Long History of the Monge-Kantorovich Transportation Problem

A. M. Vershik

The Mathematical Intelligencer

ISSN 0343-6993

Math Intelligencer
DOI 10.1007/s00283-013-9380-x

ONLINE FIRST

Springer

Long History of the Monge-Kantorovich Transportation Problem

(Marking the centennial of L.V. Kantorovich's birth!)

A. M. Vershik

Leonid Vital'evich Kantorovich (1912–1986) was one of the great mathematicians and economists of the twentieth century.

In 2012, the centenary of his birth was marked in St. Petersburg. Short histories were presented describing some of the main parts of his legacy, which continue in importance today: duality in linear programming, the so-called "Monge-Kantorovich transportation problem," and the "Kantorovich metric." Note that 2012 was also the 70th anniversary of the publication of his historic paper on the transport metric. The present article offers a somewhat expanded version of my talk on that occasion.

L. V. Kantorovich the Person

We remember Leonid Vital'evich Kantorovich for his massive contributions to foundations of mathematics, computational mathematics, and other areas, and in particular as one of the founders of mathematical economics.

He began as a child prodigy, entering Leningrad University at the age of 14. His first paper on descriptive set theory, which caught the attention of the mathematical community and in particular N. N. Lusin and A. N. Kolmogorov, appeared when he was 18. He went on to work in theory of functions, functional analysis, numerical methods, computer science, and the main contribution: linear programming and then mathematical economics, for which he was awarded the Nobel Prize in 1975. See the recent biographical article [8].

In many respects his activity in mathematics and applications reminds one of the activity of another giant of the twentieth century, John von Neumann; both were leaders in functional analysis, mathematical economics, numerical methods, and computer science, and both played important roles in the atomic projects of their countries. But their reception was quite different. The ideas of Kantorovich, and Kantorovich himself, were not appreciated in Soviet Russia. For a long time—until the end of the 1950s—his ideas on mathematical economics were considered in official circles as anti-Marxist; consequently it was prohibited and even dangerous to study and develop them. Such a story, carried much farther, is well-known in Soviet biology ("lysenkovshchina").

So, between 1947 and the late 1950s Kantorovich never mentioned his ideas about mathematical economics and linear programming. My colleagues and I heard his lectures on functional analysis (later published as a book with G. P. Akilov) but knew nothing of his economically related work. He lectured openly on the subject only after the beginning of Khrushchev's "ottepel'" ("thaw"), the liberalization of 1957–1958.

L. V. Kantorovich (LV hereinafter) has left a rich record of the defence of scientific truth. A majority of the Soviet economists of the generation of the 1960s and 1970s were pupils of LV, and many mathematicians (including the author) considered themselves his pupils.

This article will concentrate on one page of the brilliant legacy of LV in mathematics and its applications: the transport metric or Kantorovich metric. I single out this circle of ideas from his extensive activity for several reasons. First of all, because by now this area seems to be the most discussed of the discoveries of LV. Second, it is an example of deep and nontrivial relationship between fundamental science and applications, and this relationship was characteristic of the creativity of LV. And furthermore, this discovery could have been done only by someone who was at that time (late 1930s and 1940s) a leader in one of the central areas of mathematics at that time, functional analysis, and simultaneously in the area of applied and computational problems.

DOI 10.1007/s00283-013-9380-x

Published online: 30 May 2013

Figure 1. L. V. Kantorovich in his youth.

Linear programming—that is, the theory of linear extremal problems with linear constraints—was established in 1938 by LV as a response to specific practical needs (the famous "problem of Plywood trust"). Later this theory was rediscovered in the United States by George Dantzig and others. LV presented linear programming in the booklet *Mathematical Methods of Organizing and Planning of Production* [1] (1939 edition from LSU Leningrad; this booklet was reprinted several times and translated into other languages). See Fig. 2. He showed the scope of the theory by a long list of specific situations in which such a formulation applies. He mentions briefly also the transport problem, to which he returned later.

It would not be an exaggeration to compare the role of this booklet in the further development of mathematical economics—and, more generally, of the mathematical analysis of extremal problems—with the role of *The Mathematical Principles of Natural Philosophy* by Isaac Newton in mechanics. LV anticipated by 10 years not only almost all avenues of practical application of linear programming, but the main ideas of solution methods: the idea of duality, current in functional analysis of that time, and a numerical method that can be regarded as a variant of the simplex method proposed by Dantzig in the late 1940s.

The works of LV and his collaborators and students in this field did not become known in the West until the end of the 1950s, and his priority was then recognized, although not always reflected in terminology.

The idea of duality developed subsequently into a new economic theory, proposed by LV, the theory of objectively determined valuations. (In the original brochure the term was "resolving multipliers," in analogy to Lagrange multipliers.) Much more could and should be said about the dramatic later destiny of this theory. Let me just say here that the naturalness and beauty of this theory was appreciated in the USSR by mathematicians and a few brave economists, and in the West, a little later, by many scientists. In the USSR, orthodox economists, dogmatic and ill equipped to understand it, remained hostile. Their condemnation extended to LV's 1942 book about economics with limited resources.

The spokesmen of orthodoxy had the power. For them to characterize Kantorovich as a revisionist, undermining the Marxist labor theory of value, was a serious threat. The consequences for LV could have been dire indeed if his work on military research had not been so highly regarded. To give an idea of the level of the critique of LV's writings, I quote just one of his opponents (not the worst or the most obtuse): "As 'mathematical physics' does not have a subject of its own distinct from physics in general, so 'mathematical economics' does not have any subject matter distinct from political economics, and political economics necessarily includes its quantitative side." LV protested that his theory does not contradict Marx's, that it seeks only to optimize use of limited resources, and that in that context it is just as objective; but for a long time his arguments went unheard.

Clearly what was at stake was not only the purity of out-of-date dogmas, but the conservation of the influence of an élite claiming monopoly on the truth. From 70 years' hindsight, setting aside reservations no longer needed, we can say that the theory of LV significantly extends, and in some respects supersedes, Marx's theory of value. There is no need to argue this here. Even at the time there were in the West several competent articles about this, see for one [17].

Birth of the Transport Metric

In the 1939 booklet and subsequently, LV singled out the transportation problem from other problems of linear programming. Soon after, he began writing together with his

AUTHOR

A. M. VERSHIK is Head of the Laboratory of Representation Theory at the St. Petersburg Department of the Steklov Institute of Mathematics, and is also a Professor at St. Petersburg State University. He says he enjoys "everything in mathematics involving motion–dynamics, asymptotics, perturbation, randomness, representations,… but this is almost all of mathematics."

Steklov Institute of Mathematics
27 Fontanka
St. Petersburg 191023
Russia
e-mail: vershik@pdmi.ras.ru

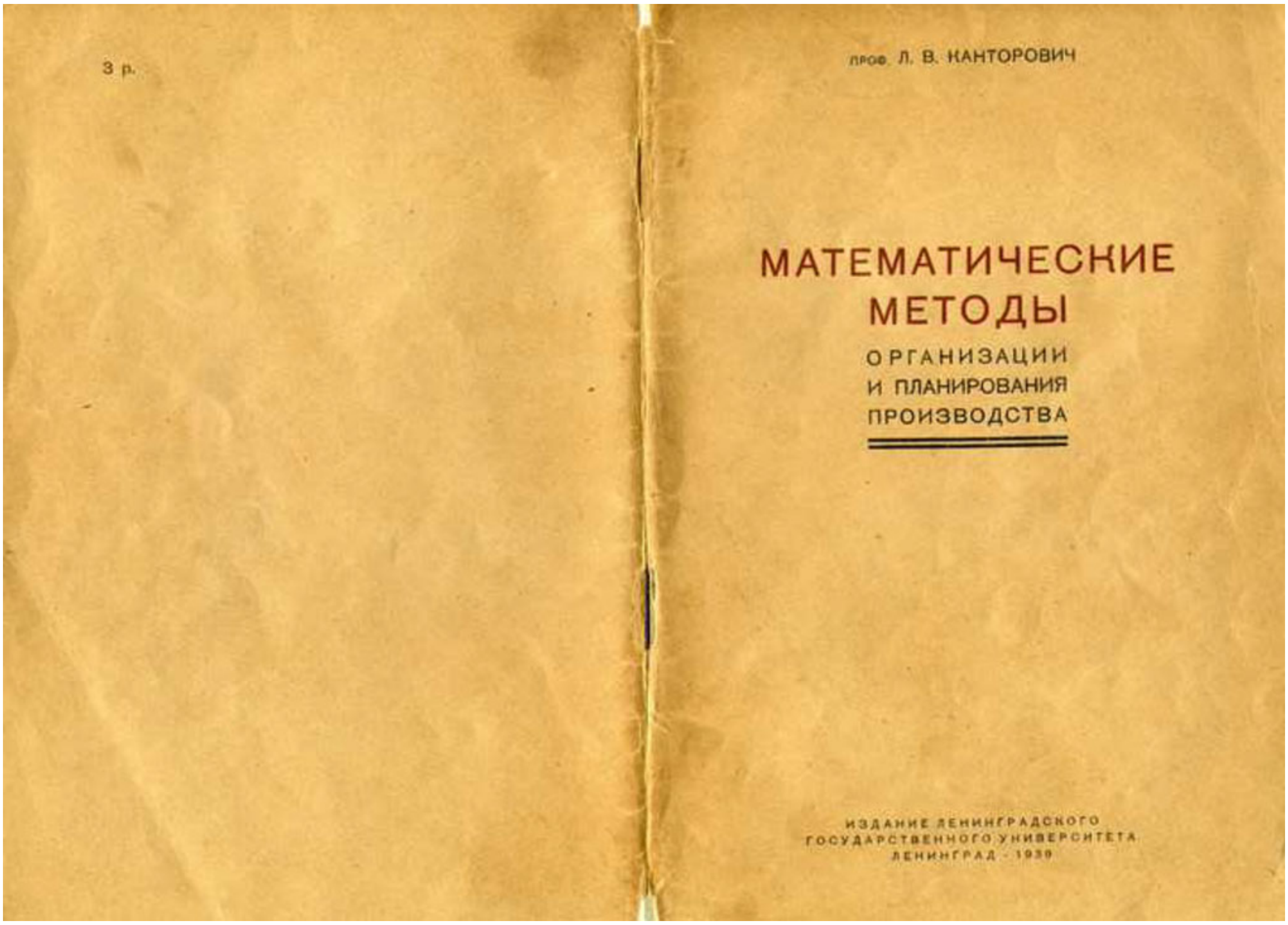

Figure 2. The first presentation of linear programming, 1939.

disciple M. K. Gavurin on a special method for solving a linear transport problem—the potential method [4]. It is an implementation of the general method of duality, and the visual interpretation leads immediately to the analogy with the theory of fluid dynamics and flows in networks, which was later much developed. The article with Gavurin was addressed to transport engineers and planners, but it was rejected by several serious journals in the field and remained unpublished for almost 10 years. Not waiting for its publication, LV wrote his "On mass transfer" [3]. I want to say about this work the same that I said about the booklet. This is a classic in all respects: it contains a profound idea that goes beyond those examples studied previously, it is brief and self-contained, one feels that there is nothing more to be said, just as there is nothing to be added to the second law of Newton, and finally, it includes a program of future research, one that was followed at first very slowly but proceeds especially quickly today.

Before I go into the content, let me say a few words about the intellectual mystery story of how this work, followed by a number of other works by LV and his school, became known around the world. This tale is told in more detail in the introductory article to the new edition of the booklet by the son of LV, V. L. Kantorovich [2].

It is clear from the year of publication that in wartime the *Doklady* note could not be immediately known either in the Soviet Union or in the West. (*Doklady* and other journals were not translated into English then, and there were only a few in the library.) The following circumstance, important in itself, helped to speed up the process.

In 1946, the world marked the bicentenary of the birth of the French mathematician and physicist Gaspard Monge (1746–1818), belonging to the brilliant constellation of founders of modern mathematics. He was involved not only in mathematics, but also its many applications; he was the founder and head of the famous *École Polytechnique*, and he was responsible for many inventions. In particular, he formulated in 1781 a mathematical problem of "Excavation and embankments" (*Les déblais et les remblais*)—how to transport the soil during the construction of the building of forts and roads with minimal transport expenses. His hypothesis of normality of optimal traffic paths to some surfaces was later proved in a 200-page work by Appel (1884).

In Leningrad in 1947 the Commission on History of Mathematical and Physical Sciences of the USSR Academy, headed by Academician V. I. Smirnov, held a public session dedicated to G. Monge, where the well-known Moscow geometer B. N. Delaunay spoke on the geometric works of Monge and drew attention in particular to this problem. The proceedings of the session were published in 1947 [5]. LV, apparently on seeing this book, drew attention to this problem of Monge and saw how its solution was connected with his work. He gave a talk to the Moscow Mathematical Society (22 December 1947), details of

Figure 3. Street signs honoring Gaspard Monge in Paris, and honoring L. V . Kantorovich in Rishon Le Zion, Israel. (Thanks to G. Thouvenot and J. Romanovsky for the photographs.) There is as yet no Kantorovich Street in Russia.

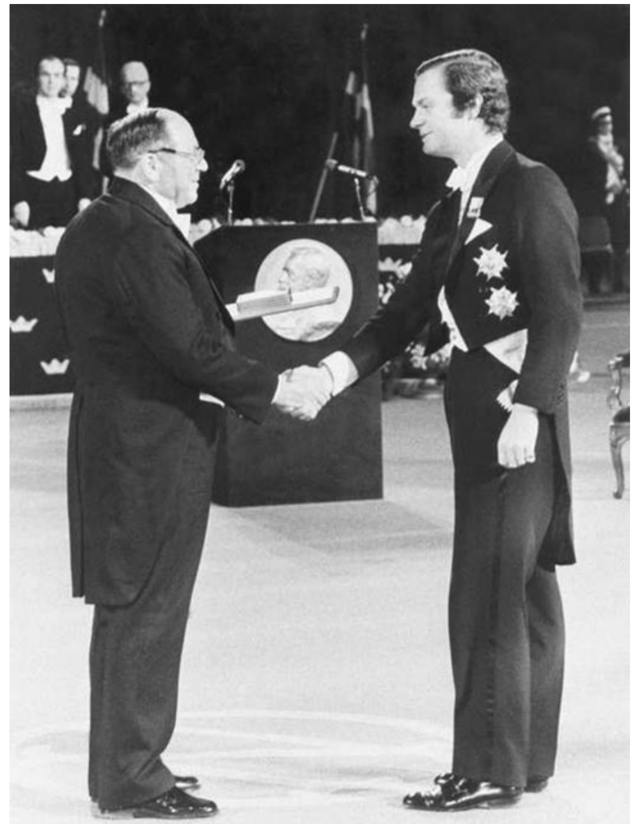

Figure 4. Presentation of the Nobel Prize in Economics to L. V. Kantorovich, 1975.

which were published in *Uspekhi* [6]. In this note it was asserted (with no details) that the surfaces Monge spoke of were just the level surface of the potentials defined in the *Doklady* note.

This very short article caught the eye of several American mathematicians and economists (*Uspekhi* also was not translated at the time, but was more widely known in the West than *Doklady*), and they began to look up the publications of LV. One of them was Tjalling C. Koopmans, who had independently studied the transportation problem in classified wartime applied work, in the finite-dimensional case, and who, for his 1949 *Econometrica* paper, was awarded the 1975 Nobel Prize together with LV.

About the end of the 1950s, LV's main work on the subject (not only economics) became known in the West, more so than in the USSR. Since that time, some scholars in the West speak of "the Monge–Kantorovich transport problem," which seems to me pretty fair terminology.

I turn now in particular to the article of 1942. Its main content as compared with the previous cycle of finite problems of linear programming is in a natural generalization of the situation—here the transportation is a probabilistic measure on a compact metric space X with the metric r, namely a measure Φ_1 is the initial one that needs to be transported, and the second measure Φ_2 is the final one, that is, the desired distribution after the transport.

A transportation, carriage, or transport plan as it is called by LV is also a probability measure Ψ on the Cartesian product of the compact space with itself $X \times X$, whose projection Pr_1 on the first factor and Pr_2 on the second (marginal distributions) are the given measures Φ_1 and Φ_2. The transportation price for a given plan is an integral of the metric as a function of two variables, and the task is to minimize this value over all admissible transport plans: in obvious notation it takes the form

$$\inf\left\{\int_X\int_X r(x,y)d\Psi(dx,dy); \Psi : \mathrm{Pr}_1\Psi = \Phi_1, \;\; \mathrm{Pr}_2\Psi = \Phi_2\right\}$$
$$\equiv K_r(\Phi_1, \Phi_2).$$

If the space X is finite, then the measures Φ_1, Φ_2 are probability vectors, and the measure Ψ is a matrix (a generalized doubly stochastic matrix). Note that in this case, r is the matrix of the cost of transporting a unit of cargo from one point to another, but it is not necessarily a metric.

However—and here the naturalness of LV's formulation of the problem for continuous spaces appears—we should assume the function r to be a metric; in the particular Monge case it is the Euclidean metric. This characterizes the fruitful generalizations of finite problems to the continuous case. For finite problems, it is not so important that the objective function be a metric.

It is easy to see that feasible plans always exist, and the problem is well posed. The indicated infimum K_r satisfies the triangle inequality as a function of the measures Φ_1, Φ_2, and we obtain *a metric on the space of probability measures on the compact space.*

Just this K_r as a function of two probability measures on the compact (X, r) should be called the **Kantorovich metric** on the product space.

Many mathematicians, not knowing of LV's work, used such a metric in particular cases (see details in [15]). However, the point of LV's treatment is not the definition, but the main theorem, which the subsequent rediscoverers did not prove, or even formulate.

Theorem 1

$$K_r(\Phi_1, \Phi_2) = \sup\left\{ \int_X u(x)(\Phi_2 - \Phi_1)(dx) : u(x) - u(y) \leqslant r(x,y) \right\}$$

Here the supremum is taken over all Lipschitz functions with constant 1 relative to the given metric; LV calls these Lipschitz functions *potentials* in the transportation problem. The transport plan Ψ is called a *potential plan* if for almost all point pairs x, y in the sense of the measure Ψ, and for a certain Lipschitz function $U(\cdot)$ (the optimal potential), we have

$$U(x) - U(y) = r(x,y);$$

in other words, the Lipschitz inequality $U(x) - U(y) \leqslant r(x, y)$ reduces to equality on a set of full Ψ measure.

Corollary 1 *A transport plan Ψ is optimal if and only if it is potential.*

The corollary follows immediately from the theorem. The optimal potential provides the maximum of the right-hand integral. The theorem is the duality theorem of linear programming, which states that the value of the inf in the original problem is the same as the sup in the second, dual, problem. All these statements imply the existence of an optimal transport plan (for a compact space).

At that time there was no duality theory for infinite-dimensional programming problems; it appeared only later. So in the article LV proved the theorem by a direct approximation. Note that the sufficiency (the inequality inf $\geqslant$ sup) is trivially true, and the approximation is needed to prove the necessity; that is, the nontrivial part is that sup $\geqslant$ inf. This example became a model for the development of the general theory of infinite-dimensional duality of extremal problems, which nowadays includes a large number of classical and nonclassical problems, and it also became the basis of computational methods for finite problems. Curiously, the inventor of the simplex method, G. Dantzig, wrote in his memoirs that he understood the connection of his method with the theory of duality only after a conversation with von Neumann, to whom he had come to show his invention, and who immediately explained all the implications and connections.

Those connections are extensive. The duality theorem of finite-dimensional linear programming is, in a different formulation, based on the fundamental theorem (von Neumann) of the theory of matrix games, and it is also Weyl's duality theorem in the theory of convex polyhedra, and a theorem about the solution of systems of linear inequalities, etc. In hindsight we see these equivalences fall out in a single line, and this, with all of those connections, based on the theory of duality, is how I organized the course "Extremal problems," which I taught for 20 years at Leningrad State University (1973–1993).

LV was the first to give a nontrivial example of an infinite-dimensional duality.

Further Development

The Kantorovich–Rubinstein Norm

In his article LV emphasizes the utility of studying the metric introduced in the space of probability measures on a compact space. It is obvious that the Kantorovich metric is functorial (that is, it is preserved under isometries of the compact space), but so are a lot of other metrics on the space of probability measures. To see what distinguishes it among all such metrics, we turn to the important work of LV and his student G. Sh. Rubinstein [7]. This work complements the main result of the note of LV, and its result follows directly from the fundamental theorem there.

Theorem 2 *Consider the linear space $V_0(X)$ of alternating measures of bounded variation with zero charge on the compact space (X, r).*

Every nonzero measure of this type is presented uniquely as the difference of two nonnegative and mutually disjoint finite measures: $\mu = \mu_+ - \mu_-$ in the space of all b.v. signed measures $V(X)$; it is obvious that $\mu_+ = \lambda\mu'_+$; $\mu_- = \lambda\mu'_-$ with the same positive factor λ, where μ'_+, μ'_- are probability measures.

Define for any such μ

$$|\mu| \equiv \lambda K_r(\mu'_+, \mu'_-).$$

Then $|\cdot|$ is a well-defined norm in $V_0(X)$, under which it is a separable normed space, which has as its dual the space of all Lipschitz functions modulo the constants.

Thus, the Kantorovich metric extends to a norm on the signed measures of bounded variation. This norm on the space of measures is called the Kantorovich-Rubinstein norm

$$|\mu|_{KR}.$$

Curiously, before this paper [7] it was not known whether the space of Lipschitz functions is conjugate to a Banach space; convergence of bounded sequences in this norm is weak convergence of measures. The space $(V_0(X), |\cdot|)$ is in general incomplete, and the completion includes rather complicated objects that are not measures.

In my student years I had the good fortune to be on friendly terms with G. Sh. Rubinstein, and I was perhaps the first whom GSh told about this norm (1957). He modestly noted that LV liked this "Kantorovich norm," in other words he recognized that it had already been implicit in the article by LV alone; I agree with this.

Behind Rubinstein's remark is the fact that *the Kantorovich metric is translation-invariant*. From this one can readily infer the existence of the extension. Indeed, by homogeneity any metric on the simplex of probability measures extends to the cone generated by the simplex; but each signed measure of finite variation can be represented as a difference of two nonnegative measures, thus if we have a norm on the cone we have norm on the linear space of measures. Finally, a metric on the cone is a norm iff it is invariant under translations. The last property of the Kantorovich metric trivially follows from the linearity of the classical transport problem.

But this remark was very important because if we have a norm we can apply the techniques of the theory of Banach spaces, which they did. This property is not enjoyed by the wider class of new metrics; see below.

Characterization of the KR-Norm

We can now formulate an axiomatic definition of the Kantorovich metric and the norms of Kantorovich-Rubinstein. In this case, we drop the requirement of compactness of the space. Indeed, the definition of the Kantorovich metric and the norm of Kantorovich-Rubinstein makes sense for noncompact spaces, but not for all measures, only for measures μ with a finite first moment: $\int r(x,y)d\mu(x)d\mu(y) < \infty$. This includes in particular the class of discrete measures with finite support.

Consider, then, a separable metric space X with metric r; the simplex $S(X)$ of all probability Borel measures with finite first moment on X; and the vector space $V_0(X)$ over $\mathbb{R}$ consisting of all finite formal real linear combinations of points of X with coefficients summing to zero. One can interpret the space $V_0(X)$ as the set of discrete signed measures on X with zero charge. By an elementary charge I will mean a difference of two delta measures: $\varepsilon_{x,y} = \delta_x - \delta_y \in V_0(X)$; the set of all elementary charges generates the space $V_0(X)$ as a linear space.

Definition 1 *A seminorm $|\cdot|$ on $V_0(X)$ (resp. a metric R on the simplex of Borel probability measures $S(X)$) is called compatible with the metric r, or admissible, if $|\varepsilon_{x,y}| = r(x,y)$ (resp. $R(\delta_x, \delta_y) = r(x, y)$).*

In general, for a given metric space (X, r) there are many norms on $V_0(X)$ compatible with the metric r, and consequently, many isometric embeddings of X into Banach spaces. For example, the well-known isometric embedding $T_{HK} : V_0(X) \to \bar{C}(X)$, the space of bounded continuous functions on X with sup norm, was considered by F. Hausdorff and in detail by K. Kuratowski:

$$T_{HK}(\varepsilon_{x,y}) = F_{x,y}(.), \quad F_{x,y}(z) = r(x,z) - r(y,z), \quad F_{x,y} \in \bar{C}(X),$$

with linear extension to V_0. Denote this norm by $|.|_{HK}$. We may examine the relationship between different embeddings, as was done in [7] and almost simultaneously in [9].

Theorem 3 ([11]) *For any metric space, the Kantorovich–Rubinstein norm in the space $V_0(X)$ is the largest admissible norm; that is, for any admissible seminorm $|\cdot|$ and any element a of the space $V_0(X)$ we have $|a| \leqslant |a|_{KR}$.*

It is surprising that this simple corollary of that same duality theorem is not stated until [11], where it plays an important role. Even for the finite-dimensional case these norms deserve further study from the geometrical point of view. Consider the finite set of n points, with metric function constant. The unit ball in that norm coincides with the convex hull of the roots in the Lie algebra $SL(n)$.

Now it may seem at first that, for example, the norms on $V_0(X)$—$|.|_{KR}$ and $|.|_{HK}$—are in general drastically different, as well as the isometric embeddings of X into $V_0(X)$, $|.|_{KR}$ and into $\bar{C}(X)$. But it turns out that there are metric spaces (the so-called linear rigid spaces) for which there is only one admissible norm on V_0, and thus the Kantorovich–Rubinstein norm coincides with the Hausdorff–Kuratowski norm and any admissible norm—see [11])! This means that up to linear isometry there is only one Banach space in which our metric space can be embedded isometrically as a subset whose linear hull is dense. The main example of such a space is the Urysohn universal metric space (see [11]).

Other Admissible Metrics

Any admissible norm $|.|$ on the space $V_0(X)$, restricted to the simplex $S(X)$ of Borel probability measures, gives a metric: $|\mu_1 - \mu_2| \equiv R_{|.|}(\mu_1, \mu_2)$, where $\mu_1, \mu_2 \in S(X)$. This metric $R_{|.|}$ on $S(X)$ evidently has the following property, which is invariant of the metric under translation:

$$R_{|.|}(\lambda\mu_1 + (1-\lambda)\nu, \lambda\mu_2 + (1-\lambda)\nu) = \lambda R_{|.|}(\mu_1, \mu_2); \quad \lambda \in (0,1).$$

The Kantorovich metric of course satisfies this condition. Because translation-invariance is a sufficient condition for the metric to be given by a norm, the opposite assertion is also true:

Theorem 4 *The admissible metric R on the simplex $S(X)$ can be extended to an admissible norm on the space $V_0(X)$—that is, $R(\mu_1, \mu_2) = |\mu_1 - \mu_2|$, when $\mu_1, \mu_2 \in S(X)$—iff it is invariant under translation.*

Thus the previous theorem can be restated as follows.

The Kantorovich metric is the largest admissible metric on the simplex of probability measures in the class of all translation-invariant admissible metrics.

But there are many admissible metrics which are not translation-invariant and consequently do not extend to an admissible norm. One such metric is the Lévy–Prokhorov metric. Here are some others.

Consider the *Kantorovich power metrics* on the simplex metric space (wrongly called Vasershtein metrics in [14], see below), defined as follows:

$$K_{r,p}(\mu_1, \mu_2) = \inf\left\{\left\{\int_X\int_X r(x,y)^p \Psi(dx,dy)\right\}^{1/p}; \; \Psi : \mathrm{Pr}_1\Psi = \Phi_1 \;\; \mathrm{Pr}_2\Psi = \Phi_2\right\}.$$

When $p = 1$ this is the Kantorovich metric. These power metrics have been considered only in recent years; they are described in detail in the book [14] and in references therein.

It is clear that if $p > 1$, then this norm is not invariant under translations.

Perhaps the analysis of the space of all transport problems or even linear problems must be studied more carefully. Recent studies have shown the particular importance of the Kantorovich quadratic metric ($p = 2$) for a range of applications (see [16]). The main result is that the simplex of probability measures on a compact Riemannian manifold equipped with the quadratic Kantorovich metric is a (possibly infinite) Riemannian manifold.

The interrelation of all admissible metrics, in particular of all Kantorovich power metrics, is more complicated. Power metrics, as L^p-norms of difference of measures, increase monotonically with increasing p, so the Kantorovich metric ($p = 1$) is the smallest in the class. Correspondingly, in the class of all admissible metrics, the Kantorovich metric is a sort of saddle point (see below).

The fact that the power r^p ($p > 1$) of a metric, is, generally speaking, not a metric (the triangle inequality may fail) explains the failure of the extension property. If for a metric r and a number $p > 1$ it happens that r^p is also a metric, then the study of the metric $K_{r,p}$ is reduced to the study of the Kantorovich metric K_r^p for a different metric space, in which extension to the Kantorovich–Rubinstein norm is possible.

It is interesting that in the case $p = \infty$, the power Kantorovich metric coincides with the well-known Lévy-Prokhorov metric–see [10], [19], [20].

Two Classes of General Transport Metrics

Let us look more closely at two kinds of definition of admissible metrics that motivated the Kantorovich duality theorem.

1. Direct posing of the transport problems (nonlinear in general) and the corresponding class of admissible metrics.

For two Borel probability measures μ_1, μ_2 on the metric space (X, r), define the class Ψ_{μ_1,μ_2} of Borel probability measures on the space $X \times X$ (transport plans) whose marginal projections are the given μ_1, μ_2. Choose a norm $N(.)$ on the space of measurable functions with respect to the measure space $(X \times X, \Psi)$ for all $\Psi \in \Psi_{\mu_1,\mu_2}$. The transportation problem is to find

$$\inf_{\psi \in \Psi_{\mu_1,\mu_2}} N(r(.,.)) \equiv R_N(\mu_1, \mu_2).$$

This infimum $R_N(.,.)$ clearly defines an admissible metric on the simplex $S(X)$.

If N is the L^p norm, we have the previous Kantorovich power metrics. In particular for $p = 1$ we obtain the Kantorovich metric, which is minimal in this family of metrics.

The general linear problem of transportation is the problem of minimizing the (double) integral of the metric (or minimum of expectation) such as the L^1-norm over a variety of plans, subject to linear constraints:

$$\inf\left\{\int_X \int_X r(x,y)d\psi(dx,dy) : \psi \in \mathcal{L}\right\},$$

where $\mathcal{L}$ is the set of plans with given marginals and satisfying the given constraints.

2. The dual posing of the transport problems and corresponding class of admissible metrics.

Define *Lipschitz-type admissible metrics* as

$$R^{\mathcal{L}}(\mu_1, \mu_2) = \sup_{u \in \mathcal{L}} \left| \int u(x)d(\mu_1 - \mu_2)\right| \equiv ||\mu_1 - \mu_2||_{\mathcal{L}},$$

where $\mathcal{L}$ is an appropriate class of Lipschitz functions on the metric space (X, r).

Theorem 5 ([11]) *The class of Lipschitz-type admissible metrics on $S(X)$ coincides with the class of translation-invariant metrics, and consequently with the class of metrics admitting extension to admissible norms on $V_0(X)$.*

If $\mathcal{L}$ is the space of all Lipschitz functions, we obtain the Kantorovich metric, which is the largest metric in this class.

It seems likely that the Kantorovich metric is the unique metric that belongs to both classes of defined metrics. This gives it a sort of "saddle point position" in the set of all admissible metrics.

The Monge Problem (Paths of Transportation)

Actually the original Monge problem was not only to find a plan of mass transportation in the sense we have been considering, but to choose a transport routing. This question is more delicate: in contrast to LV's problem, here, generally speaking, there need not be a unique solution for best transportation plan, as is shown by trivial examples already in the finite case.

One of the first accurate statements of the problem, to my knowledge, was in my work [15] (see also [16]): to find a measurable transformation T of a metric space X into itself, taking the initial probability Borel measure μ_1 to the final one μ_2 and minimizing the single integral:

$$\inf\left\{\int_X r(x,Tx)d\mu_1(x) : T\mu_1 = \mu_2.\right\}$$

This problem was attacked by V. Sudakov, who solved it for Euclidean compact sets and Lebesgue measures—see later corrections in [12, 18] and references therein.

Here there is a connection to variational problems on infinite groups, in this case the group of measure-preserving transformations (see [16]). The following variational problem is closer in nature to the original formulation of Monge: to find a one-parameter group (or semigroup) $T_t; t \in \mathbb{R}(\mathbb{R}_+)$ of measurable transformations of the metric space X minimizing the (time average) integral:

$$\inf\left\{\int_0^t \int_X r(x, T_s x)d\mu_1(x)ds : \quad T_s\mu_1 = \mu_s,\ s \in [0,t],\ t \in [0,T]\right\}.$$

Variational problems on infinite groups (in particular Lie groups) have still not been studied enough. However, striking achievements in the problem of optimal transportation, applications to the Monge-Ampère problem in partial-differential equations, the theory of the Ricci flow, and others testify to the flourishing of ideas in this theory, which was first stated in the proper generality 70 years ago.

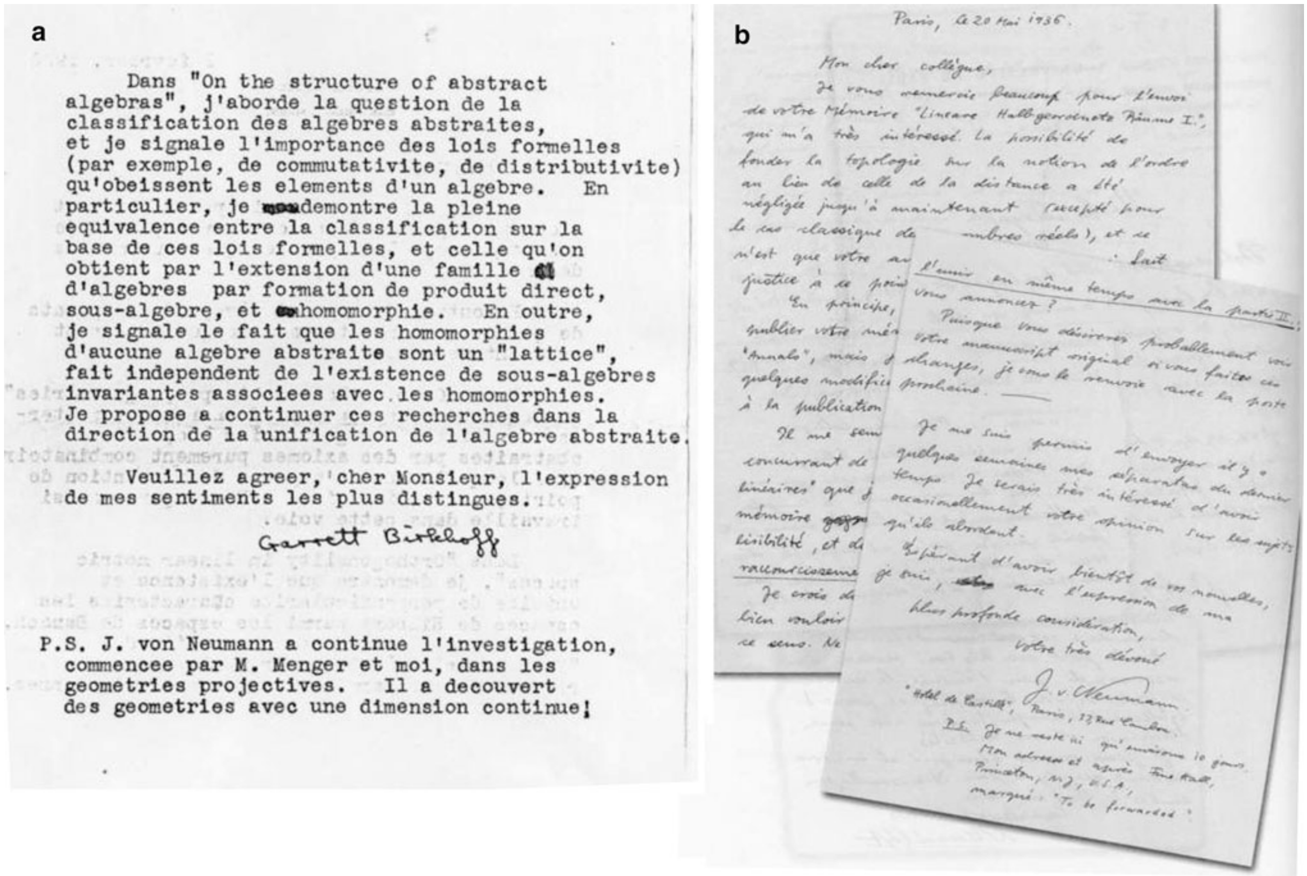

Figure 5. Communication was not completely broken between the East and the West in the mid-30's! Here are letters to Kantorovich from Garrett Birkhoff and John von Neumann.

What Name to Give These Things

Previously I used what I regard as the only correct attribution of concepts and theorems, namely "Monge–Kantorovich problem," "Kantorovich metric," etc. There is no doubt that the Kantorovich metric was defined by L. V. Kantorovich, and his name should be on it. Similarly for the Kantorovich–Rubinstein norm.

Scientific cooperation between the USSR and the Western allies was inadequate (we could almost say non-existent) in those days; even within a country there was poor awareness of developments—all of this meant that the metric was repeatedly rediscovered by other authors, but not in such a fundamental way as in LV's work, and, most importantly, without his main result on duality. In the list of those who have used similar ideas, some stand out for their importance: the $\bar{d}$-metric of D. Ornstein in ergodic theory; the method of coupling in probability, statistics, and statistical physics; the theory of polymorphisms; and much, much more. The importance and popularity of the topic first studied by LV in the late 1930s led to the reappearance of the metric under new names.

It is especially ironic to find the Kantorovich metric called the **Vasershtein metric**. Leonid Vasershtein is a famous mathematician specializing in algebraic K-theory and other areas of algebra and analysis, and my good friend—and he is absolutely not guilty of this distortion of terminology, which occurs primarily in Western literature. It so happened that my colleague and friend R. L. Dobrushin, head of the laboratory where L. N. Vasershtein worked, with understandable enthusiasm spread the word mostly among probabilists and statistical physicists about the "new" metric and its spectacular applications. I spoke to Dobrushin in 1975 and told him that what he called the Vasershtein metric in the report is the Kantorovich metric. After some discussion, he agreed fully and even said so in one of his later works. But it was too late, the wrong name stuck.

Vasershtein's interesting article [13] was very brief (it seems to me that few people who refer to it have looked at it), it does contain in passing a definition of the Kantorovich metric and applies it to the behavior of Markov fields. But there is no definition of power metrics, although in the literature [14] those are also called Vasershtein metrics. Undoubtedly, the work of Vasershtein is worthy of mention in this connection, but I think we should restore the correct terminology out of respect for L. V. Kantorovich, to the teachers and pioneers in our science.

ACKNOWLEDGMENTS

Supported by the grants RFBR 11-01-12092-ofi-m and RFBR 11-01-00677-a. The author expresses his gratitude to Joseph Romanovsky for his help in translation of the article and to Vl. Kantorovich for photos from the family archive.

REFERENCES

[1] Kantorovich, L. V., 1939: *Mathematical methods in the organization and planning of production*. Leningrad Univ., 1939. [English translation: *Management Science*, 6, 4 (1960), 363–422.]

[2] Kantorovich, L. V., 2012: *Mathematical methods in the organization and planning of production*. Reprint edition of the book, published in 1939, with introductory paper of L. V. Kantorovich. St. Petersburg, Publishing House of St. Petersburg Univ., 96 pp.

[3] Kantorovich, L. V., 1942: On translocation of masses. *USSR AS Doklady*. New Serie. vol. 37, 7–8, 227–229 (in Russian). [English translation: *J. Math. Sci.*, 133, 4 (2006), 1381–1382.]

[4] Kantorovich, L. V., and Gavurin, M. K., 1949: Application of mathematical methods to problems of analysis of freight flows. *Problems of raising the efficiency of transport performance*, Moscow-Leningrad, 110–138 (in Russian).

[5] Delaunay, B. N., 1947: Gaspard Monge as a mathematician. *Gaspard Monge. Collection of papers to bicentenary of his birthday*, USSR Academy of Science, 7, 1–7 (in Russian).

[6] Kantorovich, L. V., 1948: On a problem of Monge. *Uspekhi Mat. Nauk*, **3**, 225-226 (in Russian). [English translation: *J. Math. Sci.,* 133, 4 (2006), 1383.]

[7] Kantorovich L. V., Rubinshtein G. S., 1958: On a space of totally additive functions. *Vestn. Leningrad. Univ.,* 13, 7 (1958), 52–59 (in Russian).

[8] Vershik, A. M., Kutateladze, S. S., and Novikov, S. P. 2012: Leonid Vital'evich Kantorovich (on the 100th anniversary of his birth). *Russian Math. Surveys*, 67, 3, 589 (in Russian).

[9] Arens R., Eells J., 1956: On embedding uniform and topological spaces. *Pacif. J. Math*. 6, 397–403.

[10] Rachev, S., Rüschendorf, L., 1996: *Mass Transportation Problems*, Vol. I: Theory 1998, XXV, Springer. 540 pp.; Vol. II. Applications. 450 pp.

[11] Melleray, J., Petrov, F., and Vershik, A., 2008: Linearly rigid metric spaces and the embedding problem. *Fund. Math*. 199, 2, 177–194.

[12] Bogachev, V. I., and Kolesnikov, A. V., 2012: The Monge-Kantorovich problem: achievements, connections, and perspectives *Russian Mathematical Surveys*, 67, 5, 3–110.

[13] Vasershtein, L. N. 1969: Markov processes over denumerable products of spaces describing large system of automata, *Problems Inform. Transmission*, 5, 3, 47–52.

[14] Villani, C., 2006: *Optimal Transport, Old and New,* Springer, 635 pp.

[15] Vershik, A., 1970: Some remarks on infinite dimensional problems of linear programming, *Uspekhi Mathematicheskih nauk*, 25, 2, 117–124 (in Russian). [English translation: *Russ. Math. Surveys,* 25, 5 (1970), 117–124.]

[16] Vershik, A., 2004: Kantorovich metric: initial history and little-known applications, *Zapiski nauch. sem.* POMI, 312, 69,1–7 (in Russian). [English translation: *J. Math. Sci,* 133:4 (2006), 1410–1417.]

[17] Campbell, Robert W., 1961: Marx, Kantorovich, Novozhilov, Stoimost versus reality, *Slavic Review,* 20, No. 3, 402–418.

[18] Ambrosio, L., 2003: Lecture Notes on Optimal Transport Problems. Lecture Notes in Mathematics ''Mathematical aspects of evolving interfaces'' (Cime Series, Madeira (PT), 2000) **1812**, P. Colli and J. F. Rodriques, eds., 1–52.

[19] Strassen, V., 1965: The existence of probability measures with given marginals, *Annals of Math. Statist.* 36, 2, 423–439.

[20] Vershik, A., 2013: Two ways of defining compatible metrics on the simplex of measures, *Zapiski nauchn.sem.* POMI, 411 (in Russian). [English translation: *J. Math. Sci.*]