

Mean Absolute Deviations about the Mean, the Cut Norm and Taxicab Correspondence Analysis

Vartan Choulakian¹, Ghassan Abou-Samra²

¹Department of Mathematics and Statistics, universit  de Moncton, Moncton, NB, Canada

²Faculty of Engineering, universit  de Moncton, Moncton, NB, Canada

Email: vartan.choulakian@umoncton.ca, ghassan.abou-samra@umoncton.ca

How to cite this paper: Choulakian, V. and Abou-Samra, G. (2020) Mean Absolute Deviations about the Mean, the Cut Norm and Taxicab Correspondence Analysis. *Open Journal of Statistics*, 10, 97-112. <https://doi.org/10.4236/ojs.2020.101008>

Received: January 4, 2020

Accepted: February 22 2020

Published: February 25, 2020

Copyright   2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Optimization has two faces, minimization of a loss function or maximization of a gain function. We show that the mean absolute deviation about the mean, d , maximizes a gain function based on the power set of the individuals; and nd , where n is the sample size, equals twice the value of the cut-norm of the deviations about the mean. This property is generalized to double-centered and triple-centered data sets. Furthermore, we show that among the three well known dispersion measures, standard deviation, least absolute deviation and d , d is the most robust based on the relative contribution criterion. More importantly, we show that the computation of each principal dimension of taxicab correspondence analysis (TCA) corresponds to balanced 2-blocks seriation. These ideas are applied on two data sets.

Keywords

Mean Absolute Deviations about the Mean, Cut Norm, Balanced 2-Blocks Seriation, Taxicab Correspondence Analysis

1. Introduction

Optimization has two faces, minimization of a loss function or maximization of a gain function. The following two well known dispersion measures, the variance (s^2) and mean absolute deviations about the median (LAD), are optimal because each minimizes a different loss function

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} \leq \frac{\sum_{i=1}^n (y_i - c)^2}{n} \quad (1)$$

and

$$LAD = \frac{\sum_{i=1}^n |y_i - median|}{n} \leq \frac{\sum_{i=1}^n |y_i - c|}{n}, \tag{2}$$

where y_1, y_2, \dots, y_n and c represent a sample of $(n + 1)$ values. To our knowledge, no optimality property is known for the mean absolute deviations about the mean defined by

$$d = \frac{\sum_{i=1}^n |y_i - \bar{y}|}{n}, \tag{3}$$

even though it has been studied in several papers for modeling purposes by, see among others, [1] [2] [3]. [1] [2] essentially compare the dispersion measures d and s in the statistical literature, with their preference clearly oriented towards d for its simple interpretability. While the authors in [3] compare the statistics d and LAD with the Gini dispersion measure and conclude that “The downside of using (d and LAD) is that robustness is achieved by omitting the information on the intra-group variability”.

The following inequality $LAD \leq d \leq s$ is well known and is a corollary to Lyapounov inequality, see for instance [4]: the first part $LAD \leq d$ follows from (2) and the second part $d \leq s$ follows from $ns^2 = \sum_{i=1}^n w_i^2 - n\bar{w}^2 \geq 0$, where $w_i = |y_i - \bar{y}|$.

d is the measure of dispersion used in taxicab correspondence analysis (TCA), an L_1 variant of correspondence analysis (CA), see [5]. An explanation for the robustness of d is the boundedness of the relative contribution of a point, see [6]-[11]. However, this paper provides further details on d , relating it to cut-norm and balanced 2-blocks seriation for double-centered data. [11] argued that often sparse contingency tables are better visualized by TCA; here, we present an analysis of a 0-1 affinity matrix, where TCA produces a much more interpretable map than CA, by comparing **Figure 3** and **Figure 4**. We see that repetition and experience play an indispensable and illuminating role in data analysis.

This paper is organized as follows: In Section 2, we show the optimality of the d , s^2 and LAD statistics based on maximizing gain functions, but d beats s^2 and LAD with respect to the property of relative contribution of a point (a robustness measure used in French data analysis circles based on geometry): this results from Lemma 1, which states the fact that for a centered vector nd equals twice its cut-norm; Sections 3 and 4 generalize the optimality result of the d to double-centered and triple-centered arrays; and we conclude in Section 5. Balanced 2-blocks seriation of a matrix with application to TCA is discussed in Section 3.

2. Optimality of d

We consider the centered vector $x = y - \bar{y}\mathbf{1}_n$, where $\mathbf{1}_n$ is composed of n ones. Let $I = \{1, 2, \dots, n\}$ and $I = S \cup \bar{S}$ a binary partition of I . We have

$$\sum_{i \in I} x_i = 0 = \sum_{i \in S} x_i + \sum_{i \in \bar{S}} x_i;$$

from which we deduce

$$\sum_{i \in S} x_i = -\sum_{i \in \bar{S}} x_i. \tag{4}$$

We define the cut-norm of a centered vector \mathbf{x} to be

$$\|\mathbf{x}\|_{\square} = \max_S \sum_{i \in S} x_i = \sum_{i \in S_{opt}} x_i = -\sum_{i \in \bar{S}_{opt}} x_i \text{ by}$$

where $S_{opt} = \{i : x_i \geq 0 \text{ for } i = 1, 2, \dots, n\}$. By casting the computation of d as a combinatorial maximization problem, we have the following main result describing the optimality of the d -statistic over all elements of the power set of I .

Lemma 1: (2-equal parts property) $nd = 2\|\mathbf{x}\|_{\square} \geq 2\sum_{i \in S} x_i$ for all $S \subset I$.

Proof:

$$\begin{aligned} nd &= \sum_{i=1}^n |x_i| = \sum_{i \in S_{opt}} x_i - \sum_{i \in \bar{S}_{opt}} x_i \\ &= 2\|\mathbf{x}\|_{\square} \text{ by (4)} \\ &\geq 2\sum_{i \in S} x_i \text{ for all } S \subset I. \end{aligned}$$

Corollary 1: $d \geq \mathbf{x}'\mathbf{u}/n$ for $\mathbf{u} \in \{-1, 1\}^n$.

Proof: By defining $u_{opt}(i) = 1$ if $i \in S_{opt}$ and $u_{opt}(i) = -1$ if $i \in \bar{S}_{opt}$, we get $d = \mathbf{x}'\mathbf{u}_{opt}/n \geq \mathbf{x}'\mathbf{u}/n$.

Corollary 2: $LAD \geq (\mathbf{y} - \text{median}\mathbf{1}_n)' \mathbf{u}/n$ for $\mathbf{u} \in \{-1, 1\}^n$.

Corollary 2 shows that LAD has a second optimality property. We emphasize the fact that the optimizing function in (2) is a univariate loss function of $c \in \mathbb{R}$; while the optimizing function in Corollary 2 is a multivariate gain function of $\mathbf{u} \in \{-1, 1\}^n$.

There is a similar result also for the variance in (1), based on Cauchy-Schwarz inequality stated in Lemma 2.

Lemma 2: $s = \left\| (\mathbf{y} - \bar{y}\mathbf{1}_n) / \sqrt{n} \right\|_2 \geq (\mathbf{y} - \bar{y}\mathbf{1}_n)' \mathbf{u} / \sqrt{n}$ for $\mathbf{u}'\mathbf{u} = 1$.

We note that Corollaries 1 and 2 and Lemma 2 represent particular cases of Hölder inequality, see [11].

Definition 1: We define the relative contribution of an element y_i to d , LAD and s^2 , respectively, to be

$$\begin{aligned} RC_d(y_i) &= \frac{|y_i - \bar{y}|}{nd}, \\ RC_{s^2}(y_i) &= \frac{|y_i - \bar{y}|^2}{ns^2}, \\ RC_{LAD}(y_i) &= \frac{|y_i - \text{median}|}{nLAD}. \end{aligned}$$

Then the following inequalities are true

$$\begin{aligned} 0 \leq RC_d(y_i) &\leq 0.5, \\ 0 \leq RC_{s^2}(y_i) &< 1, \\ 0 \leq RC_{LAD}(y_i) &\leq 1; \end{aligned}$$

from which we conclude that the most robust dispersion measure among the three dispersion measures, based on the relative contribution criterion, is d because it is bounded above by 0.5.

We note that the inequality, $0 \leq RC_{s^2}(y_i) < 1$, is a weaker variant of Laguerre-Samuelson inequality; see for instance [13], whose MS thesis presents nine different proofs.

We have

Definition 2: An element $x_i = y_i - \bar{y}$ is a heavyweight if $RC_d(y_i) = 0.5$; that is, $|x_i| = |y_i - \bar{y}| = nd/2$.

We note that a heavyweight element attains the upper bound of $RC_d(y_i)$, but it never attains the upper bound of $RC_{s^2}(y_i)$ and $RC_{LAD}(y_i)$.

3. 2-Way Interactions of a Correspondence Matrix

Let $P = (p_{ij})$ be a correspondence matrix; that is, $p_{ij} \geq 0$ for $i \in I$ and $j \in J = \{1, 2, \dots, m\}$ and $\sum_{j=1}^m \sum_{i=1}^n p_{ij} = 1$. As usual, we define $p_{i*} = \sum_{j=1}^m p_{ij}$ and $p_{*j} = \sum_{i=1}^n p_{ij}$. Let $P_1 = (x_{ij} = p_{ij} - p_{i*}p_{*j})$ for $i \in I$ and $j \in J$; then P_1 represents the residual matrix of P with respect to the independence model $(p_{i*}p_{*j})$. In the jargon of statistics, the cell x_{ij} represents the multiplicative 2-way interaction of the cell $(i, j) \in I \times J$. P_1 is double-centered

$$P_1 \mathbf{1}_m = \mathbf{0}_n \text{ and } P_1' \mathbf{1}_n = \mathbf{0}_m. \tag{5}$$

From (5) we get

$$\sum_{i \in S} x_{ij} = -\sum_{i \in \bar{S}} x_{ij} \text{ for } j \in J, \tag{6}$$

$$\sum_{j \in T} x_{ij} = -\sum_{j \in \bar{T}} x_{ij} \text{ for } i \in I, \tag{7}$$

for $T \subset J$. From (6) and (7), we get

$$\sum_{j \in T} \sum_{i \in S} x_{ij} = -\sum_{j \in T} \sum_{i \in \bar{S}} x_{ij} \tag{8}$$

$$= -\sum_{i \in S} \sum_{j \in \bar{T}} x_{ij} \tag{9}$$

$$= \sum_{i \in \bar{S}} \sum_{j \in \bar{T}} x_{ij}. \tag{10}$$

We define the cut-norm of P_1 to be

$$\|P_1\|_{\square} = \max_{S, T} \sum_{j \in T} \sum_{i \in S} x_{ij} = \sum_{j \in T_{opt}} \sum_{i \in S_{opt}} x_{ij}.$$

The cut-norm $\|P_1\|_{\square}$ is a well known quantity in theoretical computer science, because of its relationship to the famous Grothendieck inequality, which is based on $\|P_1\|_{\infty \rightarrow 1}$, see among others [14].

The matrix P_1 can be considered as the starting point in taxicab correspondence analysis, an L_1 variant of correspondence analysis, see [5]. The optimization criterion in TCA of P or P_1 is based on taxicab matrix norm, which is a

combinatorial optimization problem

$$\begin{aligned} \delta_1 &= \|P_1\|_{\infty \rightarrow 1} = \|P_1^{\uparrow}\|_{\infty \rightarrow 1} \\ &= \max_{u \in \mathbb{R}^m} \frac{\|P_1 u\|_1}{\|u\|_{\infty}} = \max_{v \in \mathbb{R}^n} \frac{\|P_1^{\downarrow} v\|_1}{\|v\|_{\infty}} = \max_{u \in \mathbb{R}^m, v \in \mathbb{R}^n} \frac{v' P_1 u}{\|u\|_{\infty} \|v\|_{\infty}}, \\ &= \max \|P_1 u\|_1 \text{ subject to } u \in \{-1, +1\}^m, \\ &= \max \|P_1^{\downarrow} v\|_1 \text{ subject to } v \in \{-1, +1\}^n, \\ &= \max v' P_1 u \text{ subject to } u \in \{-1, +1\}^m, v \in \{-1, +1\}^n, \tag{11} \\ &= v_1' P_1 u_1. \tag{12} \end{aligned}$$

In data analysis, the vectors v_1 and u_1 are interpreted as first taxicab principal axes and δ_1 as first taxicab dispersion. So we can compute the projection of the rows (resp. the columns) of P_1 on the first taxicab principal axis u_1 (resp. v_1) to be

$$a_1 = P_1 u_1 \tag{13}$$

$$b_1 = P_1^{\downarrow} v_1. \tag{14}$$

Equation (12) implies

$$v_1 = \text{sign}(a_1), \tag{15}$$

$$u_1 = \text{sign}(b_1), \tag{16}$$

named transition formulas, see [5] and [11]. We also note the following identities

$$\mathbf{1}'_n a_1 = 0 \text{ and } \delta_1 = \|a_1\|_1, \tag{17}$$

$$\mathbf{1}'_m b_1 = 0 \text{ and } \delta_1 = \|b_1\|_1. \tag{18}$$

Using the above results, we get the following

Lemma 3: (4-equal parts property) The norm

$$\delta_1 = \|P_1\|_{\infty \rightarrow 1} = 4 \|P_1\|_{\square} \geq 4 \sum_{j \in T} \sum_{i \in S} x_{ij}.$$

In data analysis, Lemma 3 implies balanced 2-blocks seriation of P_1 ; see example 1. The subsets T_{opt} and S_{opt} are positively associated and

$\|P_1\|_{\square} = \sum_{j \in T_{opt}} \sum_{i \in S_{opt}} x_{ij}$; similarly the subsets \bar{T}_{opt} and \bar{S}_{opt} are positively associated and

$\|P_1\|_{\square} = \sum_{j \in \bar{T}_{opt}} \sum_{i \in \bar{S}_{opt}} x_{ij}$. While the subsets \bar{T}_{opt} and S_{opt} are negatively associated and

$\|P_1\|_{\square} = -\sum_{j \in \bar{T}_{opt}} \sum_{i \in S_{opt}} x_{ij}$; similarly the subsets T_{opt} and \bar{S}_{opt} are negatively associated and

$\|P_1\|_{\square} = -\sum_{j \in T_{opt}} \sum_{i \in \bar{S}_{opt}} x_{ij}$. [15] presents an interesting overview of seriation and block seriation.

Using Definition 2, we get

Definition 3: The relative contribution of the row i to δ_1 (respectively of the column j to δ_1) is

$$RC_{\delta_1}(\text{row } i) = \frac{|a_1(i)|}{\delta_1} \text{ and } RC_{\delta_1}(\text{col } j) = \frac{|b_1(j)|}{\delta_1}.$$

We have

$$0 \leq RC_{\delta_1}(\text{row } i) \text{ and } RC_{\delta_1}(\text{col } j) \leq 0.5.$$

Definition 4: 1) On the first taxicab principal axis the row i is heavyweight if $RC_{\delta_1}(\text{row } i) = 0.5$, and the column j is heavyweight if $RC_{\delta_1}(\text{col } j) = 0.5$.

2) On the first taxicab principal axis the cell (i, j) is heavyweight if and only if both row i and column j are heavyweights; and in this case

$$RC_{\delta_1}(p_{ij} - p_{i*}p_{*j}) = \frac{|p_{ij} - p_{i*}p_{*j}|}{\delta_1} = 0.25.$$

For an application of Definitions 3 and 4 see [6].

Using Wedderburn's rank-1 reduction rule, see [11], we construct the 2nd residual matrix $\mathbf{P}_2 = \left(x_{ij} = p_{ij} - p_{i*}p_{*j} - \frac{a_1(i)b_1(j)}{\sigma_1} \right)$, which is also

double-centered, and repeat the above procedure. After $k = \text{rank}(\mathbf{P}_1)$ iterations, we decompose the correspondence matrix \mathbf{P} into $(k + 1)$ bilinear parts

$$p_{ij} = p_{i*}p_{*j} + \sum_{\alpha=1}^k \frac{a_{\alpha}(i)b_{\alpha}(j)}{\delta_{\alpha}},$$

named taxicab singular value decomposition; which can be rewritten, similar to data reconstruction formula in correspondence analysis (CA), as

$$p_{ij} = p_{i*}p_{*j} \left(1 + \sum_{\alpha=1}^k \frac{f_{\alpha}(i)g_{\alpha}(j)}{\delta_{\alpha}} \right),$$

where in TCA

$$f_{\alpha}(i) = a_{\alpha}(i)/p_{i*} \text{ and } g_{\alpha}(j) = b_{\alpha}(j)/p_{*j}. \tag{19}$$

We note that Equations (5) through (18) are valid for higher residual correspondence matrices \mathbf{P}_{α} for $\alpha = 1, \dots, k$.

CA and TCA satisfy an important invariance property: columns (or rows) with identical profiles (conditional probabilities) receive identical factor scores $g_{\alpha}(j)$ (or $f_{\alpha}(i)$). The factor scores are used in the graphical displays. Moreover, merging of identical profiles does not change the results of the data analysis: This is named the *principle of equivalent partitioning* by [16]; it includes the famous invariance property named *principle of distributional equivalence*, on which [17] developed CA.

In the next subsections we shall present two data sets, where taxicab correspondence analysis (TCA) is applied. The first data set is a small contingency table taken from [18], for which we present the details of the computation of the first two principal dimensions; in particular we highlight the balanced 2-blocks seriation of the residual data sets \mathbf{P}_{α} for $\alpha = 1, 2$ during the computation of each principal dimension. The second data set is a networks affinity matrix from [19] who applied CA to visually explore it; on this data set we compare CA and TCA maps highlighting the robustness of the TCA map to rare observations on the second principal dimension.

The theory of CA can be found, among others, in [17] [20]-[25]; the recent book, authored by [18], presents a panoramic review of CA and related methods.

3.1. Selikoff's Asbestos Data Set

Table 1, taken from [18], is a contingency table Y of size 5×4 cross-classifying 1117 New York workers with occupational exposure to asbestos; the workers are classified according to the number of exposure in years (five categories) and the asbestos grade diagnosed (four categories). **Figure 1** and **Figure 2** display the maps obtained by CA and TCA: almost no difference between them. Here, we present the details of the computation for TCA. **Table 2** presents the residual correspondence table P_1 with respect to the independence model, where we see diagonal 2-blocks seriation of P_1 with: $S_{opt} = \{\text{row1, row2}\}$ is positively associated with $T_{opt} = \{\text{column1}\}$ and the cut-norm $\|P_1\|_{\square} = (0.1181 + 0.0151) = 0.1332$; similarly, $\bar{S}_{opt} = \{\text{row3, row4, row5}\}$ is positively associated with $\bar{T}_{opt} = \{\text{column2, column3, column4}\}$ and $\|P_1\|_{\square} = \sum_{j \in \bar{T}_{opt}} \sum_{i \in \bar{S}_{opt}} x_{ij} = 0.0087 + \dots + 0.0202 = 0.1332$. Note that the elements in the positively associated diagonal blocks have in majority positive values; while the elements in the negatively associated diagonal blocks have in majority negative values. The last three columns and the last three rows of **Table 2** display principal axes (v_1 and u_1), coordinates of the projected points (a_1 and b_1) and coordinates of TCA factor scores (f_1 and g_1).

Table 3 shows the 2nd residual correspondence matrix P_2 , where we note that its first column is zero, because by Definition 4a column 1 is heavyweight in P_1 : $RC_{\delta_1}(G0) = 0.5$, see [6]. We see that columns (3 and 4) are positively associated with rows (1 and 5); similarly column 2 is positively associated with rows (2, 3 and 4). It is difficult to interpret the diagonal balanced 2-blocks seriation in **Table 3**; however, the map in **Figure 2** is interpretable, it shows a Guttman effect known as horseshoe or parabola.

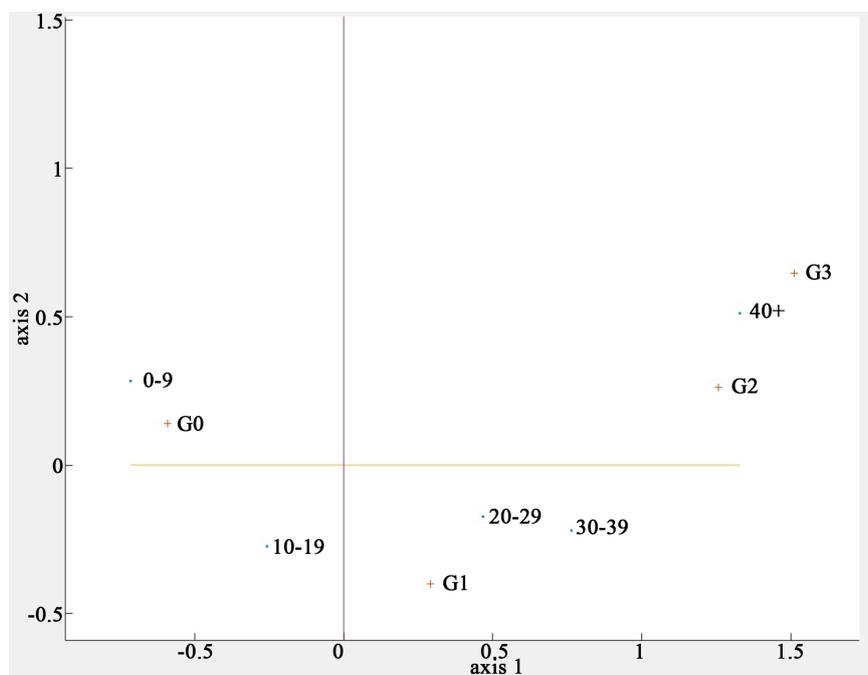


Figure 1. CA map of asbestos exposure data.

Table 1. Selikoff's Asbestos contingency table of size 5×4 .

Exposure in years	Asbestos Grade Diagnosed				Total	$p_{i\cdot}$
	None (G0)	Grade 1 (G1)	Grade 2 (G2)	Grade 3 (G3)		
0 - 9	310	36	0	0	346	0.3098
10 - 19	212	158	9	0	379	0.3393
20 - 29	21	35	17	4	77	0.0689
30 - 39	25	102	49	18	194	0.1737
40+	7	35	51	28	121	0.1083
Total	575	366	126	50	1117	1
$p_{\cdot j}$	0.5148	0.3277	0.1128	0.0448	1	

Table 2. Balanced 2-blocks seriation of $P_1 = (p_{ij} - p_{i\cdot}p_{\cdot j})$ of size 5×4 .

Exposure in years	Asbestos Grade Diagnosed				Total	v_1	a_1	f_1
	None (G0)	Grade 1 (G1)	Grade 2 (G2)	Grade 3 (G3)				
0 - 9	0.1181	-0.0693	-0.0349	-0.0139	0	-1	-0.2362	-0.7624
10 - 19	0.0151	0.0303	-0.0302	-0.0152	0	-1	-0.0303	-0.0892
20 - 29	-0.0167	0.0087	0.0074	0.0005	0	1	0.0334	0.4841
30 - 39	-0.0670	0.0344	0.0243	0.0083	0	1	0.1340	0.7718
40+	-0.0495	-0.0042	0.0334	0.0202	0	1	0.0990	0.9138
Total	0	0	0	0	0		$\delta_1 = 4 \times 0.1332$	
u_1	-1	1	1	1			$\ P_1\ _{\square} = 0.1332$	
b_1	-0.2664	0.0780	0.1302	0.0582			0.1332	$ -0.1332 $
g_1	-0.5175	0.2380	1.1553	1.2981			$ -0.1332 $	0.1332

Table 3. Balanced 2-blocks seriation of $P_2 = \left(p_{ij} - p_{i\cdot}p_{\cdot j} - \frac{a_{1i}b_{1j}}{\delta_1} \right)$ of size 5×4 .

Exposure in years	Asbestos Grade Diagnosed				Total	v_2	a_2	f_2
	None (G0)	Grade 1 (G1)	Grade 2 (G2)	Grade 3 (G3)				
0 - 9	0	-0.0347	0.0228	0.0119	0	1	0.0694	0.2241
40+	0	-0.0186	0.0092	0.0094	0	1	0.0372	0.3443
10 - 9	0	0.0347	-0.0228	-0.0119	0	-1	-0.0694	-0.2046
20 - 29	0	0.0038	-0.0007	-0.0031	0	-1	-0.0076	-0.1121
30 - 39	0	0.0148	-0.0085	-0.0063	0	-1	-0.0296	-0.1703
Total	0	0	0	0	0		$\delta_2 = 4 \times 0.0533$	
u_2	± 1	-1	1	1			$\ P_2\ _{\square} = 0.0533$	
b_2	0	-0.1066	0.0640	0.0426			$ -0.0533 $	0.0533
g_2	0	-0.3257	0.5681	0.9521			0.0533	$ -0.0533 $

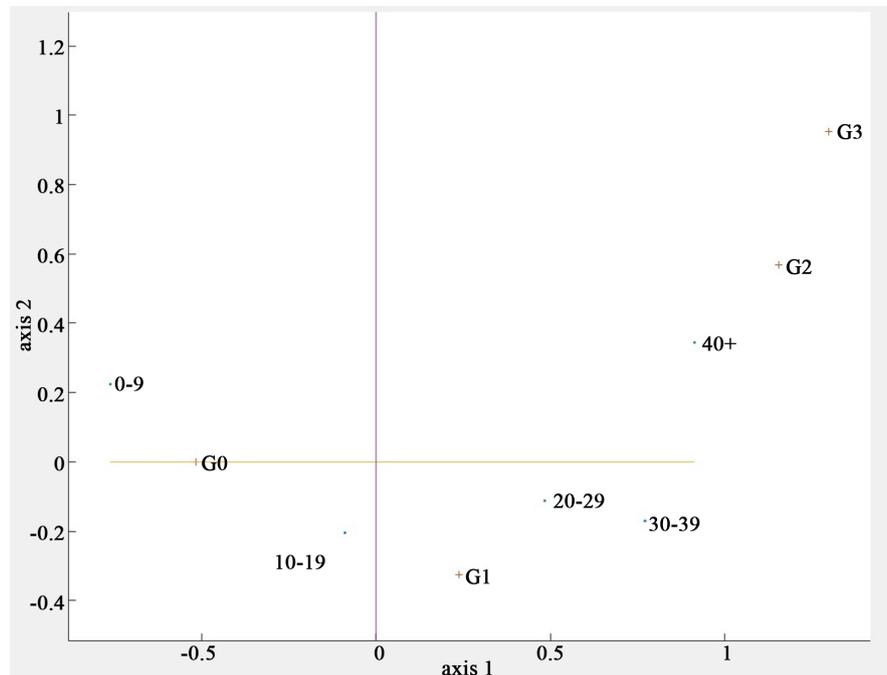


Figure 2. TCA map of asbestos exposure data.

3.2. Western Hemisphere Countries and Their Memberships in Trade and Treaty Organizations

Table 4 presents a two-mode affiliation network matrix $Z = (z_{ij})$ of size 22×15 taken from [19]. The 22 rows represent 22 countries and the 15 columns the regional trade and treaty organizations, described in Appendix A. The country i is a member of the organization j if $z_{ij} = 1$; and $z_{ij} = 0$ means the country i is not a member of the organization j . [19] visualized this data by correspondence analysis, see **Figure 3**, which is quite cluttered. She interpreted the first two principal dimensions by examining the factor scores of the countries and summarized the results in 3 points:

1) The first dimension contrasts South American countries and organizations on the one hand, and Central American countries and organizations on the other hand.

2) The second dimension clearly distinguishes Canada and the United States (both North American countries) along with NAFTA from other countries and organizations. In CA, the relative contribution of Canada (resp. US) to the second axis is $RC_{\sigma_2^2}(\text{Canada}) = RC_{\sigma_2^2}(\text{US}) = 0.409$, and $RC_{\sigma_2^2}(\text{NAFTA}) = 0.821$, where σ_2^2 is the variance, also named inertia, of the second principal dimension.

3) Organizations (SELA, OAS, and IDB) are in the center because they have membership profiles that are similar to the marginal profile: almost all countries belong to (SELA, OAS, and IDB), see **Table 4**.

Figure 4 provides the TCA map, which is much more interpretable than the corresponding CA map in **Figure 3**; where we see that, additionally to the three points mentioned by [19], the south american countries are divided into two

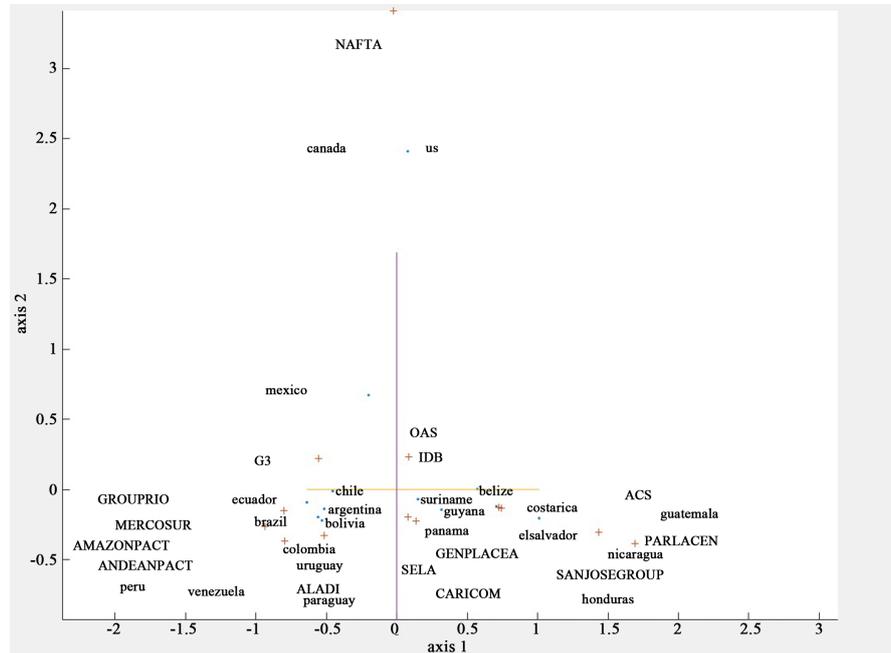


Figure 3. CA map of Western Hemisphere affinity network.

Table 4. Sociomatrix of American countries and their memberships.

Countries	Regional Trade and Treaty Organizations															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Sum
Argentina	0	1	0	0	0	1	1	0	1	1	0	1	0	0	1	7
Belize	1	0	0	0	1	0	0	0	1	0	0	1	0	0	1	5
Bolivia	0	1	1	1	0	1	1	0	1	0	0	1	0	0	1	8
Brazil	0	1	1	0	0	1	1	0	1	1	0	1	0	0	1	8
Canada	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	3
Chile	0	1	0	0	0	0	1	0	1	0	0	1	0	0	1	5
Colombia	1	1	1	1	0	1	1	1	1	0	0	1	0	0	1	10
Costa Rica	1	0	0	0	0	1	0	0	1	0	0	1	0	1	1	6
Ecuador	0	1	1	1	0	1	1	0	1	0	0	1	0	0	1	8
El Salvador	1	0	0	0	0	1	0	0	1	0	0	1	1	1	1	7
Guatemala	1	0	0	0	0	1	0	0	1	0	0	1	1	1	1	7
Guyana	1	0	1	0	1	1	0	0	1	0	0	1	0	0	1	7
Honduras	1	0	0	0	0	1	0	0	1	0	0	1	1	1	1	7
Mexico	1	1	0	0	0	1	1	1	1	0	1	1	0	0	1	9
Nicaragua	1	0	0	0	0	1	0	0	1	0	0	1	0	1	1	6
Panama	1	0	0	0	0	1	0	0	1	0	0	1	0	1	1	6
Paraguay	0	1	0	0	0	0	1	0	1	1	0	1	0	0	1	6
Peru	0	1	1	1	0	1	1	0	1	0	0	1	0	0	1	8
Suriname	1	0	1	0	0	0	0	0	1	0	0	1	0	0	1	5
United States	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	3
Uruguay	0	1	0	0	0	1	1	0	1	1	0	1	0	0	1	7
Venezuela	1	1	1	1	0	1	1	1	1	0	0	1	0	0	1	10
Sum	12	11	8	5	2	16	11	3	22	5	3	22	3	6	20	149

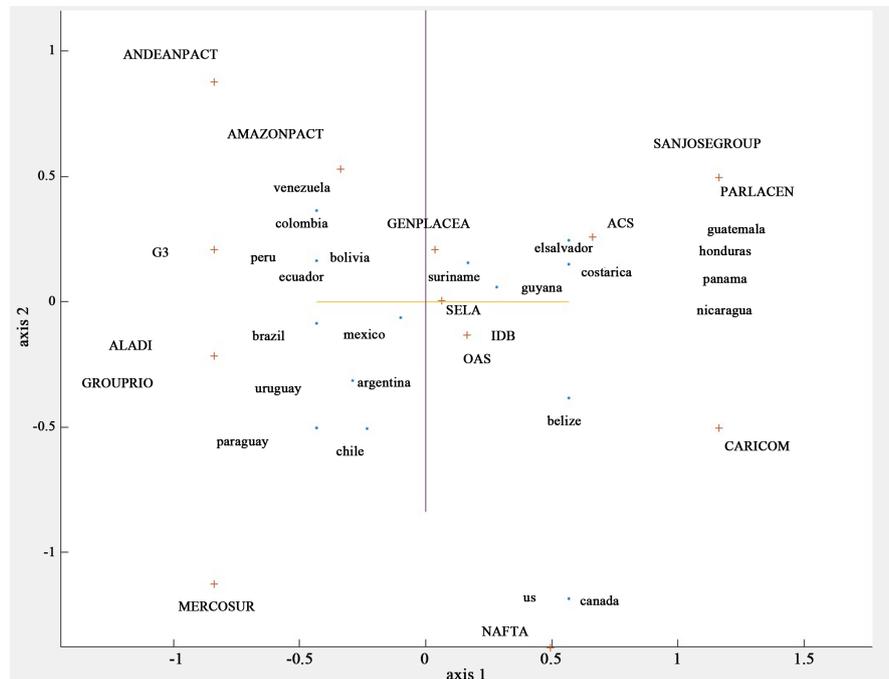


Figure 4. TCA map of Western Hemisphere affinity network.

groups, northern (Venezuela, Bolivia, Peru and Ecuador) and southern countries (Brazil, Uruguay, Argentina, Paraguay and Chile). Furthermore, the contributions of the points Canada, the United States, and NAFTA to the second axis are not substantial compared to CA: $RC_{\delta_2}(\text{Canada}) = RC_{\delta_2}(\text{US}) = 0.088$, and $RC_{\delta_2}(\text{NAFTA}) = 0.10$. This shows the robustness of TCA due to the robustness of the δ statistic following Definition 1.

It is well known that, CA is very sensitive to some particularities of a data set; further, how to identify and handle these is an open unresolved problem. However, for contingency tables [12] enumerated three under the umbrella of sparse contingency tables: rare observations, zero-block structure and relatively high-valued cells. It is evident that this data set has specially three rare observations (NAFTA, CANADA and USA), which determine the 2nd dimension of CA. A row or a column category is considered rare, if its marginal probability is quite small.

3.3. Maximal Interaction Two-Mode Clustering of Continuous Data

[26] discussed maximum interaction two-mode clustering of continuous data. By generalizing their objective function, we want to show that the results of this section can be considered a particular robust L_1 variant of their approach. Let $Y = (y_{ij})$ be a 2-way array for $i \in I, j \in J$. As usual, we define, for instance, $\bar{y}_{*j} = \sum_{i=1}^n \frac{y_{ij}}{n}$ and $\bar{y}_{**} = \sum_{j=1}^m \sum_{i=1}^n \frac{y_{ij}}{mn}$. Let $X = (x_{ij})$ be the additive double-centered array, where

$$x_{ij} = y_{ij} - \bar{y}_{i*} - \bar{y}_{*j} + \bar{y}_{**}.$$

In the jargon of statistics, the cell x_{ij} represents the additive 2-way interaction of the cell $(i, j) \in I \times J$. The matrix \mathbf{X} is double-centered, and it satisfies Equations (6) through (10). Let $I = U_{\alpha=1}^r S_\alpha$ be an r -partition of I and $J = U_{\beta=1}^c T_\beta$ be a c -partition of J . We consider the following maximization of the overall interaction problem for $p \geq 1$

$$f_p(S_\alpha, T_\beta : \alpha = 1, \dots, r \text{ and } \beta = 1, \dots, c) = \sum_{\alpha=1}^r \sum_{\beta=1}^c |S_\alpha| |T_\beta| g_p(\alpha, \beta),$$

where $|S_\alpha|$ is the cardinality of the set S_α and

$$g_p(\alpha, \beta) = \left(\frac{\sum_{i \in S_\alpha} \sum_{j \in T_\beta} x_{ij}}{|S_\alpha| |T_\beta|} \right)^p.$$

When $p = 2$, then maximizing $f_2(S_\alpha, T_\beta : \alpha = 1, \dots, r \text{ and } \beta = 1, \dots, c)$, named maximal overall interaction, is the criterion computed in [26]. When $p = 1, r = c = 2$, then maximizing $f_1(S_\alpha, T_\beta : \alpha = 1, 2 \text{ and } \beta = 1, 2) = \|\mathbf{X}\|_{\infty \rightarrow 1} = 4\|\mathbf{X}\|_{\square}$ by Lemma 3, which is the criterion computed in TCA.

4. Triple-Centered Arrays

To motivate our subject, we start with an example. Let $\mathbf{Y} = (y_{ijk})$ be a 3-way array for $i \in I, j \in J$ and $k \in K = \{1, 2, \dots, t\}$. As usual, we define, for instance,

$$\bar{y}_{ij*} = \sum_{k=1}^t y_{ijk} / t, \quad \bar{y}_{*j*} = \sum_{k=1}^t \sum_{i=1}^n \frac{y_{ijk}}{tn} \quad \text{and} \quad \bar{y}_{***} = \sum_{k=1}^t \sum_{j=1}^m \sum_{i=1}^n \frac{y_{ijk}}{tmn}.$$

Let $\mathbf{X} = (x_{ijk})$ be the triple-centered array, where

$$x_{ijk} = y_{ijk} - \bar{y}_{ij*} - \bar{y}_{i*k} - \bar{y}_{*jk} + \bar{y}_{i**} + \bar{y}_{*j*} + \bar{y}_{**k} - \bar{y}_{***}.$$

In the jargon of statistics, the cell x_{ijk} represents the additive 3-way interaction of the cell $(i, j, k) \in I \times J \times K$. The tensor \mathbf{X} is triple-centered; that is,

$$\sum_{k=1}^t x_{ijk} = \sum_{j=1}^m x_{ijk} = \sum_{i=1}^n x_{ijk} = 0.$$

A generalization of Lemma 3 is

Lemma 4: (8-equal parts property) The tensor norm

$$\begin{aligned} \|\mathbf{X}\|_{(\infty, \infty) \rightarrow 1} &= \max \sum_{k \in K} \sum_{j \in J} \sum_{i \in I} w_k v_j u_i x_{ijk} \quad \text{subject to } \mathbf{u} \times \mathbf{v} \times \mathbf{w} \in \{-1, +1\}^{m \times n \times t} \\ &= 8 \sum_{k \in W_{opt}} \sum_{j \in T_{opt}} \sum_{i \in S_{opt}} x_{ijk} \geq 8 \sum_{k \in W} \sum_{j \in T} \sum_{i \in S} x_{ijk}, \end{aligned}$$

where $W \subset K$.

The proof is similar to the proof of Lemma 3.

Lemma 4 can easily be generalized to higher-way arrays.

5. Conclusions

This essay is an attempt to emphasize the following two points.

First, we showed the optimality and robustness of the mean absolute deviations about the mean, its interpretation, and its generalization to higher-way ar-

rays. A key notion in describing its robustness is that the relative contribution of a point is bounded by 50%.

Second, within the framework of TCA, we showed that the following three identities $\delta_1 = \|\mathcal{P}_1\|_{\infty \rightarrow 1} = 4\|\mathcal{P}_1\|_{\square}$ reveal three different but related aspects of TCA: 1) δ_1 , computed in (17) and (18), by (19) represents the mean absolute deviations about the mean statistic; 2) The taxicab norm $\|\mathcal{P}_1\|_{\infty \rightarrow 1}$, via (15) and (16), shows that uniform weights are affected to the columns and the rows; 3) The cut norm $4\|\mathcal{P}_1\|_{\square}$ shows that the computation of each principal dimension of TCA corresponds to balanced 2-blocks seriation, with equality of the cut norm in the 4 associated blocks.

A list of the principal used variables is provided in Appendix B.

Acknowledgements

We thank the Editor and the referee for their comments. Research of V. Choulakian is funded by the National Science and Engineering Research Council of Canada grant RGPIN-2017-05092. This support is greatly appreciated. The authors thank William Alexander Digout for help in computations.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Pham-Gia, T. and Hung, T.L. (2001) The Mean and Median Absolute Deviations. *Mathematical and Computer Modelling*, **34**, 921-936. [https://doi.org/10.1016/S0895-7177\(01\)00109-1](https://doi.org/10.1016/S0895-7177(01)00109-1)
- [2] Gorard, S. (2015) Introducing the Mean Absolute Deviation ‘Effect’ Size. *International Journal of Research & Method in Education*, **38**, 105-114. <https://doi.org/10.1080/1743727X.2014.920810>
- [3] Yitzhaki, S. and Lambert, P.J. (2013) The Relationship between the Absolute Deviation from a Quantile and Gini’s Mean Difference. *Metron*, **71**, 97-104. <https://doi.org/10.1007/s40300-013-0015-y>
- [4] Rothagi, V.K. (1976) An Introduction to Probability Theory and Mathematical Statistics. John Wiley and Sons, New York.
- [5] Choulakian, V. (2006) Taxicab Correspondence Analysis. *Psychometrika*, **71**, 333-345. <https://doi.org/10.1007/s11336-004-1231-4>
- [6] Choulakian, V. (2008) Taxicab Correspondence Analysis of Contingency Tables with One Heavyweight Column. *Psychometrika*, **73**, 309-319. <https://doi.org/10.1007/s11336-007-9041-0>
- [7] Choulakian, V. (2008) Multiple Taxicab Correspondence Analysis. *Advances in Data Analysis and Classification*, **2**, 177-206. <https://doi.org/10.1007/s11634-008-0023-6>
- [8] Choulakian, V. and de Tibeiro, J. (2013) Graph Partitioning by Correspondence Analysis and Taxicab Correspondence Analysis. *Journal of Classification*, **30**, 397-427. <https://doi.org/10.1007/s00357-013-9145-4>

- [9] Choulakian, V., Allard, J. and Simonetti, B. (2013) Multiple Taxicab Correspondence Analysis of a Survey Related to Health Services. *Journal of Data Science*, **11**, 205-229.
- [10] Choulakian, V., Simonetti, B. and Gia, T.P. (2014) Some Further Aspects of Taxicab Correspondence Analysis. *Statistical Methods and Applications*, **23**, 401-416. <https://doi.org/10.1007/s10260-014-0259-6>
- [11] Choulakian, V. (2016) Matrix Factorizations Based on Induced Norms. *Statistics, Optimization and Information Computing*, **4**, 1-14. <https://doi.org/10.19139/soic.v4i1.160>
- [12] Choulakian, V. (2017) Taxicab Correspondence Analysis of Sparse Two-Way Contingency Tables. *Statistica Applicata-Italian Journal of Applied Statistics*, **29**, 153-179.
- [13] Jensen, S.T. (1999) The Laguerre-Samuelson Inequality with Extensions and Applications in Statistics and Matrix Theory. McGill University, Quebec. https://doi.org/10.1007/978-94-011-4577-0_10
- [14] Khot, S. and Naor, A. (2012) Grothendieck-Type Inequalities in Combinatorial Optimization. *Communications on Pure and Applied Mathematics*, **65**, 992-1035. <https://doi.org/10.1002/cpa.21398>
- [15] Liiv, I. (2010) Seriation and Matrix Reordering Methods: An Historical Overview. *Statistical Analysis and Data Mining*, **3**, 70-91. <https://doi.org/10.1002/sam.10071>
- [16] Nishisato, S. (1984) Forced Classification: A Simple Application of a Quantification Method. *Psychometrika*, **49**, 25-36. <https://doi.org/10.1007/BF02294203>
- [17] Benzécri, J.P. (1973) L'Analyse des Données: Vol. 2: L'Analyse des Correspondances. Dunod, Paris.
- [18] Beh, E. and Lombardo, R. (2014) Correspondence Analysis: Theory, Practice and New Strategies. Wiley, New York. <https://doi.org/10.1002/9781118762875>
- [19] Faust, K. (2005) Using Correspondence Analysis for Joint Displays of Affiliation Networks. In: Carrington, P.J., Scott, J. and Wasserman, S., Eds., *Models and Methods in Social Network Analysis*, Cambridge University Press, Cambridge, 117-147. <https://doi.org/10.1017/CBO9780511811395.007>
- [20] Benzécri, J.P. (1992) Correspondence Analysis Handbook. Marcel Dekker, New York. <https://doi.org/10.1201/9780585363035>
- [21] Greenacre, M. (1984) Theory and Applications of Correspondence Analysis. Academic Press, London.
- [22] Gifi, A. (1990) Nonlinear Multivariate Analysis. Wiley, New York.
- [23] Le Roux, B. and Rouanet, H. (2004) Geometric Data Analysis. From Correspondence Analysis to Structured Data Analysis. Kluwer-Springer, Dordrecht. <https://doi.org/10.1007/1-4020-2236-0>
- [24] Murtagh, F. (2005) Correspondence Analysis and Data Coding with Java and R. Chapman & Hall/CRC, Boca Raton, FL. <https://doi.org/10.1201/9781420034943>
- [25] Nishisato, S. (2007) Mutidimensional Nonlinear Descriptive Analysis. Chapman & Hall/CRC, Boca Raton, FL. <https://doi.org/10.1201/9781420011203>
- [26] Schepers, J., Bock, H.-H. and Van Mechelen, I. (2017) Maximal Interaction Two-Mode Clustering. *Journal of Classification*, **34**, 49-75. <https://doi.org/10.1007/s00357-017-9226-x>

Appendix A: List of Western Hemisphere Organizations

- 1) Association of Caribbean States (ACS): Trade group sponsored by the Caribbean Community and Common Market (CARICOM).
- 2) Latin American Integration Association (ALADI): Free trade organization.
- 3) Amazon Pact: Promotes development of Amazonian territories.
- 4) Andean Pact: Promotes development of members through economic and social integration.
- 5) Caribbean Community and Common Market (CARICOM): Caribbean trade organization; promotes economic development of members.
- 6) Group of Latin American and Caribbean Sugar Exporting Countries (GEPLACEA): Sugar-producing and exporting countries.
- 7) Group of Rio: Organization for joint political action.
- 8) Group of Three (G-3): Trade organization.
- 9) Inter-American Development Bank (IDB): Promotes development of member nations.
- 10) South American Common Market (MERCOSUR): Increases economic cooperation in the region.
- 11) North American Free Trade Agreement (NAFTA): Free trade organization.
- 12) Organization of American States (OAS): Promotes peace, security, economic, and social development in the Western Hemisphere.
- 13) Central American Parliament (PARLACÉN): Works for the political integration of Central America.
- 14) San José Group: Promotes regional economic integration.
- 15) Latin American Economic System (SELA): Promotes economic and social development of member nations.

Appendix B: A List of principal used variables

Mean absolute deviations about the mean of a sample $d = \frac{\sum_{i=1}^n |y_i - \bar{y}|}{n}$

Mean absolute deviations of a sample about the median

$$LAD = \frac{\sum_{i=1}^n |y_i - median|}{n}$$

Variance of a sample $s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}$

Cut norm of a centered sample $\|\mathbf{y} - \bar{y}\mathbf{1}_n\|_{\square} = \max_S \sum_{i \in S} (y_i - \bar{y})$, where $S \subset I = \{1, 2, \dots, n\}$

Taxicab operator norm of a double centered matrix

$$\delta_{\alpha} = \|\mathbf{P}_{\alpha}\|_{\infty \rightarrow 1} = \max_{\mathbf{u} \in \mathbb{R}^m} \frac{\|\mathbf{P}_{\alpha}\mathbf{u}\|_1}{\|\mathbf{u}\|_{\infty}}$$

Cut norm of a double centered matrix $\|\mathbf{P}_{\alpha}\|_{\square} = \max_{S,T} \sum_{j \in T} \sum_{i \in S} P_{\alpha}(i, j)$, where $T \subset J = \{1, 2, \dots, m\}$

δ_{α} is the dispersion value of α th taxicab principal axis

$f_{\alpha}(i)$ is taxicab principal factor score of row i on the α th principal axis and

$$\delta_\alpha = \sum_{i=1}^n p_{i*} |f_\alpha(i)|$$

$g_\alpha(j)$ is taxicab principal factor score of column j on α th principal axis and

$$\delta_\alpha = \sum_{j=1}^m p_{*j} |g_\alpha(j)|$$