Fine-Grained Analysis of Problems on Curves

Kevin Buchin[1] Maike Buchin[2] Maximilian Konzack[3] Wolfgang Mulzer[4] André Schulz[5]

Abstract

We provide conditional lower bounds on two problems on polygonal curves. First, we generalize a recent result on the (discrete) Fréchet distance to k curves. Specifically, we show that, assuming the Strong Exponential Time Hypothesis, the Fréchet distance between k polygonal curves in the plane with n edges cannot be computed in $O(n^{k-\varepsilon})$ time, for any $\varepsilon > 0$. Our second construction shows that under the same assumption a polygonal curve with n edges in dimension $\Omega(\log n)$ cannot be simplified optimally in $O(n^{2-\varepsilon})$ time.

1 Introduction

The *fine-grained complexity* of the (discrete) Fréchet distance between two curves has recently attracted a lot of attention. After a long period without major progress, Agarwal et al. devised a subquadratic $O\left(\frac{mn \log\log n}{\log n}\right)$-time algorithm for the discrete Fréchet distance on the word RAM [2]. Buchin et al. [10] gave a randomized algorithm for the continuous Fréchet distance with a running time slightly better than the classic bound of $O(n^2 \log n)$ [4]. Answering a question by Buchin et al. [10], Bringmann [6] showed that the (discrete) Fréchet distance cannot be computed in $O(n^{2-\varepsilon})$ time, for any $\varepsilon > 0$, assuming the *Strong Exponential Time Hypothesis* (SETH). This result was later refined and extended [8]. SETH states that for every $\varepsilon > 0$, there is a $k \in \mathbb{N}$ such that the satisfiability problem on k-CNF formulas with n variables and m clauses cannot be solved in time $m^{O(1)}2^{(1-\varepsilon)n}$.

Bringmann's work [6] triggered a lot of activity, leading to new conditional lower bounds for famous problems such as edit distance, dynamic-time warping, or longest common subsequence (LCS) [1,7]. For LCS, a more general bound states the non-existence of a $O(n^{k-\varepsilon})$-time algorithm for k strings over an alphabet of size $O(k)$. Our first result generalizes the lower bound on the discrete Fréchet distance to k curves.

Theorem 1 *For any $\varepsilon > 0$, the discrete Fréchet distance of k planar point sequences of length n cannot be decided in $O(n^{k-\varepsilon})$ time, unless SETH fails.*

The Fréchet distance between k curves was considered previously by Rote and Dumitrescu [12], who provide a near-quadratic time 2-approximation for the problem. Measuring the distance and analyzing the similarity between a set of parameterized curves in this way is also relevant for movement data analysis. For instance, it can be used in the analysis of collective movement, e.g., within a flock of birds, or to detect clusters of similar movement sequences [9].

Our second result is on simplifying a d-dimensional polygonal curve $\langle a_0, \dots, a_n \rangle$. We consider the common variant [13] where the vertices of the simplified curve should be an ordered subsequence of the original vertices, and if a_i and a_j are consecutive in the simplification, then the distance between the subcurve $\langle a_i, a_{i+1}, \dots, a_j \rangle$ and the line segment $a_i a_j$ should be at most a given $\varepsilon > 0$. We focus on the Hausdorff distance, although the reduction also applies to the Fréchet distance. There are two common variants of the simplification problem: *min-#*, in which ε is given and the number of vertices is to be minimized, and *min-ε* in which an upper bound on the number of vertices is given and ε is to be minimized.

Algorithms for the min-ε and the min-# problems with running time $O(n^2 \log n)$ and $O(n^2)$, respectively, are known for polygonal curves in the plane [11]. For the L_1-metric, Agarwal and Varadarajan [3] presented an $O(n^{4/3+\varepsilon})$-time algorithm. For curves in $\mathbb{R}^d$, Barequet et al. [5] developed efficient algorithms. Their algorithms run in near-quadratic time for $d = 3$ and in subcubic time for $d = 4$. If distance is measured according to the L_1- or the L_∞-metric, they achieve a running time of $O(n^2)$ and $O(n^2 \log n)$ for min-ε and min-#, respectively, in any fixed dimension. In particular, for L_∞ the dependency on the dimension is only a small-degree polynomial. It is a long-standing open problem whether the (near-)quadratic running time can be improved for the Euclidean distance [3].[1]

We show that, at least in sufficiently high (non-constant) dimension, this is not possible unless SETH fails. For L_∞, our construction shows that the algorithm by Barequet et al. is essentially optimal in high dimensions, assuming SETH.

Theorem 2 *There is no algorithm that optimally, min-# or min-ε, simplifies a polygonal curve with n edges in $\mathbb{R}^d$ with $d = \Omega(\log n)$ using ε-tolerance regions*

[1] TU Eindhoven, the Netherlands k.a.buchin@tue.nl
[2] RU Bochum, Germany maike.buchin@rub.de
[3] TU Eindhoven, the Netherlands m.p.konzack@tue.nl
[4] FU Berlin, Germany mulzer@inf.fu-berlin.de
[5] FeU Hagen, Germany andre.schulz@fernuni-hagen.de

[1] See also http://cs.smith.edu/~orourke/TOPP/P24.html.

in the L_1-, L_2- or L_∞-metric that runs in $O(n^{2-\varepsilon})$ time, for any $\varepsilon > 0$, unless SETH fails.

To prove the lower bounds, we use a reduction from the k-Orthogonal Vectors problem (as stated in [1]), using the notation $[n] := \{1, \dots, n\}$.

Definition 1 (k-Orthogonal-Vectors) *Suppose we are given k lists $\{\alpha_i^1\}_{i\in[n]}$, $\{\alpha_i^2\}_{i\in[n]}$, $\dots$, $\{\alpha_i^k\}_{i\in[n]}$ of vectors in $\{0,1\}^d$. We need to decide whether there are k vectors $\alpha_{i_1}^1$, $\alpha_{i_2}^2$, $\dots$, $\alpha_{i_k}^k$ with $\sum_{h=1}^d \prod_{t\in[k]} \alpha_{i_t}^t[h] = 0$. Any such collection of vectors is called* orthogonal.

The following lemma is well known [1,14].

Lemma 3 *If there is an $\varepsilon > 0$ such that k-Orthogonal Vectors on n vectors in $\{0,1\}^d$ with $d = \Omega(\log n)$ can be solved in $O(n^{k-\varepsilon})$ time, then SETH is false.*

2 Fréchet distance between k curves

We show the lower bound on the discrete Fréchet distance between k curves by a reduction from the k-Orthogonal Vectors problem. We begin with some notation. Let $A_1, \dots, A_k$ be k sequences of points in the plane, $A_i = \langle a_1^i, \dots, a_{n_i}^i \rangle$. By $a_j^i[h]$, for $h = 1, 2$, we denote the h-th coordinate of a_j^i. We set $S = [n_1] \times [n_2] \times \cdots \times [n_k]$.

We define a *coupling* of length m on S as a sequence $\mathcal{C} = \langle \mathcal{C}_1, \dots, \mathcal{C}_m \rangle$ such that we have $\mathcal{C}_i \in S$, $\mathcal{C}_1 = (0, 0, \dots, 0)$, $\mathcal{C}_m = (n_1, n_2, \dots, n_k)$, and $\mathcal{C}_{i+1}[h] = \mathcal{C}_i[h]$ or $\mathcal{C}_{i+1}[h] = \mathcal{C}_i[h] + 1$, for all $i = 0, \dots, m-1$ and $h = 1, \dots, k$. A coupling $\mathcal{C}$ defines an alignment of the curves $A_1, \dots, A_k$, and we define the *coupled distance* as $d_{\mathcal{C}}(A_1, \dots, A_k) := \max\left\{ d(a_{\mathcal{C}_i[h]}^h, a_{\mathcal{C}_i[h']}^{h'}) \mid 0 \le i \le m, 1 \le h, h' \le k \right\}$, where d denotes the Euclidean distance. The *discrete Fréchet distance* $d_F(A_1, \dots, A_k)$ between the k curves is the minimal coupled distance over all possible couplings.

Next, we describe our reduction. Suppose we have k lists $\{\beta_i\}_{i\in[n]}$, $\{\alpha_i^t\}_{i\in[n]}$, $t \in [k-1]$, of vectors $\alpha_i^t, \beta_i \in \{0,1\}^d$. We construct k curves $B, A^1, A^2, \dots, A^{k-1}$. Their discrete Fréchet distance will be 1 if the given vector lists contain a collection of k orthogonal vectors, and strictly larger than 1, otherwise. The coordinates of the vectors are encoded by *coordinate gadgets* (CG), see Figure 1. Set $\delta := 1/100$, and for $i = 1, \dots, k-1$, let $CG_i(0) := \langle(-0.5 - \delta, 0), (0.5, 0), (-0.5 - \delta, 0), \dots, (0.5, 0), (-0.5 - \delta, 0)\rangle$ be a curve with $2k-1$ vertices. We define $CG_i(1)$ to have the same vertices as $CG_i(0)$, except that the $2i$-th vertex is replaced by $(0.5 + \delta, 0)$. Further we define $CG_B(0) := \langle(-0.5, 0), (0.5, 0), (-0.5, 0), \dots, (0.5, 0), (-0.5, 0)\rangle$ with $2k-1$ vertices and $CG_B(1)$ in the same way but with only $2k-3$ vertices. We call the vertices at $(0.5, 0)$ *short spikes* and at $(0.5 + \delta, 0)$ *long spikes.*

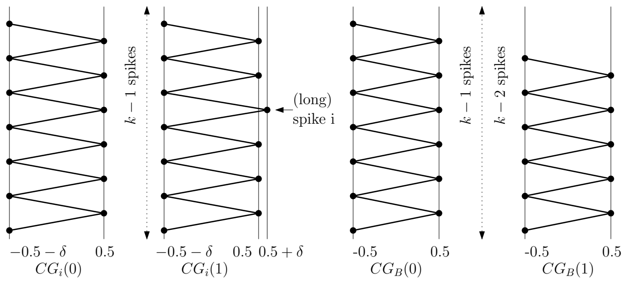

Figure 1: Coordinate gadgets (distorted vertically for the purpose of illustration).

Suppose that there were a coupling of $CG_1(1)$, $CG_2(1), \dots, CG_{k-1}(1)$, $CG_B(1)$ achieving a distance of at most 1. Then, $k-1$ spikes of $CG_1(1), \dots, CG_{k-1}(1)$ need to be coupled, but there is one long spike in each coupled spike. We need to couple every long spike with a different spike of $CG_B(1)$. This is not possible, since $CG_B(1)$ has only $k-2$ spikes. Thus, $d_F(CG_1(1), \dots, CG_B(1)) > 1$. If we replace any $CG_*(1)$ with a respective curve $CG_*(0)$, the distance becomes 1.

Next, we encode the vectors and the vector lists. To "synchronize" coordinates, we will use the point $c := (0, 0.8661)$. The start of vectors will be demarcated by $v_A := (-0.499, -1)$ and $v_B := (0, -0.8661)$. Additionally, we will use the points $t_A = (0.48, -0.01)$ and $t_B = (0.57, 1.005)$ to mark a successful synchronized traversal, and $s = (-0.499, 0)$ as a point that is close to all except t_B, see Figure 2.

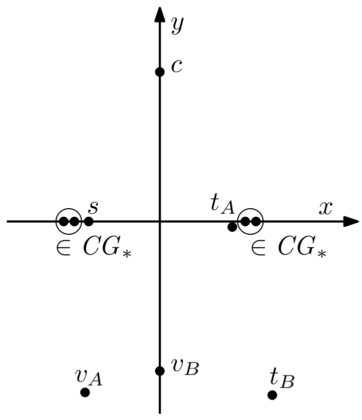

Figure 2: The points used as vertices of the curves.

Two points are *close* if their distance is at most 1: s is close to all points except t_B, and t_A is close to all except v_A. The point c is close only to s and t_A (and itself); t_B is close only to t_A and v_B; v_A is close only to s and v_B; v_B only to s, t_A and t_B.

Let $A_i^j := s \circ v_A \circ \bigcirc_{h=1}^d (CG_j(\alpha_i^j[h]) \circ c) \circ t_A$, where we use $\circ$ to denote the operation of adding a vertex to a curve or of concatenating curves. We set $A^j := \left(\bigcirc_{i=1}^{n} A_i^j\right) \circ s$. Furthermore, we define $B_i := v_B \circ \bigcirc_{h=1}^{d}(CG_j(\beta_i[h]) \circ c)$ and $B := s \circ v_A \circ \bigcirc_{i=1}^{n} B_i \circ t_B \circ s$.

First, we argue that k vectors $\alpha^1_{i_1}, \alpha^2_{i_2}, \ldots, \alpha^{k-1}_{i_{k-1}}, \beta_{i_k}$ are orthogonal if and only if the corresponding concatenated coordinate gadgets have Fréchet distance at most 1. If the vectors are orthogonal, then in each coordinate, at least one vector has a 0-entry, and a coupling of distance at most 1 is possible. On the other hand, if the vectors are not orthogonal, there is one coordinate in which all vectors have 1-entries. The c vertices force us to traverse all coordinates simultenously, so that we will have to couple k one-coordinate gadgets, giving a Fréchet distance larger than 1.

Now, let us consider the vector lists and the complete curves. If the vector lists contain a k-tuple $\alpha^1_{i_1}, \alpha^2_{i_2}, \ldots, \alpha^{k-1}_{i_{k-1}}, \beta_{i_k}$ of orthogonal vectors, then the curves $A^1, \ldots, A^{k-1}, B$ have Fréchet distance at most 1. This can be seen by the following coupling: first, A^1 walks to the first point s of $A^1_{i_1}$, while all other curves wait at s. Then, A^2 walks to the first point s of $A^2_{i_2}$, while all other curves wait at s, etc. Finally, B walks to the first point v_B of B_{i_k}, while all other curves wait at s. Since s is close to all points except for t_B, the distance so far is 1. Then, the A^j curves simultaneously jump to v_A while B waits at v_B, and then the coordinate gadgets are traversed simultaneously. Next, the A^j curves wait at t_A while B walks to the last point s. The A^j then simultaneously go to the next s, and finish the walk to the final vertex one after another while the other curves wait at s.

Next, suppose that the curves have Fréchet distance larger than 1. We argue that then there is a k-tuple of orthogonal vectors. Indeed, suppose that no such k-tuple exists, and consider the first time that B reaches t_B. Since t_B is close only to t_A and v_B, at this point, all A^j must be at t_A. It follows that before that, all A^j's must have been simultaneously at v_A, because on the A^j's, v_A comes before t_A, and v_A is close only to s and v_B. For the same reason, at this point, B also must be at v_B. Then, the coordinate gadgets of a k-tuple of vectors are traversed simultaneously, leading to Fréchet distance larger than 1, as all k-tuples are non-orthogonal. Theorem 1 follows.

Our construction also rules out a faster polynomial-time approximation scheme unless SETH fails. The coordinates were computed by hand and could be optimized to prove a specific approximation lower bound.

3 Curve Simplification

In this section, we reduce the 2-Orthogonal Vectors problem to the curve simplification problem. Given two lists of 0/1-vectors $\{\alpha_i\}_{i\in[n]}$ and $\{\beta_i\}_{i\in[n]}$ in dimension d, we interpret each vector as a point in dimensions $d+1$, as follows: we define $\hat{\alpha}_i[h] := \alpha_i[h]$

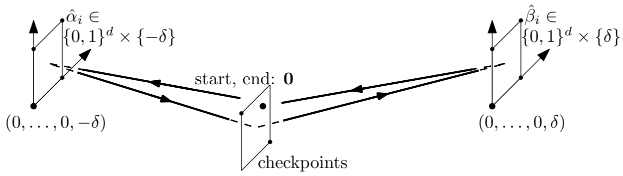

Figure 3: Construction for simplification lower bound.

for $1 \le h \le d$ and $\hat{\alpha}_i[d+1] := -\delta$ with $\delta = 2d^2$. We define $\hat{\beta}_i[h]$ analogously, except that $\hat{\beta}_i[d+1] := \delta$.

The idea of the reduction is illustrated in Figure 3. We construct a curve that moves from a starting point through all $\hat{\alpha}_i$, then through d "checkpoints", through all $\hat{\beta}_i$, and finally to an endpoint. The threshold ε for the simplification is chosen such that all points $\hat{\alpha}_i$ have pairwise distance smaller than ε, and similarly for the points $\hat{\beta}_i$. Thus, the intended simplification uses the starting point, one point $\hat{\alpha}_i$, one point $\hat{\beta}_j$, and the endpoint. The checkpoints will have distance at most ε to the edge from $\hat{\alpha}_i$ to $\hat{\beta}_j$ exactly if the two corresponding vectors are orthogonal.

For $1 \le i \le d$, let $q_i \in \mathbb{R}^{d+1}$ be defined as $q_i[i] = -\delta'$, $q_i[d+1] = 0$, and $q_i[\cdot] = 1/4$, otherwise, where δ' will be chosen later depending on the metric. We define a curve $A = \langle a_0, \ldots, a_m \rangle$ with $m = 2n+2+d$ vertices by $a_0 = a_m = (0,\ldots,0)$, $a_i = \hat{\alpha}_i$, for $1 \le i \le n$, $a_{n+i} = q_i$, for $1 \le i \le d$, and $a_{n+d+i} = \hat{\beta}_i$, for $1 \le i \le n$.

We first consider a simplification under the L_∞-metric. We set $\varepsilon = 1$ and $\delta' = 1/2$. By the choice of ε and δ, the simplification needs to include at least a_0, one point $\hat{\alpha}_i$, one point $\hat{\beta}_j$, and a_m. Assume there are orthogonal vectors α_i and β_j. Let $\ell(t)$ be the line segment between $\hat{\alpha}_i$ and $\hat{\beta}_j$ parameterized by t in the $(d+1)$-th coordinate. For the midpoint $\ell(0)$ of the segment we have $\ell(0)[h] = (\hat{\alpha}_i + \hat{\beta}_j)/2 \in \{0, 1/2\}$, for $1 \le h \le d$ (and $\ell(0)[d+1] = 0$). Hence all q_i have distance less than 1 to $\ell(0)$ and are therefore within distance ε to the segment. In contrast, let us assume α_i and β_j are nonorthogonal. In this case there is a coordinate $1 \le h \le d$ such that $\hat{\alpha}_i[h] = \hat{\beta}_j[h] = 1$. It follows that $\ell(t)[h] = 1$ for all $t \in [-\delta, \delta]$, and therefore $d_\infty(\ell(t), q_h) \ge 1 - q_h[h] > 1 = \varepsilon$. Thus, if we choose this segment, q_h has distance larger than ε to the segment. Consequently, if there is no pair of orthogonal vectors, a simplification for distance ε requires at least 5 vertices.

For the L_1-metric we set $\varepsilon = d$ and $\delta' = 3/4d - 1/4$. By the same argument as for L_∞, we get that if there are orthogonal vectors, then they induce a simplification with 4 vertices, since $d_1(q_i, \ell(0)) \le (d-1)/4 + \delta' + 1/2 = d = \varepsilon$. Now again consider the case that all α_i and β_j are nonorthogonal, so there is a coordinate $1 \le h_0 \le d$, such that $\hat{\alpha}_i[h_0] = \hat{\beta}_j[h_0] = 1$. We show that $d_1(\ell(t), q_h) > \varepsilon$ for all $t \in [-\delta, \delta]$. We can restrict our attention to $t \in [-\varepsilon, \varepsilon]$ due to the $(d+1)$-th coordinate. Now consider a coordinate $h \neq h_0, d+1$. If $\alpha_i[h] = \beta_j[h] = 0$, then $\ell(t)[h] = 0$. Otherwise $\ell(0)[h] \geq 1/2$ and $\ell(t)[h] \geq \ell(0)[h](1 - \varepsilon/\delta)$ for $t \in [-\varepsilon, \varepsilon]$. Consequently, for any t we get that $d_1(\ell(t), q_h) \geq (d-1)(1/2(1-\varepsilon/\delta) - 1/4) + 1 - \delta' = [(d-1)/4 + \delta' + 1/2] + 1/2 - (d-1)\varepsilon/\delta = \varepsilon + 1/2 - (d-1)\varepsilon/\delta > \varepsilon$. Thus, there is a simplification using 4 vertices exactly if there is an orthogonal pair.

For the L_2-metric we set $\varepsilon = \sqrt{d}$. Further we fix $\delta' = -1/2 + \sqrt{15d+1}/4$, which implies that $\delta' > 0$ and that $\sqrt{(d-1)/4 + (1/2 + \delta')^2} = \varepsilon$. By the choice of δ' orthogonal vectors, we induce points with all q_i having distance at most ε to the segment. Now again consider a pair of nonorthogonal vectors with $\alpha_i[h_0] = \beta_j[h_0] = 1$. It is sufficient then to prove that $d_2(\ell(t), q_h)^2 > \varepsilon^2 = d$ for $t \in [-\varepsilon, \varepsilon]$. Using the same derivation as for L_1, we obtain $d_2(\ell(t), q_h)^2 \geq (1+\delta')^2 + (d-1)(1/4 - \varepsilon/\delta/2)^2$. The first summand is larger than $15/16d + 1/16 + 1/4$ while the second is larger than $(d-1)/16 - (d-1)\varepsilon/\delta/4 > (d-1)/16 - 1/8$. Hence, q_h has a distance larger than ε to the segment.

As a result of this, we can reduce the 2-Orthogonal Vectors problem in dimension d to curve simplification, min-# or min-ε, in dimension $d+1$ for the L_1, L_2 and L_∞ metrics. Theorem 2 follows.

4 Conclusion and Open Problems

We have extended the recent conditional fine-grained hardness results for the Fréchet distance to the case of k curves and to the curve simplification setting, showing that any significant improvement on known methods would result in a major breakthrough in satisfiability algorithms.

We find the curve simplification result particularly intriguing, since in seems to offer a qualitative difference from the previous work: in the curve simplification setting, we have only one input object that needs to be compared to itself. Are there problems of similar flavor where analogous conditional hardness results can be obtained?

Acknowledgements. KB and MK are supported by the Netherlands Organisation for Scientific Research (NWO) under project no. 612.001.207. WM supported in part by DFG project MU 3501-1.

References

[1] A. Abboud, A. Backurs, and V. V. Williams. Tight hardness results for LCS and other sequence similarity measures. In *Proc. 56th Annu. IEEE Sympos. Found. Comput. Sci. (FOCS)*, pages 59–78, 2015.

[2] P. K. Agarwal, R. B. Avraham, H. Kaplan, and M. Sharir. Computing the discrete Fréchet distance in subquadratic time. *SIAM J. Comput.*, 43(2):429–449, 2014.

[3] P. K. Agarwal and K. R. Varadarajan. Efficient algorithms for approximating polygonal chains. *Discrete Comput. Geom.*, 23(2):273–291, 2000.

[4] H. Alt and M. Godau. Computing the Fréchet distance between two polygonal curves. *Internat. J. Comput. Geom. Appl.*, 5(1–2):78–99, 1995.

[5] G. Barequet, D. Z. Chen, O. Daescu, M. T. Goodrich, and J. Snoeyink. Efficiently approximating polygonal paths in three and higher dimensions. *Algorithmica*, 33(2):150–167, 2002.

[6] K. Bringmann. Why walking the dog takes time: Fréchet distance has no strongly subquadratic algorithms unless SETH fails. In *Proc. 55th Annu. IEEE Sympos. Found. Comput. Sci. (FOCS)*, pages 661–670, 2014.

[7] K. Bringmann and M. Künnemann. Quadratic conditional lower bounds for string problems and dynamic time warping. In *Proc. 56th Annu. IEEE Sympos. Found. Comput. Sci. (FOCS)*, pages 79–97, 2015.

[8] K. Bringmann and W. Mulzer. Approximability of the discrete Fréchet distance. *Journal of Computational Geometry*, 7(2):46–76, 2016.

[9] K. Buchin, M. Buchin, J. Gudmundsson, M. Löffler, and J. Luo. Detecting commuting patterns by clustering subtrajectories. *Internat. J. Comput. Geom. Appl.*, 21(3):253–282, 2011.

[10] K. Buchin, M. Buchin, W. Meulemans, and W. Mulzer. Four Soviets walk the dog - with an application to Alt's conjecture. In *Proc. 25th Annu. ACM-SIAM Sympos. Discrete Algorithms (SODA)*, pages 1399–1413, 2014.

[11] W. S. Chan and F. Chin. Approximation of polygonal curves with minimum number of line segments or minimum error. *International Journal of Computational Geometry & Applications*, 6(01):59–77, 1996.

[12] A. Dumitrescu and G. Rote. On the Fréchet distance of a set of curves. In *Canad. Conf. Comput. Geom. (CCCG)*, pages 162–165, 2004.

[13] H. Imai and M. Iri. Polygonal approximations of a curve – formulations and algorithms. In G. T. Toussaint, editor, *Computational Morphology*, pages 71–86. Elsevier Science, 1988.

[14] R. Williams. A new algorithm for optimal 2-constraint satisfaction and its implications. *Theoretical Computer Science*, 348(2):357–365, 2005.