Virtual continuity of measurable functions and its applications

A. M. Vershik, P. B. Zatitskiy, and F. V. Petrov

Abstract. A classical theorem of Luzin states that a measurable function of one real variable is ‘almost’ continuous. For measurable functions of several variables the analogous statement (continuity on a product of sets having almost full measure) does not hold in general. The search for a correct analogue of Luzin’s theorem leads to a notion of virtually continuous functions of several variables. This apparently new notion implicitly appears in the statements of embedding theorems and trace theorems for Sobolev spaces. In fact it reveals the nature of such theorems as statements about virtual continuity. The authors’ results imply that under the conditions of Sobolev theorems there is a well-defined integration of a function with respect to a wide class of singular measures, including measures concentrated on submanifolds. The notion of virtual continuity is also used for the classification of measurable functions of several variables and in some questions on dynamical systems, the theory of polymorphisms, and bistochastic measures. In this paper the necessary definitions and properties of admissible metrics are recalled, several definitions of virtual continuity are given, and some applications are discussed.

Bibliography: 24 titles.

Keywords: admissible metrics, virtual topology, bistochastic measures, trace theorems, embedding theorems.
1. Introduction. Admissible metrics

We consider a standard (Lebesgue–Rokhlin) probability space with continuous (atomless) measure, which is isomorphic to the interval $[0,1]$ with Lebesgue measure. The first author [13], [16], [19] proposed considering on a fixed standard measure space $(X, \mathcal{A}, \mu)$ different (admissible) metrics, in contrast to the usual approach, when a metric space is fixed and Borel measures on it vary. Such an approach is useful and necessary in ergodic theory and other situations. Matching the metric and measure structures leads to the notion of a metric (or admissible) triple.

**Definition 1.** A metric (or semimetric) $\rho$ which is measurable as a function of two variables on a standard measure space $(X, \mathcal{A}, \mu)$ is said to be admissible if there is a measurable subset $X_0 \subset X$ of full measure, $\mu(X_0) = 1$, such that the metric (respectively, semimetric) space $(X_0, \rho)$ is separable.

A standard measure space $(X, \mu)$ equipped with an admissible (semi)metric $\rho$ is called an admissible metric triple or just an admissible triple $(X, \mu, \rho)$.

Properties of admissible semimetrics and metrics are studied in detail in our previous papers [10], [24]. In particular, a number of equivalent definitions of admissibility are given.

Since the space is standard, the following result is easy to establish.

**Proposition 1.** If $\rho$ is an admissible metric on $(X, \mathcal{A}, \mu)$, then the completed Borel $\sigma$-algebra $\mathcal{B} = \mathcal{B}(X, \rho)$ coincides with $\mathcal{A}$, and the measure $\mu$ is inner regular with respect to the metric $\rho$, that is, for any $A \in \mathcal{A}$ one has

$$\mu(A) = \sup\{\mu(K): K \subset A, K \text{ is compact in the metric } \rho\}.$$ 

Thus, for any admissible metric the initial measure $\mu$ is a Radon measure in the space $(X, \rho)$.
Proof. Measurability of Borel sets, that is, the inclusion \( \mathcal{B} \subset \mathfrak{A} \), was proved in [24]. Let us now prove inner regularity, which implies the reverse inclusion. There is a subset \( X_0 \subset X \) with \( \mu(X_0) = 1 \) such that the metric space \((X_0, \rho)\) is separable. Denote by \( \mathfrak{A}_0 \) the restriction of \( \mathfrak{A} \) to \( X_0 \). Note that we can choose \( X_0 \) to be closed in \( X \) in the metric \( \rho \), and this implies that \( X \setminus X_0 \in \mathcal{B} \). Let \( X_1 \) be the completion of the metric space \((X_0, \rho)\). Define a measure \( \mu_1 \) on the Borel \( \sigma \)-algebra \( \mathfrak{B}_1 = \mathfrak{B}(X_1, \rho) \) of the Polish space \((X_1, \rho)\) by extending \( \mu \) from \( \mathfrak{B}_0 = \mathfrak{B}(X_0, \rho) \) and setting \( \mu_1(X_1 \setminus X_0) = 0 \). Let \( \mathfrak{B}_1 \) be the completion of the \( \sigma \)-algebra \( \mathfrak{B}_1 \) with respect to the measure \( \mu_1 \). Note that \((X_1, \mathfrak{B}_1, \mu_1)\) is a Lebesgue space as a Polish space with a Borel probability measure on the completed Borel \( \sigma \)-algebra. Moreover, the map

\[
\text{id}: (X_0, \mathfrak{A}_0, \mu) \to (X_1, \mathfrak{B}_1, \mu_1)
\]

is an injective measure-preserving map of Lebesgue spaces. By the lemma in §5 of [11], for such a map the image of any measurable set is also measurable. Thus, \( \mathfrak{A}_0 \subset \mathfrak{B}_1 \). Since \( \mathfrak{A}_0 \) is a \( \sigma \)-algebra on \( X_0 \), the restrictions of the \( \sigma \)-algebras \( \mathfrak{B} \) and \( \mathfrak{B}_1 \) to \( X_0 \) coincide, and \( X_0 \in \mathfrak{B}_1 \), it follows that \( \mathfrak{A}_0 \subset \mathfrak{B}_1 \). Since \( \mu(X \setminus X_0) = 0 \), we get that \( \mathfrak{A} = \mathfrak{B} \). Recall that any Borel probability measure on a Polish space is inner regular. Let us prove the inner regularity of the measure \( \mu \) on the (possibly not complete) metric space \((X, \rho)\). Consider any set \( A \in \mathfrak{A} \). Then \( A \cap X_0 \in \mathfrak{B}_1 \), and using the inner regularity of the measure \( \mu_1 \) on the Polish space \((X_1, \rho)\), we can find a compact set \( K \subset A \cap X_0 \) for which

\[
\mu(K) = \mu_1(K) > \mu_1(A \cap X_0) - \varepsilon = \mu(A) - \varepsilon
\]

for any fixed \( \varepsilon > 0 \), as required. \( \square \)

Gromov in [3] suggested considering arbitrary metric triples \((X, \mu, \rho)\), which he called \( mm \)-spaces, and he asked about their classification, having in mind classical situations (Riemannian manifolds and so on). It is natural to consider admissible triples in this framework. We define equivalence of admissible triples up to measure-preserving isometries: \((X, \mu, \rho) \sim (X', \mu', \rho')\) if

\[
\exists T: X \to X', \quad T\mu = \mu', \quad T^{-1}\mu' = \mu, \quad \rho'(Tx, Ty) = \rho(x, y).
\]

Here is the main result on this equivalence.

**Theorem 1** (Gromov [3], Vershik [13]). Consider the map \( F_\rho: X^\infty \times X^\infty \to M_\infty(\mathbb{R}) \) given by

\[
F_\rho(\{x_i, y_j\}_{(i,j) \in \mathbb{N} \times \mathbb{N}}) = \{\rho(x_i, y_j)\}_{(i,j) \in \mathbb{N} \times \mathbb{N}},
\]

and equip the infinite product \( X^\infty \times X^\infty \) with the product measure \( \mu^\infty \times \mu^\infty \). Let \( D_\rho \) denote the measure on the space of matrices (that is, the random matrix of distances) which is the \( F_\rho \)-image of the measure \( \mu^\infty \times \mu^\infty \), and call it the matrix distribution of the metric \( \rho \). Then it is a complete invariant of the above equivalence of admissible metrics.

In other words,

\[
(X, \mu, \rho) \sim (X', \mu', \rho') \iff D_\rho = D_{\rho'}.
\]
In [14] this result was generalized to so-called pure measurable functions of several variables.

The following lemma is useful in the theory of admissible metrics.

**Lemma 1.** Let $\rho_1$ and $\rho_2$ be admissible semimetrics on the standard measure space $(X, \mu)$, and let $\rho_1$ be a metric. Then for any $\varepsilon > 0$ there is a measurable subset $K \subset X$ such that $\mu(K) > 1 - \varepsilon$ and the semimetric $\rho_2$ (as a function of two variables) is continuous on $K \times K$ with respect to the metric $\rho_1$.

**Proof.** Consider the admissible metric $\rho = \rho_1 + \rho_2$. Choose a compact set $K$ in this metric so that $\mu(K) > 1 - \varepsilon$. Let us show that $\rho$ (hence $\rho_2$) is continuous on the metric space $(K \times K, \rho_1 \times \rho_1)$. The triangle inequality reduces this desired continuity to the following fact: given $\delta > 0$ there exists a $\sigma > 0$ such that $\rho(x, y) < \delta$ whenever $x, y \in K$ and $\rho_1(x, y) < \sigma$. If this is not true, then there exist $\delta > 0$ and two sequences $\{x_n\}, \{y_n\}$ in $K$ for which $\rho(x_n, y_n) > \delta$ but $\rho_1(x_n, y_n) \to 0$. Since $(K, \rho)$ is compact, we can suppose without loss of generality that there exist $x, y \in K$ such that

$$\rho(x_n, x) \to 0, \quad \rho(y_n, y) \to 0.$$ 

But then

$$\rho_1(x_n, x) \leq \rho(x_n, x) \to 0, \quad \rho_1(y_n, y) \leq \rho(y_n, y) \to 0.$$ 

Thus, $\rho_1(x, y) = 0$, while $\rho(x, y) \geq \delta$. This contradicts the assumption that $\rho_1$ is a metric. \qed

Lemma 1 has the following immediate consequence.

**Corollary 1.** Let $\rho_1$ and $\rho_2$ be two admissible metrics on the standard measure space $(X, \mu)$. Then for any $\varepsilon > 0$ there exists a set $K \subset X$ such that $\mu(K) > 1 - \varepsilon$ and the topologies defined by the metrics $\rho_1$ and $\rho_2$ on $K$ coincide.

2. Virtual continuity

2.1. Luzin’s theorem on measurable functions of one variable. In what follows we consider (measurable) real-valued functions, although most of our results remain true for maps to a standard Borel space, in particular, to a Polish space. The classical theorems of Egorov and Luzin on measurable functions of one variable are well known. The generalized Luzin theorem for an arbitrary admissible triple follows from the above results.

**Corollary 2** (Luzin’s theorem). Let $\rho$ be an admissible metric on the standard measure space $(X, \mu)$, and let $f$ be a measurable map from $X$ to the Polish space $(M, d)$. Then for any $\varepsilon > 0$ there is a measurable subset $K \subset X$ such that $\mu(K) > 1 - \varepsilon$ and $f$ is continuous on $K$ with respect to the metric $\rho$.

**Proof.** Let

$$\rho_1(x, y) = \rho(x, y) + d(f(x), f(y)).$$

Then $\rho_1$ is a trivial example of an admissible metric with respect to which $f$ is continuous almost everywhere. By Corollary 1 there is a subset $K$ having measure $\mu(K) > 1 - \varepsilon$ on which this continuity implies continuity with respect to $\rho$. \qed
But we shall see that this does not hold for arbitrary functions of several variables.

2.2. Definitions and first examples. Let \( f(\cdot, \cdot) \) be a measurable function of two variables. Then the analogue of Luzin’s theorem, that is, continuity on a product \( X' \times Y' \) of sets with measure \( > 1 - \varepsilon \) with respect to a given metric of the form

\[
\rho([x_1, y_1], [x_2, y_2]) = \rho_X(x_1, x_2) + \rho_Y(y_1, y_2),
\]

is not true in general. This leads to the following key notion of our survey. (The sum of metrics can be replaced by the maximum or by another metric giving the direct product topology, and to stress this we denote a general such metric by \( \rho_X \times \rho_Y \).

**Definition 2.** A measurable function \( f(\cdot, \cdot) \) on a product \((X, \mu) \times (Y, \nu)\) of standard measure spaces is said to be **properly virtually continuous** if for any \( \varepsilon > 0 \) there are sets \( X' \subset X \) and \( Y' \subset Y \) with measure at least \( 1 - \varepsilon \) and admissible semimetrics \( \rho_X \) and \( \rho_Y \) on \( X' \) and \( Y' \), respectively, such that the function \( f \) is continuous on \((X' \times Y', \rho_X \times \rho_Y)\).

A function which coincides with a properly virtually continuous function on a set of full measure in \( X \times Y \) is said to be **virtually continuous**. Virtually continuous functions of several variables are defined in the same way.

It is essential that an admissible metric with respect to which a measurable function becomes continuous is not arbitrary, but respects the direct product structure (in a more general setting it respects selected subalgebras; see below). It is easy to verify that there does not exist a universal metric of this type (that is, a metric such that virtual continuity implies continuity in this metric). This explains the non-trivial properties of the notion introduced.

It is clear that any admissible metric (as a measurable function of two variables) is virtually continuous. So is any function which is continuous with respect to a product of admissible metrics. Degenerate functions (or ‘finite-rank functions’)

\[
f(x, y) = \sum_{i=1}^{n} \varphi_i(x) \psi_i(y),
\]

where \( \varphi_i(\cdot) \) and \( \psi_i(\cdot) \), \( i = 1, \ldots, n \), are arbitrary measurable functions, are also virtually continuous. For a proof just use Luzin’s theorem for all the functions \( \varphi_i(\cdot) \) and \( \psi_i(\cdot) \), \( i = 1, \ldots, n \).

Less trivial examples of virtually continuous functions are given by functions in certain Sobolev classes and by kernels of trace-class operators. For virtually continuous functions there are well-defined restrictions to some subsets of zero measure—specifically, to the supports of (quasi-)bistochastic measures (see §3).

An easy example of a measurable function not virtually continuous on \([0, 1]^2\) is provided by the characteristic function of the triangle \( \{x \geq y\} \). In general, for functions on the square of a compact group which depend on the ratio of the two variables the criterion for virtual continuity is simple.

**Proposition 2.** Let \( G \) be a compact metrizable group and \( f \) a Haar measurable function on \( G \). Then the function \( F(x, y) := f(xy^{-1}) \) on \( G \times G \) is virtually continuous if and only if \( f \) is equivalent to a continuous function.
We stress once more that the definition of virtual continuity is not topological but purely measure-theoretic in nature, that is, it uses only the measurable structure of spaces and measures on them (in other words, this concept has to do with objects in the category of measure spaces). This concept concerns the choice of various metrics on the measure space, as noted above. Thus, the direct meaning of Proposition 2 is that the group structure and the measure-theoretic structure allow us to reconstruct the topology.

2.3. Further properties of virtually continuous functions. First and foremost, virtually continuous functions automatically satisfy stronger conditions than are required in the definition.

Using Corollary 1, we first see immediately that metrics can be fixed a priori.

**Theorem 2.** Let the function \( f(\cdot, \cdot) \) be properly virtually continuous. Then for any admissible metrics \( \rho_X \) and \( \rho_Y \) on \( X \) and \( Y \) and for any \( \varepsilon > 0 \) there are sets \( X' \subset X \) and \( Y' \subset Y \) with measure at least \( 1 - \varepsilon \) such that \( f \) is continuous on \( (X' \times Y', \rho_X \times \rho_Y) \).

On the other hand, by a suitable choice of metrics we can take the sets \( X' \) and \( Y' \) in the definition to have full measure.

**Theorem 3.** Let the function \( f(\cdot, \cdot) \) be properly virtually continuous. Then there are sets \( X' \subset X \) and \( Y' \subset Y \) of full measure and admissible semimetrics \( \rho_{X'} \) and \( \rho_{Y'} \) on \( X' \) and \( Y' \), respectively, such that \( f \) is continuous on \( (X' \times Y', \rho_{X'} \times \rho_{Y'}) \).

**Proof.** Fix admissible metrics \( \sigma_X \) and \( \sigma_Y \) on \( X \) and \( Y \) respectively. For any \( n \) use Theorem 2 and find sets \( X_n \subset X \) and \( Y_n \subset Y \) of measure at least \( 1 - 2^{-n} \) such that \( f \) is continuous on \( (X_n \times Y_n, \sigma_X \times \sigma_Y) \). We define cut semimetrics on \( X \) by

\[
\rho_{X;n}(x, x') = \begin{cases} 
0 & \text{if } x, x' \in X_n \text{ or } x, x' \notin X_n, \\
1 & \text{if } x \in X_n, x' \notin X_n \text{ or } x' \in X_n, x \notin X_n.
\end{cases}
\]

The set \( X' = \bigcup_{n=1}^{\infty} \bigcap_{k>n} X_k \) has full measure. Next, define the metric \( \rho_X = \sigma_X + \sum_{n} 2^{-n} \rho_{X;n} \) (it is easy to verify that \( \rho_X \) is an admissible metric). The set \( Y' \) and the metric \( \rho_Y \) are defined similarly.

Let us prove that the function \( f(x, y) \) is continuous on \( (X' \times Y', \rho_X \times \rho_Y) \). Consider convergent sequences \( x_n \to x_0 \) and \( y_n \to y_0 \) in \( X' \) and \( Y' \), respectively. Let \( N \) be so large that \( x_0 \in X_N \) and \( y_0 \in Y_N \). Then convergence with respect to the semimetric \( \rho_{X;n} \) implies that \( x_n \in X_N \) for all sufficiently large \( n \), and similarly \( y_n \in Y_N \) for large \( n \). Now the convergence \( f(x_n, y_n) \to f(x_0, y_0) \) follows from the continuity of \( f \) on \( (X_N \times Y_N, \sigma_X \times \sigma_Y) \). \( \square \)

We establish the following corollary.

**Proposition 3.** If properly virtually continuous functions \( f(x, y) \) and \( g(x, y) \) coincide on a set of full measure in \( X \times Y \), then there are sets \( X' \subset X \) and \( Y' \subset Y \) of full measure such that \( f(x, y) = g(x, y) \) for all points \( x \in X', y \in Y' \).
**Proof.** We apply Theorem 3 to both functions \( f \) and \( g \). We can assume that the corresponding sets \( X' \subset X \) and \( Y' \subset Y \) of full measure coincide for \( f \) and \( g \) (take the intersection of the sets for \( f \) and for \( g \)), and the same for the admissible semimetrics \( \rho_X \) and \( \rho_Y \) (add the semimetrics for \( f \) and \( g \)). Moreover, we can assume that \( X' \) and \( Y' \) are the supports of the measures \( \mu|_{X'} \) and \( \nu|_{Y'} \). Now note that the set of points \( (x, y) \in X' \times Y' \) for which \( f(x, y) \neq g(x, y) \) is open in \( X' \times Y' \), and hence it either has positive measure (but this is impossible by our assumption) or it is empty (as required). \( \square \)

Thus, all properly virtually continuous functions equivalent to a given virtually continuous function coincide on a product of sets having full measure.

It is useful to think about a function of two variables on \( X \times Y \) as a map from \( X \) to the space of functions on \( Y \) (that is, \( f(x, y) \equiv f_x(y) \)). See [20] for details on using this idea in classifying measurable functions. Virtual continuity can be expressed in these terms by the following equivalent definition.

**Theorem 4.** The following properties of a function \( f(\cdot, \cdot) \) are equivalent:

(i) \( f \) is virtually continuous;

(ii) for any \( \varepsilon > 0 \) there are sets \( X' \subset X \) and \( Y' \subset Y \) of measure at least \( 1 - \varepsilon \) and an admissible semimetric \( \rho_Y \) on \( Y' \) such that the function \( f_x(\cdot) \) is equivalent to a continuous function on \((Y', \rho_Y)\) for almost all \( x \in X' \);

(iii) for any \( \varepsilon > 0 \) there are sets \( X' \subset X \) and \( Y' \subset Y \) of measure at least \( 1 - \varepsilon \) such that the set of functions \( f_x(\cdot) \) on \( Y' \) (where the variable \( x \) runs over \( X' \)) is totally bounded (precompact) as a metric subspace of \( L_\infty(Y') \);

(iv) for any \( \varepsilon > 0 \) there are sets \( X' \subset X \) and \( Y' \subset Y \) having measure at least \( 1 - \varepsilon \) such that the set of functions \( f_x(\cdot) \) on \( Y' \) (where the variable \( x \) runs over \( X' \)) is separable as a metric subspace of \( L_\infty(Y') \).

**Proof.** Clearly, (i) implies (ii) and (iii) implies (iv).

Let us prove (iii) assuming (ii). Removing appropriate sets of zero measure from \( X' \) and \( Y' \), we can suppose that the function \( f(x, \cdot) \) is equivalent to a continuous function on \( Y' \) for any \( x \in X' \), and that the function \( f(\cdot, y) \) is equivalent to a continuous function on \( X' \) for any \( y \in Y' \). Choose a compact subset \( Y_1 \subset Y' \) such that \( \nu(Y_1) > 1 - 2\varepsilon \). Next, replace \( Y_1 \) by the support \( \text{supp}(\nu|_{Y_1}) \) of the measure \( \nu \) restricted to \( Y_1 \). Now for continuous functions on \( Y_1 \), the distances in \( C(Y_1) \) and in \( L_\infty(Y_1) \) coincide. Let \( \mathcal{S} \) be a countable family of open balls in \( Y_1 \) which form a base for the topology. For \( x \in X' \) denote by \( f'_x(\cdot) \) a continuous function on \( Y_1 \) which is equivalent to \( f(x, \cdot) \). Let us prove that the map \( \Phi: x \mapsto f'_x(\cdot) \) is measurable as a map from \( X \) to \( C(Y_1) \) (with the Borel \( \sigma \)-algebra). It suffices to check that the pre-image of the ball

\[ X_g := \{ x \in X : \forall y \in Y_1 \mid f'_x(y) - g(y) \mid \leq r \} \]

is measurable for any continuous function \( g \in C(Y_1) \) and any positive \( r \). Note that the inequality \( |f'_x(y) - g(y)| \leq r \) holds for all \( y \in Y_1 \) if and only if the inequalities

\[ \frac{1}{\nu(B)} \left| \int_B f'_x(y) \, d\nu(y) - \int_B g(y) \, d\nu(y) \right| \leq r \]

for every Borel \( B \subset Y_1 \) holds. Note that the set of \( \nu \) for which this inequality holds is precisely \( \nu(\mathcal{S}) \). Hence to prove (iii), it suffices to prove that \( \nu(\mathcal{S}) \) is positive. This follows from

\[ \frac{1}{\nu(B)} \left| \int_B f'_x(y) \, d\nu(y) - \int_B g(y) \, d\nu(y) \right| \leq r \]
hold for all balls \( B \in \mathcal{S} \). But under the integral sign the function \( f'(x, \cdot) \), can be replaced by the equivalent function \( f(x, \cdot) \), and the map

\[
x \mapsto \int_B f(x, y) \, d\nu(y)
\]

is measurable with respect to \( x \). Thus, the set \( X_g \) is measurable as a countable intersection of measurable sets.

Therefore, the \( \Phi \)-pre-image of the measure \( \mu \) on \( X' \) is a Borel measure on \( C(Y_1) \). The space \( C(Y_1) \) is complete and separable, hence this measure is inner regular, and there is a compact set \( K \subset C(Y_1) \) such that \( \mu(\Phi^{-1}(K)) > 1 - 2\varepsilon \). But \( \varepsilon \) is arbitrary, and thus the function \( f(x, y) \) satisfies (iii).

It remains to show the virtual continuity of \( f \) assuming (iv). Fix \( \varepsilon > 0 \), and choose \( X' \) and \( Y' \) as in (iv). Denote by \( K \) a separable family of functions of the form \( f(x, \cdot) \) in \( L_\infty(Y') \). Let \( K' \) be a countable dense subfamily of \( K \). It is not hard to construct an admissible semimetric \( \rho \) on \( Y' \) such that any function in \( K' \) is \( \rho \)-continuous (such a semimetric can be constructed, for example, as a uniformly convergent series of metrics constructed for these functions separately). Let \( Y_1 \subset Y' \) be a compact set with respect to \( \rho \) and having measure at least \( 1 - 2\varepsilon \). We replace \( Y_1 \) (if necessary) by the support of the restriction of the measure \( \nu \) to \( Y_1 \). Now any open ball in \( Y_1 \) has positive measure. This implies that for continuous functions on \( Y_1 \) distances in \( C(Y_1) \) and \( L_\infty(Y_1) \) coincide. Thus, any function \( f(x, \cdot), x \in X' \), is equivalent to a unique continuous function \( f'(x, \cdot) \) on \( Y_1 \). We define a semimetric on \( X' \) by

\[
\rho_{X'}(x, x') = \|f'(x, \cdot) - f'(x', \cdot)\|_{C(Y_1)}.
\]

It is admissible since \( C(Y_1) \) is separable. The function \( f' \) is continuous on \( (X' \times Y_1, \rho_{X'} \times \rho) \). Indeed,

\[
|f'(x_n, y_n) - f'(x_0, y_0)| \leq |f'(x_n, y_n) - f'(x_0, y_n)| + |f'(x_0, y_n) - f'(x_0, y_0)|,
\]

and both summands tend to 0 as \( x_n \to x_0 \) and \( y_n \to y_0 \). The function \( f' \) is properly virtually continuous on \( X' \times Y_1 \). By Fubini’s theorem it is equivalent to the initial function \( f \) on \( X' \times Y_1 \). Since \( \varepsilon > 0 \) is arbitrary, this implies that \( f \) is virtually continuous. \( \square \)

It is remarkable that the spaces \( X \) and \( Y \) (that is, the arguments of the function) play different roles in (ii)–(iv). However, a posteriori this property appears to be symmetric under a change in the order of the variables. This is another demonstration of the non-triviality of the virtual continuity concept.

The above characteristics of virtual continuity easily imply that the virtually continuous functions form a nowhere dense subset of the space of all measurable functions of two variables (with the topology of convergence in measure).

### 2.4. Virtual topology

A function is continuous if and only if for any open set its pre-image is open. Virtual continuity of a function of two variables admits a similar characterization.

**Definition 3.** A measurable set \( Z \subset X \times Y \) is said to be virtually open if for some subsets \( X' \subset X \) and \( Y' \subset Y \) of full measure the set \( Z \cap (X' \times Y') \) is a countable
union of measurable rectangles \( R_i = A_i \times B_i, \ i = 1, 2, \ldots \). A set is said to be virtually closed if its complement is virtually open.

The meaning (and also the name) of this concept is explained by the following assertion.

**Lemma 2.** 1) Let \( X' \) and \( Y' \) be sets of full measure in \( X \) and \( Y \), respectively, let \( \rho_X \) and \( \rho_Y \) be admissible semimetrics on \( X' \) and \( Y' \), and let the set \( Z \subset X \times Y \) be such that \( Z \cap (X' \times Y') \) is open in \( (X' \times Y', \rho_X \times \rho_Y) \). Then \( Z \) is virtually open.

2) Conversely, for any virtually open set \( Z \) in \( X \times Y \) there are sets \( X' \) and \( Y' \) of full measure and admissible semimetrics \( \rho_X \) and \( \rho_Y \) such that \( Z \cap (X' \times Y') \) is open in \( (X' \times Y', \rho_X \times \rho_Y) \).

**Proof.** 1) Replacing the sets \( X' \) and \( Y' \) by suitable subsets of full measure, we can suppose that \( (X', \rho_X) \) and \( (Y', \rho_Y) \) are separable semimetric spaces in which all balls are \( \mu \)- and \( \nu \)-measurable, respectively. Then the topologies of these spaces have countable bases consisting of measurable sets, and thus the direct product topology has a countable base consisting of measurable rectangles. This implies that an open subset of \( X' \times Y' \) is a countable union of measurable rectangles, and hence it is virtually open.

2) For a given countable family of measurable sets in \( X \) we can construct an admissible semimetric in which they are each open (a possible construction is to sum cut semimetrics with rapidly decaying coefficients). Doing this for the \( X \)-projections of the countably many rectangles whose union is our virtually open set, we construct the required admissible semimetric on \( X \) (strictly speaking, on \( X' \)), and similarly for \( Y \).

**Theorem 5.** A measurable function \( f(x, y) \) on \( X \times Y \) is properly virtually continuous if and only if the \( f \)-pre-image of any open set on the real line is virtually open.

**Proof.** If \( f \) is properly virtually continuous, then by Theorem 3 it is continuous on the product \( (X' \times Y', \rho_X \times \rho_Y) \), where \( X' \subset X \) and \( Y' \subset Y \) have full measure and \( \rho_X \) and \( \rho_Y \) are admissible semimetrics. Then the \( f \)-pre-images of open sets on the real line have the structure described in Lemma 2.

Assume that the \( f \)-pre-image of any interval with rational endpoints on the real line is virtually open. Then there are sets \( X' \) and \( Y' \) of full measure such that for the restriction \( f : X' \times Y' \to \mathbb{R} \) all these pre-images are countable unions of measurable rectangles in \( X' \times Y' \) (\( X' \) and \( Y' \) can be taken to be intersections of corresponding sets of full measure). Considering series of cut semimetrics, we can easily construct admissible metrics on \( X' \) and \( Y' \) such that each of these countably many rectangles is open. Then the function \( f \) is continuous with respect to a pair of constructed admissible semimetrics on \( X' \times Y' \), and hence it is virtually continuous.

### 2.5. Thickness

Let us consider the space \( X \times Y \) with the product measure \( \mu \times \nu \). In its \( \sigma \)-algebra of measurable sets it is natural to single out the two \( \sigma \)-subalgebras determined by the projections onto \( X \) and \( Y \). We write \( A \overset{\text{mod}}{\subset} B \) if \( A, B \subset X \times Y \) and \( \mu \times \nu(B \setminus A) = 0 \). Also, we write \( \overset{\text{mod}}{\leq} \) or \( \overset{\text{mod}}{\geq} \) if the corresponding inequality holds \( (\mu \times \nu) \)-almost everywhere.
Definition 4. For a measurable set $Z \subset X \times Y$ define its *proper thickness* as

$$\text{sth}(Z) = \inf \{ \mu(\tilde{X}) + \nu(\tilde{Y}) : \tilde{X} \subset X, \tilde{Y} \subset Y, Z \subset (\tilde{X} \times \tilde{Y}) \cup (X \times \tilde{Y}) \}. \quad (1)$$

The *thickness* of a set $Z$ is defined as

$$\text{th}(Z) = \inf \{ \mu(\tilde{X}) + \nu(\tilde{Y}) : \tilde{X} \subset X, \tilde{Y} \subset Y, Z \mod 0 \subset (\tilde{X} \times \tilde{Y}) \cup (X \times \tilde{Y}) \}. \quad (2)$$

In other words,

$$\text{th}(Z) = \min \{ \text{sth}(Z) : Z \mod 0 \subset Z' \}. \quad (2)$$

The minimum in (2) is always attained, because we can consider the intersection of a minimizing sequence of sets.

The sets $\tilde{X} \times \tilde{Y}$ and $X \times \tilde{Y}$ are just sets from the chosen subalgebras, so we can generalize our definition of thickness to other choices of selected subalgebras in a standard measure space.

The following properties of thickness are immediate:

- the thickness of a set does not exceed 1 and equals 0 for and only for sets of measure 0;
- the thickness of a subset does not exceed the thickness of a set;
- the thickness of a set is no less than its measure;
- the thickness of a finite or countable union of sets does not exceed the sum of their thicknesses.

The following lemma is also not difficult.

Lemma 3. If a set $Z \subset X \times Y$ is virtually open, then $\text{th}(Z) = \text{sth}(Z)$.

Proof. It suffices to prove that

$$\text{sth}(Z) \leq \mu(\tilde{X}) + \nu(\tilde{Y}) \quad \text{if } Z \mod 0 \subset (\tilde{X} \times \tilde{Y}) \cup (X \times \tilde{Y}).$$

Replacing $X$ and $Y$ by suitable subsets of full measure in them, we can suppose that $Z = \bigcup_{i=1}^{\infty} (A_i \times B_i)$. Note that if

$$A_i \times B_i \mod 0 \subset (\tilde{X} \times \tilde{Y}) \cup (X \times \tilde{Y}), \quad (3)$$

then either $A_i \mod 0 \subset \tilde{X}$ or $B_i \mod 0 \subset \tilde{Y}$. In both cases we can add sets of measure zero to $\tilde{X}$ and $\tilde{Y}$ so that $\mod 0 \subset$ in (3) becomes just $\subset$. Doing this for all $i = 1, 2, \ldots$ successively, we obtain the required inequality. □

The following lemma provides an equivalent and sometimes more useful definition of thickness.

Lemma 4. For any $Z$ consider pairs of measurable functions $f : X \to [0,1]$ and $g : Y \to [0,1]$ for which $f(x) + g(y) \geq \chi_Z(x,y)$ (respectively, $f(x) + g(y) \geq \chi_Z(x,y)$). Then the proper thickness (respectively, thickness) of the set $Z$ is the infimum of $\int_X f \, d\mu + \int_Y g \, d\nu$. Moreover, this infimum is realized, as well as the infimum in (1).
Proof. Clearly, if $\bar{X} \subset X$ and $\bar{Y} \subset Y$ are sets such that $Z \subset (\bar{X} \times Y) \cup (X \times \bar{Y})$, then $f = \chi_{\bar{X}}$ and $g = \chi_{\bar{Y}}$ satisfy the inequality $f(x) + g(y) \geq \chi_Z(x, y)$. Thus, we just need to prove that $\int_X f + \int_Y g \geq \text{sth}(Z)$ whenever $f(x) + g(y) \geq \chi_Z(x, y)$. For any $t \in [0, 1]$ let

$$X_t = \{x \in X : f(x) \geq t\} \quad \text{and} \quad Y_t = \{y \in Y : g(y) \geq t\}.$$ 

Clearly, if $f(x) + g(y) \geq 1$, then for any $t$ either $x \in X_t$ or $y \in Y_{1-t}$. Hence, for any $t$ we have $\chi_Z(x, y) \leq \chi_{X_t}(x) + \chi_{Y_{1-t}}(y)$, and so $\text{sth}(Z) \leq \mu(X_t) + \nu(Y_{1-t})$. Integrating with respect to $t$, we find that

$$\text{sth}(Z) \leq \int_0^1 (\mu(X_t) + \nu(Y_{1-t})) \, dt = \int_X f + \int_Y g.$$ 

Moreover, if $\text{sth}(Z) = \int_X f + \int_Y g$, then for almost all $t \in [0, 1]$ the infimum in (1) is realized on the pair of sets $(X_t, Y_{1-t})$.

It remains to prove that there is a minimizing pair of functions. Consider a minimizing sequence of pairs $(f_n(x), g_n(y))$, with $\int f_n + \int g_n \rightarrow \text{sth}(Z)$. By the well-known Komlós theorem [6] we can assume that the sequence $f'_n := (f_1 + \cdots + f_n)/n$ converges to some function $f$ almost everywhere in $X$, and that $g'_n := (g_1 + \cdots + g_n)/n$ converges to some $g$ almost everywhere in $Y$. (Only the simple version of the Komlós theorem, in which the functions are uniformly bounded, is used here.)

Thus, we can suppose that

$$f(x) = \limsup_{n} f'_n(x) \quad \text{for all } x \in X,$$

$$g(y) = \limsup_{n} g'_n(y) \quad \text{for all } y \in Y.$$ 

It follows that $f(x) + g(y) \geq \chi_Z(x, y)$ for all $x \in X$ and $y \in Y$, and hence the pair $(f, g)$ is minimizing. \(\square\)

Using Lemma 4, we can establish ‘continuity of thickness from below’.

**Lemma 5.** Let $\{Z_n\}$ be an increasing sequence of measurable sets, and let $Z = \bigcup_n Z_n$. Then

$$\text{th}(Z) = \lim \text{th}(Z_n), \quad \text{sth}(Z) = \lim \text{sth}(Z_n).$$

**Proof.** Clearly, $\text{th}(Z) \geq \text{th}(Z_n)$ for all $n$, hence

$$\text{th}(Z) \geq \lim \text{th}(Z_n).$$

Let us prove the reverse inequality. We start with functions $f_n : X \to [0, 1]$ and $g_n : Y \to [0, 1]$ such that

$$f_n(x) + g_n(y) \mod 0 \geq \chi_{Z_n}(x, y) \quad \text{and} \quad \int_X f_n + \int_Y g_n \leq \text{th}(Z_n) + \frac{1}{n}.$$ 

Any bounded sequence in $L^2$ contains a weakly convergent subsequence, and using this twice we can suppose that $f_n$ converges weakly to $f$ in $L^2(X, \mu)$ and $g_n$ converges weakly to $g$ in $L^2(Y, \nu)$. Then $f_n(x) + g_n(y)$ converges to $f(x) + g(y)$ weakly
in $L^2(X \times Y, \mu \times \nu)$. Since passage to the weak limit preserves inequalities, we have $f: X \to [0, 1]$ and $g: Y \to [0, 1]$. Moreover, for any $n$

\[ f(x) + g(y) \equiv \chi_{Z_n}(x, y). \]

Hence,

\[ f(x) + g(y) \equiv \chi_Z(x, y). \]

But

\[ \int_X f + \int_Y g = \lim \left( \int_X f_n + \int_Y g_n \right) \leq \lim \text{th}(Z_n), \]

and thus $\text{th}(Z) \leq \lim \text{th}(Z_n)$.

For proper thickness in the above proof we would replace weak convergence in $L^2$ by convergence almost everywhere, which follows from the Komlós theorem [6].

We remark that thickness is not continuous from above: all the sets

\[ \left\{ (x, y) : 0 < |x - y| < \frac{1}{n} \right\} \subset [0, 1]^2 \]

have thickness 1, but their intersection is empty.

Now we define convergence of functions ‘in thickness’ by analogy with convergence ‘in measure’. This is convergence in the following metrizable topology.

**Definition 5.** We define a distance $\tau(f(\cdot, \cdot), g(\cdot, \cdot))$ between two arbitrary measurable functions of two variables as the infimum of the numbers $\varepsilon > 0$ such that

\[ \text{th}((x, y) : |f(x, y) - g(x, y)| > \varepsilon) \leq \varepsilon. \]

Convergence in this $\tau$-metric implies convergence in measure (but not vice versa).

**Lemma 6.** The set of measurable functions is complete in the $\tau$-metric.

**Proof.** Let \{\(f_n(\cdot, \cdot)\)\} be a Cauchy sequence of measurable functions in the $\tau$-metric. Passing to a subsequence if necessary, we can suppose that $\|f_n - f_{n+1}\|_{\tau} < 2^{-n}$. Let

\[ Z_n = \{(x, y) : |f_n(x, y) - f_{n+1}(x, y)| > 2^{-n}\}. \]

Then $\text{th}(Z_n) \leq 2^{-n}$, and for the set $Z'_n := \bigcup_{k \geq n} Z_k$ we have $\text{th}(Z'_n) \leq 2^{1-n}$. Thus, the set $\bigcap Z'_n$ has zero thickness, while outside it the sequence $(f_n)$ converges pointwise to some function $f_0$. Moreover, outside $Z'_n$ this sequence converges uniformly, and

\[ |f_0 - f_n| \leq |f_n - f_{n+1}| + |f_{n+1} - f_{n+2}| + \cdots \leq 2^{1-n}. \]

This means that $\|f_0 - f_n\|_{\tau} \leq 2^{1-n}$, and hence $f_n$ converges to $f_0$ in the metric $\tau$. \(\square\)
Let
\[ \xi_X: X = \bigsqcup_{i=1}^{n} X_i, \quad \xi_Y: Y = \bigsqcup_{i=1}^{m} Y_i \]
be finite partitions of the spaces \( X \) and \( Y \), respectively, into measurable subsets of positive measure. Functions which are constant mod 0 on each product \( X_i \times Y_j \) will be called step functions, and finite linear combinations
\[ \sum_{i=1}^{N} a_i(x)b_i(y) \]
will be called functions of finite rank.

The next theorem connects finite-rank functions and virtual continuity.

**Theorem 6.** The \( \tau \)-closure of the set of step functions (or the set of finite-rank functions) is exactly the set of virtually continuous functions. In other words, a function \( f \) is virtually continuous on \( X \times Y \) if and only if for any \( \varepsilon > 0 \) there are families of disjoint measurable sets \( A_1, \ldots, A_n \subset X \) and \( B_1, \ldots, B_n \subset Y \), and numbers \( c_{ij}, 1 \leq i,j \leq n \), such that
\[ \sum \mu(A_i) > 1 - \varepsilon, \quad \sum \nu(B_i) > 1 - \varepsilon, \quad |F(x,y) - c_{ij}| < \varepsilon \quad \text{for almost all } x \in A_i, \ y \in B_j. \]

**Proof.** First, we show that a \( \tau \)-limit \( f(x,y) \) of step functions \( f_n(x,y) \) is virtually continuous. We choose admissible metrics \( \rho_X \) and \( \rho_Y \) on \( X \) and \( Y \), respectively, so that the step functions \( f_n \) are continuous.

Passing to a subsequence if necessary, we can suppose that \( \tau(f_n,f) < 1/2^n \). This means that
\[ \text{th}\left(\left\{ (x,y): |f(x,y) - f_n(x,y)| > \frac{1}{2^n} \right\}\right) < \frac{1}{2^n}. \]

We choose subsets \( X_n \subset X \) and \( Y_n \subset Y \) so that
\[ \left\{ (x,y): |f(x,y) - f_n(x,y)| > \frac{1}{2^n} \right\} \mod 0 \subset (X \times Y_n) \cup (X_n \times Y) \]
and \( \mu(X_n) + \nu(Y_n) < 1/2^n \). Let
\[ \tilde{X}_n = X \setminus \bigcup_{k>n} X_k \quad \text{and} \quad \tilde{Y}_n = Y \setminus \bigcup_{k>n} Y_k. \]

Clearly, \( \mu(\tilde{X}_n) > 1 - 1/2^n \) and \( \nu(\tilde{Y}_n) > 1 - 1/2^n \). We can replace the set \( \tilde{X}_n \) by one of its subsets (calling it \( \tilde{X}_n \) again) which is compact in the metric \( \rho_X \), coincides with the support of the measure \( \mu \) restricted to \( \tilde{X}_n \), and has large measure \( \mu(\tilde{X}_n) > 1 - 1/2^n \), and similarly for \( \tilde{Y}_n \). Now \( L_\infty \) and \( C \) define the same distance between continuous functions on \( \tilde{X}_n \times \tilde{Y}_n \), and hence \( (f_k) \) is a sequence of functions convergent in \( C(\tilde{X}_n \times \tilde{Y}_n) \). Thus, \( f \) is equivalent to a continuous function on \( \tilde{X}_n \times \tilde{Y}_n \). Since \( n \) is arbitrary, we conclude that \( f \) is virtually continuous.
Now we prove the converse. Let \( f \) be virtually continuous. We need to approximate it by step functions in the \( \tau \)-metric. Fixing any \( \varepsilon > 0 \), we find sets \( \tilde{X} \subset X \) and \( \tilde{Y} \subset Y \) with measures \( > 1 - \varepsilon \) and admissible metrics \( \rho_X \) and \( \rho_Y \) such that \( f \) is equivalent to a function \( \tilde{f} \) which is continuous on \( \tilde{X} \times \tilde{Y} \). Passing to subsets if necessary, we can also suppose that \( \tilde{X} \) and \( \tilde{Y} \) are compact in the respective metrics. Using the uniform continuity of \( \tilde{f} \) on a compact metric space, we partition \( \tilde{X} \) and \( \tilde{Y} \) into small enough parts so that \( \tilde{f} \) is constant to within \( \varepsilon \) on products of partition elements. This provides a step function which is \( \varepsilon \)-close to \( f \) in the \( \tau \)-metric. \( \square \)

Theorem 6 also shows the purely measure-theoretic nature of virtual continuity and the possibility of generalizing it for other pairs of \( \sigma \)-subalgebras. Closely related notions are discussed in [12].

We apply Theorem 6 for proving Proposition 2.

**Proof of Proposition 2.** If a function \( f \) is equivalent to a continuous function, then the function \( F \) on \( G \times G \) is equivalent to a continuous function (in the metric of \( G \), which is admissible with respect to Haar measure). Hence, \( F \) is virtually continuous.

Let us prove the converse. Fix \( \varepsilon > 0 \). We choose any point \( g_0 \in G \) and establish that in a sufficiently small neighbourhood of it the essential variation of \( f \) does not exceed \( \varepsilon \). Since \( g_0 \) and \( \varepsilon > 0 \) are arbitrary, we see that \( f \) coincides almost everywhere with its essential upper limit, which is a continuous function itself.

Using the virtual continuity of \( f \), we find families of disjoint measurable subsets \( A_1, \ldots, A_n \) and \( B_1, \ldots, B_n \), and numbers \( c_{ij}, 1 \leq i, j \leq n \), such that

\[
\sum |A_i| > \frac{1}{2}, \quad \sum |B_i| > \frac{1}{2},
\]

\[
|F(x, y) - c_{ij}| < \frac{\varepsilon}{3} \quad \text{for almost all} \ x \in A_i, \ y \in B_j.
\]

By the pigeonhole principle we find indices \( i, j \) such that \( |C| > 0 \), where \( C = A_i \cap g_0 B_j \).

The continuity in the mean of translation implies that for some neighbourhood \( \Delta \) of unity we have \( |C \cap sC| > |C|/2 \) for all \( s \in \Delta \). It suffices to prove that in the neighbourhood \( \Delta g_0 \) of \( g_0 \) the inequality \( |f(z) - c_{ij}| < \varepsilon/3 \) holds for almost all points \( z \). Let \( Sg_0 \), where \( S \subset \Delta \), be the set of \( z \) for which this is not so. Then for almost all \( x \in C \subset A_i \) and \( y \in g_0^{-1}C \subset B_j \) we have \( xy^{-1} \notin Sg_0 \). In other words, the following integral vanishes:

\[
0 = \int \chi_C(x)\chi_{g_0^{-1}C}(y)\chi_{Sg_0}(xy^{-1}) \, dx \, dy
\]

\[
= \int \chi_C(x)\chi_tC(x)\chi_S(t) \, dx \, dt = \int_S |C \cap tC| \, dt \geq |S| \cdot \frac{|C|}{2},
\]

and hence \( |S| = 0 \) as required (the second equality corresponds to the change of variables \( (x, y) \mapsto (x, t), t = xy^{-1}g_0^{-1} \)). \( \square \)

Measurable functions \( f(\cdot, \cdot) \), as we have seen, are classified by matrix distributions, that is, by measures on the space of infinite matrices \( (a_{ij})_{i,j=1}^\infty \) induced by the
map \( f \mapsto (a_{ij} = f(x_i, y_j)) \), where the points \( x_i \) in \( X \) and \( y_i \) in \( Y \), \( i = 1, 2, \ldots \), are chosen independently. Virtual continuity can also be characterized in these terms.

**Theorem 7.** Let \( x_1, x_2, \ldots \) (respectively, \( y_1, y_2, \ldots \)) be independent random points in \( X \) (respectively, in \( Y \)). Virtual continuity for a measurable function \( f(x, y) \) is equivalent to each of following two conditions.

(i) For any \( \varepsilon > 0 \) there is a positive integer \( N \) such that the probability of the following event tends to 1 as \( n \) tends to \( +\infty \): if points \( x_1, \ldots, x_n \) are chosen independently at random in \( X \) and \( y_1, \ldots, y_n \) are chosen independently at random in \( Y \), then there are partitions

\[
\{1, \ldots, n\} = \bigsqcup_{i=0}^{N} A_i = \bigsqcup_{i=0}^{N} B_i,
\]

such that

\[
|A_0| < \varepsilon n, \quad |B_0| < \varepsilon n, \quad |f(x_s, y_t) - f(x_r, y_p)| < \varepsilon
\]

for \( s, r \in A_i \) and \( p, t \in B_j \), \( 0 < i, j \leq N \).

(ii) For any \( \varepsilon > 0 \) there is a positive integer \( N \) for which the probability of the following event equals 1: if points \( x_1, x_2, \ldots \) are chosen independently at random in \( X \) and \( y_1, y_2, \ldots \) are chosen independently at random in \( Y \), then there are two partitions of the natural numbers

\[
\{1, 2, \ldots \} = \bigsqcup_{i=0}^{N} A_i = \bigsqcup_{i=0}^{N} B_i,
\]

such that the upper density of the set \( A_0 \cup B_0 \) (that is, \( \limsup |(A_0 \cup B_0) \cap [1, n]|/n \)) is less than \( \varepsilon \) and

\[
|f(x_s, y_t) - f(x_r, y_p)| < \varepsilon \quad \text{for} \quad s, r \in A_i, \quad p, t \in B_j, \quad 0 < i, j \leq N.
\]

**Proof.** A virtually continuous function can be approximated in the \( \tau \)-metric by step functions, and hence satisfies (i) and (ii) by the law of large numbers.

We deduce (i) from (ii). A set of upper density less than \( \varepsilon \) contains fewer than \( 2\varepsilon n \) elements from 1 to \( n \) for all sufficiently large \( n \). This means that with probability 1 for all sufficiently large \( n \) there are partitions of the set \( \{1, \ldots, n\} \) which satisfy (i). This certainly implies that the probability of the specific event tends to 1 as a function of \( n \).

It remains to deduce virtual continuity from (i). We can and do suppose that both \( X \) and \( Y \) are the unit intervals \([0, 1]\) with Lebesgue measure. We need the following standard lemma on large deviations for the \( U \)-statistic.

**Lemma 7.** Let \( Z \subset X \times Y \) be a measurable subset of \( X \times Y \). Consider the following event: the number of points \((x_i, y_j)\), \(1 \leq i, j \leq n \), in \( Z \) differs from \( n^2|Z| \) by more than \( n^{9/5} \). Then its probability does not exceed \( 2n^{-3/5} \).

**Proof.** We introduce the function \( g(x, y) = \chi_Z(x, y) - |Z| \) (with zero mean) and the random variables \( \xi_{i,j} = g(x_i, y_j) \). We have \( n^2 \) centred random variables taking
values in $[-1, 1]$. Most of them are independent, which allows us to estimate the variance:

$$
E\left( \sum_{i,j} g(x_i, y_j) \right)^2 = E\left( \sum_{i,j} g(x_i, y_j)^2 + \sum_{i,j,k: k \neq j} g(x_i, y_j)g(x_i, y_k) \right) + \sum_{i,j,k: k \neq i} g(x_i, y_j)g(x_k, y_j) \leq 2n^3.
$$

Now the required estimate follows from Chebyshev’s inequality. □

Let $\mathcal{N} \subset \mathbb{N}$ be a subset such that

$$
\sum_{n \in \mathcal{N}} n^{-3/5} < +\infty.
$$

Then by the Borel–Cantelli lemma, for any measurable set $Z \subset [0, 1]^2$ and for almost all random pairs of sequences $(\{x_i\}, \{y_i\})$ we have

$$
\frac{n^2}{2} |Z| \leq \# \{(i, j): 1 \leq i, j \leq n, (x_i, y_j) \in Z\} \leq 2n^2 |Z| \quad (4)
$$

for sufficiently large $n \in \mathcal{N}$.

Applying this argument to a countable family of sets of the form $Z = R \setminus f^{-1}(\Delta)$, where $R$ runs over the rectangles with rational vertex coordinates and $\Delta$ runs over the intervals of the real line with rational endpoints, we get that for almost every random pair of sequences $(\{x_i\}, \{y_i\})$ the inequality (4) holds for any $Z$ of the indicated form and for sufficiently large $n$ (how large depends on the random pair and on $Z$). In what follows we consider only pairs of sequences with this property (this includes almost all random pairs of sequences).

We fix an $\varepsilon > 0$ and find $N$ from (i). Consider a random pair of sequences of independent points $x_1, x_2, \ldots \in X$ and $y_1, y_2, \ldots \in Y$. With probability 1 this sequence satisfies (i) for each $n$. For fixed $n \in \mathcal{N}$ we consider the empirical distributions

$$
\mu_j(n) = n^{-1} \sum_{i \in A_j^n} \delta(x_i)
$$

on $X$ and the analogous empirical distributions on $Y$. Passing to a subsequence if necessary, we can assume that the sequences of measures $\{\mu_j(n)\}_{n \in \mathcal{N}}$ converge weakly to some measures $\mu_j(\infty)$ as $n \to +\infty$, and the same for the corresponding measures on $Y$. Note that

$$
\sum_{j=1}^N \mu_j(\infty) = \mu
$$

with probability 1, hence with probability 1 all the measures $\mu_j(\infty)$ are absolutely continuous with respect to $\mu$. Then they have non-negative measurable densities $\varphi_j$ by the Radon–Nikodym theorem. We have $\int \varphi_0 \leq \varepsilon$, and thus the measure of the set of $x \in X$ satisfying $\varphi_0(x) \geq 1/2$ does not exceed $2\varepsilon$. For any other
\(x\) we have \(\sum_{j=1}^{N} \varphi_j(x) \geq 1/2\), hence \(\varphi_j(x) \geq 1/(2N)\) for some \(j\). Therefore, we can partition \(X\) into sets \(X_0, X_1, \ldots, X_N\) such that \(\mu(X_0) < 2\varepsilon\) and \(\varphi_j \geq 1/(2N)\) on \(X_j\). We can construct an analogous partition of \(Y\). Let us prove that with probability 1 the function \(f\) has essential variation (the essential supremum minus the essential infimum) at most \(2\varepsilon\) on \(X_j \times Y_k\) for \(j, k \geq 1\). Indeed, if the essential variation exceeds \(2\varepsilon\), then there are intervals \(\Delta_1\) and \(\Delta_2\) with rational endpoints and at a distance at least \(3\varepsilon/2\) from each other such that the sets \(f^{-1}(\Delta_i), \ i = 1, 2\), have intersection with \(X_j \times Y_k\) of positive measure. We fix some small \(\delta > 0\) (to be chosen later depending on \(\varepsilon\) and \(N\)). There are rectangles \(R_i, i = 1, 2\), with rational vertices such that

\[
|R_i \cap (X_j \times Y_k) \cap f^{-1}(\Delta_i)| > (1 - \delta)|R_i|.
\]

From the weak convergence of the measures, the inequalities \(\varphi_j \geq 1/(2N)\) on \(X_j\), and the analogous inequalities on \(Y_k\) it follows that as \(n \to \infty, n \in \mathcal{N}\), we have

\[
(\mu_j(n) \times \nu_j(n))(R_i) \to (\mu_j(\infty) \times \nu_j(\infty))(R_i)
\]

\[
\geq (\mu_j(\infty) \times \nu_j(\infty))(R_i \cap (X_j \times Y_k)) \geq \frac{1}{4N^2}|R_i \cap (X_j \times Y_k)|,
\]

and therefore for sufficiently large \(n \in \mathcal{N}\) the number of points \((x_s, y_t) \in R_i\) with \(s \in A_j^n\) and \(t \in B_k^n\) is not less than

\[
\frac{1 - \delta}{10N^2}|R_i|n^2.
\]

The property (i) guarantees that either none of those points (for \(i = 1\) and \(i = 2\) together) lie in \(f^{-1}(\Delta_1)\), or none of them lie in \(f^{-1}(\Delta_2)\). But for each of the sets \(Z = R_i \setminus f^{-1}(\Delta_i)\) of measure at most \(\delta|R_i|\) and for sufficiently large \(n \in \mathcal{N}\) the number of points \((x_s, y_t), 1 \leq s, t \leq n,\) in \(Z\) does not exceed \(2\delta|R_i|n^2\). This contradicts the above lower estimate when \(\delta\) is small enough.

Thus, with probability 1 the function \(f\) can be approximated by a step function (respecting the constructed partitions) in the \(\tau\)-metric. It remains to use Theorem 6 and the arbitrariness of \(\varepsilon\). \(\square\)

### 2.6. Bistochastic measures and polymorphisms

From the measure-theoretic point of view, a function of \(k\) variables on a product of standard measure spaces with a continuous measure is nothing but a function on a standard measure space with a continuous measure (due to the isomorphism of all such spaces). In order to deal with it as a function of \(k\) variables, we have to introduce a different category other than just measure spaces.

Namely, consider the following structure: a measure space \((\mathcal{X}, \mathcal{A}, m)\), with \(k\) selected \(\sigma\)-subalgebras \(\mathcal{A}_1, \ldots, \mathcal{A}_k\) of \(\mathcal{A}\). It is natural to assume that these subalgebras together generate the whole \(\sigma\)-algebra \(\mathcal{A}\). We identify the algebras \(\mathcal{A}_i\) with subalgebras of \(\mathcal{A} = \prod \mathcal{A}_i\) by multiplying them by the trivial subalgebras of the other factors. In other words, a function \(f(x_1, \ldots, x_k)\) on \(\mathcal{X}\) is \(\mathcal{A}_1\)-measurable if and only if \(f\) depends only on the \(i\)th variable \(x_i\) \((i = 1, \ldots, k)\). Functions depending on any smaller number of variables are defined similarly. In this way a measurable function
of several variables, and the variables themselves, are defined in a measure space with a distinguished family of σ-subalgebras which together generate the whole σ-algebra.

**Definition 6.** A measurable function on $\mathcal{X}$ with $k$ selected subalgebras will be called a general measurable function of $k$ variables.

In the classical case these subalgebras are independent and the variables themselves are said to be independent (although this is just a lucky coincidence of the probabilistic and analytic meanings of independence), but many facts about measurable functions remain true in the general case as well.

Let $\lambda$ be a measure on the σ-algebra $\mathcal{A}$. It can be restricted to the σ-subalgebras $\mathcal{A}_i$, $i = 1, \ldots, k$. We shall consider measures $\lambda$ such that their restrictions are absolutely continuous with respect to the restrictions of $m$ to $\mathcal{A}_i$. If the restrictions of $\lambda$ to $\mathcal{A}_i$ coincide with $m$ for $i = 1, \ldots, k$, then such a measure $\lambda$ is said to be multistochastic with respect to the given subalgebras (bistochastic for $k = 2$), but if the restrictions are just equivalent to $m$ for $i = 1, \ldots, k$, then we say that $\lambda$ is almost multistochastic. Finally, if $\lambda(U) \leq m(U)$ for any $U \in \mathcal{A}_i$, $i = 1, \ldots, k$, then we say that $\lambda$ is submultistochastic.

Of course, a bistochastic measure on $X \times Y$ can be singular with respect to the product measure. For instance, in the case of a direct product of intervals $(X, \mu) = (Y, \nu) = [0, 1]$ there is a bistochastic measure $\lambda$ concentrated on the diagonal $\{x = y\}$ (the push forward of $\mu$ by the map $x \mapsto (x, x)$).

Throughout what follows we suppose for simplicity that $k = 2$, that is, we consider functions of two variables, though there is no serious difference for $k > 2$. We do not consider only independent variables: most of the notions can be defined for a general pair of σ-algebras. But even the case of independent variables is often usefully treated as a general case.

Bistochastic measures on a direct product of spaces define a so-called polymorphism of the space $(X, \mu)$ into the space $(Y, \nu)$ (see [17]), that is, a ‘multivalued map’ with an invariant measure. The case of identified variables $(X, \mathcal{A}, \mu) = (Y, \mathcal{B}, \nu)$ is of special interest: a polymorphism in this case generalizes the concept of an automorphism of a measure space. Almost bistochastic measures define a polymorphism with a quasi-invariant measure. A bistochastic (or almost bistochastic) measure $\lambda$ also defines a bilinear ($k$-linear in the general case) form

$$(f(x), g(y)) \mapsto \int f(x)g(y) \, d\lambda(x, y)$$

which corresponds to a so-called Markov (respectively, quasi-Markov) operator acting in the corresponding function spaces. We note that this operator $U_\lambda$ is a contraction, that is, has norm at most 1, and preserves the cone of non-negative functions. In the case of a bistochastic measure this operator (as well as the adjoint operator) preserves constants: $U_\lambda 1 = 1$.

See [15], [17], [18] about many connections in the theory of polymorphisms (Markov operators, joinings, couplings, correspondences, Young measures, bibundles, and so on). Bistochastic measures play a key role in the intensively developing theory of continuous graphs [7].
We remark that for a quasi-bistochastic measure $\lambda$ on $X \times Y$ all sets of zero proper thickness are measurable and have measure 0. Hence, all virtually open sets are $\lambda$-measurable, and therefore properly virtually continuous functions are measurable. Equivalent (with respect to the measure $\mu \times \nu$) properly virtually continuous functions are also equivalent with respect to $\lambda$ in view of Proposition 3. Thus, to any equivalence class of measurable virtually continuous functions there corresponds a unique $\lambda$-equivalence class of $\lambda$-measurable functions. In what follows, we formulate this observation as an embedding theorem for normed spaces.

2.7. A norm on virtually continuous functions. The convergence in the $\tau$-metric defined above is an analogue of convergence in measure for virtually continuous functions. There are analogues of familiar Banach spaces of measurable functions.

A measurable function $h(\cdot, \cdot)$ on the space $(X \times Y, \mu \times \nu)$ is said to be subbistochastic if the measure with $(\mu \times \nu)$-density $|h(\cdot, \cdot)|$ is subbistochastic. Denote by $\mathcal{S}$ the set of subbistochastic functions.

We say that a function of the form $f(x, y) = a(x) + b(y)$ is additively separable. The following construction defines a norm (the so-called regulator norm) of a function of two variables, where the regulator is an additively separable function and the norm is taken in $L^1$. We define a finite or infinite norm of a measurable function $f(\cdot, \cdot)$ by

$$\|f\|_{SR^1} := \inf\left\{ \int_X a(x) \, d\mu(x) + \int_Y b(y) \, d\nu(y) : a(x) \geq 0, b(y) \geq 0, |f(x, y)| \leq a(x) + b(y) \right\}.$$ 

The connection between the $SR^1$-norm and the $\tau$-metric is as follows.

**Lemma 8.** For any function $f$

$$\tau(0, f) \leq \sqrt{2\|f\|_{SR^1}}.$$ 

**Proof.** If $\|f\|_{SR^1} = \infty$, then the assertion is trivial. Assume that $\|f\|_{SR^1} < t^2/2$ for some $t > 0$. Then there are non-negative functions $a: X \to \mathbb{R}$ and $b: Y \to \mathbb{R}$ such that

$$|f(x, y)| \leq a(x) + b(y) \quad \text{and} \quad \int_X a + \int_Y b < \frac{t^2}{2}.$$ 

Then Chebyshev’s inequality implies that

$$\mu\left\{ x : a(x) \geq \frac{t}{2} \right\} + \nu\left\{ y : b(y) \geq \frac{t}{2} \right\} < t.$$ 

But

$$(x, y) : f(x, y) \geq t \quad \text{mod} 0 \quad \subseteq \quad \left\{ x : a(x) \geq \frac{t}{2} \right\} \times Y \cup \left\{ y : b(y) \geq \frac{t}{2} \right\},$$

and hence $\tau(0, f) < t$, as required. $\square$
Corollary 3. Convergence in the $\text{SR}^1$-norm implies convergence in the $\tau$-metric.

The next theorem is an analogue of Kantorovich’s well-known duality theorem [4] in the mass transportation problem (specifically, of the duality between the space of measures with the Kantorovich metric and the space of Lipschitz functions (see also [21])).

Theorem 8.

$$\|f\|_{\text{SR}^1} = \sup \left\{ \int_{X \times Y} |f(x,y)| h(x,y) \, d\mu(x) \, d\nu(y) : h \in \mathcal{H} \right\}. \tag{5}$$

Proof. We show first that the left-hand side of (5) is no less than the right-hand side. Indeed, if $h \in \mathcal{H}$, and the functions $a : X \to \mathbb{R}$ and $b : Y \to \mathbb{R}$ satisfy $|f(x,y)| \mod 0 \leq a(x) + b(y)$, then

$$\int_{X \times Y} |f(x,y)h(x,y)| \, d\mu(x) \, d\nu(y) \leq \int_{X \times Y} |h(x,y)|(a(x) + b(y)) \, d\mu(x) \, d\nu(y)$$

$$\leq \int_X a(x) \int_Y |h(x,y)| \, d\nu(y) \, d\mu(x) + \int_Y b(y) \int_X |h(x,y)| \, d\mu(x) \, d\nu(y)$$

$$\leq \int_X a(x) \, d\mu(x) + \int_Y b(y) \, d\nu(y).$$

Taking the infimum over admissible pairs of functions $a$ and $b$, we obtain the desired inequality.

It remains to verify that the right-hand side of (5) is no less than the left-hand side. Let $\theta > 0$ be arbitrary and consider the following two convex subsets of $L^1(X \times Y, \mu \times \nu)$:

$$A = \left\{ a(x) + b(y) : a, b \geq 0, \int_X a(x) \, d\mu(x) + \int_Y b(y) \, d\nu(y) \leq \|f\|_{\text{SR}^1} - \theta \right\},$$

$$B = \left\{ g(x,y) : g(x,y) \mod 0 \geq |f(x,y)| \right\}.$$

By the definition of the norm these two sets are disjoint. Let us check that the distance between them in $L^1$ is positive. If not, then there are sequences of non-negative functions $a_n$, $b_n$, and $g_n$ such that $a_n(x) + b_n(y) \in A$, $g_n \in B$, and

$$\|a_n(x) + b_n(y) - g_n(x,y)\|_{L^1(X \times Y)} \to 0.$$

Passing to a subsequence if necessary, we see that

$$a_n(x) + b_n(y) - g_n(x,y) \to 0 \quad (\mu \times \nu)\text{-almost everywhere.}$$

Using the Komlós theorem [6], we pass to a subsequence such that

$$\frac{1}{N} \sum_{k=1}^N a_k(x) \to a(x) \quad \mu\text{-almost everywhere,}$$

$$\frac{1}{N} \sum_{k=1}^N b_k(y) \to b(y) \quad \nu\text{-almost everywhere.}$$
for some non-negative functions $a \in L^1(X)$ and $b \in L^1(Y)$. Consider the function 

$$g(x, y) := a(x) + b(y).$$

Then

$$\frac{1}{N} \sum_{k=1}^{N} g_n(x, y) \to g(x, y) \quad (\mu \times \nu)\text{-almost everywhere.}$$

Clearly, this implies that $g \in B$. On the other hand,

$$\int a + \int b \leq \limsup_{N} \frac{1}{N} \sum_{k=1}^{N} \left( \int a_k + \int b_k \right) \leq \|f\|_{SR^1} - \theta$$

due to the lower semicontinuity of the integral of non-negative functions with respect to convergence almost everywhere. Hence, $g \in A$, a contradiction.

Now we use the Hahn–Banach separability theorem. Since $A$ contains 0, there is a function $h \in L^\infty(X \times Y)$ such that $\int gh < 1$ for any $g \in A$ and $\int gh > 1$ for any $g \in B$. And since $B$ is a translate of the non-negative cone and $\int gh > 1$ for any $g \in B$, we get that $h \geq 0$ almost everywhere. For any set $X_1 \subset X$ put

$$a(x) = \frac{\|f\|_{SR^1} - \theta}{\mu(X_1)} \chi_{X_1}(x).$$

Then $a(x) \in A$, hence

$$\frac{1}{\mu(X_1)} \int_{X_1} \int_{Y} (\|f\|_{SR^1} - \theta) h(x, y) \, d\nu(y) \, d\mu(x) \leq 1.$$ 

We interchange the variables and obtain the analogous inequalities, from which we can conclude that the function $\tilde{h} = (\|f\|_{SR^1} - \theta) h$ belongs to $\mathcal{S}$. But $|f| \in B$, hence

$$\int |f| \tilde{h} = (\|f\|_{SR^1} - \theta) \int |f| h \geq \|f\|_{SR^1} - \theta.$$ 

Since $\theta$ is arbitrary, the theorem is proved. □

Here is another theorem about the $SR^1$-norm.

**Theorem 9.** For any measurable function $f : X \times Y \to \mathbb{R}$,

$$\frac{1}{4} \|f\|_{SR^1} \leq \int_{0}^{\infty} \text{th}\{|f| \geq \lambda\} \, d\lambda \leq 2\|f\|_{SR^1}. \quad (6)$$

**Proof.** If $|f(x, y)| \mod 0 \leq a(x) + b(y)$, then

$$\{(x, y) : |f(x, y)| \geq \lambda\} \mod 0 \subset \left( \left\{ x : a(x) \geq \frac{\lambda}{2} \right\} \times Y \right) \cup \left( X \times \left\{ y : b(y) \geq \frac{\lambda}{2} \right\} \right),$$

and hence

$$\text{th}\{|f| \geq \lambda\} \leq \mu\left\{ a \geq \frac{\lambda}{2} \right\} + \nu\left\{ b \geq \frac{\lambda}{2} \right\}.$$
Integrating with respect to $\lambda$, we have

$$
\int_0^\infty \text{th}\{|f| \geq \lambda\} \, d\lambda \leq \int_0^\infty \left( \mu\left\{ a \geq \frac{\lambda}{2} \right\} + \nu\left\{ b \geq \frac{\lambda}{2} \right\} \right) \, d\lambda = 2\left( \int_X a + \int_Y b \right).
$$

Taking the infimum over pairs of functions $a$ and $b$, we obtain the right-hand inequality in (6).

Let us prove the left-hand inequality. Since $\text{th}\{|f| \geq \lambda\}$ decreases monotonically with respect to $\lambda$,

$$
\int_0^\infty \text{th}\{|f| \geq \lambda\} \, d\lambda \geq \sum_{k \in \mathbb{Z}} 2^{k-1} \text{th}\{|f| \geq 2^k\}. \quad (7)
$$

For any $\varepsilon > 0$ choose sets $A_k$ and $B_k$ so that

$$
\{|f| \geq 2^k\} \mod 0 \subset (A_k \times Y) \cup (X \times B_k)
$$

and

$$
\mu(A_k) + \nu(B_k) \leq (1 + \varepsilon) \text{th}\{|f| \geq 2^k\}.
$$

Let $a(x) = \sum 2^{k+1} \chi_{A_k}(x)$ and $b(y) = \sum 2^{k+1} \chi_{B_k}(y)$. It is easy to check that $|f(x,y)| \mod 0 \leq a(x) + b(y)$, hence

$$
\|f\|_{SR^1} \leq \int_X a + \int_Y b \leq 4(1 + \varepsilon) \sum_{k \in \mathbb{Z}} 2^{k-1} \text{th}\{|f| \geq 2^k\}.
$$

The last inequality combined with (7) (and the arbitrariness of $\varepsilon$) completes the proof. □

This theorem has a useful corollary.

**Corollary 4.** If $\|f\|_{SR^1} < \infty$, then the function $f$ can be approximated in the $SR^1$-norm by its cutoffs.

**Proof.** Let $f_N$ be the two-sided cutoff of $f$ at level $N$. Then $\{|f - f_N| \geq \lambda\} \subset \{|f| \geq N + \lambda\}$, and therefore

$$
\int_0^\infty \text{th}\{|f - f_N| \geq \lambda\} \, d\lambda \leq \int_0^\infty \text{th}\{|f| \geq N + \lambda\} \, d\lambda = \int_0^\infty \text{th}\{|f| \geq \lambda\} \, d\lambda \to 0
$$

as $N \to \infty$. By Theorem 9, $\|f - f_N\|_{SR^1} \to 0$. □

**Theorem 10.** The closure of the set of step functions in the $SR^1$-norm consists exactly of all virtually continuous functions having finite norm (in particular, each bounded virtually continuous function belongs to this closure).

**Proof.** Corollary 3 and Theorem 6 imply that the $SR^1$-limit of a sequence of step functions is virtually continuous.

Now we have to approximate any virtually continuous function with finite $SR^1$-norm by step functions. Assume that $\|f\|_{SR^1} < \infty$ and let $\varepsilon > 0$. By Corollary 4 the two-sided cutoff $f_N$ of $f$ approximates $f$: $\|f - f_N\|_{SR^1} < \varepsilon$ for sufficiently
large $N$. Fix such an $N$. Next, the function $f_N$ is virtually continuous, hence it can be approximated in the metric $\tau$ by step functions in view of Theorem 6. Then $\tau(g, f_N) < \varepsilon/N$ for some step function $g$, and we can assume that the absolute value of $g$ is bounded by the same number $N$ (otherwise replace $g$ by a cutoff of it). Since $\tau(g, f_N) < \varepsilon/N$, there are sets $X_0 \subset X$ and $Y_0 \subset Y$ such that $\mu(X_0) < \varepsilon/N$, $\nu(Y_0) < \varepsilon/N$, and $|f_N - g| < \varepsilon/N$ almost everywhere on $(X \setminus X_0) \times (Y \setminus Y_0)$. But then

$$|f_N(x, y) - g(x, y)| \leq \frac{\varepsilon}{N} + 2N\chi_{X_0}(x) + 2N\chi_{Y_0}(y),$$

and hence

$$\|f_N - g\|_{SR^1} \leq \frac{\varepsilon}{N} + 2N(\mu(X_0) + \nu(Y_0)) \leq 5\varepsilon.$$

Thus, $\|f - g\|_{SR^1} \leq 6\varepsilon$ and the theorem is proved. $\Box$

Denote by $VC^1$ the space of all virtually continuous functions with finite $SR^1$-norm. It is an analogue of the space $L^1$ for virtually continuous functions and is a pre-dual for the space of polymorphisms with bounded densities of projections.

**Theorem 11.** The space dual to $VC^1$ is the space $QB^\infty$ of quasi-bistochastic signed measures $\eta$ on $X \times Y$ with finite norm

$$\|\eta\|_{qb} = \max\left\{ \left\| \frac{\partial P^x}{\partial \mu} |\eta| \right\|_{L^\infty(X,\mu)}, \left\| \frac{\partial P^y}{\partial \nu} |\eta| \right\|_{L^\infty(Y,\nu)} \right\},$$

where $P^x$ and $P^y$ are the projections onto $X$ and $Y$, respectively, and $|\eta|$ is the total variation of the signed measure $\eta$. The coupling between an $\eta \in QB^\infty$ and an $f(x, y) \in VC^1$ is defined as $\int \tilde{f} \, d\eta$, where $\tilde{f}$ is a properly virtually continuous function equivalent to $f$.\footnote{The remark at the end of §2.6 guarantees that the function $\tilde{f}$ is $\eta$-measurable and the value of the integral does not depend on the choice of $\tilde{f}$ for fixed $f$.}

In order to prove Theorem 11 we need a lemma.

**Lemma 9.** Let $K$ be a compact metric space and $F$ a continuous linear functional on the space $C(K)$. Assume that the continuous functions $f_1, f_2, \ldots$ on $K$ have uniformly bounded norms and disjoint supports. Then the series $\sum F(f_i)$ converges absolutely. If the (pointwise defined) function $f = \sum f_i$ is continuous, then $F(f) = \sum F(f_i)$.

**Proof.** By Riesz’s theorem our functional $F$ is integration with respect to a signed Borel measure of finite total variation. The absolute convergence of the above series follows from the countable additivity of this measure and the finiteness of the total variation. The equality $F(f) = \sum F(f_i)$ follows from Lebesgue’s dominated convergence theorem. $\Box$

**Proof of Theorem 11.** Let $\eta$ be a signed measure such that $\|\eta\|_{qb} < \infty$. Note that if for a step function $h$ the estimate $|h(x, y)| \leq a(x) + b(y)$ holds $(\mu \times \nu)$-almost everywhere, then it holds almost everywhere on a product of sets of full measure,

Thus, $\|f - g\|_{SR^1} \leq 6\varepsilon$ and the theorem is proved. $\Box$
and thus $|\eta|$-almost everywhere. This lets us integrate this inequality with respect to the measure $|\eta|$, which gives

$$\left| \int h(x, y) \, d\eta \right| \leq \int |h| \, d|\eta| \leq \int (a(x) + b(y)) \, d|\eta|$$

$$\leq \|\eta\|_{q\beta} \left( \int |a(x)| \, d\mu(x) + \int |b(y)| \, d\nu(y) \right).$$

Taking the infimum with respect to $a$ and $b$ such that $|h| \equiv 0 \mod \|h\|_{SR^1}$, we get that $\left| \int h \, d\eta \right| \leq \|h\|_{SR^1} \|\eta\|_{q\beta}$, as required.

Now we need to show that any continuous linear functional $F$ on $VC^1$ has such a representation. We can suppose without loss of generality that $\|F\| = 1$. For a step set (a finite union of rectangles) $Z \subset X \times Y$ let

$$\eta(Z) := F(\chi_Z),$$

$$|\eta|(Z) := \sup \sum_{\bigcup Z_i \subset Z} |\eta(Z_i)|,$$

where the supremum is taken over all sequences of disjoint step sets $Z_1, Z_2, \ldots$ in $Z$. Obviously, the supremum can be taken just over finite families, and we can also take just rectangular sets $Z_i$. The set functions defined above are finitely additive.

For any finite family of disjoint step sets $Z_i \subset Z$ we have

$$\sum |\eta(Z_i)| = F\left( \sum (\pm \chi_{Z_i}) \right).$$

Moreover, $|\pm \chi_{Z_i}| \leq \chi_Z$. If $Z = X_1 \times Y$, then $\|\sum (\pm \chi_{Z_i})\|_{SR^1} \leq \mu(X_1)$, and hence $|\eta|(X_1 \times Y) \leq \mu(X_1)$. Similarly, $|\eta|(X \times Y_1) \leq \nu(Y_1)$. Let us check that the finitely additive set functions $\eta$ and $|\eta|$, defined on the algebra of step sets, can be extended to a signed measure and a measure on the whole $\sigma$-algebra $\mathfrak{A} \times \mathfrak{B}$ on the space $X \times Y$.

By the Kolmogorov–Hahn criterion it suffices to verify that $|\eta|(Z) = \sum |\eta|(Z_i)$ whenever the $Z_i$ are disjoint step sets and $Z = \bigcup Z_i$ is also a step set. Since $|\eta|$ is a premeasure, the inequality $|\eta|(Z) \geq \sum |\eta|(Z_i)$ is clear. It remains to prove the reverse inequality. By definition, $|\eta|(Z)$ is the supremum of the sums $\sum |\eta|(P_k)$ over all finite families of disjoint rectangles $P_k$ in $Z$, and hence it suffices to prove that

$$\sum_k |\eta|(P_k) \leq \sum_i |\eta|(Z_i).$$

Since $|\eta|$ is finitely additive, it suffices to prove that

$$|\eta|(P_k) \leq \sum_i |\eta|(Z_i \cap P_k)$$
for each rectangle $P_k$. Subdividing each set $Z_i \cap P_k$ into finitely many rectangles, we reduce it to an inequality of the form

$$|\eta(Q)| \leq \sum_i |\eta(Q_i)|,$$

where the rectangle $Q$ is the union of the disjoint rectangles $Q_i$.

Considering a series of cut semimetrics, we can easily construct admissible semimetrics $\rho_X$ and $\rho_Y$ such that the metric spaces $(X, \rho_X)$ and $(Y, \rho_Y)$ are precompact, and the projection of each side of each of the rectangles $Q_i$ and $Q$ is at a positive distance from its complement. In this case all the functions $\chi_{Q_i}$ and $\chi_Q$ are uniformly continuous on $(X \times Y, \rho_X \times \rho_Y)$ and therefore can be extended continuously to its completion (as 1 to the closure of a corresponding rectangle and as 0 to the closure of its complement). The supports of the extended functions are still disjoint. The space of continuous functions on the completion of $X \times Y$ embeds into $VC^1$ with norm at most 1, and thus $F$ acts as a continuous functional on it. Applying Lemma 9 to the sequence $\chi_{Q_i}$, we get that

$$\eta(Q) = F(\chi_Q) = \sum_i F(\chi_{Q_i}) = \sum_i \eta(Q_i),$$

which completes the proof. \(\square\)

**Corollary 5.** For virtually continuous functions in the space $VC^1$ (in particular, for bounded virtually continuous functions) there are well-defined integrals not only over sets of positive measure, as for all integrable functions, but also integrals with respect to bistochastic (singular) measures such as the Lebesgue measure on the diagonal $\{x = y\} \subset [0, 1]^2$ or with respect to a measure concentrated on the graph of a transformation with a quasi-invariant measure. Thus, virtually continuous functions have a ‘trace (restriction) on the diagonal’ in the sense of trace theorems.

As an application we prove a variant of the continuous Hall lemma, the Borel version of which is given in the appendix [8] to the book [7].

**Theorem 12.** Let $Z \subset X \times Y$ be a virtually closed set. Then the following two conditions are equivalent:

(i) there is a bistochastic measure $\lambda$ such that $\lambda(Z) = 1$;

(ii) the proper thickness $sth(Z)$ of the set $Z$ equals 1.

**Proof.** (ii) follows immediately from (i). Let us prove the reverse implication. Without loss of generality the set $Z$ can be assumed to be closed in the metric $d = \rho_X \times \rho_Y$, where $\rho_X$ and $\rho_Y$ are admissible metrics on $X$ and $Y$. Then the function $f(x, y) := \exp\{-d((x, y), Z)\}$ is properly virtually continuous.

We assert that its $VC^1$-norm equals 1. If not, then there are non-negative functions $a(x)$ and $b(y)$ such that $a(x) + b(y) \geq f(x, y)$ almost everywhere and

$$\int a + \int b = e^{-2\varepsilon}, \quad \varepsilon > 0.$$
Consider the $\varepsilon$-neighbourhood $Z_\varepsilon$ of the set $Z$ in the metric $d$. It is a virtually open set, hence by Lemma 3 its proper thickness coincides with its thickness (and therefore equals 1). We have
\[ a(x) + b(y) \geq e^{-\varepsilon} \chi_{Z_\varepsilon} \]
for almost all $x \in X$ and $y \in Y$, and thus
\[ \text{th}(Z_\varepsilon) \leq e^\varepsilon \left( \int a + \int b \right) = e^{-\varepsilon}, \]
a contradiction.

For an element with unit norm in a Banach space there is a linear functional of norm 1 which attains its norm on this element. By Theorem 11 this functional corresponds to a subbistochastic signed measure $\lambda$. But
\[ \left| \int f \, d\lambda \right| \leq \int |f| \, d|\lambda| = \int_0^1 |\lambda((f^{-1}[t, 1]))| \, dt \leq 1, \]
and since all the inequalities are just equalities, we have $\lambda(Z) = \lambda(f^{-1}\{1\}) = 1$. \(\square\)

As usual, Hall’s lemma has a standard formulation.

**Corollary 6.** For virtually closed $Z$ the equality $\max \lambda(Z) = \text{sth}(Z)$ always holds, where the maximum is taken over all bistochastic measures $\lambda$.

**Proof.** We use a typical trick: to the spaces $X$ and $Y$ we add spaces $X_0$ and $Y_0$ of measure $1 - \text{sth}(Z)$, and we normalize the measures on $X \sqcup X_0$ and $Y \sqcup Y_0$. Consider the set $Z \sqcup X_0 \times Y_0 \sqcup X_0 \times Y \sqcup X \times Y_0$ in $(X \sqcup X_0) \times (Y \sqcup Y_0)$. It has proper thickness 1. Apply Lemma 12 to this set and find the corresponding bistochastic measure. The part of it in $X \times Y$ is what we are searching for. \(\square\)

### 3. Applications: optimal transport, embedding theorems, traces of nuclear operators, restrictions of metrics

In this section we mention some applications of the concept of virtual continuity.

#### 3.1. Kantorovich duality in the optimal transportation problem.

We connect Theorem 11 on the space $QB^\infty$ of quasi-polymorphisms which is dual to the space $VC^1$ of virtually continuous functions, with the classical theorem of Kantorovich on duality in the continuous linear transportation problem. We recall this.

Consider a metric space $(X, \rho)$ and two Borel probability measures $\mu_1$ and $\mu_2$ on it.\(^2\) The following infimum is to be found:
\[
\inf \left\{ \int_X \int_X \rho(x, y) \, d\Psi(x, y) : \Psi \in QBS_{\mu_1, \mu_2}^\infty \right\}, \tag{8}
\]
where $QBS_{\mu_1, \mu_2}^\infty$ is the set of measures on $X \times X$ with (marginal) projections $\mu_1$ and $\mu_2$ (another name for such measures $\Psi$ is a polymorphism from $(X, \mu_1)$ to

\(^2\)In the classical paper [4] this space is compact, but in what follows we need only that it is a complete separable metric (=Polish) space.
virtual continuity, or a transportation scheme, or coupling, or joining, or Young measure, and so on).

The main facts covered under the general name of the duality theorem assert the following (we use our terminology).

1) The infimum in (8) is attained on some non-negative element $\Psi_0$ of the set $QBS_{\mu_1,\mu_2}^\infty$ (and is not attained, in general, on the space of absolutely continuous measures of the form $d\Psi = p(x,y)d\mu_1(x)d\mu_2(y)$, where $p$ is an integrable measurable function).

2) This infimum can be considered as the norm $\|\mu_1 - \mu_2\|$ of the signed measure in a certain space of signed Borel measures with finite total variation on the space $(X, \rho)$.

3) There is a dual definition of the norm:

$$\|\mu_1 - \mu_2\| = \sup \left\{ \int_X u(x) d(\mu_1 - \mu_2)(x) : u \in Lip_1(\rho) \right\},$$

where $Lip_1(\rho)$ is the unit ball in the space of Lipschitz functions with the usual Lipschitz norm. The supremum is also realized on some Lipschitz function $u_0$, and $u(x) - u(y) = \rho(x, y)$ holds $\Psi_0$-almost everywhere.

The main sense of the above assertions is that the norm of an element in a Banach space can be calculated using a functional in the dual space, and this reduces the problem to searching for the dual space.

The above statement is known as the duality theorem (or optimality criterion) in the optimal transportation problem and was formulated in the pioneering paper [4]. In fact, what is used is that the Lipschitz space is the Banach dual to the Kantorovich–Rubinstein space of measures. We stress that this is a duality theorem for functions of ‘one variable’, and we supplement it (more precisely, ‘cover’ it) by a duality theorem for functions of two variables.

Below we show how to apply Theorem 11, which is an assertion about the dual space of the space $VC^1$ of virtually continuous functions, to Kantorovich duality. Here it is more convenient to talk about transportation between two different spaces (of course, this is equivalent to the above problem on transportation in the same space).

Thus, we obtain yet another proof of the duality theorem, but the main feature is that our scheme includes spaces of metrics and spaces of schemes, unlike Kantorovich’s original approach. The choice of the spaces $VC^1$ and $QB^\infty$ is natural in the sense that smaller spaces are not enough (see the remark above) and admissible metrics are virtually continuous functions.

The two-level duality theorem in our specific situation leads, in turn, to the following general two-level duality theorem, and we hope that it has other applications. This is why we start with our general statement and then later explain how to apply it to optimal transportation.

---

3This observation, made in [5], established the convention of calling this the Kantorovich–Rubinstein norm, and the metric the Kantorovich metric.
Assume the following condition on $Z$:

Let $W$ and $Z = W^*$ denote two other real linear spaces, and consider a linear operator $A: W \to X$ and the adjoint operator $B = A^*: Y \to Z$:

$$W \xrightarrow{A} X, \quad X^* \xrightarrow{A^*} W^*. $$

Fix a positive element $\rho \in K \subset X$ and define a (finite or infinite) quasi-norm on $Z$ by

$$\|z\|_\rho = \inf \{(y, \rho): By = z, \ y \geq 0\}. $$

Assume the following condition: the space $X$ is the sum of the cone $K$ and the space $A(W)$. Then

$$\|z\|_\rho = \|z\|'_\rho := \sup\{(z, u): Au \leq \rho\}. $$

Moreover, if $\|z\|_\rho < \infty$, then there is a non-negative continuous functional $y \in X^*$ such that $(y, \rho) = \|z\|_\rho$. 

**Remark 1.** In the case when $X$ is a Banach space and the cone $K$ is closed and generating $(K - K = X)$, the classical Kakutani theorem says that a non-negative functional $y$ on $X$ is automatically norm bounded on $X$.

**Proof.** Obviously, $\|z\|'_\rho \leq \|z\|_\rho$. Indeed, for any element $y \in Y$ such that $y \geq 0$ and $By = z$ and for any element $u \in W$ such that $Au \leq \rho$ we have

$$(z, u) = (By, u) = (y, Au) = (y, \rho) - (y, \rho - Au) \leq (y, \rho), $$

with equality if $(y, \rho - Au) = 0$.

Assume that

$$C := \|z\|'_\rho = \sup\{(z, u): Au \leq \rho\} < \|z\|_\rho. $$

Note that finiteness of $C$ implies that $(z, w) = 0$ whenever $Aw = 0$ (otherwise consider $u = \lambda w$ for real $\lambda$ of appropriate sign, and then the coupling $(z, u)$ is unbounded). This means that $z \in B(Y)$, since the range of the adjoint operator is just the annihilator of the kernel of the direct operator. Moreover, if $Aw \geq 0$, then $(z, w) \geq 0$ (otherwise consider $u = \lambda w$ with negative $\lambda$). If $\rho = Au_0$ for some $u$, then for any $y$ with $By = z$ we have

$$(z, u_0) = (By, u_0) = (y, Au_0) = (y, \rho), $$

as required. Now let $\rho \notin A(W)$. We have to find a functional $y \in Y = X^*$ such that $(y, Au) = (z, u)$ for all $u \in W$ (this just means that $z = A^* y = By$), $y \geq 0$, and $(y, \rho) = C$. Such a $y$ is already defined on the linear span of the space $A(W)$ and the element $\rho$, and it is non-negative on this linear span. The last assertion holds for non-negative elements of the form $\rho - Au$ by the definition of the number $C$, and thus we should check it on non-negative elements of the form $Au - \rho$. Fix $\varepsilon > 0$ and find an element $w \in W$ such that $Aw \leq \rho$ and $(y, Aw) = (z, w) \geq C - \varepsilon$. We have

$$(y, Au - \rho) = (y, Au) - C \geq (y, Au) - (y, Aw) - \varepsilon = (y, A(u - w)) - \varepsilon \geq -\varepsilon$$
because \( A(u - w) = (Au - \rho) + (\rho - Aw) \geq 0 \), and \( y \) is non-negative on \( A(W) \). Since \( \varepsilon > 0 \) was arbitrary, \( (y, Au - \rho) \geq 0 \).

The space \( X \) is the span of the space \( A(W) \) and the cone \( K \), so Riesz’s theorem on extending a non-negative functional allows us to extend \( y \) to a non-negative functional on \( X \). \( \square \)

In our situation \( X = VC^1(\Omega_1 \times \Omega_2), Y \supset QB^\infty(\Omega_1 \times \Omega_2) \), where \( (\Omega_i, \mu_i) \) is a standard probability space for \( i = 1, 2 \), \( W = L^1(\Omega_1) \oplus L^1(\Omega_2), Z = L^\infty(\Omega_1) \oplus L^\infty(\Omega_2) \), the operator \( A \) maps a pair of functions \( u = (w_1(t), w_2(t)) \in W \) into the function \( w_1(t_1) + w_2(t_2) := (Au)(t_1, t_2) \in X \), and the restriction of \( B \) to \( QB^\infty \) maps the quasi-bistochastic signed measure \( \eta \) into the pair of its projections on \( X \) and \( Y \):

\[
L^1(\Omega_1) \times L^1(\Omega_2) \xrightarrow{A} VC^1(\Omega_1, \Omega_2),
\]

\[
(w_1(x), w_2(y)) \xmapsto{A} w_1(x) + w_2(y);
\]

\[
QB^\infty(\Omega_1, \Omega_2) \xrightarrow{A^*} L^\infty(\Omega_1) \times L^\infty(\Omega_2),
\]

\[
\mu \xmapsto{A^*} (\Pr_{\Omega_1}, \mu, \Pr_{\Omega_2} \mu).
\]

The element \( \rho(t_1, t_2) \) is understood as the price of transportation from a point \( t_1 \in \Omega_1 \) to a point \( t_2 \in \Omega_2 \). We take the element \( z \in Z \) equal to the pair of constant functions \( (1, 1) \) (without loss of generality: for other functions we just change the measures \( \mu_1 \) and \( \mu_2 \) to equivalent measures). Note that by the definition of the space \( VC^1 \) each function in this space can be represented as the sum of a non-negative function and an additively separable function \( w_1(t_1) + w_2(t_2) \). Thus, the hypotheses of Theorem 13 hold. Remark 1 for that theorem (the cone of non-negative functions is clearly closed and generating in \( VC^1 \)) guarantees that this functional is norm bounded, and hence it corresponds to some polymorphism in \( QB^\infty \). The norm \( \|(1, 1)\|_\rho \) is the infimum of the scheme prices for transporting \( \mu_1 \) to \( \mu_2 \) with price function \( \rho \). Thus, Theorem 13 implies the existence of an optimal scheme.

If we replace \( VC^1(\Omega_1 \times \Omega_2) \) by \( L^1(\Omega_1 \times \Omega_2) \), then both the assumptions and the assertion of Theorem 13 fail. In this case non-negative bounded functionals correspond to non-negative bounded functions (not just to polymorphisms), and an optimal scheme need not exist.

3.2. Sobolev spaces and trace theorems.

**Theorem 14.** Let \( \Omega_1 \) and \( \Omega_2 \) be domains of dimensions \( d_1 \) and \( d_2 \), respectively, and suppose that \( pl > d_2 \) or \( p = 1 \) and \( l = d_2 \). Then the functions in the Sobolev space \( W^l_p(\Omega_1 \times \Omega_2) \) (whose \( l \)th generalized derivatives are integrable to the power \( p \)) are virtually continuous as functions of the two variables \( x \in \Omega_1 \) and \( y \in \Omega_2 \). The embedding of \( W^l_p(\Omega_1 \times \Omega_2) \) into \( VC^1(\Omega_1, K) \) is continuous for any compact subset \( K \) of the domain \( \Omega_2 \).

**Proof.** Using the theorem on embedding Sobolev spaces into the space of continuous functions (see, for instance, [1], [9]), we have the following estimate for functions \( h(y) \in W^l_p(\Omega_2) \):

\[
\|h\|_{C(K)} \leq c(\Omega_2, K)\|h\|_{W^l_p(\Omega_2)}.
\]
Let \( f(x, y) \in W^l_p(\Omega_1 \times \Omega_2) \) be a smooth function, and let
\[
a(x) := \|f(x, \cdot)\|_{W^l_p(\Omega_2)}.
\]
Then by Fubini’s theorem \( a \in L^1(\Omega_1) \) and
\[
\int |a| \leq c(\Omega_1, \Omega_2)\|f\|_{W^l_p(\Omega_1 \times \Omega_2)}.
\]
The following estimate holds on \( \Omega_1 \times K \):
\[
|f(x, y)| \leq \|f(x, \cdot)\|_{C(K)} \leq c(\Omega_2, K)a(x).
\]
Summarizing, we have
\[
\|f\|_{VC^1(\Omega_1, K)} \leq c(\Omega_1, \Omega_2, K)\|f\|_{W^l_p(\Omega_1 \times \Omega_2)}.
\]  \hfill (9)
Each function in the class \( W^l_p(\Omega_1 \times \Omega_2) \) is the limit of a sequence of smooth functions, and by (9) it is a limit in \( VC^1 \) as well. \( \Box \)

Therefore, under the conditions of this embedding theorem we can integrate functions with respect to any quasi-bistochastic measures. This generalizes the usual theorems about traces on submanifolds.

3.3. Nuclear operators acting in Hilbert space. It is well known that the space of nuclear (trace-class) operators acting in the Hilbert space \( L^2 \) is a projective tensor product of Hilbert spaces. Their kernels are measurable functions of two variables, which are difficult to describe directly. The following theorem asserts that the kernels of nuclear integral operators are virtually continuous as functions of two variables. We remark that the kernels of Hilbert–Schmidt operators are not virtually continuous in general.

**Theorem 15.** Let \( (X, \mu) \) and \( (Y, \nu) \) be standard probability spaces. The space of kernels of nuclear operators from \( L^2(X) \) to \( L^2(Y) \) (with the Schatten–von Neumann norm) embeds continuously in \( VC^1 \).

**Proof.** Let \( K(x, y) \) be the kernel of a finite-rank integral operator from \( L^2(X) \) to \( L^2(Y) \) with nuclear norm 1. Then there is a finite-sum representation
\[
K(x, y) = \sum s_k a_k(x) b_k(y),
\]
where the \( s_k \) are singular values of the operator, \( (a_k) \) and \( (b_k) \) are orthonormal systems, and \( \sum |s_k| \leq 1 \). Almost everywhere we have
\[
|K(x, y)| \leq \frac{1}{2} \sum |s_k| |a_k^2(x)| + \frac{1}{2} \sum |s_k| |b_k^2(y)|.
\]
The right-hand side has the form \( A(x) + B(y) \), and
\[
\int |A(x)| \, dx + \int |B(y)| \, dy \leq 1.
\]
Thus, the norm of \( K(x, y) \) in the space \( VC^1 \) does not exceed 1. It remains to note that any nuclear operator can be approximated in the nuclear norm by operators of finite rank, and by the above estimate this is an approximation in \( VC^1 \) as well. \( \Box \)
This implies that such kernels can be integrated not only over the diagonal when \( X = Y \), which is well known, but also with respect to bistochastic measures. But the space \( VC^1 \) is broader than just the kernels of nuclear operators. If we view \( VC^1 \) as the space of kernels of integral operators acting in suitable \( L^p \) spaces, then it is not invariant with respect to conjugation by nuclear operators, in contrast to the Schatten–von Neumann spaces and, in particular, to spaces of nuclear operators. Indeed, the definition of \( VC^1 \) essentially uses known \( \sigma \)-subalgebras which do not have the necessary invariance. A closely related question was considered in [2]. See more on the traces of nuclear operators and virtual continuity in [23].

3.4. Restrictions of metrics. The following problem was one of the origins of this paper. Let \( (X, \mu) \) be a standard measure space with a continuous measure. Assume that \( \rho \) is an admissible metric and \( \xi \) is a measurable partition of \( (X, \mu) \) with parts of null measure (say, \( \xi \) is the partition into the level sets of a function which is not constant on sets of positive measure). Can we restrict our metric (as a function of two variables) correctly to elements of this partition?

This is not immediately clear, since the metric is a priori just a measurable function. But any admissible metric is virtually continuous, and so for our goal it suffices to define a bistochastic measure to whose support we have to restrict the metric. Suppose for simplicity that \( X = [0, 1]^2 \), \( \mu \) is Lebesgue measure, and \( \xi \) is the partition into vertical line segments. Then we talk about the restriction of a virtually continuous function defined on \( X^2 = [0, 1]^4 \) to the three-dimensional submanifold \( \{ (x_1, x_2, x_3, x_4): x_1 = x_3 \} \). It is easy to see that such a submanifold equipped with three-dimensional Lebesgue measure defines a bistochastic measure on \( X \times X \).

We are grateful to L. Lovász, who sent us his recent monograph [7], in which related questions are discussed, and to A. Logunov for drawing our attention to the possibility of a dual definition of thickness.

Bibliography

[1] R. A. Adams and J. J. F. Fournier, *Sobolev spaces*, 2nd ed., Pure Appl. Math. (Amst.), vol. 140, Elsevier/Academic press, Amsterdam 2003, xiv+305 pp.

[2] M. Denker and M. Gordin, “Limit theorems for von Mises statistics of a measure preserving transformation”, *Probab. Theory Related Fields* 160:1-2 (2014), 1–45; 2013, 47 pp., arXiv:1109.0635v2.

[3] M. Gromov, *Metric structure for Riemannian and non-Riemannian spaces*, Progr. Math., vol. 152, Birkhäuser Boston, Inc., Boston, MA 1999, xx+585 pp.

[4] Л.В. Канторович, “О перемещении масс”, Докл. АН СССР 37:7-8 (1942), 227–229; перепеч. в сб.:, Теория представлений, динамические системы. XI, Специальный выпуск, Зап. науч. сем. ПОМИ, 312, ПОМИ, СПб. 2004, с. 11–14; English transl., L. V. Kantorovich, “On the translocation of masses”, *J. Math. Sci. (N. Y.)* 133:4 (2006), 1381–1382.

[5] Л.В. Канторович, Г. Ш. Рубинштейн, “Пространство вполне аддитивных функций”, Вестн. Ленград. ун-та 13:7 (1958), 52–59. [L. V. Kantorovich and G. Sh. Rubinstein, “The space of completely additive functions”, *Vestnik Leningrad. Univ.* 13:7 (1958), 52–59.]

[6] L. V. Kantorovich and G. Sh. Rubinstein, “The space of completely additive functions”, *Vestnik Leningrad. Univ.* 13:7 (1958), 52–59.
A. M. Vershik, P. B. Zatitskiy, and F. V. Petrov

[6] J. Komlós, “A generalization of a problem of Steinhaus”, *Acta Math. Acad. Sci. Hungar.* **18**:1-2 (1967), 217–229.

[7] L. Lovász, *Large networks and graph limits*, Amer. Math. Soc. Colloq. Publ., vol. 60, Amer. Math. Soc., Providence, RI 2012, xiv+475 pp.

[8] L. Lovász, *Coupling measure concentrated on a given set*, 2 pp., http://www.cs.elte.hu/~lovasz/book/homnotes-A-3-4b.pdf.

[9] В. Г. Мазья, *Пространства Соболева*, Изд-во ЛГУ, Л. 1985, 416 с.; English transl., V.G. Maz’ja, *Sobolev spaces*, Springer Ser. Soviet Math., Springer-Verlag, Berlin 1985, xix+486 pp.

[10] А. М. Вершик, П. Б. Затицкий, и Ф. В. Петров, “Geometry and dynamics of admissible metrics in measure spaces”, *Cent. Eur. J. Math.* **11**:3 (2013), 379–400.

[11] В. А. Рохлин, “Об основных понятиях теории меры”, *Matem. сб.* **25**(67):1 (1949), 107–150. [V. A. Rohlin, “Basic notions of measure theory”, *Mat. Sb.* **25**(67):1 (1949), 107–150.]

[12] K.-T. Sturm, *The space of spaces: curvature bounds and gradient flows on the space of metric measure space*, 2012, 75 pp., arXiv:1208.0434.

[13] А. М. Вершик, “Универсальное пространство Урысона, метрические тройки Громова и случайные метрики на натуральном ряде”, *УМН* **53**(5)(323) (1998), 57–64; English transl., A. M. Vershik, “The universal Urysohn space, Gromov metric triples and random metrics on the natural numbers”, *Russian Math. Surveys* **53**:5 (1998), 921–928.

[14] А. М. Вершик, “Классификация измеримых функций нескольких аргументов и инвариантно распределенные случайные матрицы”, *Функц. анализ и его прил.* **36**:2 (2002), 12–27; English transl., А. М. Вершик, “Classification of measurable functions of several variables and invariantly distributed random matrices”, *Funct. Anal. Appl.* **36**:2 (2002), 93–105.

[15] А. М. Вершик, “Random and universal metric spaces”, *Dynamics and randomness* II, Nonlinear Phenom. Complex Systems, vol. 10, Kluwer Acad. Publ., Dordrecht 2004, pp. 199–228.

[16] А. М. Вершик, “Случайные метрические пространства и универсальность”, *УМН* **59**:2(356) (2004), 65–104; English transl., А. М. Вершик, “Random metric spaces and universality”, *Russian Math. Surveys* **59**:2 (2004), 259–295.

[17] А. М. Вершик, “Полиморфизмы, Марков процессы, и квазисопоставимость”, *Discrete Contin. Dyn. Syst.* **13**:5 (2005), 1305–1324.

[18] А. М. Вершик, “Как выглядит типичный марковский оператор?”, *Алгебра и анализ* **17**:5 (2005), 91–104; English transl., А. М. Вершик, “What does a generic Markov operator look like?”, *St. Petersburg Math. J.* **17**:5 (2006), 763–772.

[19] А. М. Вершик, “Dynamics of metrics in measure spaces and their asymptotic invariants”, *Markov Process. Related Fields* **16**:1 (2010), 169–185.

[20] А. М. Вершик, “О классификации измеримых функций нескольких переменных”, *Теория представлений, динамические системы, комбинаторные методы*. XXI, Зап. науч. сем. ПОМИ, **403**, ПОМИ, СПб. 2012, с. 35–57; English transl., А. М. Вершик, “On classification of measurable functions of several variables”, *J. Math. Sci. (N. Y.)* **190**:3 (2013), 427–437.

[21] А. М. Вершик, “Long history of Monge–Kantorovich transportation problem”, *Math. Intelligencer* **35**:4 (2013), 1–9.

[22] А. М. Вершик, П. Б. Затицкий, Ф. В. Петров, “Виртуальная непрерывность измеримых функций многих переменных и теоремы вложения”, *Функц. анализ и его прил.* **47**:3 (2013), 1–11; English transl., А. М. Вершик, П. Б. Затицкий, и F. V. Petrov, “Virtual continuity of measurable functions of several variables and embedding theorems”, *Funct. Anal. Appl.* **47**:3 (2013), 165–173.
Virtual continuity 1063

[23] A. M. Vershik, P. B. Zatitskiy, F. V. Petrov, “Integrating virtually continuous functions with respect to bistochastic measures and a trace formula for nuclear operators”, *Algebra i Analiz* (to appear).

[24] П. Б. Затицкий, Ф. В. Петров, “Об исправлении метрик”, *Теория представлений, динамические системы, комбинаторные методы. XX, Зап. науч. сем. ПОМИ, 390*, ПОМИ, СПб. 2011, c. 201–209; English transl., P. B. Zatitskiy and F. V. Petrov, “Correction of metrics”, *J. Math. Sci. (N. Y.)* 181:6 (2012), 867–870.

Anatolii M. Vershik
St. Petersburg Department of Russian Academy of Sciences; St. Petersburg State University; Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute)
E-mail: avershik@gmail.com

Pavel B. Zatitskiy
Chebyshev Laboratory at St. Petersburg State University; St. Petersburg Department of Steklov Mathematical Institute of Russian Academy of Sciences
E-mail: paxa239@yandex.ru

Fedor V. Petrov
St. Petersburg Department of Steklov Mathematical Institute of Russian Academy of Sciences; St. Petersburg State University
E-mail: fedyapetrov@gmail.com