KANTOROVICH-TYPE INEQUALITIES FOR OPERATORS VIA D-OPTIMAL DESIGN THEORY

L PRONZATO, H P WYNN, AND A ZHIGLJAVSKY

Abstract. The Kantorovich inequality is $z^T Az\, z^T A^{-1}z \leq (M+m)^2/(4mM)$, where A is a positive definite symmetric operator in $\mathbb{R}^d$, z is a unit vector and m and M are respectively the smallest and largest eigenvalues of A. This is generalised both for operators in $\mathbb{R}^d$ and in Hilbert space by noting a connection with D-optimal design theory in mathematical statistics. Each generalised bound is found as the maxima of the determinant of a suitable moment matrix.

1. Background

1.1. **Kantorovich inequality.** Let A be a positive definite symmetric operator in $\mathbb{R}^d$ with minimum and maximum eigenvalues m and M $(0 < m < M)$ respectively, and let z be a generic vector in $\mathbb{R}^d$. The Kantorovich inequality takes the form:

$$z^T Azz^T A^{-1}z \leq \frac{(m+M)^2}{4mM} \parallel z \parallel^4 . \tag{1.1}$$

Attributed to L.V. Kantorovich, [1], the inequality has built up a considerable literature. This typically comprises generalisations. Examples are [2], [3], [4]. Operator versions are developed in [5] and [6]. All the generalisations in this paper also have operators versions. Multivariate versions have been useful in statistics to assess the robustness of least squares: see [7] and [8] and the references therein.

We shall prefer to write (1.1) as

$$\max_{\|z\|=1}\{z^T Azz^T A^{-1}z\} = \frac{(M+m)^2}{4mM}. \tag{1.2}$$

This is then reduced to a one-dimensional problem by a spectral resolution of A

$$A = \sum_{i=1}^{d} \lambda_i u_i u_i^T ,$$

where the $m = \lambda_1 \leq \ldots \leq \lambda_d = M$ are the ordered eigenvalues and u_i $(i = 1, \ldots, d)$ are corresponding orthogonal unit eigenvectors.

Define $\xi_i = (u_i^T z)^2 \geq 0$ and note that $\parallel z \parallel= 1$, $\parallel u_i \parallel= 1$ $(1 = 1, \ldots, d)$ and the u_i being orthogonal forces $\sum \xi_i = 1$. Thus, $\xi = \{\xi_i, \lambda_i\}$ can be considered as a discrete probability distribution with masses ξ_i at the support points λ_i, respectively. Therefore, the equality (1.2) can be written as

$$\max_{\xi} \left\{ \sum_{i=1}^{d} \lambda_i \xi_i \sum_{i=1}^{d} \lambda_i^{-1} \xi_i \right\} = \frac{(M+m)^2}{4mM} . \tag{1.3}$$

With det denoting determinant, this equality can be written as

$$\max_{\xi} \det(\Gamma(\xi)) = \frac{(M-m)^2}{4mM}\,, \tag{1.4}$$

where

$$\Gamma(\xi) = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$$

and

$$m_{11} = \sum \lambda_i^{-1}\xi_i,\;\; m_{12} = m_{21} = \sum \lambda_i^{-1/2}\lambda_i^{1/2}\xi_i = \sum \xi_i = 1,\;\; m_{22} = \sum \lambda_i \xi_i\,.$$

We see that $\Gamma(\xi)$ is the moment matrix

$$\Gamma(\xi) = \sum_i f(\lambda_i) f(\lambda_i)^T \xi_i$$

in the special case $f(x) = (x^{-1/2}, x^{1/2})^T$.

This is the point at which the generalisations described here begin. We simply look at any vector of functions $f(x) = (f_1(x), f_2(x))^T$ with $f_1(x), f_2(x) > 0,\;\; x \in [m, M]$ and seek an upper bound:

$$\det(\Gamma(\xi)) = \sum_i f_1(\lambda_i)^2\xi_i \sum_i f_2(\lambda_i)^2\xi_i - \left(\sum_i f_1(\lambda_i) f_2(\lambda_i)\xi_i\right)^2 \le \max_{\xi} \det(\Gamma(\xi))\,. \tag{1.5}$$

The maximum is taken over all non-negative (probability) measures on $[m, M]$, that is

$$\xi(dx) \ge 0 \;\text{ on } [m, M],\;\; \int_m^M \xi(dx) = 1\,, \tag{1.6}$$

(although it is achieved for discrete measures). Note that the lower bound for $\det(\Gamma(\xi))$ in (1.5)

$$\min_{\xi} \left\{ \sum f_1(\lambda_i)^2\xi_i \sum f_2(\lambda_i)^2\xi_i - \left(\sum f_1(\lambda_i) f_2(\lambda_i)\xi_i\right)^2 \right\} \ge 0$$

is just the Cauchy-Schwartz inequality, as pointed out by many authors.

In section 1.3 we shall cover the maximum determinant problem, which in mathematical statistics is called the D-optimality problem. In order to generalise the Kantorovich inequality while retaining some of its simplicity we shall first study the special case when $f(x) = (x^p, x^q)^T$.

1.2. **The Hilbert space case.** All bounds in this paper carry over to the Hilbert space case. We consider a Hilbert space with inner product $< \cdot, \cdot >$ and A to be a positive bounded self adjoint operator with spectrum ξ in $[m, M]$, where $0 < m < M < \infty$. We replace the quadratic form $z^T A z$ by the inner product:

$$< zA, z >= \int_m^M x\xi(dx) = \mu(1)$$

and similarly

$$< zA^{\alpha}, z >= \int_m^M x^{\alpha}\xi(dx) = \mu(\alpha)\,.$$

Since in the D-optimality results we take the maximum over all probability measures ξ, we interpret this as taking the supremum over all bounded self adjoint operators with spectral range $[m, M]$.

This connection between D-optimality and moments problem more generally arose during work by the authors on renormalised steepest descent problems [9].

1.3. **D-optimal design theory.** Consider a set of continuous functions $\{f_1, \dots, f_k\}$ on compact set X in $\mathbb{R}^d$. In linear regression theory the aim is to fit a model with expected response of the form:

$$\eta = \mathrm{E}(Y) = \sum_{j=1}^{k} \theta_j f_j(x)\,.$$

A set of N points in X, namely an experimental design, is idealised to a probability measure ξ on X. (We shall use the word "measure", for short.) One can think of this as a normalisation which avoids the use of the sample size, N. Following a statistical justification, which we ignore, the D-optimality criterion is

$$\max_{\xi} \det\{\Gamma(\xi)\}\,,$$

where

$$\Gamma(\xi) = \int_X f(x) f(x)^T \xi(dx)$$

is the $k \times k$ moment matrix for the $f(x) = (f_1(x), \dots, f_k(x))^T$. We call a measure ξ^* which achieves this maximum D-optimal and the fact that it is achieved derives from the continuity of the f_i's and the compactness of X. We shall also need the "variance function"

$$d(x, \xi) = f(x)^T \Gamma(\xi)^{-1} f(x)$$

which, statistically, is the normalised version of the variance of prediction of η at a point x (under standard regression assumptions).

We state without proof the General Equivalence theorem (GET), see [10].

Theorem 1.1. *The following three statements are equivalent for a measure ξ^* on the compact set X with continuous function $f = (f_1, \dots, f_k)^T$.*

- **(i):** *ξ^* is D-optimal: achieves* $\max_\xi \det\{\Gamma(\xi)\}$
- **(ii):** $\min_\xi \max_{x \in X} d(x, \xi) = \max_{x \in X} d(x, \xi^*)$
- **(iii):** $\max_{x \in X} d(x, \xi^*) = k$.

Here are some useful lemmas, closely connected to the GET.

Lemma 1.2. *For any measure ξ on X and functions $f = (f_1, \dots f_k)^T$ with non-singular moment matrix $\Gamma(\xi)$ we have*

$$\int_X d(x, \xi)\xi(dx) = k\,.$$

Lemma 1.3. *If the D-optimal design is supported at discrete points x_i $(i = 1, \dots, n)$, then*

$$d(x_i, \xi^*) = k \ (i = 1, \dots, n).$$

Lemma 1.4. *If a D-optimal measure is supported at k point, where k is the number of functions f_j, then the masses at the support point are $\xi_i = 1/k$ $(i = 1, \dots, k)$.*

Lemma 1.2 follows from

$$\int_X d(x, \xi)\xi(dx) = \int_X f(x)\Gamma(\xi)^{-1} f(x)^T \xi(dx)$$

$$= \mathrm{trace}\{\Gamma(\xi) \int_X f(x) f(x)^T \xi(dx)\} = \mathrm{trace}\{\Gamma(\xi)^{-1}\Gamma(\xi)\} = k.$$

Lemma 1.3 follows from Lemma 1.2 and the GET by contradiction, as follows. Suppose there is a x_i such that $d(x_i, \xi^*) < k$. The GET (iii) says that $d(x, \xi^*) \le k$ for all $x \in X$ and so holds in particular for the x_i. The last two statements together contradict Lemma 1.2. Lemma 1.4 follows by noting that in that case $\Gamma(\xi)$ is square so that $\prod \xi_i$ is a factor of $\det(\Gamma(\xi))$, which is maximised, subject to $\sum \xi_i = 1$, by all $\xi_i = 1/k$.

One further, "dual" result, gives important geometric intuition [11].

Lemma 1.5. *Define the set in* $\mathbb{R}^k$

$$F = \left\{ (f_1(x), \ldots, f_k(x))^T : \, x \in X \right\} ,$$

then ξ^* *is* D*-optimal if and only if the ellipsoid given by*

$$z^T \Gamma(\xi^*)^{-1} z = k$$

is the minimum volume ellipsoid, centred at the origin, which contains F*.*

As pointed out, any D-optimal design problem with $X = [m, M]$ and $k = 2$ leads to a simple generalisation of the Kantorovich inequality. But, of course, when $k > 2$ we have another kind of generalisation based on

$$\det\{\Gamma(\xi)\} \leq \det\{\Gamma(\xi^*)\} ,$$

where ξ^* is D-optimum. We return to this discussion after our special example in the next section.

2. Examples

2.1. **The case** $f(x) = (x^p, x^q)^T$**.** The purpose of this section is to give a simple generalisation of the original Kantorovich inequality as stated in version (1.2).

Theorem 2.1. *Let* A *be a positive definitive matrix with* m, M *as the minimum and maximum eigenvalues. Then, if* p *and* q *have opposite signs* ($p\,q < 0$)*,*

$$\sup_{\|z\|=1} \{ z^T A^{2p} z \, z^T A^{2q} z - \left(z^T A^{p+q} z \right)^2 \} = \frac{1}{4} \left(m^p M^q - m^q M^p \right)^2 . \tag{2.1}$$

If $p, q > 0$

$$\sup_{\|z\|=1} \left\{ z^T A^{2p} z \, z^T A^{2q} z - \left(z^T A^{p+q} z \right)^2 \right\} = \frac{1}{4} \left(v^p M^q - v^q M^p \right)^2 , \tag{2.2}$$

where

$$v = \max \left\{ m, \left(\frac{q}{p} \right)^{1/(p-q)} M \right\} .$$

Proof. We first reduce by the spectral decomposition as explained above for the Kantorovich case. We next exhibit the D-optimal solution and check that Theorem 1 (iii) holds. There are two cases, separated by the critical point $x^* = (q/p)^{1/(p-q)} M$, in the theorem.

When p and q are of opposite signs or when $p, q > 0$ and $m \geq x^*$ the D-optimal design ξ^* is supported with mass $1/2$ on each of the points m, M. Then $d(x, \xi^*)$ achieves a maximum value of 2 at $\{m, M\}$.

When $p, q > 0$ and $m \leq x^*$, ξ^* places mass $1/2$ at each of x^* and M. It is verified that $d(x, \xi^*)$ achieves a maximum of 2 at x^* and M (see Figure 1 for typical functions $d(x, \xi^*)$).

The value x^* was found by putting x^* and M as support points with masses $1/2, 1/2$ and forcing $\partial d(x, \xi)/\partial x$ to be zero at x^*, giving a single equation for x^*. That is sufficient is consequence of the fact that $\max_{x \in X} d(x, \xi^*)$ is achieved at every support point, by Lemma 1.3, and that, for this example, any $d(x, \xi)$ has at most two turning points in $[m, M]$.

From Theorem 1.1 (GET) we infer that in each case ξ^* is D-optimum. The maximum values, $\det\{\Gamma(\xi^*)\}$, are given in the right-hand sides of (2.1) and (2.1). Since the ξ^* are discrete measures the maximum value is achieved by the two point measure ξ^* in the first case, but in the second the bound can be strict and only achieved when there is an eigenvalue of A equal to x^*. $\square$

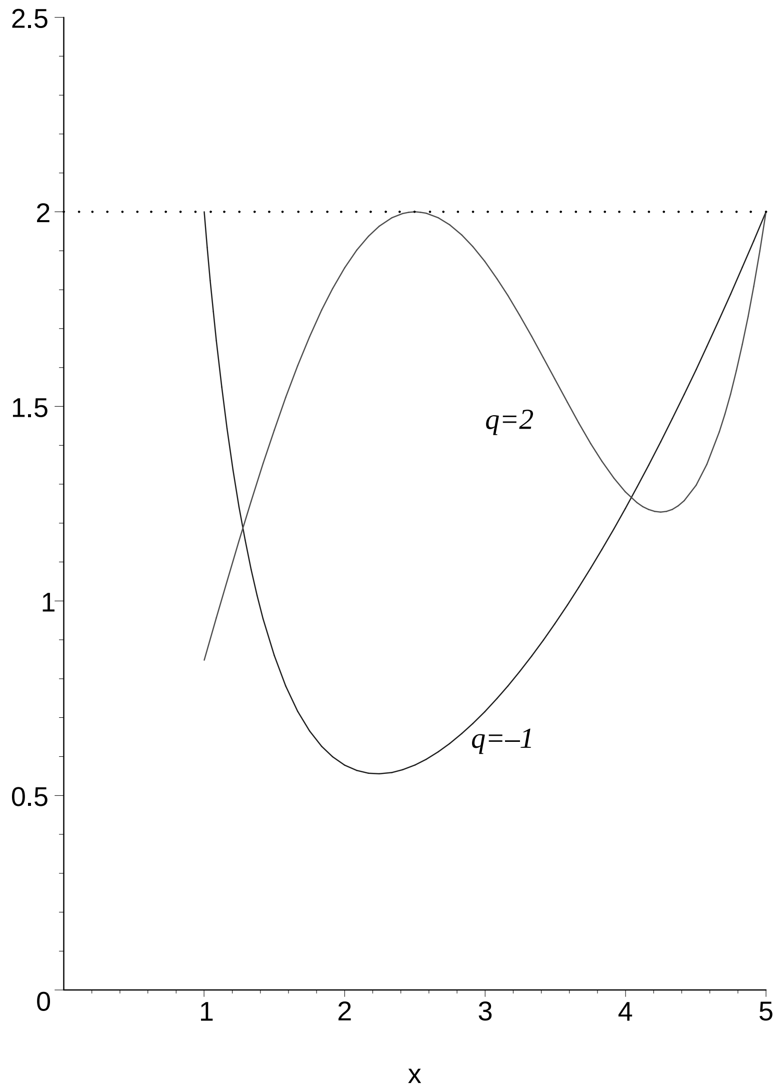

FIGURE 1. The function $d(x, \xi^*)$ for the case $k = 2$, $m = 1, M = 5$, $f(x) = (x, x^q)^T$ with $q = -1$ and $q = 2$

2.2. **Other $k = 2$ examples.** In the example just described the D-optimal measure in each of the two cases has a two-point support. We are able to claim, by Lemma 1.3, that the masses are equal on the design points. In general a D-optimal design can be found for a problem with k parameters which has a maximum of $s = k(k+1)/2$ support points. This follows using Caratheodory's theorem and the fact that the set of all $\Gamma(\xi)$ is a convex set in s dimensions and a D-optimal solution can be found on the boundary. Thus, when $d = 2$ we can expect in general up to 3 support points. If for $k = 2$ the solution has three points ξ^* is somewhat harder to find because there is no reason to expect it to have uniform measure (Lemma 1.5 does not apply).

A simple case in which there is always a 2-point support for the D-optimal design is when the locus

$$F = \{(f_1(x), f_2(x)) \;\; x \in [m, M]\}$$

is a convex and decreasing arc in $\mathbb{R}^2$, considering f_2 as a function of f_1 (or vice-versa). This is most easily seen from the dual version in Lemma 1.4. It is clear, in this case, that the minimal volume ellipse containing F must intersect F only at $(f_1(m), f_2(m))$ and $(f_1(M), f_2(M))$ and that therefore the D-optimal measure is uniform $\{1/2, 1/2\}$ on $\{m, M\}$. In this case, the bound, namely, the maximum value of $\det(\Gamma(\xi))$ is given by

$$\frac{1}{4}(f_1(m)f_2(M) - f_1(M)f_2(m))^2.$$

The Kantorovich inequality is the special case where the arc is given by $f_1 f_2 = 1$, and the case when p, q has different signs in Theorem 1.1 is also in the class.

In the proof of Theorem 2.1 we used the number of turning points of $d(x, \xi^*)$ (2 in that case) as part of the proof. We can give a number of elementary results using this idea. They are specialisations to the case $k = 2$ of the following.

Lemma 2.2. *Let $f = (f_1, \ldots, f_k)^T$ be positive continuous functions on $[m, M]$ with the property that for any measure ξ on $[m, M]$ for which the moment matrix $\Gamma(\xi)$ is non-singular and $d(\xi, x)$ is differentiable the function $\partial d(x, \xi)/\partial x$ has r zeros in $[m, M]$. Assume also that at least one such non-singular $\Gamma(\xi)$ exists. Then the D-optimum measure has at most $(r+2)/2$ points when r is even and at most $(r+3)/2$ points when r is odd.*

We can strengthen the condition concerning turning points in Lemma 2.2 to require that for any positive definite matrix B, the derivative of $f^T B f$ has r zeros. Or we can weaken the condition to require that the derivative of $d(x, \xi^*)$ itself has r zeros.

2.3. **Two $k = 3$ examples.** We give a couple of inequalities which follow from using $k = 3$. We give them in the moment form using the notation $\mu(\alpha) = \int_X x^\alpha \xi(dx)$ which is $\sum x_i^\alpha \xi_i$, in the discrete case. Both examples are cases where $k = 3$ and the D-optimal measure is supported at 3 points and therefore, by Lemma 1.4, $\xi_1 = \xi_2 = \xi_3 = 1/3$. The proof essentially uses Lemma 2.2 with $k = 3, r = 3$.

(i) Taking $f = (1, x, x^2)^T$ the D-optimal design is uniform $\{1/3, 1/3, 1/3\}$ on $\{m, (m + M)/2, M\}$ and we obtain the bound

$$\det(\Gamma(\xi)) = \mu(2)\mu(4) - \mu(3)^2 - \mu(1)^2\mu(4) + 2\mu(1)\mu(2)\mu(3) - \mu(2)^3 \leq \frac{(M - m)^6}{432}\,.$$

(ii) Taking $f = (1, 1/x, x)^T$ the D-optimal design is supported, again uniformly, at $\{m, (mM)^{1/2}, M\}$ which gives

$$\begin{aligned} \det(\Gamma(\xi)) &= \mu(-2)\mu(2) - \mu(-1)^2\mu(2) - \mu(1)^2\mu(-2) + 2\mu(-1)\mu(1) - 1 \\ &\leq \tfrac{1}{27m^2M^2}(M - m)^2(m^{1/2} - M^{1/2})^4\,. \end{aligned}$$

For the original quadratic form versions of (i) and (ii) we put $\mu(\alpha) = z^T A^\alpha z$, $\| z \| = 1$.

One may notice that the same bounds are obtained for a two dimensional model ($k = 2$). Indeed, one can easily check that $f = (x - 1, x^2)^T$ and $f = (x - 1, 1/x)^T$ respectively give the same bounds that in (i) and (ii).

3. Conclusion

The paper shows that the Kantorovich inequality for operators and in $\mathbb{R}^d$ can be reduced to a moment bound in one dimension for spectral measures over the spectral range $[m, M]$. D-optimal design theory in statistics is a rich source of such bounds and indeed the Kantorovich bound can be written as a simple D-optimal design problem. Essentially, any D-optimal design problem leads to a special Kantorovich-type bound and some small examples are given. If the Kantorovich bound is considered as as the converse bound to the Cauchy Schwarz bound so general "upper" moments types bounds arising from D-optimality and elsewhere are converses to the "lower" moment bounds which might arise, for example, by requiring A to be non-negative definite.

Extensions and alternatives to D-optimality are quite numerous: linear optimality, D_s-optimality, c-optimality, weighted D-optimality, ϕ_p-optimality, and so on, each producing a moment bound of some kind. Moreover, most of these reduce to special optimality problems in moment space with the theory being most attractive because what is being maximised is a convex matrix functional. So, in summary, the rather beautiful Kantorovich bound is perhaps the simplest case of a vast range of bounds based on optimising a functional on the spaces of spectral moments.

References

1. L. V. Kantorovich, *Functional analysis and applied mathematics (Russian)*, Uspekhi Mat. Nauk. **3** (1948), 89–185.
2. Baksalary J. K. and S. Puntanem, *Generalized matrix versions of the Cauchy Schwartz and Kantorovich inequalities*, Aequationes Mathematicae **41** (1991), 103–110.
3. Greub W. and Rheinboldt W., *On a generalistion of an inequality of L.V. Kantorovich*, Proc. Amer. Math. Soc **10** (1959), 407–415.
4. W. G. Strang, *On the Kantorovich inequality*, Proc. Amer. Math. Soc **22** (1959), 2904.
5. P. G. Spain, *Operator versions of the Kantorovich inequality*, Proc. Amer. Math. Soc. **124** (1996), 2813–2819.
6. R Bhatia and C Davis, *More operator verions of the Schwarz inequality*, Commun. Math. Phys **215** (2000), 239–244.
7. Bloomfield P. and Watson G. S., *The inefficiency of least squares*, Biometrika **62** (1975), 121–128.
8. Liu S. and Neudecker H., *Kantorovich inequalities and the efficiency comparisons for several classes of estimators in linear models*, Staistica Neerlandica **51** (1997), 345–355.
9.
10. Kiefer, J. and Wolfowitz, J., *The equivalence of two extremum problems,* Canad. J. Math., **12**, (1960), 363–366.
11. Titterington D M (1975) Optimal design: some geometrical aspects of D-optimality, vol 62 pp 313-320.