Effective Sample Size in Spatial Modeling

Vallejos, Ronny
Universidad Técnica Federico Santa María, Department of Mathematics
Avenida España 1680
Valparaíso 2340000, Chile
E-mail: ronny.vallejos@usm.cl

Moreno, Consuelo
Universidad Técnica Federico Santa María, Department of Mathematics
Avenida España 1680
Valparaíso 2340000, Chile
E-mail: consuelo.moreno@alumnos.usm.cl

INTRODUCTION

Approaches to spatial analysis have developed considerably in recent decades. In particular, the problem of determining sample size has been studied in many different contexts. In spatial statistics, it is well known that as spatial autocorrelation latent in geo-referenced data increases, the amount of duplicated information contained in these data also increases. This property has many implications for the posterior analysis of spatial data. For example, Clifford, Richardson and Hémon (1989) used the notion of effective degrees of freedom to denote the equivalent number of degrees of freedom for spatially independent observations. Similarly, Cressie (1993, p. 14-15) illustrated the effect of spatial correlation on the variance of the sample mean using an AR(1) correlation structure with spatial data. As a result, the new sample size (the effective sample size) could be interpreted as the equivalent number of independent observations.

This paper addresses the following problem: if we have n data points, what is the effective sample size (ESS) associated with these points? If the observations are independent and a regional mean is being estimated, given a suitable definition, the answer is n. Intuitively speaking, when perfect positive spatial autocorrelation prevails, ESS $= 1$. With dependence, the answer should be something less than n. Getis and Ord (2000) studied this kind of reduction of information in the context of multiple testing of local indices of spatial autocorrelation. Note that the general approach to addressing the question above does not depend on the data values. However, it does depend on the spatial locations of the points on the range for the spatial process. It also depends on the spatial dimension. In this article, we suggest a definition of effective spatial sample size. Our definition can be explored analytically given certain special assumptions. We conduct this sort of exploration for patterned correlation matrices that commonly arise in spatial statistics (considering a single mean process with intra-class correlation, a single mean process with an AR(1) correlation structure, and CAR and SAR processes). Theoretical results and examples are presented that illustrate the features of the proposed measure for effective sample size. Finally, we outline some strands of research to be addressed in future studies.

PRELIMINARIES AND NOTATION

Consider a set of n locations in an r-dimensional space, e.g., $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n \in D \subset \mathbb{R}^r$, such that the covariance matrix of the variables $Y(\boldsymbol{s}_1)$, $Y(\boldsymbol{s}_2)$, $\cdots, Y(\boldsymbol{s}_n)$ is Σ. The effective sample size can be characterized by the correlation matrix $R_n = (\sigma_{ij})/(\sigma_{ii}\sigma_{jj}) = A^{-1}\Sigma A^{-1}$, where $A = diag(\sigma_{11}^{1/2}, \sigma_{22}^{1/2}, \cdots, \sigma_{nn}^{1/2})$. For example, there are many reductions of R_n to a single number and many appropriate but arbitrary transformations of that number to the interval $[1, n]$. Our goal is to find a function ESS $=$ ESS(n, R_n, r) that satisfies $1 \leq$ ESS $\leq n$.

For the case in which $A = I$, one illustrative reduction is provided by the Kullback-Leibler distance from $N(\mu\mathbf{1}, R)$ to $N(\mu\mathbf{1}, I)$ where $\mathbf{1}$ is a n-dimensional column vector of ones. Straightforward calculations indicate that KL $= \frac{1}{2}\left(log|R| + tr(R^{-1} - I)\right)$. For an isotropic spatial process with spatial variance σ^2 and an exponential correlation function $\rho(\boldsymbol{s}_i - \boldsymbol{s}_j) = \exp(-\phi||\boldsymbol{s}_i - \boldsymbol{s}_j||)$, $\phi > 0$, KL needs to be inversely scaled to $[1, n]$ and decreases in ϕ. Another way to avoid making an arbitrary choice of transformation is to use the relative efficiency of $\overline{Y}$, the sample mean, to estimate the constant mean μ under the process compared with estimating μ under independence. Scaling by n readily indicates this quantity to be

(1) $\quad n^2(\mathbf{1}^t R_n \mathbf{1})^{-1}.$

At $\phi = 0$, the expression (1) equals 1, and as ϕ increases to ∞, (1) increases to n. (1) is attractive in that it assumes no distributional model for the process. The existence of $Var(\overline{Y})$ is implied by the assumption of an isotropic correlation function. A negative feature of this process, however, is that for a fixed ϕ, the effective sample size need not increase in n.

Creating an alternative to the previous suggestions regarding effective sample size, Griffith (2005) suggested a measure of the size of a geographic sample based on a model with a constant mean given by $\boldsymbol{Y}(\boldsymbol{s}) = \mu\mathbf{1} + \boldsymbol{e}(\boldsymbol{s}) = \mu\mathbf{1} + \Sigma^{-1/2}\boldsymbol{e}^*(\boldsymbol{s})$, where $\boldsymbol{Y}(\boldsymbol{s}) = (Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_n))$, $\boldsymbol{e}(\boldsymbol{s})$ and $\boldsymbol{e}^*(\boldsymbol{s})$, respectively, denote $n \times 1$ vectors of spatially autocorrelated and unautocorrelated errors such that $Var(\boldsymbol{e}(\boldsymbol{s})) = \sigma^2_{e^*}\Sigma^{-1}$ and $Var(\boldsymbol{e}^*(s)) = \sigma^2_{e^*}I_n$. This measure is

(2) $\quad n^* = \mathrm{tr}(\Sigma^{-1})n/(\mathbf{1}^t\Sigma^{-1}\mathbf{1}),$

where tr denotes the trace operator. Later, Griffith (2008) used this measure (2) with soil samples collected from across Syracuse, NY.

In another alternative reduction, one can compare the reciprocal of the variance of the BLUE unbiased estimator of μ under R_n, which is readily shown to be $\mathbf{1}^t R_n^{-1}\mathbf{1}$. As ϕ increases to ∞, this quantity increases to n. Again, no distributional model for the process is assumed. However, this expression does arises as the Fisher information about μ under normality. In fact, for $\boldsymbol{Y}(\boldsymbol{s}) \sim N(\mu\mathbf{1}, \sigma^2 R_n)$, $I(\mu) = \mathbf{1}^t R_n^{-1}\mathbf{1}/\sigma^2$, yielding the following definition.

Definition 1. Let $\boldsymbol{Y}(\boldsymbol{s})$ be a $n \times 1$ random vector with expected value $\mu\mathbf{1}$ and correlation matrix R_n. The quantity

(3) $\quad \mathrm{ESS} = \mathrm{ESS}(n, R_n, r) = \mathbf{1}^t R_n^{-1}\mathbf{1}$

is called the effective sample size.

Remark 1. Hining (1990, p. 163) pointed out that spatial dependency implies a loss of information in the estimation of the mean. One way to quantify that loss is through (3). Moreover, the asymptotic variance of the generalized least squares estimator of μ is 1/ESS.

Remark 2. The definition of effective sample size can be generalized if we consider the Fisher information for other multivariate distributions. In fact, consider a spatial elliptical random vector $\boldsymbol{Y}(\boldsymbol{s})$ with density function

$$f_{\boldsymbol{Y}(\boldsymbol{s})} = \frac{c_n}{|\Sigma|^{\frac{1}{2}}} g_n((\boldsymbol{Y}(\boldsymbol{s}) - \mu\mathbf{1})^t\Sigma^{-1}(\boldsymbol{Y}(\boldsymbol{s}) - \mu\mathbf{1})),$$

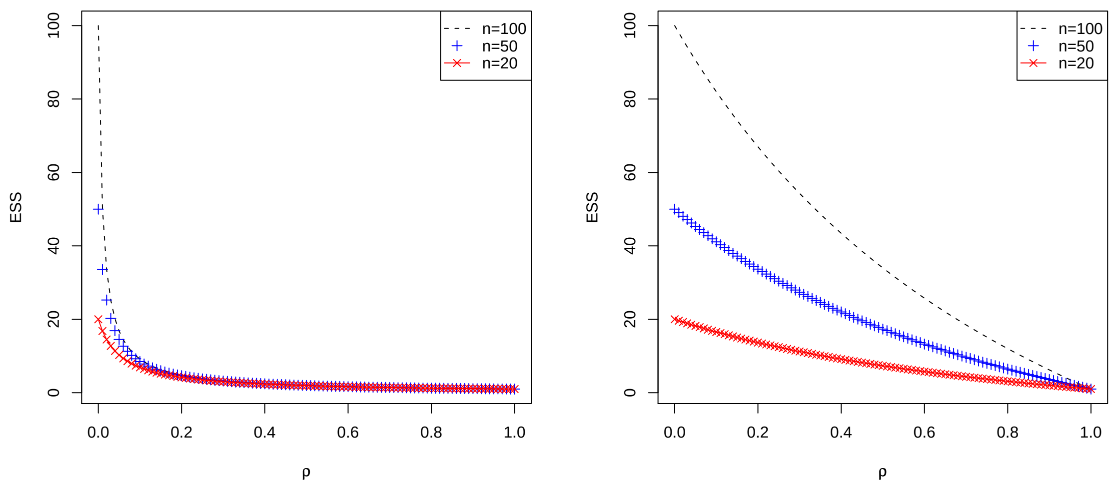
where $\mu\mathbf{1}$ and Σ are the location and scale parameters, g_n is a positive function, and c_n is a normalizing constant. If the generating function $g_n(u) = \exp(-|u|), u \in \mathbb{R}$ the distribution is known as the Laplace distribution, and when Σ is known, the Fisher information is

$$I(\mu) = E\left[4(\boldsymbol{Y}(\boldsymbol{s})^t\Sigma^{-1}\mathbf{1} - \mu\mathbf{1}^t\Sigma^{-1}\mathbf{1})^2\right] = 4\mathbf{1}^t\Sigma^{-1}\mathbf{1}.$$

Remark 3. If the n observations are independent and $R_n = I$, then ESS $= n$. If perfect positive spatial correlation prevails, then $R_n = \mathbf{1}\mathbf{1}^t$. Thus, $rank(R_n) = 1$, and ESS $= 1$.

Example 1. Let us consider the intra-class correlation structure with $\boldsymbol{Y}(\boldsymbol{s}) \sim (\mu\mathbf{1}, R_n)$, where $R_n = (1-\rho)\boldsymbol{I}+\rho\boldsymbol{J}$, $\boldsymbol{J}$ is an $n\times n$ unit matrix, and $-1/(n-1) < \rho < 1$. Then $\text{ESS}_I = n/(1+(n-1)\rho)$. Notice from Figure 1 (a) that the reduction that takes place in this case is quite severe. For example, for $n = 100$ and $\rho = 0.1$, ESS $= 9.17$, and for $n = 100$ and $\rho = 0.5$, ESS $= 1.98$. In general, such noticeable reductions in sample size are not expected. However, the intra-class correlation does not take into account the spatial association between the observations. With more rich correlation structures, the effective sample size is better at reducing the information from R.
Figure 1.

(a) ESS for the intra-class correlation; (b) ESS for the a Toeplitz correlation.



(a) (b)

Example 2. In the case of a Toeplitz correlation matrix, consider the vector $\boldsymbol{Y}(\boldsymbol{s})$ with the mean $\mu\mathbf{1}$ and correlation matrix R_n, where for $|\rho| < 1$, (see Graybill, 2004)

$$R_n = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix}, \quad R_n^{-1} = \begin{cases} 1/(1-\rho^2), & \text{if } i = j = 1, n, \\ (1+\rho^2)/(1-\rho^2), & \text{if } i = j = 2, \cdots, n-1, \\ -\rho/(1-\rho^2), & \text{if } |j-i| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$\text{ESS}_T = (2 + (n-2)(1-\rho))/(1+\rho).$$

Furthermore, straightforward calculations show that for $0 < \rho < 1$ and $n > 2$,

$$\text{EES}_I < \text{ESS}_T.$$

Hence, the reduction in R under the Toeplitz structure is not as severe as in the intra-class correlation case. Based on Figure 1 (a) and (b), we see that in both cases, ESS is decreasing in ρ.

SOME RESULTS

Proposition 1. Let $s_1, s_2, \ldots, s_n$ be n locations in $D \subset \mathbb{R}^r$, with r fixed. Consider a random spatial vector $\boldsymbol{Y}(\boldsymbol{s}) = (Y(\boldsymbol{s}_1),\ Y(\boldsymbol{s}_2),\ \cdots, Y(\boldsymbol{s}_n))^t$ with the expected value $\mu\mathbf{1}_n$ and correlation matrix R_n. ESS is increasing in n for a fixed value of r.

Proposition 2. Under the same conditions as in Proposition 1, $1 \leq \text{ESS} \leq n$.

Now, consider a CAR model of the form

$$(4) \quad Y(\boldsymbol{s}_i) \mid Y(\boldsymbol{s}_j), j \neq i \sim N(\mu_i + \rho \sum_j b_{ij} Y(\boldsymbol{s}_j), \tau_i^2), i = 1, 2, \cdots, n.$$

where ρ determines the direction and magnitude of the spatial neighborhood effect, b_{ij} are weights that determine the relative influence of location j on location i, and τ_i^2 is the conditional variance. If n is finite, we form the matrices $B = (b_{ij})$ and $D = diag(\tau_1^2, \tau_2^2, \cdots, \tau_n^2)$. According to the factorization theorem,

$$\boldsymbol{Y}(\boldsymbol{s}) \sim (\boldsymbol{\mu}, (I - \rho B)^{-1} D).$$

We assume that the parameter ρ satisfies the necessary conditions for a positive definite matrix (See Banerjee et. al, p.79-82). One common way to construct B is to use a defined neighborhood matrix W that indicates whether the areal units associated with the measurements $Y(s_1), Y(s_2), \cdots, Y(s_n)$ are neighbors. For example, if $b_{ij} = w_{ij}/w_{i+}$ and $\tau_i^2 = \tau^2/w_{i+}$, then

$$(5) \quad \boldsymbol{Y}(\boldsymbol{s}) \sim (\boldsymbol{\mu}, \tau^2 (D_w - \rho W)^{-1}),$$

where $D_w = diag(w_{i+})$. Note that $\Sigma_{\boldsymbol{Y}}^{-1} = (D_w - \rho W)$ is nonsingular if $\rho \in (1/\lambda_{(1)}, 1/\lambda_{(n)})$ where $\lambda_{(1)}$ and $\lambda_{(n)}$ are the smallest and largest eigenvalues of $D_w^{-1/2} W D_w^{-1/2}$,, respectively.

Proposition 3. For a CAR model with $\Sigma = \tau^2 (D_w - \rho W)^{-1}$ where $\sigma_i = \Sigma_{ii}^{1/2}$ and $C = diag(\sigma_1, \sigma_2, \ldots, \sigma_n)$

$$(6) \quad \text{ESS} = \frac{1}{\tau^2} \left[\sum_i \sigma_i^2 w_{i+} - \rho \sum_i \sum_j \sigma_i \sigma_j w_{ij} \right],$$

where $w_{i+} = \sum_j w_{ij}$.

Now, let us consider a SAR process of the form

$$\begin{aligned} \boldsymbol{Y}(\boldsymbol{s}) &= X(\boldsymbol{s}) + \boldsymbol{e}(\boldsymbol{s}) \\ \boldsymbol{e}(\boldsymbol{s}) &= B\boldsymbol{e}(\boldsymbol{s}) + \boldsymbol{v}(\boldsymbol{s}) \end{aligned}$$

where B is a matrix of spatial dependency, $\mathbb{E}[\boldsymbol{v}(\boldsymbol{s})] = 0$, and $\Sigma_{\boldsymbol{v}(\boldsymbol{s})} = diag[\sigma_1^2, \ldots, \sigma_n^2]$. Then, $\Sigma = Var[Y(\boldsymbol{s})] = (I - B)^{-1} \Sigma_{\boldsymbol{v}(\boldsymbol{s})} (I - B^t)^{-1}$. Then, we can state the following results.

Proposition 4. For a SAR process with $B = \rho W$ where W is any contiguity matrix, $\sigma_{\boldsymbol{v}(\boldsymbol{s})} = \sigma^2 I$, $\sigma_i = \Sigma_{ii}^{1/2}$ and $C = diag(\sigma_1, \sigma_2, \ldots, \sigma_n)$, the effective sample size is given by

$$(7) \quad \text{ESS} = \frac{1}{\sigma^2} \left[\sum_i \sigma_i^2 - 2\rho \sum_i \sum_j \sigma_i \sigma_j w_{ij} + \rho^2 \sum_i \sum_j \sum_k \sigma_i \sigma_j w_{ki} w_{kj} \right].$$

The proofs of Propositions 1-4 are in the Appendix.

FUTURE RESEARCH

There are several ways to study effective sample size. One line of research involves studying the effect of dimension r on ESS. This can be done by considering the unit sphere centered at the origin with the radius constant over r, e.g., $1/2$. This makes the spaces comparable in terms of their dimensions with regard to the maximum distance. Our conjecture is that ESS is increasing in r, assuming a uniform distribution of the locations. Another line of research involves the estimation of ESS. Let us consider a model of the form

$$(8) \quad \boldsymbol{Y}(\boldsymbol{s}) = X(\boldsymbol{s})\beta + \boldsymbol{\epsilon}(\boldsymbol{s}),$$

where $\boldsymbol{Y}(\boldsymbol{s}) = (Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \cdots, Y(\boldsymbol{s}_n))^t$, $\boldsymbol{\epsilon}(\boldsymbol{s}) = (\epsilon(\boldsymbol{s}_1), \epsilon(\boldsymbol{s}_2), \ldots, \epsilon(\boldsymbol{s}_n))^t$, and $X(\boldsymbol{s})$ is a design matrix compatible with the dimensions of the parameter $\boldsymbol{\beta}$. Let us assume that $\epsilon(\boldsymbol{s}) \sim N(\boldsymbol{0}, \Sigma(\boldsymbol{\theta}))$. This notation emphasizes the dependence of Σ on $\boldsymbol{\theta}$. Notice that the model for which ESS was defined in (3) is a particular case of (8) when $X(\boldsymbol{s})\beta = \boldsymbol{1}\mu$. Thus, we can rewrite the effective sample size to emphasize its dependence on the unknown parameter $\boldsymbol{\theta}$ as follows:

$$\text{ESS} = \boldsymbol{1}^t R_n^{-1}(\boldsymbol{\theta})\boldsymbol{1}.$$

To estimate ESS, it is necessary to estimate $\boldsymbol{\theta}$. Cressie and Lahiri (1993) studied the asymptotic properties of the restricted maximum likelihood (REML) estimator of $\boldsymbol{\theta}$ in a spatial statistics context. We find it necessary to study the asymptotic properties of the estimation

$$\widehat{\text{ESS}} = \boldsymbol{1}^t R_n^{-1}(\boldsymbol{\theta}_{reml})\boldsymbol{1}.$$

The limiting value of $\boldsymbol{1}R_n^{-1}\boldsymbol{1}$ has been studied in the context of Ornstein-Uhlenbeck processes (Xia, et al., 2006). More specifically, the single mean model with intra-class correlation can be studied in detail following the inference developed in Paul (1990).

ACKNOWLEDGEMENTS

This research was supported in part by the UTSFM under grant 12.10.03, by the Centro Científico y Tecnológico de Valparaíso FB 0821 under grant FB/01RV/11, and by CMM, Universidad de Chile.

REFERENCES

Banerjee, S., Carlin, B., Gelfand, A. E., 2004, Hierarchical Modeling and Analysis for Spatial Data. Chapman Hall/CRC, Boca Raton.
Clifford, P., Richardson, S., Hémon, D., 1989. Assessing the significance of the correlation between two spatial processes. Biometrics 45, 123-134.
Cressie, N., 1993. Statistics for Spatial Data. Wiley, New York.
Cressie, N., Lahiri, S. N., 1993, Asymptotic distribution of REML estimators. Journal of Multivariate Analysis 45, 217-233.
Getis, A., Ord, J., 2000. Seemingly independent tests: Addressing the problem of multiple simultaneous and dependent tests. Paper presented at the 39th Annual Meeting of the Western Regional Science Association, Kanuai, HI, 28 February.
Graybill, F., 2001. Matrices with Applications in Statistics. Duxbury Classic series.
Griffith, D., 2005. Effective geographic sample size in the presence of spatial autocorrelation. Annals of the Association of American Geographers 95, 740-760.
Griffith, D., 2008. Geographic sampling of urban soils for contaminant mapping: how many samples and from where. Environ. Geochem. Health 30, 495-509.

Harville, D., 1997. Matrix algebra from a statistician's perspective. Springer, New York.

Hining, R., 1990. Spatial Data Analysis in the Social Environmental Sciences. Cambridge University Press, Cambridge.

Paul, R. R. 1990. Maximum likelihood estimation of intraclass correlation in the analysis of familial data: Estimating equation approach. Biometrika 77, 549-555.

APPENDIX

Proof of Proposition 1

Proof. It is enough to show that $\text{ESS}_{n+1} - \text{ESS}_n \geq 0$, for all $n \in \mathbb{N}$.

First, we define the matrix

$$R_{n+1} = \begin{pmatrix} R_n & \boldsymbol{\gamma} \\ \boldsymbol{\gamma}^t & 1 \end{pmatrix},$$

where $\boldsymbol{\gamma}^t = (\gamma_1, \gamma_2, \cdots, \gamma_n), 0 \leq \gamma_i \leq 1$, for all i. Since R_{n+1} is positive definite, the Schur complement $(1 - \boldsymbol{\gamma}^T R_n^{-1} \boldsymbol{\gamma})$ of R_n is positive definite (Harville 1997, p. 244). Thus $(1 - \boldsymbol{\gamma}^T R_n^{-1} \boldsymbol{\gamma}) > 0$.

Now, writing R_{n+1} as a partitioned matrix we get

$$ESS_{n+1} = \mathbf{1}_{n+1}^t R_{n+1}^{-1} \mathbf{1}_{n+1} = \mathbf{1}_{n+1}^t \begin{pmatrix} R_n & \boldsymbol{\gamma} \\ \boldsymbol{\gamma}^t & 1 \end{pmatrix}^{-1} \mathbf{1}_{n+1} = \text{ESS}_n + \frac{(\mathbf{1}_n^t R_n^{-1} \boldsymbol{\gamma})^2 - 2\mathbf{1}_n^t R_n^{-1} \boldsymbol{\gamma} + 1}{1 - \boldsymbol{\gamma}^t R_n^{-1} \boldsymbol{\gamma}},$$

where $\mathbf{1}_{n+1}^t = (\mathbf{1}_n \ 1)^t$. Since the function $f(x) = x^2 - 2x + 1 = (x-1)^2 \geq 0$, for all x, we have that $\text{ESS}_{n+1} - ESS_n \geq 0$, for all $n \in \mathbb{N}$. □

Proof of Proposition 2

Proof. To prove that $1 \leq$ ESS it is enough to use the Cauchy- Schwartz inequality for matrices. ESS $\leq n$ can be proved by induction over n. □

Proof of Proposition 3

Proof. For $\Sigma = \tau^2 (D_w - \rho W)^{-1}$ where $\sigma_i = \Sigma_{ii}^{1/2}$ and $C = diag(\sigma_1, \sigma_2, \ldots, \sigma_n)$, it is easy to see that

$$\text{ESS} = \frac{1}{\tau^2} \left[\sum_i \sigma_i^2 w_{i+} - \rho \sum_i \sum_j \sigma_i \sigma_j w_{ij} \right].$$

□

Proof of Proposition 4

Proof. Equation (7) can be derived from the following facts: $R_{SAR}^{-1} = C \Sigma_{SAR}^{-1} C = \frac{1}{\sigma^2} C (I - \rho W^t)(I - \rho W) C = \frac{1}{\sigma^2} (I - \rho W - \rho W^t + \rho^2 W^t W)$, $\rho \mathbf{1}^t C W C \mathbf{1} = \rho \mathbf{1}^t C W^t C \mathbf{1} = \rho \sum_i \sum_j \sigma_i \sigma_j w_{ij}$, and $\rho^2 \mathbf{1}^T C W^T W C \mathbf{1} = \rho^2 \sum_i \sum_j \sum_k \sigma_i \sigma_j w_{ki} w_{kj}$. □