Алгебра и анализ
Том 28 (2016), № 6

St. Petersburg Math. J.
Vol. 28 (2017), No. 6, Pages 783–787
http://dx.doi.org/10.1090/spmj/1473
Article electronically published on October 2, 2017

CORRECTING CONTINUOUS HYPERGRAPHS

F. PETROV

Abstract. A general result in the spirit of the continuous hypergraph removal lemma is stated and proved: if a "closed" property of values of a measurable function on $[0,1]^n$ holds almost everywhere, then the function may be changed on a set of measure 0 so that this property holds everywhere. It is also shown that in some situations a discrete version fails.

§1. Positive statement

Our result generalizes some specific lemmas of [2, 3, 5]. Statements of this type are used in the theory of continuous graphs, see the book [4] and the references therein. Discrete versions have various applications in combinatorics.

We introduce the necessary notation. Let K be a compact metric space, and let X be a standard continuous measure space (say, $X = [0,1]$ with Lebesgue measure). Next, let k be a positive integer and $f(x_1, x_2, \dots, x_k)\colon X^k \to K$ a measurable function (with respect to the Borel σ-algebra on K). We denote by $\mathcal{N}_k$ the set of all ordered k-tuples $I = (i_1, \dots, i_k)$ of mutually distinct positive integers. For any such k-tuple I and any points $y_1, y_2, \dots$ in X, we denote $f_I(y_1, y_2, \dots) := f(y_{i_1}, \dots, y_{i_k})$. So, f induces a map $\widetilde{f}$ from $X^{\mathbb{N}}$ to Tychonoff's compact set $K^{\mathcal{N}_k}$ that sends a point $(y_1, y_2, \dots) \in X^{\mathbb{N}}$ to the function $I \to f_I(y_1, y_2, \dots,)$ on $\mathcal{N}_k$.

Theorem 1. 1. *Let M be a fixed closed subset of $K^{\mathcal{N}_k}$. Assume that for almost all $y_1, y_2, \dots$ in X the value $\widetilde{f}(y_1, y_2, \dots)$ belongs to M. Then there exists a measurable function g on X^k equivalent to f and such that $\widetilde{g}(y_1, y_2, \dots)$ belongs to M for all mutually different $y_1, y_2, \dots$ in X.*

2. *Assume additionally that f is "almost symmetric", i.e.,*

$$f(x_1, \dots, x_k) = f(x_{\pi_1}, \dots, x_{\pi_k}) \tag{1}$$

for almost all $x_1, \dots, x_k$ in X and any permutation π of the set $\{1, 2, \dots, k\}$. Then there exists a symmetric measurable function g on X^k equivalent to f and such that $\widetilde{g}(y_1, y_2, \dots)$ belongs to M for all (not necessary different) $y_1, y_2, \dots$ in X.

Before passing to the proof, we mention some examples.

Example 1. Suppose $k = 2$, $K = \{0, 1\}$, and for almost all y_1, y_2, y_3 we have

$$f(y_1, y_2) = f(y_2, y_1), \quad f(y_1, y_2) \cdot f(y_2, y_3) \cdot f(y_1, y_3) = 0.$$

(This corresponds to some explicit closed set M.) So, f determines a graphon that has almost no triangles. Then the claim is that we may save almost all edges so that there would be no triangle at all, i.e., a continuous version of the triangle removal lemma holds true. Similarly we get the hypergraph removal and induced hypergraph removal lemmas.

2010 *Mathematics Subject Classification.* Primary 05C65.

Key words and phrases. Removal lemma, continuous hypergraph, Ramsey theorem.

Supported by St. Petersburg State University grants nos. 6.38.223.2014 and 6.37.208.2016 and by RFBR grant no. 14-01-00373a.

In [1] it was shown by using ultralimits how these lemmas allow one to get discrete counterparts. In the next example a discrete version turns out to be false (see §2).

Example 2 ([5])**.** Suppose $k = 2$, $K = [0, \infty]$, and for almost all y_1, y_2, y_3 we have

$$f(y_1, y_2) = f(y_2, y_1), \quad f(y_1, y_2) + f(y_2, y_3) \geq f(y_1, y_3).$$

Again, this corresponds to an appropriate closed set M, which is cylindrical, like in the previous example. In this case we deal with an "almost metric space", which therefore may be "corrected" by changing distances on a null set of pairs in X^2 to a genuine semimetric space. The values $g(x, x)$ may be redefined to 0 if needed. Also, *a priori* infinite distances may occur. But in fact almost all distances should be finite, and hence on some set X' of full measure all distances are finite. We may identify the complement $X \setminus X'$ of this set with one of the points $x_0 \in X'$, which makes all distances finite.

The proof of part 1 (nonsymmetric version) consists of two ingredients: the Lebesgue density theorem and the Tychonoff compactness theorem. In part 2 (symmetric version) we need also the following standard variant of the Ramsey theorem.

Theorem 2 (Ramsey theorem)**.** *Given $c < \infty$ colors, positive integers $\nu, k_1, \ldots, k_\nu$, and positive integers $N_1, \ldots, N_\nu$, there exist positive integers $R_1, \ldots, R_\nu$ such that for disjoint finite sets $A_1, \ldots, A_\nu$ of cardinalities $|A_i| = R_i$, $1 \leq i \leq \nu$, the following statement holds true.*

Assume that each array $(B_1, \ldots, B_\nu)$, where $B_i \subset A_i$ and $|B_i| = k_i$, is colored in one of our c colors. Then there always exist sets $C_i \subset A_i$, $|C_i| = N_i$, so that the colors of the arrays satisfying $B_i \subset C_i$ are all the same.

We identify X with $[0, 1)$ equipped by the Lebesgue measure μ, and for $x \in X$ denote by $\Delta_m(x)$ a unique half-interval $[s/m, (s+1)/m)$ containing x ($s = 0, 1, \ldots$).

We need the following variant of the Lebesgue density theorem.

Theorem 3. *For almost all $x_1, \ldots, x_k$ in X and any open set $U \subset K$ containing $f(x_1, \ldots, x_k)$ we have*

$$\lim_m m^k \cdot \mu \left(f^{-1}(U) \cap \prod_{i=1}^{k} \Delta_m(x_i) \right) = 1.$$

Proof. Consider a countable base of the topology on K. It suffices to take open sets U from this base. For each of them, the claim is simply the usual Lebesgue density theorem for the set $f^{-1}(U)$. □

Denote by $Y \subset X^k$ a set of full measure for which the condition of Theorem 3 is fulfilled.

For a positive integer ν, we define a metric on K^ν by

$$\operatorname{dist}\big((x_1, \ldots, x_\nu), (y_1, \ldots, y_\nu)\big) := \max_{1 \leq i \leq \nu} \operatorname{dist}_K(x_i, y_i).$$

Proof of Theorem 1. We start with part 1.

First, we require that f and g coincide on Y. This already implies that g is measurable and equivalent to f.

Now we need to define the values of g on $X^k \setminus Y$ so that g satisfy the condition of Theorem 1. Denote by Φ the set of K-valued functions g on X^k such that $g = f$ on Y. This set Φ is a closed subset of the Tychonoff compact space K^{X^k}, which may naturally be identified with $K^{X^k \setminus Y}$.

The closed set M is an intersection of closed cylindrical sets, say, $M = \bigcap_\alpha M_\alpha$. For fixed α and fixed mutually different points $y_1, y_2, \dots$ in X, the condition

$$\widetilde{g}(y_1, y_2, \dots) \in M_\alpha \tag{2}$$

gives rise to a closed subset of Φ. We must prove that all such closed subsets of have Φ a point in common. So, it suffices to prove that any finite collection of such subsets has a common point. Fix such a collection. It only deals with a finite number of k-tuples in X^k. Denote by $A = \{x_1, \dots, x_n\}$ the finite set of all points in those k-tuples. Then conditions (2) hold simultaneously if and only if

$$\widetilde{g}(x_1, \dots, x_n, \dots) \in M', \tag{3}$$

where M' denotes the closed cylindrical set determined by some k-tuples of different indices not exceeding n. We write $\widetilde{g}(x_1, \dots, x_n)$ for the left-hand side of (3), because the further arguments of $\widetilde{g}$ are of no importance.

We need to define g on all k-tuples $(z_1, \dots, z_k) \in A^k$ so that

(i) $\widetilde{g}(x_1, \dots, x_n) \in M'$; and

(ii) g coincides with f on $Y \cap A^k$.

Fix arbitrary $\varepsilon > 0$. Assume that we have succeeded to define g so that

(i$-\varepsilon$) $\widetilde{g}(x_1, \dots, x_n)$ is ε-close to M'; and

(ii$-\varepsilon$) $g(y_1, \dots, y_k)$ and $f(y_1, \dots, y_k)$ are ε-close in K provided that $(y_1, \dots, y_k) \in Y \cap A^k$.

Then we let ε tend to 0 and choose a convergent subsequence of values of g on all k-tuples in A. Clearly this limit function satisfies (i) and (ii), as desired.

For finding g satisfying (i$-\varepsilon$) and (ii$-\varepsilon$), we take m large and replace any point $z \in A$ with a random point $z' \in \Delta_m(z)$. Then we define

$$g(z_1, \dots, z_k) = f(z_1', \dots, z_k')$$

for any points $z_1, \dots, z_k$ in A.

Then (i) (and, hence, also (i$-\varepsilon$)) holds with probability 1. Due to Theorem 3, condition (ii$-\varepsilon$) holds with probability arbitrarily close to 1, provided that m is sufficiently large. So, with positive probability such g works.

Now we pass to proving part 2 of the theorem. First, we change M so as to ensure that any function g satisfying $\widetilde{g}(y_1, y_2, \dots) \in M$ be symmetric. For this, we intersect M with the sets defined by $g(y_1, \dots, y_k) = g(y_{\pi_1}, y_{\pi_2}, \dots, y_{\pi_k})$. Of course, still we have $\widetilde{f}(y_1, y_2, \dots) \in M$ for almost all $y_1, y_2, \dots \in X$.

We follow the lines of the proof of part 1. In particular, we find a finite subset $A = \{x_1, \dots, x_n\}$. The difference is that now we need to check not merely that $\widetilde{g}(x_1, \dots, x_n) \in M'$ for appropriate M', but that

(i′) $\widetilde{g}(y_1, \dots, y_n) \in M'$ for all $y_1, \dots, y_n$ in A (not necessary different).

For this, at the last step, instead of replacing each point $z \in A$ with a random point $z' \in \Delta_m(z)$, we fix a large number R to be specified later (R depends only on K, ε and, $n = |A|$, but does not depend on m) and for each $z \in A$ choose R independent random points in $\Delta_m(z)$. They form a random set $\Omega(z)$ (of course, the sets $\Omega(z), z \in A$, are disjoint sets of cardinality R with probability 1). For arbitrary points $z_1, \dots, z_k$ in A we require $g(z_1, \dots, z_k) = f(z_1', \dots, z_k')$ for some $z_i' \in \Omega(z_i)$. Then for sufficiently large m condition (ii-ε) holds with probability almost 1 for all possible choices $z_i' \in \Omega(z_i)$.

We split K into a finite number c of parts so that each part has diameter less then ε. Let us think that these parts correspond to c different colors.

For any set $S = \{y_1, \dots, y_k\} \subset \cup_{z \in A} \Omega(z)$, we define its color as a part containing $f(y_1, \dots, y_k)$ (with probability 1 this is consistent, i.e., does not depend on the order of elements in S). Let the type of a set S be defined as the function $z \mapsto |I \cap \Omega(z)|$

on A. The number of possible types depends only on $|A|$ and k. Applying Theorem 2 repeatedly for sufficiently large R, we can find subsets $\Xi(z) \subset \Omega(z)$, $|\Xi_i| = n$, so that all subsets $S \subset \cup_{z\in A}\Xi(z)$ of the same type have the same color. Now we are ready to define $g(z_1, \dots, z_k)$ for all (not necessary distinct) points $z_1, \dots, z_k$ in A. We require that $g(z_1, \dots, z_k) := f(z'_1, \dots, z'_k)$, where $z'_i \in \Xi(z_i)$ are mutually different points, and that g be symmetric on A^k. Of course, both conditions may be satisfied. Note that for other possible choices of z'_i the value of g moves by a distance of at most ε.

We need to check (i'). Choose mutually different points $y'_i \in \Xi(y_i)$, $i = 1, 2, \dots, n$. With probability 1 we have $\widetilde{f}(y'_1, \dots, y'_n) \in M'$. Our construction of g guarantees that $\widetilde{g}(y_1, \dots, y_n)$ and $\widetilde{f}(y'_1, \dots, y'_n)$ are ε-close in $K^{n(n-1)\dots(n-k+1)}$, as desired. □

§2. Counterparts

It is easily seen that the requirement that f be symmetric is essential in part 2. Say, if $K = \{0, 1\}$, $k = 2$, and the condition on f is $|f(x, y)-f(y, x)| = 1$ for *almost all* $x, y \in X$ (there are many such functions), this cannot be satisfied for *all* x, y. The reason is that a Ramsey type theorem fails for directed graphs.

Also, we cannot replace the subset I by a multiset even in the symmetric case. Say, if $k = 2$, $K = \{0, 1\}$, and $|f(x, y) - f(x, x)| = 1$ for almost all x, y (which is true for $f(x, y) = \chi_{x\neq y}$), this cannot be made true for all x, y.

In what follows we need the next elementary statement.

Lemma 1. *Let ABC be a triangle on the Euclidean plane with side lengths $AB = c$, $BC = a$, $CA = b$ and with altitude length $CH = h$. Assume that $a + b - c \le 1/2$. Then $4h^2 \le c + 1/4$*

Proof. Let D be the point symmetric to B with respect to the line parallel to AB and passing through C. Then $CD = CB = a$, $c+1/2 \ge a+b = AC+CD \ge AD = \sqrt{c^2 + 4h^2}$, and the claim follows. □

Now we show that a naive discrete version of Example 2 (correction of metrics) is false. Namely, consider a large finite set X equipped by a function $d(x, y)\colon X \times X \to [0,\infty)$ ("almost a metric") such that

(i) $d(x, y) = 0$ if and only if $x = y$;

(ii) $d(x, y) = d(y, x)$ for all $x, y \in X$;

(iii) $d(x, y) + d(y, z) \ge d(x, z)$ for all but $o(|X|^3)$ triples $(x, y, z) \in X^3$.

Do these properties imply that

(iv) d coincides with a semimetric $\widetilde{d}$ on X for all but $o(|X|^2)$ pairs $(x, y) \in X^2$?

The answer is in the negative. A possible counterexample is the following. Let X be a set of integral points inside a Euclidean disk $\{x : \|x\| \le R\}$ of large radius R on the plane; then $|X|$ grows as πR^2. Let $d(x, y) = \|x - y\| - 1/2$ for $x \ne y$; $d(x, x) = 0$. Note that most triangles xyz satisfy $\|x - y\| + \|y - z\| \ge \|x - z\| + 1/2$, which yields (iii). Indeed, if we fix a largest side xz, by Lemma 1 the locus of points y satisfying the reverse inequality $\|x - y\| + \|y - z\| < \|x - z\| + 1/2$ is contained in a strip of width $O(\sqrt{R})$, so that the number of such points y is $O(R^{3/2}) = o(|X|)$. It remains to sum up over all possible pairs xz. On the other hand, for any ordered pair (x, y) of distinct integral points x, y lying in the disk $\{x : \|x\| < R/2\}$ we can consider a point $z \in X$ such that y is a midpoint of xz; any metric $\widetilde{d}$ must differ from d for at least one side of the triangle xyz (and each side is counted at most 6 times). This implies that (iv) fails.

There is a general machinery of Elek and Szegedy [1], allowing to get discrete versions from continuous ones. But the above example shows that sometimes a continuous version is true while the discrete one is wrong. It would be interesting to understand when such things happen and when not. Say, what is the answer for almost metrics d that satisfy the additional restriction $1 \leq d(x, y) \leq 100$ for $x \neq y$?

Acknowledgments

I would like to thank Pavel Zatitskiĭ, Petr Naryshkin, and Ilya Shkredov for fruitful discussions, and especially László Lovász and Svante Janson for motivating questions and comments, which led to a sequence of generalizations of the main statement and to improvements of the exposition.

References

[1] G. Elek and B. Szegedy, *A measure-theoretic approach to the theory of dense hypergraphs*, Adv. Math. **231** (2012), no. 3-4, 1731–1772,
arXiv:0810.4062v2. MR2964622

[2] S. Janson, *Quasi-random graphs and graph limits*, Europ. J. Combin. **32** (2011), no. 7, 1054–1083. MR2825535

[3] ———, *Graph limits and hereditary properties*, Europ. J. Combin. **52** (2016), pt. B, 321–337. MR3425983

[4] L. Lovász, *Large networks and graph limits*, Amer. Math. Colloq. Publ., vol. 60, Amer. Math. Soc., Providence, RI, 2012. MR3012035

[5] P. B. Zatitskiĭ and F. V. Petrov, *Correction of metrics*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **390** (2011), 201–209; English transl., J. Math. Sci. (N.Y.) **181** (2012), no. 6, 867–870, arXiv:1112.2380. MR2870235

St. Petersburg Branch, Steklov Mathematical Institute, Russian Academy of Sciences, Fontanka 27, 191023 St. Petersburg, Russia

E-mail address: fedyapetrov@gmail.com

Received 19/JUL/2016

Translated by F. PETROV