer perhaps is that the testing theory deals with *analytical* properties of functions that *nearly* satisfy the tests, while the algebraic theory of CSPs deals with *algebraic* properties of functions that keep *all* relations. Our main contribution is that we positively demonstrate, that it is worthwhile to take an analytic approach to functions that keep *all* relations. When these functions are examined both from analytic and algebraic viewpoints, nontrivial conclusions like the Hell-Nešetřil theorem can be obtained.

The connection has another great benefit, namely it lends more sense to rewriting algebraic identities into analytic form. Let us explain: Bulatov, Jeavons and Krokhin essentially conjectured that CSP(Γ) is tractable iff there is a compatible operation which can be distinguished from the projections by its identites. E.g a majority operation satisfies $f(y, x, x) = f(x, x, x), f(x, y, x) = f(x, x, x), f(x, x, y) = f(x, x, x)$, the i^{th} identity shows that this can not be a projection to the i^{th} coordinate, since this coordinate is x on one side and y on the other side. The above is just a special case.

Before getting closer to algebra we have to use two technical assumptions. First, we deal with the case when Γ is a core, i.e. every homomorphism $\Gamma \to \Gamma$ is an automorphism. Every structure has a unique core (up to isomorphism), and a structure and its core define the same CSP language. This is always assumed in the literature to make algebraic methods to work. Secondly, we only consider idempotent operations: idempotency means that $f(x, \ldots, x) = x$ for every $x \in D$. The simple reason is that the complexity of a CSP problem for cores depends only on its idempotent operations, and on the other hand this assumption simplifies the algebraic theory a lot.

By a result of McKenzie and Maróti, if there is a compatible operation which can be distinguished from the projections by its identites, then there is also a special type, called weak near-unanimity (WNU) term. An idempotent operation f is a WNU if for every $x, y \in X$ it satisfies $f(y, x, \ldots, x) = f(x, y, x, \ldots, x) = \cdots = f(x, \ldots, x, y)$.

The following theorem uses the WNU condition of Maróti and McKenzie [35], while condition (1) is stated in the combinatorial terminology of Nešetřil, Siggers and Zádori [36].

Theorem 7. *For any constraint family Γ the following are equivalent.*

(1) Γ *is block-projective, i.e. there exist no disjoint subsets S_1, S_2 of D such that for every compatible, idempotent operation f there exists k such that $f(x_1, \ldots, x_n) \in S_i$ iff $x_k \in S_i$ for $i = 1, 2$.*

(2) *There exists a compatible WNU term operation.*

In the next sections we add to the above equivalent conditions a new one: There exists a compatible WNU term operation iff there exists an operation whose powers, as they grow, become arbitrarily resilient to small noise. This offers an analytic approach to investigating NP-hardness of CSPs.

It may occur that in $Pol(\Gamma)$ there is no WNU operation, but CSP(Γ) is NP-complete, however this cannot happen when Γ is a core. So we restrict ourselves to cores as promised.

Theorem 8. *If Γ is a core and has no compatible WNU operation then* CSP(Γ) *is NP-complete.*

The Dichotomy Conjecture of Bulatov, Jeavons and Krokhin states that Theorem 8 can be reversed in the following sense:

Conjecture 9 (Algebraic Dichotomy Conjecture [12]). Let Γ be a core. If Γ admits compatible WNU operation then CSP(Γ) is tractable, else it is NP-complete.

Example 10. This gives a remark for Dinur's test: her test implies the NP-hardness of CSP($\neg a = b,\ a \wedge b \wedge c$) by Theorem 8. While this is not earth-shattering, the Algebraic Dichotomy Conjecture also immediately suggests that the $\neg a = b$ folding is essential for the test to work.

Bulatov, Jeavons and Krokhin used the term *Polymorphism* for functions (of arbitrary number of variables) that are compatible with all relations in Γ, and they denoted this set of functions by $Pol(\Gamma)$. They have proved that $Pol(\Gamma)$ determines the complexity of Γ. The approach has been applied in several contexts, in particular, this is how Bulatov solved the problem for $|D| = 3$. Another application of their theory by Bulatov proves dichotomy, when Γ is a set of list homomorphisms [6]. The original goal of the algebraic theory was to deal with decision problems, though it proved to be successful in other cases. Bulatov an Dalmau proved a dichotomy theorem for counting the solutions of CSPs [10], Bodirsky and Nešetřil [4] managed to extend the theory to (omega-categorical) countably infinite target structures, Chen partly managed for quantified CSPs [13].

In the last section we are going to look at the class of CSPs with "no ability to count." This class is defined algebraically and its complexity is not yet known. We give an analytic definition of the class, which seems to yield a fresh insight into its tractability.

5. Resilience

Let D be a finite domain, $f : D^n \to D$ and $\mu_1, \ldots, \mu_n$ be distributions on D. By $f(\mu_1, \ldots, \mu_n)$, or shortly by $f(\vec{\mu})$, we denote the distribution on D that we obtain by plugging independent D-valued random variables into f such that the ith variable is distributed as μ_i. μ. We define

$$Resil(f, l, \mu) = \sup_{\mu_1,\ldots,\mu_n} \delta(f(\mu, \mu, \ldots, \mu), f(\mu_1, \ldots, \mu_n)),$$

where δ refers to the statistical difference, $\delta(\mu, \nu) = \frac{1}{2}\sum_{x\in D} |\mu(x) - \nu(x)|$, of distributions and $\mu_1, \ldots, \mu_n$ runs through all sequences of distributions on D with the properties that at most l of the μ_is are different from μ and the support of each μ_i is contained in the support of μ. We call $Resil(f, l, \mu)$ the *resilience* of f.

6. Asymptotic resilience

The iterates of a function $f : D^n \to D$ are $f^1 = f$, $f^{i+1} = f(f^i, \ldots, f^i)$ for $i > 1$. The arity of f^k is n^k, and we can visualize it as an n-ary tree of depth k built of fs. We say that a function $f : D^n \to D$ is *asymptotically resilient* if for every distribution μ on D and every $l, \varepsilon > 0$ we have $Resil(f^k, l, \mu) < \varepsilon$ for any sufficiently large k. Most functions are asymptotically resilient, but e.g. projections are not. Instead of giving further examples we describe *all* asymptotically resilient idempotent functions.(With a little extra effort one can give a similar characterization without the idempotency condition.)

Theorem 11. *Let $f : D^n \to D$ be idempotent. The following are equivalent:*

1. *f is asymptotically resilient.*
2. *$Resil(f^k, 1, \mu)$ goes to zero as k goes to infinite for every fixed μ.*
3. *f generates a WNU (including that itself is a WNU).*
4. *There do not exist pairs of disjoint subsets $S_1, S_2 \subseteq D$ and $1 \leq k \leq n$ such that $f(x_1, \ldots, x_n) \in S_i$ iff $x_k \in S_i$ for $i = 1, 2$.*

Proof. (1) implies (2) by the definition of asymptotic resilience. We will show (2) → (1) in the Appendix. The equivalence of (3) and (4) comes from algebra. It is easy to see that (1) implies (4): in fact if (4) does not hold then $Resil(f^k, 1, \mu) = 1$ for every k. Now we prove that (3) implies (2). For the sake of simplicity we assume that f is a WNU itself, of arity n (if f only *generates* a WNU, the proof needs only a minor adjustment). For our argument we fix μ. Let $\mu_k = f^k(\mu^{n^k})$ (recall that the arity of f^k is n^k). We would like to estimate the statistical difference of μ_k and $f^k(\mu^{i-1}\nu\mu^{n^k-i})$ for any $1 \leq i \leq n^k$ and any ν, whose support is contained in the support of μ. Let $\alpha_k = \max_{i,\nu} \delta(\mu_k, f^k(\mu^{i-1}\nu\mu^{n^k-i}))$. What we need to show is that $\alpha_k \to 0$. By the following propositions and its corollary it is straightforward that α_k is non-increasing:

Proposition 12. *The variation distance of two distributions cannot increase under any map $F : X \to Y$.* $\left(\delta(\mu,\nu) = \frac{1}{2}\sum_{y\in Y} |\mu(y) - \nu(y)| \geq \frac{1}{2}\sum_{x\in X} \left|\sum_{y\in F^{-1}(x)}(\mu(y) - \nu(y))\right| = \delta(F(\mu), F(\nu)).\right)$

Corollary 13. *Let $f : D^n \to D$ be arbitrary and $\mu_1, \dots, \mu_n$, $\nu_1, \dots, \nu_n$, be two sequences of distributions on X. Then $\delta(f(\mu_1, \dots, \mu_n), f(\nu_1, \dots, \nu_n)) \leq \sum_{i=1}^n \delta(\mu_i, \nu_i)$.*

Proof. The corollary follows from $\delta(\prod_{i=1}^n \mu_i, \prod_{i=1}^n \nu_i) \leq 1 - \prod_{i=1}^n (1 - \delta(\mu_i, \nu_i)) \leq \sum_{i=1}^n \delta(\mu_i, \nu_i)$. □

We now want to go a step further and to show that α_{k+1}/α_k is is upper bounded by a constant (i.e. independent of k) less than 1. It is easy to see that Proposition 12 can be strengthened if we find an $x \in X$, $y_0, y_1 \in F^{-1}(x)$ such that $\mu(y_0) - \nu(y_0) \geq 0$, $\nu(y_1) - \mu(y_1) \geq 0$:

$$\delta(F(\mu), F(\nu)) \leq \delta(\mu,\nu) - \min\{\mu(y_0) - \nu(y_0), \nu(y_1) - \mu(y_1)\}. \tag{5}$$

At this point we exploit that f is a WNU, and certain identities hold for its output. Before describing what we get from this (proof in the Appendix) we need a technical definition:

Definition 14. Let μ and ν be probability distributions on X. We define

$$\min \frac{\mu}{\nu} = \min_{x:\nu(x)\neq 0} \frac{\mu(x)}{\nu(x)}.$$

Lemma 15. *For every WNU term f and probability distributions μ and ν on D:*

$$\delta(f(\mu^{i-1}\nu\mu^{n-i}), f(\mu^n)) \leq \left(1 - \frac{\delta(\mu,\nu)^{n-1}}{|D|^n} \min\frac{\mu}{\nu}\right)\delta(\mu,\nu)$$

for every $1 \leq i \leq n$.

Lemma 15 gives that $\alpha_{k+1}/\alpha_k \leq 1 - \frac{\alpha_k^{n-1}}{|D|^n} \min_{\tilde{\mu}_k} \min \frac{\mu_k}{\tilde{\mu}_k}$, where $\tilde{\mu}_k$ ranges among distributions of the form $f^k(\mu^{i-1}\nu\mu^{n^k-i})$. Indeed, use that $f^{k+1}(\mu^{i-1}\nu\mu^{n^{k+1}-i})$ can be written as f on many copies of μ_k and one copy of $f^k(\mu^{i'-1}\nu\mu^{n^k-i'})$. An easy analysis shows that this improvement is sufficient, because $\min_{\tilde{\mu}_k} \min \frac{\mu_k}{\tilde{\mu}_k}$ remains bounded from below by $\min \frac{\mu}{\nu}$. This follows from the more general:

$$\min \frac{\prod_i \mu_i}{\prod_i \nu_i} = \prod_i \min \frac{\mu_i}{\nu_i}; \quad \min \frac{F(\mu)}{F(\nu)} \geq \min \frac{\mu}{\nu}.$$

□

7. The Hell-Nešetřil theorem

Recall that the Hell-Nešetřil theorem [24] states that for a simple, connected, undirected graph H the complexity of CSP(H) is polynomial if H is bipartite and NP-complete otherwise. The first part of the theorem is trivial: every bipartite graph with at least one edge has a retraction to into any of its edges. This implies that a graph G has a homomorphism into H iff it has a homomorphism into a single edge, i.e. G is bipartite. The interesting, and combinatorially quite involved part is the NP-completeness of CSP(H) when H is non-bipartite. This was the first dichotomy theorem with a really sophisticated proof using gadget reductions. Later Bulatov [7] streamlined the proof using the algebraic theory, though his proof still has some ad hoc part. Barto, Kozik and Niven [2] extended the theorem proving dichotomy for digraphs with no sink and source.

Here we give a proof based on our notion of asymptotic resilience. According to Theorem 8 in [12] and using the fact that the core of a non-bipartite graph is also non-bipartite it is sufficient to prove that:

Lemma 16. *Let H be a simple, connected, graph, which is not bipartite. Then $Pol(H)$ has no asymptotically resilient term.*

Proof. (of Lemma 16) Set $\varepsilon = \frac{1}{7}$. We assume for a contradiction that we have an asymptotically resilient compatible operation f on H. Let l be the integer given by Lemma 2 (to ε). An appropriate iteration of our asymptotically resilient operation is a homomorphism $G^n \to G$ such that the followings holds: for any l coordinates $1 \le a_1 < a_2 < \cdots < a_l \le n$ and $u, v \in G^l$ the variational distance of the distributions is small: $\delta(f(u,\nu), f(v,\nu)) < \varepsilon$. (Here $g = f(u,\nu)$ denotes the distribution where $g_{a_i} = u_i$ and else the distribution of g_j is the stationary distribution corresponding to ν. The coordinates of g are independent.) We use now Lemma 2: this gives a mapping $s : H^n \to P(H)$ that is an "almost homomorphism", "almost covers f" and depends on l coordinates, we might assume that these are the first l coordinates. Let $v \in G^l$ and $f_v : V(G)^{n-l} \to V(G)$ denote the subfunction of f we get when fixing the first l variables to v. We call a vertex $v \in G^l$ bad if $\nu(\{t \in G^{n-l} : f_v(t) \notin s(t)) \ge \frac{1}{3}$. The measure of bad vertices is $\le 3\varepsilon$ by Markov's inequality, since $\mathrm{Prob}_{v\in\mu_H^n}(f(v) \notin s(v)) \le \varepsilon$. On the other hand if $\varepsilon < \frac{1}{3}$ then for any two good vertices $u, v \in G^l$ we have $s(u) \cap s(v) \ne \emptyset$. Since G has no loop, the pair $(s(u), s(v))$ is not an edge of the graph $P(G)$. But the stationary measure of bad vertices is $\le 3\varepsilon$, so these cover edges of measure $\le 6\varepsilon$. All other edges are mapped to non-edges by s. At least $1 - 6\varepsilon > \varepsilon$ edges are mapped to non-edges by s, contradicting that $\mathrm{Prob}_{(v,w)\in\mu_E^n}((s(v), s(w))$ is not an edge in $P(H)) \le \varepsilon$. □

8. No Ability to Count and Local vs Global

The harder part of the algebraic dichotomy conjecture is the tractable part: how does an algebraic condition lead to tractability? Jeavons, Cohen and Gyssens [15] proved that the existence of a semilattice operation implies tractability, Cohen, Cooper and Jeavons [14] proved it in case of the existence of a so-called near-unanimity operation (a generalization of majority operations), Bulatov and Dalmau [9] in case of the existence a so-called Maltsev term (what shows that the algebra is "somewhat grouplike"): the algorithms are generalizations of the ones solving Horn-formulas, 2-SAT and linear system of equations, respectively. But to solve a general tractable CSP problem we needs to combine these algorithms (and we might need essentially different ones). There are very few results combining such algorithms of different nature: Bulatov [6]did this when proving dichotomy for list homomorphism problems, Dalmau [18] for CSPs that have an operation on every pair behaving like a group or a majority operation, his result was generalized in a "truly algebraic" manner in [3]. Feder and Vardi have studied CSP problems that no linear system of equations (over a finite field) can be reduced to using gadget reductions: they called these CSP problems without the ability to count. Unlike in the case of linear equations, one might expect here some local algorithms to solve the problem. We will denote this class by Λ. It turned out in the work of Larose, Valeriote and Zádori [34] show that Λ can be well understood in algebraic terms. They use a branch of algebra called Tame Congruence Theory, a localization theory for finite algebras. The localization process of this theory corresponds to gadget reductions of CSP problems. The algebraic characterization of CSP problems reducible to 3-SAT using this theory has led to the algebraic dichotomy conjecture. The work of Larose, Zádori and Valeriote is more involved: they manage to characterize Λ in terms of having locally no algebra that has only group operations and no algebra with only projections.

A recurring theme in combinatorics and computer science is whether consistent local solutions can be patched together into global solution. The notion of *bounded width* intends to capture those CSPs for which local solutions can be made global.

A partial assignment σ with support $X \subseteq \mathbb{N}$ assigns a value from D to each variable x_i, $i \in X$. We say that σ with support on X and σ' with support on Y are *consistent* if they assign the same values to variables in $X \cap Y$. A CSP instance is *satisfied* by a partial assignment σ with support on X, if σ satisfies all constraints that take variables only from X.

Definition 17. An instance of CSP(Γ) is k-consistent if there exists a set Ξ of partial solutions such that:

(1) Every $\sigma \in \Xi$ has support size $k+1$;
(2) Every $\sigma \in \Xi$ satisfies the instance;
(3) For every $|X| = k+1$, $x \in X$ and $y \notin X$, and $\sigma \in \Xi$ with support X, there exists a partial assignment $\sigma' \in \Xi$ with support $(X \setminus \{x\}) \cup \{y\}$ that is consistent with σ.

Definition 18 (Width k)**.** CSP(Γ) has width k if and only if the existence of a set of k-consistent solution guarantees a (global) satisfying assignment.

The notion of local consistency emerged independently in graph theory [26], finite model theory [29] and algebra [15]. This was a successful direction of research in the last years: Foniok, Nešetřil and Tardif [37, 22] studied CSP problems with good characterizations in the category of relational structures with homomorphisms (with finitely many obstructions, these are called finite dualities). Rossman [40] proved the well-known Homomorphism Preservation Theorem in model theory. Dalmau, Kolaitis and Vardi [17, 28, 29] have found the connection with logic, Datalog and existential pebble games, see also Atserias [1]. Hell, Nešetřil and Zhu [26] proved that the k-consistency of a given input can be characterized by obstructions of treewidth at most $(k+1)$. Some of the above authors and others (Larose and Zadori) believe that the only reason for a tractable CSP having a local, but not having a global solution, is that it can solve linear equations:

Conjecture 19. Every problem in Λ has bounded width.

If the Conjectures 9 and 19 are correct, then whenever CSP(Γ) $\in \Lambda$, $Pol(\Gamma)$ should contain a WNU, since it is immediate that bounded width instances are tractable. More is true: we can characterize Λ completely via its compatible WNUs. A non-obvious algebraic theorem of Maróti and McKenzie implies:

Theorem 20. [35] CSP(Γ) $\in \Lambda$ *if and only if there is an* $k_0 \in \mathbb{N}$ *such that for all* $k \geq k_0$ $Pol(\Gamma)$ *contains a WNU of arity* k*.*

Note that the "only if" direction relatively painlessly follows from the fact that term operations that are compatible with linear equations over mod p cannot have arity divisible by p.

Can we exploit Theorem 20 to prove Λ has bounded width? There are only a few cases known where complexity results are shown via *general* WNUs. Considering our proof, a key idea was to "boost" the power of a WNU to obtain a term that has statistically noticeable properties.

In the case of Theorem 20 we have to face one more challenge: Its criterion talks about an infinite number of terms instead of one. Relying on the concept of *influence* we will be able to formulate an analytical condition, called *strong resilience*, that captures Λ with a single intuitive criterion, which, we believe, brings us closer to proving Conjecture 19.

Definition 21 (Influence)**.** Let D be a finite domain and μ be a measure on D. The influence of the i^{th} variable of $f : D^n \to D$ is $\text{Inf}_i(f) = Prob_{\mu^{n+1}}(f(x) \neq f(x'))$, where x, x' runs through all random input-pairs that differ only in the ith coordinate: μ^{n+1} gives a natural measure on such pairs. The maximal influence, max inf(f) is $\max_i \mathrm{Inf}_i(f)$.

Definition 22 (Strong Resilience)**.** A function $f : D^n \to D$ is strongly resilient if for every measure μ on D: max inf$(f^k) \to 0$ when $k \to \infty$.

Theorem 23. *Let* Γ *be a core. Then* CSP$(\Gamma) \in \Lambda$ *if and only if* $Pol(\Gamma)$ *has a strongly resilient term.*

Although it follows from our previous theorems, one can also directly see that:

Proposition 24. *If* $f : D^n \to D$ *is strongly resilient then it is also resilient.*

Indeed, for measure μ and function f it holds that max inf$(f) \leq O(Resil(f, 1, \mu))$. Putting it informally: Functions, whose variables have small influence will stay constant with high probability, when we change a given coordinate of a random input in a random manner. On the other hand, resilient functions only show resilience in a *statistical* sense: the *statistics of their output* is not influenced, when we fix a coordinate *before* randomly setting all other coordinates. Take the parity function as an example: Its variables have high influence, because no matter how we set all but one variable, the un-set variable has full influence on the output. On the other hand, the resilience is zero: no matter how we fix any coordinate, the output statistics remains (1/2, 1/2).

We will prove Theorem 23 only in the longer version of the paper, here we only give a glimpse into the proof. The hard part is to show that $Pol(\Gamma)$ has a strongly resilient term if CSP$(\Gamma) \in \Lambda$. In order to create this term we use the WNU characterization of Λ in Theorem 20. First we create a term with a weaker property:

Definition 25 (1-immune)**.** $f : D^n \to D$ is *immune* with respect to variable i if there are $c_1, \ldots, c_m \in D$ such that the (one-variate) expression $f(c_1, \ldots, c_{i-1}, x, c_{i+1}, \ldots, c_m)$ does not depend on x. f is *1-immune* if it is immune with respect to all of its variables.

We would like to create a 1-immune term from one of the WNUs provided by Theorem Theorem 20. Undoubtedly, the best case is when we find a WNU w, which is a majority form, i.e. $w(x, y \ldots y) = \ldots w(y \ldots y, x) = y$ for all $x, y \in D$. In this case we do not have to go farther, since w is already 1-immune. This might make us think that the worst case is when we find only *minority* forms, i.e. WNUs for which $w(x, y \ldots y) = \ldots w(y \ldots y, x) = x$ for all $x, y \in D$. Luckily, there is a solution in this case too, but, if w is on m variables, we need another (arbitrary) WNU on $m - 1$ variables:

Lemma 26. *Let* w *be a minority WNU on* m *variables and* w' *be an arbitrary WNU on* $m - 1$ *variables in an algebra* $\mathbf{A}$ *that also contains all projections. Then there is a 1-immune expression in* $\mathbf{A}$.

Proof. Consider

$$w(w'(x_2, x_3, \ldots, x_m), w'(x_1, x_3, \ldots, x_m), \ldots, w'(x_1, x_2, \ldots, x_{m-1}))$$

with $x_1, \ldots, x_m$ as variables! ☐

The above single line proof is the essence of turning sequences of WNUs provided by Theorem 20 into 1-immune expressions. The full proof is, of course, contains a little more details, since the existence of a minority form among the WNUs is not given. Finally we need to show:

Theorem 27. *Let* $\mathbf{A}$ *be a finite idempotent algebra. The following are equivalent:*

(1) *In all sub-algebras of* $\mathbf{A}$ *there is a 1-immune term operation.*
(2) *There is a strongly resilient term in* $\mathbf{A}$.

References

[1] A. Atserias, *On Digraph Coloring Problems and Treewidth Duality*, in: Proceedings of the Twentieth Annual, IEEE Symp. on Logic in Computer Science, LICS 2005, 2005, pp. 106–115.

[2] L. Barto, M. Kozik, T. Niven, *CSP dichotomy holds for digraphs with no sink and source*, SIAM J. on Computing, accepted, 2008.

[3] J. Berman, P. Idziak, P. Marković, R. McKenzie, M. Valeriote, R. Willard, *Tractability and learnability arising from algebras with few subpowers*, Proceedings of the 22nd IEEE Symposium on Logic in Computer Science (LICS 2007), (2007), 213–224.

[4] M. Bodirsky, J. Nešetřil: *Constraint Satisfaction with Countable Homogeneous Templates.* J. Log. Comput. 16(3): 359-373 (2006).

[5] A. Bulatov, *A dichotomy theorem for constraints on a three-element set*, Journal of the ACM, 53(1), 2006, 66-120.

[6] A. Bulatov, *Tractable conservative constraint satisfaction problems*, Proceedings of the 18th IEEE Symposium on Logic in Computer Science (LICS 2003), (2003), 321–330.

[7] A. Bulatov, *H-coloring dichotomy revisited*, Theoret. Sci Comp. Sci. 349,1 (2005), 31–39.

[8] A. Bulatov, *A graph of a relational structure and Constraint Satisfaction Problems*, In: Proceedings of the 19th IEEE Symposium on Logic in Computer Science, (LICS04), 2004.

[9] A. Bulatov, V.Dalmau *Mal'tsev constraints are tractable*, SIAM J. on Computing, 36(1), 2006, 16-27.

[10] A. Bulatov, V. Dalmau, *Towards a dichotomy theorem for the counting constraint satisfaction problem*, Information and Computation, 205(5), 2007, pp. 651-678.

[11] A. Bulatov, P. Jeavons, *Algebraic structures in combinatorial problems*, Int. J. of Algebra and Computing, 2001, submitted.

[12] A. Bulatov, P. Jeavons, A. A. Krokhin, *Constraint satisfaction problems and finite algebras*, Automata, languages and programming (Geneva, 2001), Lecture notes in Comput. Sci., **1853**, Springer, Berlin, (2002), 272–282.

[13] H. Chen, *Quantified Constraint Satsifaction and the polynomially generated powers property*, 35th Colloquium on Automata, Languages and Programming, (ICALP), 2008.

[14] P. G. Jeavons, D. A. Cohen and M. Cooper, *Constraints consistency and closure*, Artificial Intelligence, **101** (1998) 251–265.

[15] P. G. Jeavons, D. A. Cohen and M. Gyssens, *Closure properties of constraints*, J. of the ACM **44** 1997, 527–548.

[16] V. Dalmau, *Generalized majority-minority operations are tractable*, LICS05.

[17] V. Dalmau, P. G. Kolaitis, M. Vardi, *Conjunctive query-containment and constraint satisfaction*, Journal of Computer and System Sciences, **61**(2):302–332, 2000.

[18] Irit Dinur, *The PCP Theorem by gap amplification* JACM, Proc. of 38th STOC, pp. 241-250, 2006.

[19] Irit Dinur, Ehud Friedgut, Oded Regev, *Independent Sets in Graph Powers are Almost Contained in Juntas* Geometric and Functional Analysis, accepted.

[20] Irit Dinur, Elchanan Mossel, Oded Regev, *Conditional Hardness for Approximate Coloring*, Proc. of 38th STOC, pp. 344-353, 2006.

[21] T. Feder, M. Y. Vardi, *The computational stucture of monotone monadic (SNP) and constraint satisfaction: A study through datalog and group theory*, SIAM Journal of Computing **28**, (1998), 57–104.

[22] J. Foniok, J. Nešetřil, C. Tardif, *Generalized dualities and maximal finite antichains in the homomorphism order of relational structures*, KAM-DIMATIA Series 2006-766 (to appear in European J. Comb.).

[23] J. Hastad, *Some optimal inapproximability results*, J. ACM 48(4): 798-859 (2001).

[24] P. Hell, J. Nešetřil, *On the complexity of H-coloring*, J. Combin. Theory Ser. B, **48**, (1990), 92–110.

[25] P. Hell, J. Nešetřil, *Colouring, Constraint Satisfaction and Complexity*, manuscript, 2008. `http://iti.mff.cuni.cz/series/files/iti413.pdf`

[26] P. Hell, J. Nešetřil, X. Zhu, *Duality and polynomial testing of tree homomorphisms*, Transactions of the American Mathematical Society, **348**(4):1281–1297, 1996.

[27] P. G. Jeavons, *On the algebraic structure of combinatorial problems*, Theoretical Computer Science, (1998), **200**, 185-204.

[28] P. G. Kolaitis and M. Y. Vardi, *Conjunctive query containment and constraint satisfaction*, in: Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems, Seattle, Washington, 1998, pp. 205–213.

[29] P. G. Kolaitis, M. Vardi, *A game-theoretic approach to constraint satisfaction*, In Proceedings of the 17th National (US) Conference on Artificial Intelligence, AAAI'00, 175–181, 2000.

[30] G. Kun, *Constraints, MMSNP and expander relational structures*, Combinatorica, submitted.

[31] G. Kun, J. Nešetřil: *Forbidden lifts (NP and CSP for combinatorists)*, European J. Comb. 29 (2008) 930945.

[32] R. E. Ladner: On the structure of Polynomial Time Reducibility, Journal of the ACM, 22,1 (1975), 155–171.

[33] B. Larose, L. Zádori, *Bounded width problems and algebras*, Algebra Universalis, **56**, 439–466, 2007.

[34] B. Larose, L. Zádori, M. Valeriote, *Omitting types, bounded width and the ability to count*, manuscript, 2008.
[35] R. McKenzie, M. Maróti: *Existence theorems for weakly symmetric operations*, Algebra Universalis, to appear.
[36] J. Nešetřil, M. Siggers, L. Zádori: *A Combinatorial constraint satisfaction problem dichotomy classification conjecture*, manuscript, 2007.
[37] J. Nešetřil and C. Tardif, *Duality theorems for finite structures (characterising gaps and good characterizations)*, J. Combin. Theory B **80** (2000), 80–97.
[38] J. Nešetřil and X. Zhu, *On bounded treewidth dualitiy of graphs*, J. of Graph Theory, **23**, 151–162, 1996.
[39] Y. Rabinovich, A. Sinclair, A. Wigderson: *Quadratic Dynamical Systems (Preliminary Version)* FOCS 1992: 304-313.
[40] B. Rossman, *Existential positive types and preservation under homomorphisms*, In 20th IEEE Symposium on Logic in Computer Science, 2005.
[41] T. J. Schaefer, *The complexity of satisfiability problems*, Proc. of the 10th STOC, pp. 216–226, 1978.

9. Appendix

Proof. ((2) → (1), Theorem 11) We show by induction on l that if $r(f^k, 1, \mu)$ goes to zero as k goes to infinite then so does $r(f^k, l, \mu)$ for every $l \geq 1$. Fix $l \geq 2$ and $\varepsilon > 0$, and let k', k'' be such that $r(f^{k'}, 1, \mu) < \varepsilon/l$ and $r(f^{k''}, l-1, \mu) < \varepsilon$, respectively. Let $g = f^{k'}$ and $h = f^{k''}$. For $k = k' + k''$ we have:

$$f^k = h(g, \dots, g).$$

Let $L \subseteq [n]^k$ be a subset of l inputs for f^k. We distinguish between two cases:
Case 1: Each g gets at most one input from L. The output of those that get a label from L is ε/l-close to the distribution $h(\mu^{n^{k'}})$ by the choice of k'. We then use Proposition 13.
Case 2: There is a g which gets at least two labels from L. In that case at most $l-1$ of the gs involve labels from L, and we use that $r(h, l-1, \mu) < \varepsilon$. □

Proof. (of Lemma 15) There are $x, y \in D$ such that

$$\mu(x) - \nu(x) \geq \delta(\mu, \nu)/|D| \tag{6}$$

$$\nu(y) - \mu(y) \geq \delta(\mu, \nu)/|D|. \tag{7}$$

Without loss of generality assume that $i = 1$. Define:

$$\begin{aligned}
p_1 &= Prob_{\mu^n}(y, x, \dots, x) &&= \mu(y)\mu(x)^{n-1}, \\
q_1 &= Prob_{\nu\mu^{n-1}}(y, x, \dots, x) &&= \nu(y)\mu(x)^{n-1}, \\
p_2 &= Prob_{\mu^n}(x, y, x, \dots, x) &&= \mu(x)\mu(y)\mu(x)^{n-2}, \\
q_2 &= Prob_{\nu\mu^{n-1}}(x, y, x, \dots, x) &&= \nu(x)\mu(y)\mu(x)^{n-2}.
\end{aligned}$$

From (6) and (7) we obtain that

$$\begin{aligned}
\mu(x) &\geq \frac{\delta(\mu, \nu)}{|D|}, \\
\mu(y) &\geq \frac{\delta(\mu, \nu)}{|D|} \min \frac{\mu}{\nu}, \\
p_1 - q_1, q_2 - p_2 &> 0.
\end{aligned}$$

Let $a = f(y, x, \dots, x) = f(x, y, x \dots, x)$. From (5):

$$\begin{aligned}
\delta(f(\mu^{i-1}\nu\mu^{n-i}), f(\mu^n)) &\leq \delta(\mu^{i-1}\nu\mu^{n-i}, \mu^n) - \min\{p_1 - q_1, q_2 - p_2\} \\
&\leq \delta(\mu, \nu) - \frac{\delta(\mu, \nu)^n}{|D|^n} \min \frac{\mu}{\nu}.
\end{aligned}$$

□