A SIMPLE EXPERIMENT ON VIRO'S PATCHWORKING TECHNIQUE

ALPEREN ERGÜR, ABIGAIL MARTINEZ, JESUS REBOLLO-BUENO, AND IAN SOLIS

Abstract. Viro's patchworking provides technique to create a real algebraic sets with a prescribed topology. The inverse direction of this, that is, if 'most' topology types are handled by Viro's patchworking is a widely open problem where expert intuitions have proven to be wrong. This is an experiment in the most simple case to develop some understanding of this matter.

1. Introduction

Viro's patchworking method is a beautiful technique to create real algebraic sets with prescribed topology. It lead to breakthroughs in our understanding of Hilbert's 16th problem. It was claimed that in zero-dimensional case Viro's patchworking captures essentially all potential behaviour, which was later refuted. After developing an algorithm based on Viro's technique, Ergür received the same question in emails from several experts: If we randomly create a polynomial is it going to be a patchworked zero set? Here we would like to perform an experiment in the most simple case of this and for an even more relaxed version of this question: Is the expected topology (sum of Betti numbers or number of connected components) of a randomly created degree d n-variate patchworked set is equivalent (up to constants) to the expected topology of random degree d n-variate real hypersurface? We push simplicity to extreme and do the experiment with $n = 1$.

2. Viro's Patchworking Method for Univariate Polynomials

We need to begin with a definition.

Definition 2.1 (Lower Convex Hull)**.** *Let $A = \{(0, f_1), (1, f_2), \ldots, (d, f_d)\}$ be a set of vectors in $\mathbb{R}^2$. We say $\{(i, f_i), (k, f_k)\}$ is an edge in the lower convex hull of A if for all $i < j < k$ we have*

$$\frac{f_j - f_i}{j - i} \geq \frac{f_k - f_i}{k - i}$$

Or equivalently, if for all all j with $i < j < k$ we have

$$f_j(k - i) - f_i(k - j) - f_k(j - i) \geq 0$$

We call the union of these edges the lower convex hull of A.

Let $p(x) = c_0 + c_1x + c_2x^2 + \ldots + c_dx^d$ be a polynomial where the coefficients c_i are real numbers. We use $sgn(c_i)$ to denote the sign of c_i, and we define $f_i := -\ln|c_i|$. Now consider the edges in the lower convex hull of the set $A = \{(0, f_1), (1, f_2), \ldots, (d, f_d)\}$. More precisely, we consider the following set

$$V := \{i : 0 \leq i \leq d \text{ and } i \text{ is an endpoint of an edge in the lower convex hull of A}\}$$

Note that V is a collection of integers between 0 and d. We need to attach to sign vectors to set V.

$$S_1(V) := \{sgn(c_i) : i \in V\}\ ,\ S_2(V) := \{(-1)^i sgn(c_i) : i \in V\}$$

Since $S_1(V)$ is a sign vector, we count the number of sign changes in it. Likewise we can count the number of sign changes in $S_2(V)$. Our heuristic method claims that the number

2010 *Mathematics Subject Classification.* Primary ; Secondary .

Key words and phrases. Viro's Patchworking Methods, Randomization, Real Zero Count.

of positive real zeros of p equals to the number of sign changes in $S_1(V)$, and the number of negative real zeros of p equals to the number of sign changes in $S_2(V)$.

Example 2.2. *Let us consider a polynomial with degree 4,* $p(x) = 1 - e^{-1}x + e^{-3}x^2 - e^{-6}x^3 + e^{-10}x^4$*. The corresponding set is* $A = \{(0,0),(1,1),(2,3),(3,6),(4,10)\}$*. Then the lower convex hull has four edges, and we have that* $V = \{0,1,2,3,4\}$*. Then,*

$$S_1(V) = \{+,-,+,-,+\}, S_2(V) = \{+,+,+,+,+\}$$

The number sign changes in $S_1(V)$ *is four; count the sign change from left to right starting with + and eventually ending with +. The number sign changes in* $S_2(V)$ *is zero. Clearly, the polynomial p has no negative real zeros. So the number sign changes in* $S_2(V)$ *gives the correct count. The sign changes in* $S_1(V)$ *is dubious.*

3. Models of Randomness

For a given degree d, we consider three families of randomly generated univariate polynomials.

(1) (Kac Ensemble) Let $p(x) = a_0 + a_1x + a_2x^2 + \ldots + a_dx^d$ where a_i are independent identically distributed (i.i.d.) Gaussian random variables with mean zero and variance one. Then, we know that $\mathbb{E}\,|Z_{\mathbb{R}}(p)| \sim \log(d)$. In words; the number expected real zeros of a Kac random polynomial is about $\log(d)$.
(2) (Kostlan or Elliptic Ensemble) Let $p(x) = a_0 + a_1x + a_2x^2 + \ldots + a_dx^d$ where a_i are independent Gaussian random variables with mean zero and variance $\sqrt{\binom{d}{i}}$. Then, we know that $\mathbb{E}\,|Z_{\mathbb{R}}(p)| \sim \sqrt{d}$.
(3) (Flat Ensemble) Let $p(x) = a_0 + a_1x + a_2x^2 + \ldots + a_dx^d$ where a_i are independent Gaussian random variables with mean zero and variance $\sqrt{\frac{1}{i!}}$. Then, we know that $\mathbb{E}\,|Z_{\mathbb{R}}(p)| \sim \sqrt{d}$.

4. The Experiment

For a fixed degree d, we generate 1000 data points from each three ensembles described at the beginning of this note.

(1) Using a computer algebra software we compute average number real zeros for each of these 1000 data point groups. We create three plots. In these plots we need to visualize the number of real zeros of the 1000 randomly generated polynomials. For instance, if 37 out 1000 polynomials have 1 real root then this gives 0.037 density to the 1 real root in our visualization.
(2) Using our own implementation of the heuristic idea, we count real zeros of these randomly generated polynomials and create similar data visualization.

We repeat the experiment for a moderate amount of degrees d and collect beautiful visuals. This is the end of the first part of the experiment.

Second part of the experiment is as follows. Select many degrees d, and do the following:

(1) For each degree d, generate 1000 random samples from the three ensembles. This time we only need to compute average number of real zeros of the 1000 samples for each of the three ensembles. Basically, you have three means and that's all.

(2) Using our implementation, compute the average number of real zeros predicted by the heuristic for these data sets. Basically, you have three mean numbers generated by our heuristic.
(3) Create three plots corresponding to three ensembles. In each plot we need to see the number of average real zeros vs. the degree d according to correct count coming from the computer algebra packages and also according to our heuristic idea.
(4) Depending on how the plots look, we might want to increase the sample size 1000. Computer algebra packages are relatively quick, the limitations on the sample size will mainly depend on the capabilities of our implementation.

5. Tricks to help implementation issues

Let $p(x) = a_0 + a_1x + a_2x^2 + \ldots + a_dx^d$ be a polynomial. The number of real zeros of p, and the following polynomials p_1, p_2, and p_3 are the same.

- Let $r_d > 0$ be a real number, and let
$$p_1 := r_da_0 + r_da_1x + ra_2x^2 + \ldots + r_da_dx^d$$
- Let $c_d > 0$ be a real number, and let
$$p_2 := a_0 + a_1(\frac{x}{c_d}) + a_2(\frac{x}{c_d})^2 + \ldots + a_d(\frac{x}{c_d})^d$$
- Let $r_d, c_d > 0$ be two real numbers and let
$$p_3 := r_da_0 + r_da_1(\frac{x}{c_d}) + r_da_2(\frac{x}{c_d})^2 + \ldots + r_da_d(\frac{x}{c_d})^d$$

In quick summary, for a given polynomial $p(x) = \sum_{i=0}^{d} a_ix^i$ we can pick two positive real numbers $r_d, c_d > 0$ and consider

$$q(x) = r_d \sum_{i=0}^{d} a_ic_d^{-i}x^i \tag{1}$$

with the same number of real zeros if $q(x)$ is better for implementation purposes.

Now let us recall Stirling's approximation hrefhttp://page.mi.fu-berlin.de/shagnik/notes/binomials.pdfar exposition:

$$\sqrt{2\pi k}(\frac{k}{e})^k < k! < \sqrt{2\pi k}(\frac{k}{e})^k e \tag{2}$$

and also

$$(\frac{n}{k})^k < \binom{n}{k} < (\frac{ne}{k})^k. \tag{3}$$

For Flat: With the above reasoning in mind, as considering $\frac{1}{k!}$ variances results in numbers python cannot compute, we want to use Stirling's approximation together with (1) to consider variances $\frac{\text{constant}}{k!} \sim 1$. Therefore, using (2), for all $k = 0, \ldots, d$

$$\frac{\sqrt{2\pi k}\left(\frac{k}{e}\right)^k}{k!} < 1 < \frac{\sqrt{2\pi k}\left(\frac{k}{e}\right)^k e}{k!}. \tag{4}$$

At this stage, by using the left-hand side of (4), we could consider the polynomial

$$\tilde{p}(x) = \sum_{k=0}^{d} \sqrt{2\pi k}\, a_k \left(\frac{x}{\left(\frac{e}{k}\right)}\right)^k. \tag{5}$$

Notice, we have here the amounts $r_k = \sqrt{2\pi k}$ and $c_k = \frac{e}{k}$ which are dependent of k, so $\tilde{p}$ has not the same number of zeros as p_1. However, we have the flexibility to make a choice for $k \in \{0, \ldots, d\}$. Although we do not know the optimal choice of k for now (it could potentially be $\lfloor \frac{d}{2} \rfloor$), we try for now with $k = d$. That is, we consider the polynomial

$$\tilde{p}(x) = \sum_{k=0}^{d} \sqrt{2\pi d}\, a_k \left(\frac{x}{\left(\frac{e}{d}\right)} \right)^k ,$$

choosing $r_d = \sqrt{2\pi d}$ and $c_d = \left(\frac{e}{d}\right)$. This suggests that instead of generating coefficients $a_k^\prime s$ with variances $\frac{1}{k!}$ we could generate coefficients $b_k^\prime s$ with variances

$$b_k = \frac{\sqrt{2\pi d}\, d^k}{e^k\, k!} = \sqrt{2\pi d} \prod_{j=0}^{k-1} \frac{d}{e\,(k-j)}. \tag{6}$$

University of Texas at San Antonio, One UTSA Circle, San Antonio, TX, 78249, USA
Email address: alperen.ergur@utsa.edu

University of Texas at San Antonio, One UTSA Circle, San Antonio, TX, 78249, USA
Email address: abigail.martinez@utsa.edu

University of Texas at San Antonio, One UTSA Circle, San Antonio, TX, 78249, USA
Email address: jesus.rebollobueno@utsa.edu

University of Texas at San Antonio, One UTSA Circle, San Antonio, TX, 78249, USA
Email address: iasolis418@gmail.com