Revista Colombiana de Matemáticas
Vol. XXI (1987), págs. 147-154

TWO PROOFS OF THE KANTOROVICH INEQUALITY AND SOME GENERALIZATIONS

by

Wolfgang J. BÜHLER

Abstract. Two elementary probabilistic proofs of the Kantorovich inequality are given and various generalizations and inequalities are discussed.

§1. Introduction. The Kantorovich inequality in its simplest form seems to have been found originally as early as 1914 by Schweitzer [8]. In the context of statistical applications it is usually stated in the form

$$1 \leqslant (x'Ax)\cdot(x'A^{-1}x) \leqslant (a+b)^2/4ab \qquad (1)$$

where A is a positive definite n×n matrix with eigenvalues $0 < a = \lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n = b$ and x is a unit vector.

For given A there are unit vectors x_1, x_2 for which the left or right inequality respectively is an equality. There have been various generalizations and modifications, see e. g. [6] and the references given there. My own interest in the matter comes from reading the proof in [7]. This proof starts by transforming the term $F = (x,Ax)(x'A^{-1}x)$ by diagonalizing A into

$$F = \left(\sum_{i=1}^{n} y_i \lambda_i^2\right)\left(\sum_{i=1}^{n} y_i^2 \lambda_i^{-1}\right), \text{ with } \sum_{i=1}^{n} y_i^2 = 1.$$

Letting $p_i = y_i^2$ this has the following probabilistic interpretation not used in [7]. F is the product of the expected value of a random variable Z taking values λ_i with probabilities p_i and of the expected value of 1/Z. Both proofs given in the present paper use the representation of F as EZE(1/Z).

§2. **First proof and generalization.** Let Z be a random variable, not necessarily discrete, taking values in [a,b], $0 < a < b < \infty$.

THEOREM 1. $1 \leq EZ\cdot E(1/Z) \leq (a+b)^2/4ab.$ (2)
The right inequality is an equality if and only if Z *takes each of the values* a *and* b *with probability* 1/2, *the left inequality if and only if* Z *is degenerate.*

REMARK. The special case where the random variable Z is defined as an integrable function on [0,1] is already contained in Schweitzer [8].

Proof of Theorem 1. We start from the well known bound for the covariance

$$EZ\cdot E(1/Z) \leq E[Z(1/Z)] + [V(Z)V(1/Z)]^{\frac{1}{2}}$$

the symbol V denoting the variance. Combining this with the fact that the variance of the bounded variable Z is at most $(a+b)^2/4$ (and similarly for V(1/Z)) we obatin the upper bound $1+\{(a+b)^2/4\cdot(1/a+1/b)^2/4\}^{\frac{1}{2}} = (a+b)^2/4ab$. The first bound is attained if 1/Z is a.s. a decreasing linear function of Z which allows at most two values, the second if the two values a and b have probability 1/2 each. The left inequality in (2) could be proved via essentially the same path. It is however also a simple application of Jensen's inequality.

COROLLARY. *Let* (Ω, F, μ) *be a measure space with* $\mu(\Omega) < 1$ *and let* f *be a measurable real valued function on* Ω *with* $a < f(\omega) < b$ *almost everywhere* $|\mu|$. *Then*

$$\int f d\mu \int (1/f) d\mu < (a+b)^2/4ab. \qquad (3)$$

Proof. Augment Ω by an additional point δ and μ by a point mass $1-\mu(\Omega)$ at δ to make Ω into a probability space. Define the random variable Z as f on Ω and $Z(\delta) = a$, say. Then apply Theorem 1.

Let x, y be two unit n-vectors, A a matrix as above (without loss of generality assumed to be diagonal). Then $\mu(\{i\}) = |x_i y_i|$ defines a measure on $\{1,2,...n\}$ to which we can apply the Corollary with $f(i) = \lambda_i$. The resulting inequality

$$(x'Ay)\cdot(y'A^{-1}x) \leqslant \sum \lambda_i |x_i y_i| \sum (1/\lambda_i)|x_i y_i| \leqslant (a+b)^2/4ab$$

is Strang's [9] generalization of Kantorovich's inequality.

§3. Second proof and further generalization. The second proof makes use of the convexity of $f(x) = 1/x$ and can in fact be applied for any convex f. We are thus led to

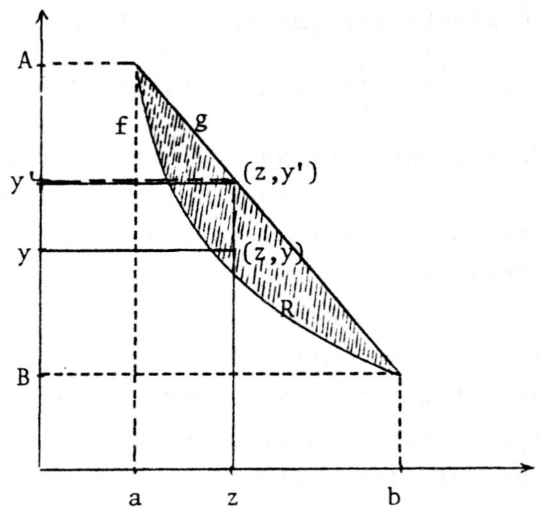
THEOREM 2. *Let* Z *be a random variable with* $P(a \leqslant Z \leqslant b) = 1$, $0 < a < b < \infty$; *let* f *be defined on* $[a,b]$ *and convex with* $f(a) = A > B = f(b)$. *Then we have*

$$EZ \cdot Ef(Z) \leqslant \max\{(Ab-Ba)^2/[4(A-B)(b-a)],\ aA,\ bB\}. \qquad (4)$$

Equality is reached for some Z *taking at most the two values* a *and* b. *If* f *is concave then the inequality is reversed, i.e.*

(4) *yields a lower bound.*

Proof. The two dimensional random variable $(Z,f(Z))$ takes its values on the graph of f. E being a convex operation the point $(EZ,Ef(Z))$ must lie in the (closed) shaded region R bounded by the graph of f and by the straight line $g(z) = A+(B-A)(z-a)/(b-a)$.



The aim is now to maximize the area zy of a rectangle with $(z,y) \in R$. Obviously $zy' > zy$ such that we can limit our search to points $(z,g(z))$. The area $h(z) = z\cdot g(z)$ is maximized when

$$0 = h'(z) = g(z) + zg'(z) = \{A(b-a) + (B-A)(z-a) + z(B-A)\}/(b-a) \ ,$$

i.e. when $z = (Ab-Ba)/2(A-B) =: z_o$ with $h(z_o) = (Ab-Ba)^2/[4(A-B)(b-a)]$. Here $z_o = \alpha a + (1-\alpha)b$ with

$$\alpha = (Ab+Ba-2Bb)/[2(b-a)(A-B)].$$

The local maximum of h at z_o being the only extremum it is obviously a global maximum over the whole line g. However not always is $a \leqslant z_o \leqslant b$. If it is, i.e. if $0 \leqslant \alpha \leqslant 1$ then the variable Z with $P(Z=a) = \alpha = 1-P(Z=b)$ will have $EZEf(Z) = h(z_o)$; if $\alpha < 0$ the Z degenerate at a will bring $EZ\cdot Ef(Z) = aA$; if $\alpha > 1$ then Z degenerate at b will attain the upper bound in the form bB.

REMARK. If in the figure we let $z = EX = \beta b + (1-\beta)a$ and $y = Ef(X)$ for some random variable X with $P(a \leqslant X \leqslant b) = 1$ then with $y' = \beta B + (1-\beta)A$ we find the estimate

$$f(EX) \leqslant Ef(X) \leqslant [(Ab-aB) - (A-B)EX]/(b-a).$$

The left inequality is of course Jensen's, the right inequality holds without the restrictions that $0 < a$ and $A > B$. In the present context, if EX is known, it is sharper than (4).

COROLLARY. *Let* Z *be as in the Theorem. Let* f, g *both be convex, one of them increasing the other decreasing,* $f(a) = A, f(b) = B$, $g(a) = C, g(b) = D$. *Then*

$$Eg(Z)\, Ef(Z) \leqslant \max\{(AD-BC)^2/[4(C-D)(B-A)], AC, BD\} \qquad (5)$$

Proof. Under the conditions stated $f \circ g^{-1}$ is convex and decreasing. Thus the theorem can be applied with $Y = g^{-1}(Z)$ in lieu of Z. (4) then takes the form (5).

§4. The inequality of Greub and Rheinboldt. The following inequality contains that given by Greub and Rheinboldt [3] as a special case.

THEOREM 3. *Let* X, Y *be random variables defined on a probability space* $(\Omega, \mathbf{F}, P)$ *with* $a \leqslant X \leqslant b$ *and* $A \leqslant Y \leqslant B$ a.s. *Then*

$$EX^2 \cdot EY^2/(EXY)^2 \leqslant (aA+bB)^2/4aAbB. \qquad (6)$$

Proof. We first remark that both sides of (6) are invariant to scale. We thus rescale X and Y to achieve $EXY = 1$ without changing the ratio X/Y. The equation $\int XYdP = EXY = 1$ displays XY as the density of a probability measure Q. Denoting by $E_Q Z = \int ZdQ = \int ZXYdP = EZXY$ the corresponding expectation we obtain

$$EX^2 \cdot EY^2 = E_Q(X/Y) \cdot E_Q(Y/X) \leqslant \frac{(a/B+b/A)^2}{4(a/B)(b/A)} = \frac{(aA+bB)^2}{4aAbB},$$

where we have applied (2) to $Z = X/Y$ as a random variable on $(\Omega,\mathcal{F},Q)$.

COROLLARY. *If* G *and* H *are commuting matrices with eigenvalues* $0 = a = \lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n = b$ *and* $0 < A < \mu_1 \leqslant \mu_2 \leqslant \ldots \leqslant \mu_n = B$, *then*

$$(x'G^2x)(x'H^2x)/(x'GHx)^2 \leqslant (aA+bB)^2/4aAbB, \tag{7}$$

Proof. G and H can simultaneously be diagonalized. The Corollary then, in an analogous way as (1), follows from Theorem 1.

§5. Examples. Let us go back to the original situation of a matrix A with eigenvalues $0 < a = \lambda_1 \leqslant \ldots \leqslant \lambda_n = b$. Then for $m,k > 1$ one has

$$(x'A^m x)(x'A^{-k}x) \leqslant \max\{a^{m-k}, b^{m-k}, (b^{m+k}-a^{m+k})^2/[4a^kb^k(b^m-a^m)(b^k-a^k)]\}$$

for all x of unit length. This is just an application of the corollary of Section 3 with $f(z) = z^m, g(z) = z^{-k}$. Except in the case $m = k$ there are situations in which a^{m-k} or b^{m-k} is the appropriate upper bound. When $m = 1$ or $k = 1$ it may be worthwhile to cancel two factors b-a to get the bound e. g. in the form

$$(x'A^2x)(x'A^{-1}x) \leqslant \max\{(a,b,(b^2+ab+a^2)^2/[4ab(b+a)]\} = U \tag{8}$$

As a further illustration we consider an application of Theorem 2. Let $0 < s < r$ and let X be a random variable with $P(a \leqslant X \leqslant b) = 1$. Put $Z = X^s$ and $f(z) = z^{-r/s}$. Then according to Theorem 2 we have

$$EX^sEX^{-r} \leqslant \max\{a^{s-r}; (a^sb^{-r}-b^sa^{-r})^2/[4(b^s-a^s)(a^{-r}-b^{-r})]\} = V. \tag{9}$$

Considering the special situation in which $EX^s = 1$ this leads to $(EX^s)^{1/s}(EX^{-r})^{1/r} = (EX^s\ EX^{-r})^{2/r} \leqslant U$ where

$$U = V^{1/r} = \max\{a^{(s-r)/r}; (b^{s+r}-a^{s+r})^{2/r}[4(b^s-a^s)(b^r-a^r)(ab)^{2r}]^{1/r}\}.$$

For the general situation $EX^s =: c^s$ the upper bound for $(EX^s)^{1/s}(EX^{-r})^{1/r}$ will be $Uc^{(r-s)/r}$ which is at most $Ub^{(r-s)/r}$.

Our approach also covers a more general situation. Let H be a Hilbert space of real valued square integrable functions with scalar product $<x,y> = \int x(t)y(t)dt$. Let the operator A be given by $(Ax)(t) = h(t)x(t)$ where $0 < a \leqslant h(t) \leqslant b$ for all t. Then $(A^{-1}x)(t) = x(t)/h(t)$ and the value U given in (8) will be an upper bound for the product $<A^2x,x><A^{-1}x,x>$, a result which obviously extends to the class of self adjoint integral operators that can be transformed into the "diagonal" form considered. Similarly the inequalities of Strang and of Greub and Rheinboldt apply in this context.

REFERENCES

[1] Beckenbach, E.F., *On the inequality of Kantorovich*. Amer. Math. Monthly, **71**, (1964) 606-619.

[2] Clausing, A., *Kantorovich-type inequalities*. Amer. Math. Monthly, **89** (1982) 314 and 327-330.

[3] Greub, W. and Rheinboldt, W., *On a generalization of an inequality of L.V. Kantorovich*. Proc. Amer. Math. Soc., **10** (1959) 407-415.

[4] Kantorovich, L.V., *Functional analysis and applied mathematics*. Uspehi Mat. Nauk., **3** (1948) 89-135 (also translated from the Russian by C.D. Benster, Nat. Bur. Standards Report Nº 1509, March 7, 1952). (1948).

[5] Khatri, C.G. and Rao, C.R., *Some extensions of Kantorovich inequality and statistical applications*. J. Multivar. Anal., 11, (1981) 498-505.

[6] Rao, C.R. and Khatri, C.G., *Some generalizations of Kantorovich inequality*. Sankhyā, Ser. A, **44** (1982) 91-102.

[7] Raghavachari, M., *A linear programming proof of Kantorovich's inequality*. The American Statistician, **40** (1986) 136-137.

[8] Schweitzer, P., *An inequality concerning the arithmetic mean* (in Hungarian). Math. Phys. Lapok, **23** (1914) 257-261.
[9] Strang, W.G., *On the Kantorovich inequality*. Proc. Amer. Math. Soc., **11** (1960) 468.

*

Fachbereich Mathematik
Johannes Gutenberg-Universität
D-6500 Mainz
República Federal de Alemania.

(Recibido en octubre de 1986).