Rose-Hulman Institute of Technology

Rose-Hulman Scholar

Mathematical Sciences Technical Reports (MSTR)

Mathematics

8-6-2007

The Barycenter of the Numerical Range of a Matrix

Sean A. Broughton
Rose-Hulman Institute of Technology, brought@rose-hulman.edu

Roger G. Lautzenheiser
Rose-Hulman Institute of Technology

Thomas Werne
Rose-Hulman Institute of Technology

Follow this and additional works at: https://scholar.rose-hulman.edu/math_mstr

 Part of the Algebra Commons, Analysis Commons, and the Applied Mathematics Commons

Recommended Citation

Broughton, Sean A.; Lautzenheiser, Roger G.; and Werne, Thomas, "The Barycenter of the Numerical Range of a Matrix" (2007). *Mathematical Sciences Technical Reports (MSTR)*. 39.
https://scholar.rose-hulman.edu/math_mstr/39

This Article is brought to you for free and open access by the Mathematics at Rose-Hulman Scholar. It has been accepted for inclusion in Mathematical Sciences Technical Reports (MSTR) by an authorized administrator of Rose-Hulman Scholar. For more information, please contact weir1@rose-hulman.edu.

The Barycenter of the Numerical Range of a Matrix

S. Allen Broughton, Roger Lautzenheiser, Thomas Werne

Mathematical Sciences Technical Report Series
MSTR 07-04

August 7, 2007

Department of Mathematics
Rose-Hulman Institute of Technology
http://www.rose-hulman.edu/math.aspx

Fax (812)-877-8333 **Phone (812)-877-8193**

The Barycenter of the Numerical Range of a Matrix

S. Allen Broughton
Rose-Hulman Institute of Technology

Roger G. Lautzenheiser
Rose-Hulman Institute of Technology

Thomas Werne
Rose-Hulman Institute of Technology

August 6, 2007

Abstract

The numerical range $W(A)$ of an $n \times n$ matrix A is the totality of the scalar products $\langle A\mathbf{x}, \mathbf{x}\rangle$ as $\mathbf{x}$ varies over all unit vectors in $\mathbb{C}^n$. The barycenter (center of mass) of the numerical range is defined geometrically as the center of mass of $W(A)$ considered as a planar lamina with variable density and also as a limit of sample averages $\frac{1}{N}\sum_{i=1}^{N}\langle A\mathbf{x}_i, \mathbf{x}_i\rangle$. Under a wide range the sampling schemes it is shown that the barycenter is the average of the spectrum (eigenvalues) of A weighted according to algebraic multiplicity which also equals $\mathrm{tr}(A)/n$. The results of this paper justifies calling $\mathrm{tr}(A)/n$ the barycenter of $W(A)$.

1 Introduction

In a second course on linear algebra offered by the second author, the third author tackled the problem of determining the centroid C of the numerical range $W(A)$ of a $n \times n$ matrix A. Proceeding by experimental investigation and planar plotting techniques the second author was led to several conjectures, one of which is that the centroid (in the 2×2 case) and the barycenter (in the $n \times n$ case), is equal to the center of mass of the spectrum, weighted by multiplicity, which in turn equals $\mathrm{tr}(A)/n$. Following a seminar presentation of the results, the first author was able to prove one of the conjectures. We present a summary of our combined efforts along the mathematical trail from experimentation to definition/conjecture and finally to proof.

In 1917, O. Toeplitz defined the numerical range of an $n \times n$ matrix A to be

$$W(A) = \{\langle A\mathbf{x}, \mathbf{x}\rangle : \mathbf{x} \in \mathbb{C}^n, \|\mathbf{x}\| = 1\}$$

and a year later, F. Hausdorff proved that $W(A)$ is convex. The fact that $W(A)$ is a convex, compact subset of the complex plane is referred to as the Toeplitz-Hausdorff Theorem. It is easy to see that $W(A)$ contains the eigenvalues of A, and $W(A) = \{\lambda\}$ if and only if $A = \lambda I$. Furthermore, by the Toeplitz-Hausdorff Theorem, interior$(W(A))$ is empty if and only if $W(A)$ is a line segment, which is true if and only if A is a normal matrix with colinear eigenvalues. In particular, $W(A)$ is a subset of $\mathbb{R}$ if and only if A is Hermitian. More generally, if A is normal then $W(A)$ is the convex hull of the eigenvalues. The numerical ranges of 2×2 matrices are ellipses or line segments in the degenerate cases, and in general, non-differentiable boundary points of $W(A)$ are reducing eigenvalues. The numerical range has been studied extensively over the years, and a good source of basic information is Horn & Johnson [HJ]. The following online articles, [PT] and [CKL], also give many basic properties and include many references.

In our search of the literature we found only one result related to centroid or barycenter though these terms are not specifically used. The result, found on page 5 of [PT] and a consequence of Theorem 1.6.1 of [HJ], states that $W(A)$ is a line segment if and only if $\mathrm{tr}(A)/n$ lies on the boundary of $W(A)$. In Corollary 4, we give this result a more geometric flavor by showing $W(A)$ is a line segment if and only if the barycenter of $W(A)$ is on the boundary of $W(A)$.

Given a set in the complex plane, an obvious attempt at finding the set's centroid would be to compute the averages $\frac{1}{N}\sum_{i=1}^{N} z_i$ for N values z_i distributed uniformly throughout the set. When using this expression

for $W(A)$ with $z_i = \langle A\mathbf{x}_i, \mathbf{x}_i\rangle$ and $\mathbf{x}_i$'s unit vectors, we made two important observations after looking at several examples. First, for various reasonable distributions used to generated the $\mathbf{x}$'s, it was clear that the corresponding z's were not necessarily uniformly distributed over $W(A)$. See section 2 for examples. Secondly, it appeared that our average $\frac{1}{N}\sum_{i=1}^{N} z_i = \frac{1}{N}\sum_{i=1}^{N}\langle A\mathbf{x}_i, \mathbf{x}_i\rangle$ was very close to $\frac{1}{n}\sum_{i=1}^{n}\lambda_i$. where λ_i are the eigenvalues of A counting multiplicities.

In general we do not have a simple description of $W(A)$, so a calculation of the centroid will need to be based on the $\mathbf{x}$'s which define $W(A)$ and the average value of $\langle A\mathbf{x}, \mathbf{x}\rangle$, and not simply on the geometry of $W(A)$. Except in the 2×2 case (see Remark 5) the centroid is not always equal to the average value of $\langle A\mathbf{x}, \mathbf{x}\rangle$ since the average corresponds to the center of mass of the planar lamina defined by $W(A)$ but with a variable density. We shall use the word *barycenter* which is a shorter but equivalent term for center of mass.

Based on the above discussion, we define the barycenter of $W(A)$ to be

$$BW(A) = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\langle A\mathbf{x}_i, \mathbf{x}_i\rangle$$

where the $\mathbf{x}_i$'s are chosen from the uniform distribution the boundary of the unit ball in $\mathbb{C}^n$. We now give an equivalent definition and ultimately, show $BW(A) = \frac{1}{n}\sum_{i=1}^{n}\lambda_i$ (or $\dfrac{\mathrm{tr}(A)}{n}$).

Let B_n be the unit ball in $\mathbb{C}^n$, and let ∂B_n be its boundary, which is a sphere of dimension $2n-1$. Choosing the $\mathbf{x}_i$'s uniformly on ∂B_n is equivalent to saying that for each closed subset U of ∂B_n,

$$\lim_{N\to\infty} \frac{\#\{i : \mathbf{x}_i \in U\}}{N} = \frac{vol(U)}{vol(\partial B_n)}, \tag{1}$$

where $vol(U)$ is the volume of U computed as a subset of the boundary sphere. Assuming (1) holds, it follows that

$$BW(A) = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\langle A\mathbf{x}_i, \mathbf{x}_i\rangle = \int_{\partial B_n} \langle A\mathbf{x}, \mathbf{x}\rangle \, d\omega. \tag{2}$$

where $d\omega$ is the normalized volume measure on the sphere.

Furthermore, it is now easy to see how to define the density function on $W(A)$. Let $f_A : \partial B_n \to \mathbb{C}$ be the function $f_A(\mathbf{x}) = \langle A\mathbf{x}, \mathbf{x}\rangle$ and for $z \in W(A)$ define

$$\delta(z) = \lim_{r\to 0} \frac{\omega(f_A^{-1}(\Delta_r(z)))}{\pi r^2}$$

where $\Delta_r(z) = \{w \in \mathbb{C} : \|w - z\| \le r\}$. Then in case $W(A)$ has non-empty interior

$$\int_{W(A)} z\delta(z)dxdy = \int_{\partial B_n} \langle A\mathbf{x}, \mathbf{x}\rangle \, d\omega = BW(A).$$

In the experimental investigations, approximations $BW(A) \approx \frac{1}{N}\sum_{i=1}^{N}\langle A\mathbf{x}_i, \mathbf{x}_i\rangle$ were also computed from distributions of points on the sphere that did not satisfy the uniformity condition given in equation (1) but rather

$$\lim_{N\to\infty} \frac{\#\{i : \mathbf{x}_i \in A\}}{N} = \mu(A), \tag{3}$$

for some different probability measure μ on the sphere. Under very mild conditions on μ the approximation $\frac{1}{N}\sum_{i=1}^{N}\langle A\mathbf{x}_i, \mathbf{x}_i\rangle \approx \frac{1}{n}\sum_{i=1}^{n}\lambda_i$ is still satisfied. The details are given in Theorem 2. We now state the two main results of the paper.

Theorem 1 *Let $d\omega$ be the unitarily invariant volume on ∂B_n, normalized as a probability measure. If the barycenter $BW(A)$ of the numerical range is defined by*

$$BW(A) = \int_{\partial B_n} \langle A\mathbf{x}, \mathbf{x}\rangle \, d\omega.$$

then

$$BW(A) = \frac{\mathrm{tr}(A)}{n} = \frac{1}{n}\sum_{i=1}^{n} \lambda_i.$$

Theorem 2 *Let $d\mu$ be a probability measure on ∂B_n, which is invariant under the following transformations*

$$U_{i,j} : (x_1, \ldots, x_i, \ldots, x_j, \ldots, x_n) \longrightarrow (x_1, \ldots, x_j, \ldots, x_i, \ldots, x_n) \tag{4}$$

for any distinct i, j and

$$V_i : (x_1, \ldots, x_i, \ldots, x_n) \longrightarrow (x_1, \ldots, -x_i, \ldots, x_n) \tag{5}$$

for any i. If the μ-barycenter of $W(A)$ is defined by

$$B_\mu W(A) = \int_{\partial B_n} \langle A\mathbf{x}, \mathbf{x}\rangle \, d\mu,$$

then we have

$$B_\mu W(A) = \frac{\mathrm{tr}(A)}{n} = \frac{1}{n}\sum_{i=1}^{n} \lambda_i.$$

2 Experimental investigation

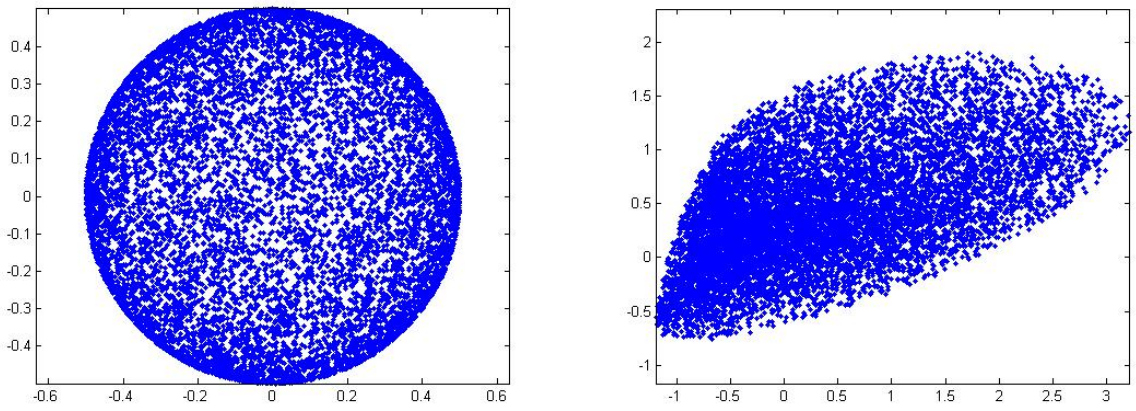
As noted in section 1, the third author began this project by plotting approximations to $W(A)$. Plots were created with Matlab using random vectors (known as "test vectors") from the unit sphere ∂B_n. Using a relatively small number of points, it becomes apparent that (for at least some probability distributions on the unit sphere) the distribution of points in $W(A)$ is not uniform. For example, it can be shown analytically that $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ yields $W(A) = \left\{z : |z| \leq \frac{1}{2}\right\}$ and, assuming test vectors are chosen uniformly, the density function on $W(A)$ is the rotationally symmetric $\delta\left(re^{i\theta}\right) = \frac{2}{\pi}\frac{1}{\sqrt{1-4r^2}}$. The observation of nonuniform point density on $W(A)$ can clearly be seen in Figure 1.

Since the eigenvalues of a matrix characterize the matrix in some sense, they were plotted along with the approximated $W(A)$. After many experimental runs, a pattern began to appear: the nonuniformity of $W(A)$ notwithstanding, the center of mass of the approximated $W(A)$ (that is, its barycenter) seemed to be the same as the average of the eigenvalues. An example of this can be seen in Figure 2.

As noted earlier, some of the nonuniformity of the approximation for $W(A)$ may be due to the distribution of test vectors on ∂B_n. Three distinct distributions were used in examining this problem:

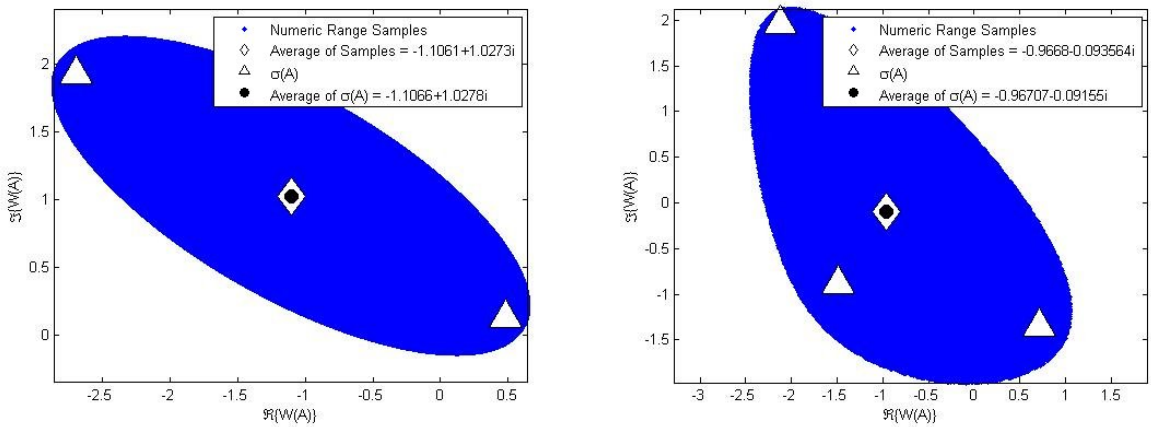
1. Generate random vectors in the box circumscribed about ∂B_n and project them onto ∂B_n.
2. For each real and imaginary part of every complex dimension, generate a random value from a normal distribution and project these random vectors onto ∂B_n.
3. For each complex dimension, choose the modulus from a normal distribution and the argument from a uniform distribution and project the entire vector onto ∂B_n.

A necessary condition for a uniform distribution on ∂B_n (which gives the density of $W(A)$) is invariance of the induced density of $W(A)$ under all unitary similarity transforms of A (the unitary similarity transform effectively just rotates the test vectors in ∂B_n, so the induced density should be unaffected if the distribution on ∂B_n is uniform). This test can be used to rule out distributions 1 and 3 above. While distribution number 2 gives the true uniform distribution on ∂B_n, these investigations indicated that the hypothesis that the barycenter of the $W(A)$ is the average of the eigenvalues held under all three distributions, that is, $\frac{1}{N}\sum_{i=1}^{N} \langle A\mathbf{x}_i, \mathbf{x}_i\rangle \approx \frac{1}{n}\sum_{i=1}^{n} \lambda_i$ resulting in Theorem 2.



(a) A $\in \mathbb{C}^{2x2}$ (b) A $\in \mathbb{C}^{3x3}$

Figure 1: Approximate W(A)



(a) A $\in \mathbb{C}^{2x2}$ (b) A $\in \mathbb{C}^{3x3}$

Figure 2: Approximate W(A), Eigenvalues, and Approximated Barycenters

3 Proof of the barycenter formula

We prove both theorems at the same time. First we shall first prove Theorem 2, and then show that $d\omega$ satisfies the hypothesis of the second theorem.

Proof. From the definitions we have.

$$B_\mu W(A) = \int_{\partial B_n} \langle A\mathbf{x}, \mathbf{x}\rangle \, d\mu = \sum_{i,j} \int_{\partial B_n} a_{i,j} x_i \overline{x_j} d\omega$$

If we can prove that

$$\int_{\partial B_n} x_i \overline{x_j} d\omega = \frac{1}{n}\delta_{i,j} \tag{6}$$

then

$$\begin{aligned} B_\mu W(A) &= \sum_{i,j} \int_{S_n} a_{i,j} x_i \overline{x_j} d\omega = \frac{1}{n}\sum_{i,j} a_{i,j}\delta_{i,j} \\ &= \frac{1}{n}\sum_i a_{i,i} = \frac{1}{n}\mathrm{tr}(A) = \frac{1}{n}\sum_i \lambda_i \end{aligned}$$

To prove equation 6 we first recall how invariance of a measure under a transformation yields an integration invariance formula. If $d\mu$ is a measure on ∂B_n and $T : \partial B_n \to \partial B_n$ is a continuous transformation then $T^*(d\mu)$ is the measure on ∂B_n defined by $T^*(d\mu)(B) = d\mu(T^{-1}(B))$ for each Borel subset $B \subseteq \partial B_n$. The measure $d\mu$ is invariant under T if $T^*(d\mu) = d\mu$. For an invariant measure we have for any continuous function f on ∂B_n

$$\int_{\partial B_n} f(\mathbf{x}) d\mu = \int_{\partial B_n} f(T(\mathbf{x}))T^*(d\mu) = \int_{\partial B_n} f(T(\mathbf{x}))d\mu$$

The first equality is just the "change of variables" integration formula; the second comes from invariance. Let $U_{i,j}$ and V_i denote the unitary matrices defined by equations 4 and 5, and define the functions

$$f_i(\mathbf{x}) = x_i\overline{x_i},\ f_{i,j}(\mathbf{x}) = x_i\overline{x_j}$$

Note that

$$\sum_i f_i(\mathbf{x}) = \sum_i x_i\overline{x_i} = \langle \mathbf{x}, \mathbf{x}\rangle = 1$$

$$f_i(U_{i,j}\mathbf{x}) = f_i(x_1, \ldots, x_j, \ldots, x_i, \ldots, x_n) = x_j\overline{x_j} = f_j(\mathbf{x}) \tag{7}$$

$$f_{i,j}(V_i\mathbf{x}) = f_{i,j}(x_1, \ldots, -x_i, \ldots, x_n) = -x_i\overline{x_j} = -f_{i,j}(\mathbf{x}) \tag{8}$$

It follows by the invariance assumption that

$$\int_{\partial B_n} x_i\overline{x_i} d\mu = \int_{\partial B_n} f_i(\mathbf{x}) d\mu = \int_{\partial B_n} f_i(U_{i,j}\mathbf{x}) d\mu = \int_{\partial B_n} f_j(\mathbf{x}) d\mu = \int_{\partial B_n} x_j\overline{x_j} d\mu$$

and so

$$n\int_{\partial B_n} x_i\overline{x_i} d\mu = \int_{\partial B_n} \sum_j x_j\overline{x_j} d\mu = \int_{\partial B_n} 1 d\mu = 1$$

proving $\int_{\partial B_n} x_i\overline{x_i} d\omega = \frac{1}{n}$. Now assuming $i \neq j$, it follows from equation 8 that

$$\int_{\partial B_n} x_i\overline{x_j} d\mu = \int_{\partial B_n} f_{i,j}(\mathbf{x}) d\mu = \int_{\partial B_n} f_{i,j}(V_i\mathbf{x}) d\mu = \int_{\partial B_n} -f_{i,j}(\mathbf{x}) d\mu = -\int_{\partial B_n} x_i\overline{x_j} d\mu.$$

and hence $\int_{\partial B_n} x_i\overline{x_j} d\mu = 0$. Therefore,

$$\int_{\partial B_n} x_i\overline{x_j} d\omega = \frac{1}{n}\delta_{i,j}.$$

Theorem 1 follows from 2 easily since the unitary invariance of $d\omega$ simply means that $d\omega$ is invariant under every unitary transformation: $U : \partial B_n \to \partial B_n$, which includes $U_{i,j}$ and V_i. ■

Corollary 3 *The barycenter of the numerical range of a matrix lies in $W(A)$. Moreover $BW(A)$ is not an extreme point of $W(A)$ unless $W(A)$ is a single point.*

Proof. Since $W(A)$ is convex and $\lambda_i \in W(A)$, it follows that $BW(A) = \frac{1}{n}\sum_{i=1}^{n} \lambda_i \in W(A)$. Now suppose that $BW(A)$ is an extreme point of $W(A)$. Since $BW(A)$ is a non-degenerate convex linear combination of the eigenvalues, $BW(A)$ can be an extreme point only if there is a single eigenvalue. Since the numerical ranges of unitarily equivalent matrices are the same, we may assume $A = \lambda I + N$ where N is a nilpotent matrix and $BW(A) = \lambda$. If N is not zero, let $k > 1$ be the smallest integer such that $N^k = 0$. Choose a non-zero $\mathbf{y} \in N^{k-1}(\mathbb{C}^n)$ and $\mathbf{x} \in \mathbb{C}^n$ with $N\mathbf{x} = \mathbf{y}$, and $N\mathbf{y} = 0$. Applying the Gram-Schmidt process to the span of $\{\mathbf{y}, \mathbf{x}\}$ yields orthogonal unit vectors $\mathbf{x}_1, \mathbf{y}_1$ such that $N\mathbf{y}_1 = 0$ and $N\mathbf{x}_1 = a\mathbf{y}_1$ with $a > 0$. Therefore $\mathbf{x}_{\theta,\eta} = \cos(\theta)e^{i\eta}\mathbf{x}_1 + \sin(\theta)\mathbf{y}_1$ is a unit vector satisfying

$$\langle N\mathbf{x}_{\theta,\eta}, \mathbf{x}_{\theta,\eta}\rangle = a\cos(\theta)\sin(\theta)e^{i\eta}\langle \mathbf{y}_1, \mathbf{y}_1\rangle = \frac{a}{2}\sin(2\theta)e^{i\eta}$$

which implies $\lambda + \frac{a}{2}\sin(2\theta)e^{i\eta}$ lies in $W(A)$. As we vary θ and η we trace out a disc of radius $a/2$ in $W(A)$ with λ at the midpoint giving us a contradiction. Hence $N = 0$ and $A = \lambda I$. ■

Corollary 4 *The numerical range of a matrix A is a line segment if and only if the barycenter lies on the boundary.*

Proof. As noted in the introduction $W(A)$ has empty interior if and only if $\mathrm{tr}(A)/n \in \partial W(A)$. This is equivalent to the statement of the Corollary. ■

Remark 5 *If $n = 2$, $W(A)$ is a (possibly degenerate) ellipse with the eigenvalues at the foci [PT]. Thus the centroid and barycenter are obviously the same, although the distribution of $\langle A\mathbf{x}, \mathbf{x}\rangle$ on W(A) is not necessarily uniform (see the left panels of Figures 1 and 2). If $n = 3$, and A is a normal matrix with non-collinear eigenvalues $W(A)$ is a triangle with vertices at the eigenvalues. In this case the barycenter and the centroid are the same, and it can be shown that the distribution on $W(A)$ is uniform. For $n > 2$ there are lots of examples where the centroid and barcyenter are not the same. Since $W(A)$ is the convex hull of the eigenvalues if A is a normal matrix, simply chose A to be any matrix where one of the eigenvalues is a linear combination of the others but distinct form the centroid of $W(A)$. For example the 3×3 diagonal matrix with $0, 0,$and 1 as its diagonal elements or the 4×4 diagonal matrix with $-1, 0, 1, i$ as eigenvalues.*

Remark 6 *The invariance under the $U_{i,j}$ could easily be replaced by insisting that the means $\int_{\partial B_n} x_i\overline{x_i}d\mu$ are all equal by assuming, for example, that the x_i's are identically distributed.*

Remark 7 *The transformations $U_{i,j}$ and V_i generate a finite group whereas the invariance group of $d\omega$ is a compact Lie group of dimension n^2. Thus there are many more measures satisfying the hypotheses of Theorem 2 than Theorem 1.*

Acknowledgements The authors would to thank Ilya Spikovsky for his advice on preliminary results in this work and Panagiotis Psarrakos for some help on references.

References

[HJ] R.A. Horn and C.R. Johnson, *Topics in Matrix Analysis, Cambridge University Press, Cambridge*

[PT] P. Psarrakos and M. Tsatsomeros, *Numerical Range: (in) a Matrix Nutshell, Parts 1/2.* Mathematical Notes from Washington State University, Vol 45, No. 2, 2002 / Vol 46, No. 1, 2003.

[CKL] C.K Li, *Numerical Ranges and Numerical Radii*, http://www.math.wm.edu/~ckli/wonra.html.