Bayes Quadratic Unbiased Estimator of Spatial Covariance Parameters

Dissertation zur
Erlangung des akademischen Grades
des Doktors der Naturwissenschaften (Dr. rer. nat.)
am Fachbereich Mathematik/Informatik
der Universität Kassel

vorgelegt von
Younis M. Fathy
Kassel im Juli 2006

Abstract

This work presents Bayes invariant quadratic unbiased estimator, for short BAIQUE. Bayesian approach is used here to estimate the covariance functions of the regionalized variables which appear in the spatial covariance structure in mixed linear model. Firstly a brief review of spatial process, variance covariance components structure and Bayesian inference is given, since this project deals with these concepts. Then the linear equations model corresponding to BAIQUE in the general case is formulated. That Bayes estimator of variance components with too many unknown parameters is complicated to be solved analytically. Hence, in order to facilitate the handling with this system, BAIQUE of spatial covariance model with two parameters is considered.

Bayesian estimation arises as a solution of a linear equations system which requires the linearity of the covariance functions in the parameters. Here the availability of prior information on the parameters is assumed. This information includes a priori distribution functions which enable to find the first and the second moments matrix. The Bayesian estimation suggested here depends only on the second moment of the prior distribution. The estimation appears as a quadratic form $\mathbf{y}'A\mathbf{y}$, where $\mathbf{y}$ is the vector of filtered data observations. This quadratic estimator is used to estimate the linear function of unknown variance components. The matrix A of BAIQUE plays an important role. If such a symmetrical matrix exists, then Bayes risk becomes minimal and the unbiasedness conditions are fulfilled. Therefore, the symmetry of this matrix is elaborated in this work. Through dealing with the infinite series of matrices, a representation of the matrix A is obtained which shows the symmetry of A. In this context, the largest singular value of the decomposed matrix of the infinite series is considered to deal with the convergence condition and also it is connected with Gerschgorin Discs and Poincare theorem.

Then the BAIQUE model for some experimental designs is computed and compared. The comparison deals with different aspects, such as the influence of the position of the design points in a fixed interval. The designs that are considered are those with their points distributed in the interval $[0, 1]$. These experimental structures are compared with respect to the Bayes risk and norms of the matrices corresponding to distances, covariance structures and matrices which have to satisfy the convergence condition. Also different types of the regression functions and distance measurements are handled. The influence of scaling on the design points is studied, moreover, the influence of the covariance structure on the best design is investigated and different covariance structures are considered. Finally, BAIQUE is

applied for real data. The corresponding outcomes are compared with the results of other methods for the same data. Thereby, the special BAIQUE, which estimates the general variance of the data, achieves a very close result to the classical empirical variance.

Zusammenfassung

Diese Arbeit präsentiert den Bayesschen, invarianten, quadratischen und unverfälschten Schätzer, kurz BAIQUE. Die Bayessche Annäherung wird hier verwendet, um die Kovarianzfunktion regionalisierter Variablen zu schätzen, die in der räumlichen Kovarianzstruktur in gemischten linearen Modellen auftreten. Zunächst wird ein kurzer Überblick über räumliche Prozesse, die Varianz- und Kovarianzkomponentenstruktur und den Bayesschen Rückschluss gegeben, da dieses Projekt diese Konzepte behandelt. Dann wird das lineare Schätzungsmodell, dem BAIQUE im allgemeinen Fall entspricht, formuliert. Es ist kompliziert Bayes-Schätzer für Varianzkomponenten mit zu vielen unbekannten Parametern zu lösen. Um den Umgang mit diesem System zu vereinfachen, wird daher BAIQUE für räumliche Kovarianz-Modelle mit zwei Parametern betrachtet.

Die Bayessche Schätzung ergibt sich als Lösung eines linearen Gleichungssystems, das die Linearität der Kovarianzfunktionen in den Parametern erfordert. Dabei wird die Verfügbarkeit von a priori Informationen über die Parameter vorausgesetzt. Diese Informationen beinhalten die a priori Verteilungsfunktionen, die es ermöglichen, die Matrix der ersten und zweiten Momente zu finden. Die Bayessche Schätzung, auf die hier eingegangen wird, beruht nur auf den zweiten Momenten der a priori Verteilung. Die Schätzung besitzt die quadratische Form $\mathbf{y}'A\mathbf{y}$, wobei $\mathbf{y}$ der Vektor der gefilterten Beobachtungsdaten ist. Dieser quadratische Schätzer wird verwendet, um die lineare Funktion der unbekannten Varianzkomponenten zu schätzen. Hierbei spielt die Matrix A von BAIQUE eine wichtige Rolle, denn wenn eine solche symmetrische Matrix existiert, wird das Bayes-Risiko minimal und die Anforderungen der Erwartungstreue sind erfüllt. Deshalb wird die Symmetrie dieser Matrix in dieser Arbeit sorgfältig ausgearbeitet. Durch das Arbeiten mit den unendlichen Reihen von Matrizen wird eine Darstellung der Matrix A erhalten, von der die Symmetrie abgelesen werden kann. In diesem Zusammenhang wird der größte Eigenwert der zerlegten Matrix der unendlichen Reihen berücksichtigt, um die Konvergenzbedingung zu behandeln, und dieser Wert wird mit den Gerschgorinschen Kreisscheiben und dem Poincare Theorem in Verbindung gesetzt.

Anschließend wird das BAIQUE-Modell für einige Versuchspläne berechnet und verglichen. Der Vergleich behandelt verschiedene Aspekte, wie beispielsweise den Einfluss der Position der Punkte des Versuchsplans in einem festen Intervall. Hierbei sind die betrachteten Pläne solche, deren Punkte im Intervall $[0, 1]$ liegen. Diese Versuchsstrukturen werden in Hinsicht auf das Bayes-Risiko und die Normen der Ma-

trizen für Distanz und Kovarianzstruktur und Matrizen, die die Konvergenzbedingung erfüllen, verglichen. Ebenso werden verschiedene Typen von Regressionsfunktionen und Distanzmessungen bearbeitet. Ferner wird der Einfluss der Skalierung der Versuchspunkte studiert, der Einfluss der Kovarianzstruktur des besten Versuchsplans untersucht und verschiedene Kovarianzstrukturen betrachtet. Schließlich wird BAIQUE auf reale Daten angewandt und die entsprechenden Ergebnisse werden mit den Resultaten anderer Methoden für die selben Daten verglichen. Dabei erzielt der spezielle BAIQUE, der die allgemeine Varianz schätzt, ein sehr ähnliches Ergebnis wie die klassische empirische Varianz.

Acknowledgements

I would like to express my appreciation to my advisor Prof. Dr. Christine Müller for her guidance and patient assistance throughout this work.

I would like to thank my parents and my family for their support.

I extend my special thanks to all my friends for their encouragement.

I want to thank DAAD (Deutscher Akademischer Austausch Dienst) for the scholarship to accomplish my PhD studies.

I deeply thank Almighty God for his generosity.

Contents

1 Introduction

This work deals with Bayesian estimation of spatial covariance functions which appear in the spatial covariance structure in mixed linear models. Bayesian philosophy is one of the most important mathematical methodologies. Bayesian approaches are considered to be powerful methods which propose logical procedures to accomplish the reasonable statistical inference and decisions. The selection of priors including noninformative priors is one of the main subjects in Bayesian analysis. The principle of the Bayes rule is to minimize the posterior expected loss under a chosen loss function. The resulting estimator is called the Bayes estimator.

Spatial statistics deals with the regionalized variables. In general, regionalized variables are spatially distributed and spatially structured data. Traditional geostatistics is based on the spatial correlations. Different measures of spatial correlations are widely used covariance functions or variograms to describe similarity or dissimilarity between two separated points in space. Geostatistics and the probabilistic approach in general is particularly suited to the study of natural phenomena. The geostatistical school has made important contributions to the linear estimation of spatial variables, including the popularizing of the variogram and the generalized covariance function. This approach is well accepted among practitioners because it is a realistic approach to solve problems encountered in practices using statistical concepts.

Variance covariance components play an important role in many fields of sciences. There are several procedures to estimate variance components. A first solution to the problem was provided by Helmert (1924), who proposed a method for unbiased variance estimates. Rao (1971) has introduced the minimum norm quadratic estimation (MINQUE) method. In this approach a quadratic estimator is sought that satisfies the minimum norm criterion. MINQUE has been implemented and proven useful in various applications. In this context, Marshall and Mardia (1985) have presented this procedure through dealing with a nested spatial covariance function. Stein (1987) has investigated this method with the variogram model and Witkovsky (1996) has considered MINQUE approach with first order autoregressive disturbances. Quadratic estimation of variance components and its properties has been considered by many researchers such as Seely (1971), LaMotte (1973), Gnot and Kleffe (1983), Kubacek (1985) and Sjöberg (1995).

Using the least squares method there is an unified procedure for the derivation of estimators for variance components in the linear model. These least squares

estimators are unbiased. The invariant and unbiased least squares estimators for variance components are given by the MINQUE estimators as Verdooren (1988) has shown. Pelletier et al. (2004) have studied fitting the linear model of coregionalization by generalized least squares. In this context, Volaufova and Witkovsky (1991) have dealt with the quadratic invariant estimators of the linear functions of variance components in mixed linear models and different types of least squares estimators.

Another common approach for variance components estimation arises through analysis of variance (ANOVA). This includes traditional ANOVA, various correlation and regression methodologies. Additionally methodologies with modification to the ANOVA and sum of squares of unbalanced data such as the four methods of Henderson, Henderson(1984). This approach has a nice feature that the estimators are unbiased regardless of whether the data are normally distributed and it is especially useful in the analysis of controlled experiments when the structure of the data has regular form. There are many contributions in this field. For example, Townsend and Searle (1971) has considered best quadratic unbiased estimation of variance components from unbalanced data. Hartely et al. (1978) have investigated this approach with a new algorithm.

Maximum likelihood (ML) is one of the famous methods to estimate variance components. One of the basic requirements of maximum likelihood estimation is to assume an underlying probability distribution for the data. ML was first introduced by Hartley and Rao (1967) for estimating variance components. Conceptually, ML attempts to identify the values of the parameters of the distribution that maximize the likelihood of the observed data. As in the case for almost all maximum likelihood methods, this distribution is assumed to be multivariate normal. The maximum likelihood approach for spatial data was apparently first proposed by Kitanidis (1983) for linear covariance functions. Mardia and Marshall (1984) and Kitanidis and Lane (1985) have extended ML to general covariance functions. Zimmerman and Harville (1990), Mardia and Watkins (1989) and Pardo-Iguzquiza (1998) have presented a nice general discussion of maximum likelihood.

On the other hand, the method of restricted (residual) maximum likelihood (REML) involves finding maximum likelihood estimates of variance components from the distribution of the residuals. This allows for estimation of the variance components without the complication of the fixed effects. The first who developed REML for the mixed model generally was Patterson and Thompson (1974). This approach has used by many researchers such as, Pardo-Iguzquiza and Dowd (1998) have considered REML with second order stationary universal kriging model. This

approach has handled the spatial linear models by Christensen et al. (1993) and Berke (1999). Stein et al. (2004) have dealt with approximating likelihood for large spatial data set. This procedure was used with systematically sampled data on oil by Lark and Cullis (2004).

Recently, some methodologies have been developed based on Bayesian approach. This approach also requires that the distribution of the vector of observation is specified as with the maximum likelihood methods but the difference with this approach is that it requires some prior knowledge about the vector of variance covariance components in form of a prior probability density function. Koch (1987) has used Bayesian inference to derive the posterior density function of variance covariance components. As prior density, he has assumed the noninformative prior based on invariance property established by Jeffreys (1961). Approximative Bayes estimation in general linear model has considered by Ou (1991). Bayesian kriging approach has handled by several works such as Le and Zidek (1992), Handcock and Stein (1993), Hjort and Omre (1994) and De Waal and Groenewald (1995). Other contributions have investigated Bayesian approach in fields of agricultural experiments such as Besag and Higdon (1999) or in the assessment of environmental models such as Fuentes et al. (2003).

This work presents Bayes invariant quadratic unbiased estimator, for abbreviation one writes BAIQUE. Bayesian technique is used here to estimate the covariance functions of the regionalized variables which appear in the spatial covariance structure in mixed linear model. The BAIQUE requires the linearity of the covariance functions in the parameters. We assume the availability of the prior information on the parameters. This information includes a priori distribution function which enables to find the first and the second moments matrix. The Bayesian estimation suggested here depend on the second moment of the prior distribution of the parameters. The solution obtained in the sense minimizing the quadratic criterion is still optimal even if the data are not gaussian, Kitanidis (1985).

The problem at hand is to estimate a linear function of the unknown vector variance components $\theta = (\theta_1, \theta_2, \ldots, \theta_r)$, say $\mathbf{b}'\theta = \sum_{i=1}^{r} b_i\theta_i$ by a quadratic form $Y'AY$, where the coefficients b_i are known, Y is the vector of the filtered data and A is a matrix obtained by solving the partitioned equations system of BAIQUE. So we deal with the general case of BAIQUE under unbiased condition and minimum variance with respect to all unbiased quadratic estimators. Since Bayes estimator of variance components is complicated to be solved analytically when there are too

many unknown parameters, numerical integration must be performed and in such case problems with computational difficulty arise. Therefore, in order to facilitate the handling with the model, the spatial covariance model with two parameters is considered here. The general form of the least square solution corresponding to this case is found and the form of the solution which represents matrix A is obtained. This matrix plays an important role since BAIQUE depends on it. Moreover, if a symmetric matrix exists, then the Bayes risk becomes minimal and the unbiasedness conditions are fulfilled. Therefore, the symmetricity of this matrix is presented in more detail. In this manner, the infinite series matrices are used to get the general form of the solution. An outline of the essential structure of this thesis is discussed below.

The second section gives a brief summary of three topics which are related with this work. Its first subsection treats the spatial process and gives some properties of regionalized variables. The second subsection deals with spatial covariance structure and variance covariance components models. The third subsection is concerned with Bayesian inference, a brief review about Bayesian statistics and the point estimators resulting from quadratic loss are introduced.

Section three contains the model of BAIQUE, the meaning of BAIQUE and the formulation of the general case which appears by means of partitioned matrices. This estimator is restricted with respect to translation invariance, unbiasedness and minimizes Bayes risk function. Then the general case is reduced to the case when the number of the variance components is equal to two. Here some examples are introduced in order to understand more the model. Through that the problem of the singularity of this system arises. Thus, the matrix A in BAIQUE form satisfies the symmetry condition only when the pseudoinverse is used to solve the BAIQUE equations system. In the last subsection, the partitioned least squares solution corresponding to this model is presented by means of Kronecker product and vec operation. This form is useful to investigate the symmetry of the matrix in Bayes solution. Therefore, the symmetry of this matrix is proved. For this purpose some lemmas and conditions are given. In particular, it is shown that a sufficient condition for symmetry is the exchangeability of sums of infinite series of matrices.

Section four is assigned to the comparison of the designs. The comparisons deal with some aspects, such as the influence of the position of the design points in a fixed interval. The designs that were considered are those with their points distributed in the interval $[0, 1]$. These experimental structures are compared with respect to the Bayes risk and norms of the matrices corresponding to distances, co-

variance structures and matrices which have to satisfy the convergence condition of the finite series sum. Also different types of the regression functions and distance measurements are handled. In another subsection the influence of scaling on the design points is studied. The interesting here is to find out the effect of scaling on the norm of the covariance function. In a separate subsection, the influence of the covariance structure on the best design is investigated and different covariance structures are considered. Finally, in order to make the results of this work more useful and have practical aspects, BAIQUE is applied for real data. The corresponding outcomes have been compared with other methods outcomes for the same data. The BAIQUE achieves very close results to the actual data outcomes. We mention here that BAIQUE has been programmed in Maltlab language and all the results in this work have been computed by using Matlab program.

The last section is devoted to appendices, some lemmas, theorems, properties which are connected to the concepts of the suggested project. In separate appendices, some aspects and qualities of Kronecker product and vec operation, generalized inverse matrices, infinite series representation of the matrix inverse and Neumann theorem, distribution of quadratic forms and finally non-negative matrices are presented .

2 Preliminaries

2.1 Spatial process

The outcome $z(x)$ of a random variable $Z(x)$ is considered as the observed value at a data point x. The mean $\mu(x)$ of $Z(x)$ is called the drift. The values $z(x)$ at points where no measurements have been made are well defined even though they are unknown. They can also be thought of as being the outcomes (or realizations) of the corresponding random variable $Z(x)$. Spatial data can be considered as a realization of a stochastic process (random field), $Z(x)$; $x \in \mathcal{D} \subset \mathbb{R}^d$. Here x is a location in $\mathcal{D}$. Most often d, the dimension of the space, is 1, 2 or 3. We also assume that the variance of $Z(x)$ exists for all $x \in \mathcal{D}$.

In mathematical term, the family of all these random variables is called a random function (stochastic process, random field). The distribution of a random function is characterized by its finite dimensional distributions, i.e. by the joint distributions of any set of variables $Z(x_1), Z(x_2), \ldots, Z(x_k)$, for all k, and for all points $x_1, x_2, \ldots, x_k$. It would be difficult to work with such model unless we are prepared to make some assumptions about the characteristics of these distributions.

The process Z is said to be Gaussian if, for any $k \geq 1$ and locations $x_1, x_2, \ldots, x_k$, the vector $(Z(x_1), Z(x_2), \ldots, Z(x_k))$ has a multivariate normal distribution with $E(Z(x)) = \mu(x)$, $Var(Z(x)) = \sigma^2(x)$ and covariance function $Cov(h, x) = C(Z(x), Z(x+h))$, where $\| h \|$ is the Euclidean distance between x and $x+h$. The process Z is said to be strictly stationary if the joint distribution of $(Z(x_1), Z(x_2), \ldots, Z(x_k))$ is the same as that of $(Z(x_1 + h), Z(x_2 + h), \ldots, Z(x_k + h))$ for any k spatial points $x_1, x_2, \ldots, x_k$ and any $h \in \mathbb{R}^d$, provided only that all of $x_1, x_2, \ldots, x_k, x_1 + h, x_2 + h, \ldots, x_k + h$ lie within the domain $\mathcal{D}$, see Armstrong (1998). A stationary random process has some properties such as isotropic and positive definite covariance structure, as in the following.

Some properties of regionalized variables

It is often in statistical design to assume that the variable is *stationary*, i.e. its distribution is invariant under translation. In the same manner a stationary random function is homogeneous and self-repeating in space. For any increment h, the distribution of $Z(x_1), Z(x_2), \ldots, Z(x_k)$ is the same as that of $Z(x_1+h), Z(x_2+h), \ldots, Z(x_k+h)$. This makes statistical inference possible on a single realization. In its strictest sense stationary requires all the moments to be invariant under translation, but since this cannot be verified from the limited experimental data, we usually require only the first two moments. This is called "*weak*" or "*second order stationary*". In

other words, the expected value (or mean) of $Z(x)$ must be constant for all points x. That is,

$$E(Z(x)) = \mu(x) = \mu.$$

On the other hand, the covariance function between any two points x and $x + h$ depends on the vector h but not on the point x. That is,

$$E[Z(x)Z(x + h)] - \mu^2 = C(h).$$

For a zero distance, i.e. $C(0)$ there is no need to make an assumption about the variance because it turns out to be equal to the covariance.

In practice, it often happens that the assumptions of second order stationary are not satisfied. Clearly when there is a marked trend the mean value cannot be assumed to be constant. Another branch of geostatistics has been developed to handle "*nonstationary*" regionalized variables, by further assuming that the mean is constant but by weakening the assumption of constant covariance. This is why Matheron (1963) developed the "*intrinsic hypothesis*" . It assumes that the increments $Z(x + h) - Z(x)$ exist and are independent of the points

$$E[Z(x + h) - Z(x)] = 0 \quad ; \quad Var[Z(x + h) - Z(x)] = 2\gamma(h).$$

The function $\gamma(h)$ is called the $semi - variogram$ ($variogram$ for short). It is the basic tool for the structural interpretation of spatial phenomena as well as for estimation. For any second-order stationary process there is an interested relation between the variogram $\gamma(h)$ and covariance function $C(h)$ which is expressed by

$$\gamma(h) = C(0) - C(h),$$

where $C(0)$ is the variance function. In general, but not always, when the distance h increases, then the mean quadratic deviation between the two variables $Z(x)$ and $Z(x + h)$ tends to increase and so $\gamma(h)$ increases from its initial zero value. Then it becomes stable beyond some distance $a = |h|$ called the "range". At this distance a typical variogram reaches a limit called its "*sill*". In opposite the covariance function tends to decrease, see Journal and Huijbregts (1978.)
$\gamma(h)$ does not tend to zero as h tends to 0. This means that the variable is highly irregular at short distances, the variogram of most geological variables have this discontinuity at the origin which indicates erratic short scale behaviour called a $nugget\ effect$.

For stationary and intrinsic variables, the mean of $Z(x+h) - Z(x)$ is zero and so the variogram $\gamma(h)$ is just the mean square difference. Consequently,

$$\gamma(h) = \frac{1}{2}E[Z(x+h) - Z(x)]^2.$$

Moreover, to estimate the vaiogram from given data we can use

$$\hat{\gamma}(h) = \frac{1}{2n(h)}\sum_{i=1}^{n(h)}[z(x_i) - z(x_i+h)]^2.$$

If the data points are considered in pairs it is the variance per point, $n(h)$ is the number of pairs of data points separated by the particular lag vector h.
Often the second order properties of a process can be assumed to depend only on the distance between two points and not on the direction between them. Let $u, v \in \mathcal{D} \subset R$, a second order stationary process is *isotropic* if

$$C(u-v) = C(|u-v|).$$

An intrinsically stationary process is isotropic if

$$\gamma(u-v) = \gamma(|u-v|),$$

a process that is not isotropic is said to be *anisotropic*, see Christensen (1991).
Since the common estimators are linear combinations of the data, we need to be able to calculate their variance. Consider $C(h)$, the covariance of the stationary variable $Z(x)$. Suppose the linear combination Z^* be $Z^* = \sum_i \lambda_i Z(x_i)$, where λ_i are the weights and x_i are the sample locations. If μ is the mean of $Z(x)$, $E(Z^*) = \mu\sum_i \lambda_i$. Then the variance can be defended as

$$\begin{aligned} Var(Z^*) &= Var(\sum_i \lambda_i Z(x_i)) \\ &= E(\sum_i \lambda_i (Z(x_i) - \mu))^2 \\ &= \sum_i \sum_j \lambda_i \lambda_j \, C(x_i - x_j) \geq 0, \end{aligned} \tag{2.1.1}$$

This must be non-negative for any choice of points and weights. A function $C(h)$ which satisfies this condition is said to be *positive definite*, see Armstrong(1998).

2.2 Spatial covariance structure and the variance covariance components models

The general linear model corresponds to the common assumption that the regionalized variable is a realization of a random function

$$z(\mathbf{x}) = \sum_{i=1}^{p} f_i(\mathbf{x})\beta_i + e(\mathbf{x}), \tag{2.2.1}$$

where $\mathbf{x}$ is the vector of the spatial coordinates of the point, z is sampled, β_i, $i = 1, \ldots, p$ are generally unknown parameters, $f_i(\mathbf{x})$, $i = 1, \ldots, p$ are known functions of the spatial coordinates, such as polynomial, and $e(\mathbf{x})$ is a zero mean spatial random function. The term $\sum_{i=1}^{p} f_i(\mathbf{x})\beta_i$ on the right hand side in Eq. (2.2.1) represents a $drift$ ($trend$) and the second term represents a zero mean random field ($stochastic\ part$). The covariance function $C(u, v; \theta)$ of the stochastic part is defined by

$$E[e(u)e(v)] = C(u, v; \theta), \tag{2.2.2}$$

where θ is an $m \times 1$ vector of parameters. From Eq.(2.2.1) and Eq.(2.2.2) one may obtain as special cases the commonly used models. For example, Eq.(2.2.1) represents a stationary isotropic field with mean equal to β_1 when $p = 1$, $f_i(\mathbf{x}) = 1$ and $C(u, v) = C(|u - v|)$. In other cases, polynomials or periodic functions may be used to represent the drift.

Suppose that Z is a vector of n measurements. In such situations Eq. (2.2.1) yields the following general relation

$$Z = X\beta + \mathbf{e}, \tag{2.2.3}$$

where X is a known $n \times p$ matrix of the known functions $f(x)$ of the spatial coordinates, β is the $p \times 1$ vector of drift coefficients, and $\mathbf{e}$ is a random vector with zero mean and covariance matrix V which is a known function of the parameter vector θ. The essence of spatial statistics is spatial correlation, and consequently it is important to model this aspect of the problem adequately. To guarantee that the covariance matrix is positive definite, the spatial covariance matrix V is assumed to be of some parameteric form and expressed as $V(\theta)$. Spatial statistics focuses on the choice of the covariance function and estimation of its parameters. See Cressie (1993).

The covariance function describes the spatial association between the random effect at any two locations in space, say u and v

$$Cov(e(u), e(v)) = C(u, v; \theta). \tag{2.2.4}$$

We may write $C(u, v) = \sigma_e^2 \rho_e(u, v; \theta)$ where ρ_e is the correlation function, being scaled by the variance component σ_e^2. The covariance function allows one to fill-in the elements of $V(\theta)$. Thus, given covariance function C, spatial parameters θ, and any two locations in space such as u and v, the covariance between $e(u)$ and $e(v)$ may be determined. Typically, assumptions are imposed on the process to facilitate the estimation of the parameters. The two usual assumptions are second order stationary and isotropy, the former being translation invariance of the second-moment structure of $\mathbf{e}$. Normally, the stationary assumption would imply a similar constraint on the first moment structure, but we have assumed $\mathbf{e}$ to have mean zero, accommodating any mean nonstationarity in $X\beta$. Thus under these assumptions, the covariance between any two points is only a function of the distance separating them

$$Cov(e(u), e(v)) = \sigma_e^2 \rho_e(\parallel u - v \parallel; \theta), \tag{2.2.5}$$

where $\parallel u - v \parallel$ is the distance between points u and v, say Euclidean, geographic distance, or other. This simplified correlation structure conveniently dictates the covariance between observed and unobserved values of z for which predictions are desired. See Wackernagel (2003).

The polynomial generalized covariance function including the nugget effect term can be represented as

$$g(|h|) = \sum_{i=1}^{r} \theta_i g_i(|h|). \tag{2.2.6}$$

Where r is the number of unknown parameters, θ_i are the parameters to be inferred subject to sufficient constraints to ensure that the right-hand side of (2.1.1) will have the properties of a variance. Moreover, $g_i(|h|)$ are known functions, the term $\theta_r g_r(|h|) = \theta_r \delta$ represents small-scale variability referred to as the *nugget effect*, where δ denotes Kronecker's delta, i.e.

$$\delta = \begin{cases} 0 & \text{if } |h| > 0 \\ 1 & \text{if } |h| = 0, \end{cases} \tag{2.2.7}$$

If the random field $Z(x)$ is an stationary with ordinary stationary covariance function $C(h)$, then $C(h)$ is clearly also a generalized covariance. An important practical advantage of the polynomial generalized covariance function given by (2.2.6) is that it is linear in the parameters. In such cases the statistical inference from

available data is somewhat facilitated. In particular, if the criterion is least square, the parameters estimation problem can be solved in computationally very efficient ways which are similar to linear regression. See Kitanidis (1987) and Starks and Fang(1982).

Consider the general mixed linear model in Eq. (2.2.3). The $n \times 1$ vector $\mathbf{e}$ can be expressed as

$$\mathbf{e} = U_1\zeta_1 + \ldots + U_2\zeta_2 + U_l\zeta_l, \tag{2.2.8}$$

where U_i denotes a $n \times r_i$ known matrix and ζ_i a $r_i \times 1$ unknown and unobservable random vector with $E(\zeta_i) = 0$ and $cov(\zeta_i, \zeta_j) = \sigma_{ij}R_{ij}$ with $\sigma_{ij} = \sigma_{ji}$ and $R'_{ij} = R_{ji}$ for $i, j \in \{1, \ldots, l\}$. Then the covariance matrix V of the observations Z follows with Theorem (5.24), see Appendix (5.4), by

$$V = \sigma_1^2 U_1 R_{11} U_1' + \sigma_{12}(U_1 R_{12} U_2' + U_2 R_{21} U_1') + \sigma_{13}(U_1 R_{13} U_3' + U_3 R_{31} U_1') + \ldots + \sigma_l^2 U_l R_{ll} U_l'.$$

By putting $U_l R_{11} U_l' = V_1, U_1 R_{12} U_2' + U_2 R_{21} U_1' = V_2, \ldots, U_l R_{ll} U_l' = V_k$, where $k = l(l+1)/2$ and the matrices V_m with $m \in \{1, \ldots, k\}$ are symmetric, one obtains

$$V = E(\mathbf{e}\mathbf{e}') = \sigma_1^2 V_1 + \sigma_{12} V_2 + \ldots + \sigma_l^2 V_k. \tag{2.2.9}$$

Note that when $l = 1$ and $V_1 = I_n$, where I_n stands for the $n \times n$ identity matrix, the model (2.2.3) reduces to the linear regression model. Also such models as 1-way classification model and 2-way classification model could be regarded as special cases of the model (2.2.3).

Let matrix V in Eq. (2.2.9) be positive definite. This model is called the Gauss-Markoff model with k unknown variance and covariance components σ_i^2 and σ_{ij} with $i \in \{1, \ldots, l\}$, $i < j \leq l$ and $l \leq k \leq l(l+1)/2$, and the variance components have to be estimated. See Koch (1999) and Rao and Kleffe (1988).

Example 2.1. Let the vector Z_1 of observations with the weight matrix P_1 contain the measurements of distances and the vector Z_2 with the weight matrix P_2 contain the measurements of angles. So the covariance matrix of Z_1 and Z_2 with the unknown variance component σ_1^2 for the distance and the unknown variance component σ_2^2 for the angles follows as

$$Cov\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \sigma_1^2 \begin{pmatrix} P_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \sigma_2^2 \begin{pmatrix} 0 & 0 \\ 0 & P_2^{-1} \end{pmatrix}.$$

If the sets of different observations are dependent and the covariances known except for common factors, then the factors have to be estimated as covariance components. One finds for this example when R_{12} is the covariance matrix of the distances and angles with the unknown covariance component σ_{12} and $R_{21} = R'_{12}$ the covariance matrix

$$Cov\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \sigma_1^2 \begin{pmatrix} P_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \sigma_{12} \begin{pmatrix} 0 & R_{12} \\ R_{21} & 0 \end{pmatrix} + \sigma_2^2 \begin{pmatrix} 0 & 0 \\ 0 & P_2^{-1} \end{pmatrix},$$

see Koch (1999). □

Example 2.2. Consider the mixed model in the form

$$Z = X\beta + U_1\zeta_1 + U_2\zeta_2 + \epsilon$$

With $E(\zeta_i) = 0$, $E(\zeta_i\zeta_i') = \sigma_i^2 I_i$, $E(\zeta_i\zeta_j') = 0$ and $Cov(\zeta_i, \epsilon) = 0$ for $i \neq j$, let $n = 8$, $p = 4$ and the matrix U_1 of type 8×2 and matrix U_2 of the type 8×4, then the vectors ζ_1, ζ_2 and ϵ are random of dimension 2, 4 and 8 respectively with zero mean and covariance matrices $\sigma_1^2 I_{2\times 2}$, $\sigma_2^2 I_{4\times 4}$ and $\sigma_3^2 I_{8\times 8}$ respectively. Then

$$Cov(Z) = \sigma_1^2 U_1 U_1' + \sigma_2^2 U_2 U_2' + \sigma_3^2 I_{8\times 8}.$$

□

2.3 Bayesian inference

The Bayesian approach to statistics provides a theoretical and practical framework to view many statistical problems. Bayesian methods have become increasingly popular because they provide solutions to many intractable problems. In the context of parameters estimation, the available information about the parameters can be used. In Bayesian inference, one treats all unknown quantities and the parameters as random variables and constructs a joint probability distribution for all of them. The information about the parameters is specified by their *prior distribution* $p_\Theta(\theta)$. The information provided by the data on θ is contained in the *likelihood* $p_{X|\Theta=\theta}(x)$, where x is the $n \times 1$ vector of n experimental data. The Bayes' theorem is used to combine both pieces of information. After observing the data, one constructs the conditional distribution of Θ given $X = x$ containing all the information about the

parameters which is called the *posterior distribution*

$$p_{\Theta|X=x}(\theta) = \frac{p_\Theta(\theta)p_{X|\Theta=\theta}(x)}{p_X(x)} = \frac{p_\Theta(\theta)p_{X|\Theta=\theta}(x)}{\int p_\Theta(\theta)p_{X|\Theta=\theta}(x)d\theta} \ ,$$

where $p_X(x)$ is the *marginal distribution* of x. The entire distribution of the parameters θ is obtained through the Bayesian approach.

The *Bayes risk* $r(\hat{\theta})$ of the decision function $\hat{\theta}$ can be defined as the expectation of *risk function* $R(\Theta, \hat{\theta})$ over all possible values of θ, thus,

$$r(\hat{\theta}) = ER(\Theta, \hat{\theta}) = \int R(\Theta, \hat{\theta})p_\Theta(\theta)d\theta \ . \tag{2.3.1}$$

It seems sensible to minimize one's losses and accordingly a Bayes decision function $\hat{\theta}$ is defined as one which minimizes the Bayes risk $r(\hat{\theta})$

$$\begin{aligned} r(\hat{\theta}) &= \int R(\Theta, \hat{\theta})p_\Theta(\theta)d\theta \\ &= \int\int L(\Theta, \hat{\theta}(x))p_{X|\Theta=\theta}(x)p_\Theta(\theta)dxd\theta \\ &= \int\int L(\Theta, \hat{\theta}(x))p_{(X,\Theta)}(x,\theta)dxd\theta \\ &= \int[\int L(\Theta, \hat{\theta}(x))p_{\Theta|X=x}(\theta)d\theta]p(x)dx. \end{aligned} \tag{2.3.2}$$

See Lee (1997) and Box and Tiao (1973). Therefore, *Bayes estimator* can be defined as follows.

Definition 2.3. Suppose that $\hat{\theta}(x)$ is an estimator of θ. Let $L(\Theta, \hat{\theta}(x))$ be the loss function of $\hat{\theta}(x)$ and $p_{\Theta|X=x}(\theta)$ be the posterior distribution of θ. Then the posterior mean loss function of $\hat{\theta}(x)$ is defined as

$$B(\hat{\theta}(x), x) = E_{\Theta|X=x}[L(\Theta; \hat{\theta}(x))] = \int_\Theta L(\Theta, \hat{\theta}(x))p_{\Theta|X=x}(\theta)d\theta. \tag{2.3.3}$$

The above mean loss function is also known as the posterior risk function. The decision rule $\hat{\theta}(x)$ such that Eq. (2.3.3) attains its minimum is called the *Bayes estimator*, see Wang and Chow(1994).

Point estimators resulting from quadratic loss

In the case of point estimation a Bayes decision rule is referred to as a Bayes estimator. In such situations it is convenient to work with *quadratic loss*, i.e. with a squared-error loss function $L(\theta, \hat{\theta}(x)) = (\theta - \hat{\theta}(x))^2$. There are a number of reasons why it is often considered in evaluating decision rules. It is originally used in estimation problems when unbiased estimators of θ are considered, since the risk function $R(\theta, \hat{\theta}(x)) = E_\theta L(\theta, \hat{\theta}(X)) = E_\theta[\theta - \hat{\theta}(X)]^2$ would then be the variance of the estimator. A second reason for its popularity is due to its relationship to classical least squares theory. The similarity between them makes squared-error loss seems familiar to statisticians. Finally, the use of the quadratic loss for most decision analyses makes the calculations relatively easy. In this case $B(\hat{\theta}(x), x)$ in Eq. (2.3.3) is the mean square error, thus,

$$\begin{aligned} B(\hat{\theta}(x), x) &= \int [\theta - \hat{\theta}(x)]^2 p_{\Theta|X=x}(\theta) d\theta \\ &= \int [\theta - E(\Theta|X=x) + (E(\Theta|X=x) - \hat{\theta}(x))]^2 p_{\Theta|X=x}(\theta) d\theta \\ &= \int (\theta - E(\Theta|X=x))^2 p_{\Theta|X=x}(\theta) d\theta + 2(E(\Theta|X=x) \\ &\quad - \hat{\theta}(x)) \int (\theta - E(\Theta|X=x)) p_{\Theta|X=x}(\theta) d\theta + (E(\Theta|X=x) - \hat{\theta}(x))^2. \end{aligned}$$

The second term vanishes, therefore

$$B(\hat{\theta}(x), x) = var(\Theta|X=x) + (E(\Theta|X=x) - \hat{\theta}(x))^2$$

which attains its minimum when $\hat{\theta}(x) = E(\Theta|X=x)$, so that a Bayes estimator $\hat{\theta}(x)$ is the posterior mean of θ and in this case $B(\hat{\theta}(x), x)$ is the posterior variance of θ. See Lee(1997) and Berger(1985).

3 Bayes invariant quadratic unbiased estimator

3.1 Formulation of BAIQUE

The observations $z(x)$ in the spatial statistics are interpreted as realizations of a stochastic process. The parameters of this process are not known beforehand and have to be estimated from the observed values. Then it is possible to predict unobserved values using these process parameters. The main idea of Bayesian kriging is to assume the unknown model parameters such as (trend, covariance) to be random. Then these parameters are distributed according to some prior distribution. The prior knowledge has to be expressed in a probabilistic form specifying this prior distribution.

This work deals with Bayesian estimation of spatial covariance functions. Bayes invariant quadratic unbiased estimator(BAIQUE) firstly has to be formulated. The procedure here regards the estimation of linear functions of the variance components from a Bayesian point of view. The restriction on Bayes quadratic unbiased estimator has some advantages. Firstly, for general linear model this procedure is rather practicable without additional assumptions. Also the BAIQUE approach depends only on the first and second moments of the prior distribution. Finally, the method gives a simple style to use with the conception of invariance and to deal with the problem of the existence of uniformly best quadratic unbiased estimator. See Kleffe and Pincus(1974).

Consider the linear model such as

$$\begin{cases} \mathrm{Z} \sim N(X\beta, V), \quad X \in \mathbb{R}^{n\times p}, \beta \in \mathbb{R}^p \\ Var(Z) = \theta_1 U_1 + \theta_2 U_2 + \ldots + \theta_r U_r = Var(e) = \sum_{i=1}^{r} \theta_i U_i = V(\theta), \end{cases} \tag{3.1.1}$$

where V the covariance matrix of the observations and $\theta = (\theta_1, \ldots, \theta_r)' \in R^r$ is unknown parameters vector. In the basic form of the Gaussian geostatistical model, the stochastic spatial process is a stationary Gaussian process with mean μ, variance σ^2 and covariance function $C(h) = Cov(e(u), e(v))$, where $h = \parallel u - v \parallel$, the Euclidean distance measure between **u** and **v** as it was shown in Eq. (2.2.4) and (2.2.5), $\mathbf{e} \sim N(0, \tau^2)$ as well. Then the distribution of Z is multivariate Gaussian $Z \sim N(\mu\mathbf{1}, \sigma^2 V + \tau^2 I)$, where **1** denotes an n-element vector of ones, I is the $n \times n$ identity matrix and V is the $n \times n$ matrix with $(i, j)^{th}$ element $C(h_{ij})$ where $h_{ij} = \parallel u_i - v_j \parallel$.

Example 3.1. In many applications one may assume the covariance matrix of the observations as $\Sigma = \kappa V(\theta)$, where κ is an unknown scale parameter and $V(\theta)$ is a

vector of standardized covariances determined by the unknown parameter vector θ. For example, consider the exponential variogram structure equivalent to a covariance function

$$Cov(Z(x), Z(x+h)) = \begin{cases} c_1 \exp(-|h|/a) & \text{if } x \neq x+h \\ c_1 + c_0 & \text{if } x = x+h, \end{cases}$$

we may define $\kappa = c_1$ and $\omega = c_0/(c_1 + c_0)$ which represents the ratio between the nugget and the sill, so $\theta = (\omega, a)$ where the parameter a stands for the range of the spatial process. Let $V(\theta)$ denote the matrix whose diagonal entries are all $1/(1-\omega)$ and off-diagonal entries are of the form $v_{ij} = \exp(-h_{ij}/a)$, where h_{ij} is the distance between the i^{th} and j^{th} sampling points.

In this context, suppose the covariance model as

$$Cov(z) = \sigma_1 V_1 + \sigma_2 I,$$

where $V_1 = \exp(-h_{ij}/a)$ and I is an identity matrix. One can suppose that $\kappa = \sigma_2$ and $\omega = \frac{\sigma_1}{\sigma_2}$. Thus one has

$$Cov(z) = \sigma_2(I + \omega V_1) = \kappa V.$$

Then it could be written corresponding to model (3.1.1) as a general case $Y \sim N(X\beta; \sigma_1 V + \sigma_2 I)$. □

Example 3.2. The 1-way classification model with n observations in each of m classes is given by

$$y_{ij} = \mu + c_i + e_{ij} \quad i = 1, \ldots, m, \quad \textit{and} \;\; j = 1, \ldots, n$$

where μ is a general mean, $\mathbf{c} = (c_1, \ldots, c_m)$ is the vector of effects, $\mathbf{e} = (e_{11}, \ldots, e_{1n}, e_{21}$ $\ldots, e_{mn})$ is the vector of residual errors, $\text{Var}(c_i) = \sigma_c^2$, $\text{Var}(e_{ij}) = \sigma_e^2$ and the vectors $\mathbf{e}$ and $\mathbf{c}$ are uncorrelated. This model is given by Searle et al.(1992) as

$$y = (1_m \bigotimes 1_n)\mu + (I_m \bigotimes 1_n)\mathbf{c} + \mathbf{e},$$

with the dispersion matrix

$$\Sigma_\theta = \sigma_{\mathbf{c}}^2(I_m \bigotimes J_n) + \sigma_{\mathbf{e}}^2(I_m \bigotimes I_n)$$

where $\theta = (\sigma^2_{\mathbf{c}}, \sigma^2_{\mathbf{e}})$ and $\mathbf{1_m}$ vector of m elements 1 and J_n is a square matrix with every element equal to one. □

In this project, the problem at hand is to estimate a linear function of the unknown vector parameter $\theta = (\theta_1, \theta_2, \ldots, \theta_r)'$, say

$$\alpha(\theta) = \mathbf{b}'\theta = \sum_{i=1}^{r} b_i\theta_i, \tag{3.1.2}$$

where the coefficients b_i are known, by using the quadratic form $\hat{\alpha}(Z) = Z'\mathcal{Q}Z$. Here $\mathcal{Q}$ is $n \times n$ matrix which has to be symmetric and we are here interested to find it. We restrict our consideration to the quadratic estimate $\hat{\alpha}(Z) = Z'\mathcal{Q}Z$ which is invariant with respect to the translation $Z \to Z + X\beta$, i.e. $\hat{\alpha}(Z) = \hat{\alpha}(Z + X\beta)$, for all β. Also it is unbiased and minimizes the Bayes risk, that is,

$$\begin{aligned} r(\hat{\alpha}) &= E(E_\Theta(\hat{\alpha}(Z) - \alpha)^2) \\ &= \int_{\theta\in\Theta} E_\Theta(\hat{\alpha}(Z) - \alpha)^2 p_\Theta(\theta)d\theta \end{aligned}$$

where p_θ is a prior distribution for the vector parameter θ having the second order moments of the form

$$E(\Theta_i\Theta_j) = \int_{\theta\in\Theta} \Theta_i\Theta_j \; p_\Theta(\theta)d\theta = C_{ij} \geq 0 \; , \quad i, j = 1, \ldots, r. \tag{3.1.3}$$

in which the integral sign here represents multi-integral with respect to the number of the parameters. In this work we assume that the prior information of the parameters is such that the matrix of second order moments is positive definite. Hence one can use the square root of the second order moments matrix, thus let

$$(C_{ij})_{i,j=1,\ldots,r} = RR = (\sum_{k=1}^{r} r_{ik}r_{kj})_{i,j=1,\ldots,r} \quad , \tag{3.1.4}$$

see Corollary (5.23) in Appendix (5.4). Here Bayes invariant quadratic estimator has to define.

Definition 3.3. A quadratic form $\hat{\alpha}(Z) = Z'\mathcal{Q}Z$ is called a Bayes invariant quadratic estimate (BAIQE) if it minimizes $E(E_\Theta(\hat{\alpha}(Z)-\alpha)^2)$ subject to all invariant quadratic forms, and we call $\hat{\alpha}(Z)$ a Bayes invariant quadratic unbiased estimate (BAIQUE) if $E(E_\Theta(\hat{\alpha}(Z) - \alpha)^2)$ is minimum subject to all invariant quadratic and unbiased estimates. See Gnot and Kleffe(1983).

Now let us introduce the projection matrix M which is expressed by the following formula

$$M = I - X(X'X)^{-1}X' \,, \tag{3.1.5}$$

and satisfies the conditions

$$M = M'; \quad M^2 = M; \quad MX = 0.$$

The model (3.1.1) contains the unknown parameters β. The intention in this work is just to estimate the vector of parameters θ, so one could filter out them by multiplying the model (3.1.1) by the projection matrix M. In this context, it is more convenient to work with generalized increments of the data than to work with the original data, see Kitanidis(1983). Thus, a linear transformation of the original data vector Z into a vector Y of stationary data increments is

$$Y = MZ = MX\beta + Me = Me \,,$$

where $MX = 0$. Doing that and by using Theorem (5.24) in Appendix (5.4), the transformed model can be given by

$$\begin{cases} Y = MZ \quad ; E(Y) = 0, \\ Var(Y) = Var(MZ) = MVar(Z)M' = \theta_1 V_1 + \theta_2 V_2 + \ldots + \theta_r V_r \end{cases} \tag{3.1.6}$$

where $V_i = MU_iM' \quad ; i = 1, \ldots, r$. Clearly $E(Y) = 0$, because of

$$E(Y) = E(MZ) = ME(Z) = MX\beta = 0.$$

Since $M = M'$ is a symmetric matrix, the relation (3.1.6) could be written as,

$$Var(Y) = MV(\theta)M = M(\sum_{i=1}^{r} \theta_i U_i)M = \sum_{i=1}^{r} \theta_i MU_iM = \sum_{i=1}^{r} \theta_i V_i.$$

From this, one can get the following model

$$Y = Me \quad ; E(Y) = 0 \quad ; Var(Y) = \sum_{i=1}^{r} \theta_i V_i = V(\theta) \quad . \tag{3.1.7}$$

The vector $Y = MZ$ is invariant statistic and $\hat{\alpha}(Z) = Z'\mathcal{Q}Z$ is BAIQUE for $\alpha = b'\theta$ in the model (3.1.1) if and only if $\mathcal{Q} = MAM$ and $\hat{\alpha}(Y) = Y'AY$ is Bayes quadratic unbiased estimator for α under the transformed model (3.1.7), see Stuchly (1989).

Now the unbiasedness of the estimator $Y'AY$ and minimization of Bayes risk in sense of this estimator have to be presented. For this purpose, the following two theorems are introduced.

Theorem 3.4. *(Unbiasedness of $Y'AY$), see Rao (1972).*
The quadratic estimator $Y'AY$ is unbiased estimator for the linear function $\alpha = b'\theta$ if and only if $trAV_i = b_i \ ; i = 1, \ldots, r$.

Proof. By taking the expectation of the quadratic form $Y'AY$ with trace operation to be written as tr,we can get

$$\begin{aligned} E(Y'AY) &= E(tr(Y'AY)) \\ &= E(trAYY') = trAE(YY') \\ &= trAVar(Y) \\ &= trA\sum_{i=1}^{r}\theta_i V_i \\ &= \sum_{i=1}^{r}\theta_i trAV_i \end{aligned}$$

is unbiased estimator for α if and only if

$$trAV_i = b_i \quad ; i = 1, \ldots, r \quad . \tag{3.1.8}$$

□

Theorem 3.5. *(Minimization Bayes risk), see Fathy and Qassim (2002).*
The quadratic form $Y'AY$ has minimum Bayes risk with respect to the prior information within all invariant, unbiased quadratic estimators if matrix A satisfies the constraints (3.1.8) and

$$\begin{pmatrix} vecV_1 & vecV_2 & \dots & vecV_r & \vdots & T \\ \dots & \dots & \dots & \dots & \vdots & \dots \\ 0 & 0 & \dots & 0 & \vdots & vecV_1' \\ 0 & 0 & \dots & 0 & \vdots & vecV_2' \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \vdots & vecV_r' \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_r \\ \dots \\ vecA \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \dots \\ b_1 \\ \vdots \\ b_r \end{pmatrix} \tag{3.1.9}$$

where $\lambda_1, \dots, \lambda_r$ are some Lagrange multipliers corresponding to the constraints (3.1.8), $b_1, \dots, b_r$ are constants, vec stands for vec operator as it defined in Appendix (5.1) and $T = \sum_{k=1}^{r} T_k \bigotimes T_k$, $T_k = \sum_{i=1}^{r} r_{ik} V_i$, r_{ik} are given by (3.1.4).

Proof. By using Theorem (5.22) (in Appendix (5.4)) of the cumulants of the quadratic form, one can find out

$$\begin{aligned} r(\hat{\alpha}) &= \int E_\theta(\hat{\alpha}(Y) - \alpha)^2 p_\Theta(\theta) d\theta \\ &= E(E_\theta(\hat{\alpha}(Y) - \alpha)^2) \\ &= E(Var_\Theta(\hat{\alpha}(Y))) = E(Var_\Theta(Y'AY)) \\ &= E(2trAV(\Theta)AV(\Theta)) = E(2trA(\sum_i \Theta_i V_i)A(\sum_i \Theta_i V_i)) \\ &= E(2\sum_i \sum_j \Theta_i \Theta_j trAV_iAV_j) \\ &= 2(\sum_i \sum_j E(\Theta_i \Theta_j) trAV_iAV_j) \\ &= 2(\sum_i \sum_j C_{ij} trAV_iAV_j). \end{aligned} \tag{3.1.10}$$

The prior information on the parameters are assumed to be available in the first and second moments $E(\Theta)$ and $E(\Theta\Theta')$ respectively, thus

$$C = E(\Theta\Theta') = Var(\Theta) + E(\Theta)E(\Theta)' \ ; \qquad C = \sqrt{C}\sqrt{C} = RR \tag{3.1.11}$$

where R is the square root of the second order moment matrix C which has been expressed in Eq. (3.1.4), see Corollary (5.23) in Appendix (5.4). From Eq.(3.1.11)

and (3.1.10) one can get

$$\begin{aligned} r(\hat{\alpha}) &= 2\sum_{i=1}^{r}\sum_{j=1}^{r}\sum_{k=1}^{r} r_{ik}r_{kj}trAV_iAV_j \\ &= 2\sum_{k=1}^{r} trA(\sum_{i=1}^{r} r_{ik}V_i)A(\sum_{j=1}^{r} r_{kj}V_j) \\ &= 2\sum_{k=1}^{r} trAT_kAT_k \end{aligned} \tag{3.1.12}$$

where $T_k = \sum_{i=1}^{r} r_{ik}V_i$. In order to minimize Bayes risk in Eq. (3.1.12), one has to use lagrange multiplier with the condition (3.1.8). So the Lagrangian function for this minimization becomes

$$S = 2\sum_{k=1}^{r} trAT_kAT_k + 4\sum_{i=1}^{r}\lambda_i(trAV_i - b_i), \tag{3.1.13}$$

where $4\lambda_i$ represent the r Lagrange multipliers for the constrains (3.1.8) of unbiasedness. The partial derivatives of the function S in (3.1.13) with respect to matrix A and λ_i by using Theorem (5.26) in Appendix (5.4) are

$$\frac{\partial S}{\partial A} = 4\sum_{k=1}^{r} T_kAT_k + 4\sum_{i=1}^{r}\lambda_iV_i = 0$$

$$\sum_{k=1}^{r} T_kAT_k + \sum_{i=1}^{r}\lambda_iV_i = 0 \tag{3.1.14}$$

$$\frac{\partial S}{\partial \lambda_i} = trAV_i - b_i$$

$$trAV_i = b_i \quad , \tag{3.1.15}$$

for $i = 1, \ldots, r$. The equations (3.1.14) and (3.1.15) have the unknown variables matrix A and λ_i. By using vec operation and Kronecker product (see the Appendix (5.1) for more detail) these equations could be expressed as the following

$$(\sum_{k=1}^{r} T_k \bigotimes T_k)vecA + \sum_{i=1}^{r}\lambda_i vecV_i = 0 \quad . \tag{3.1.16}$$

$$(vecV_i)'vecA = b_i. \tag{3.1.17}$$

Let

$$T = \sum_{k=1}^{r} T_k \bigotimes T_k.$$

So the equation (3.1.16) becomes

$$TvecA + \sum_{i=1}^{r} \lambda_i vecV_i = 0. \tag{3.1.18}$$

One can rewrite the equations (3.1.17) and (3.1.18) in terms of partitioned matrices and vectors leading to the following linear equations system

$$\begin{pmatrix} vecV_1 & vecV_2 & \dots & vecV_r & \vdots & T \\ \dots & \dots & \dots & \dots & \vdots & \dots \\ 0 & 0 & \dots & 0 & \vdots & vecV_1' \\ 0 & 0 & \dots & 0 & \vdots & vecV_2' \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \vdots & vecV_r' \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_r \\ \dots \\ vecA \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \dots \\ b_1 \\ \vdots \\ b_r \end{pmatrix}. \tag{3.1.19}$$

□

3.2 Computation BAIQUE of two variance components

Having the system corresponding to BAIQUE in general case, the intention now is to investigate and compute the equations system corresponding to BAIQUE of two spatial variance components. That means, one has to regard the case when the number of the parameters in Eq. (2.2.6) of the generalized covariance function equal two (i.e. $r = 2$). So one gets the parametric linear model of covariance function as follows

$$C(h) = \theta_1 u_1(h) + \theta_2 u_2(h), \tag{3.2.1}$$

where $u_1(h)$ is a function of the distance $h = \parallel u - v \parallel$ between points u and v while $u_2(h)$ refers to the nugget effect as it has mentioned before. $\theta = (\theta_1, \theta_2)'$ is a vector of unknown variance components $\Theta = \{\theta : \theta_1 > 0, \theta_2 \geq 0\}$. In spatial statistics θ_2 stands for a nugget effect and $\theta_1 + \theta_2$ represents the sill. We consider here this model to find the general solution and to show that the matrix A of the quadratic estimator is symmetric through solving the partitioned equations system

model corresponding to this case.

Each point in the multidimensional space is described by $x_i = (x_{i1}, x_{i2}, \ldots, x_{im})$, where x_{ik} is the projection of the object i in the axis k, $k = 1, \ldots, m$. This vector is the coordinates vector of the object i. The origin of coordinates is assumed to be $\mathbf{0_n} = (0_1, \ldots, 0_n)$. In this context, we might define the correlation function $u_1(h)$ in term of the distance between two sites i and j that belong to an Euclidean space corresponding to the length of the segment of their connecting line,

$$U_1 = D = (\parallel x_i - x_j \parallel)_{i,j=1,\ldots,n} = ((x_i - x_j)'(x_i - x_j))^{1/2}_{i,j=1,\ldots,n} \ . \tag{3.2.2}$$

Whereas $U_2 = I_{n\times n}$ is an identity matrix, in some references it is written by

$$U_2 = \delta(\parallel x_i - x_j \parallel)_{i,j=1,\ldots,n} \tag{3.2.3}$$

and called dirac delta or Kronecker delta function $\delta(h)$ as it has defined before in Eq. (2.2.7). So the covariance model of the transformed model $Y = Me$ takes the form

$$Var(Y) = \theta_1 V_1 + \theta_2 V_2 = V(\theta), \tag{3.2.4}$$

where $V_1 = MU_1M$ and $V_2 = MU_2M$. From the unbiased condition (3.1.15) of the estimator $Y'AY$, we have $E(Y'AY) = \alpha = b'\theta$, where $\mathbf{b} = (b_1, b_2)'$ are known constants and $\theta = (\theta_1, \theta_2)'$ parameters. So one obtains that

$$\begin{aligned} E(Y'AY) = trAVar(Y) &= trA\sum_{i=1}^{2}\theta_i V_i \\ &= \sum_{i=1}^{2}\theta_i trAV_i \end{aligned}$$

is unbiased estimator for α if and only if

$$trAV_i = b_i \quad ; i = 1, \ldots, 2 \quad . \tag{3.2.5}$$

To minimize Bayes risk, according to the relation (3.1.10), one has

$$\begin{aligned} r(\hat{\alpha}) &= 2(\sum_{i}^{2}\sum_{j}^{2} E(\Theta_i\Theta_j) trAV_iAV_j) \\ &= 2(\sum_{i}^{2}\sum_{j}^{2} C_{ij} trAV_iAV_j). \end{aligned} \tag{3.2.6}$$

Where

$$C = RR = (\sum_{k=1}^{2} r_{ik}r_{kj})_{i,j=1,\ldots,2} \quad . \tag{3.2.7}$$

and $T_k = \sum_{i=1}^{2} r_{ik}V_i$. Hence the Lagrangian function to minimize Bayes risk (3.2.6) with condition (3.2.5) becomes

$$S = 2\sum_{k=1}^{2} trAT_kAT_k + 4\sum_{i=1}^{2} \lambda_i(trAV_i - b_i). \tag{3.2.8}$$

Doing that the equations in (3.1.17) and (3.1.18) can be put in the following forms respectively

$$(vecV_i)'vecA = b_i, \quad i = 1, 2, \tag{3.2.9}$$

$$TvecA + \sum_{i=1}^{2} \lambda_i vecV_i = 0, \tag{3.2.10}$$

where $T = \sum_{k=1}^{2} T_k \bigotimes T_k$. Therefore the reducing model for two variance components can be expressed in the following partitioned form

$$\begin{pmatrix} vecV_1 & vecV_2 & \vdots & T \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \vdots & vecV_1' \\ 0 & 0 & \vdots & vecV_2' \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \cdots \\ vecA \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \ldots \\ b_1 \\ b_2 \end{pmatrix}. \tag{3.2.11}$$

Hence to estimate θ_1 one has to take $b = (1, 0)$, while for θ_2 one regards $b = (0, 1)$ and to find an estimate for $\theta_1 + \theta_2$ substitute $b = (1, 1)$. The solution of system (3.2.11) gives the matrix A which leads to the estimate of $\mathbf{b}'\theta$ by the quadratic form $Y'AY$. Let us represent the above model by $GO = W$, where G is the partitioned

matrix on the left-hand side of Eq.(3.2.11). Clearly, this matrix is of dimension $((n^2+2)\times(n^2+2))$. Vector W represents the right-hand side of Eq.(3.2.11) which contains n^2 zeros plus the vector $\mathbf{b}=(b_1,b_2)$, while O stands for the unknown vector containing the Lagrange multipliers λ_1, λ_2 and the elements of matrix A which we want to find for the purpose of getting the quadratic estimation $Y'AY$. If the inverse of matrix G exists, then the solution is given by $O=G^{-1}W$. Unfortunately, this inverse does not exist as we will show later in this subsection. That is why, the aim of the next subsection is to find the general solution of this system. This solution shall be presented in terms of ordinary inverse which in the same time satisfies conditions of Moore-Penrose inverse, thus such a solution is unique.

Nonsensical negative estimates
Variance estimates should be positive. They are based on a sum of squares and the square of any value, positive or negative, must also be positive. Still, nonsensical negative outcomes can occur because one estimates and corrects for error variance in a variety of statistical procedures. This variance is what one might expect on average, but it can be considerable greater or less in any specific situation. When there is less error than expected but one corrected based on what is expected, these nonsensical estimates tend to happen, see Steel and Kammeyer-Mueller (2003).

When the estimate is negative, this is not viewed as negative for the method. Negative variance component estimates should be allowed by the estimators used, because negative estimates may indicate that the statistical model is misrepresenting the true model, see Slanger (1996). In this context, Kitanidis (1985) has proposed to set simply $\hat{\theta}=0$ if the used procedure gives $\hat{\theta}<0$.

The solution obtained in the sense minimizing the quadratic criterion is still optimal even if the data are not gaussian. The estimates may not have minimum variance but are unaffected by fitted drift coefficients and are generally unbiased. See Kitanidis (1985).

Example 3.6. Consider the general linear model as follows

$$Z=\mathbf{1_n}\mu+\mathbf{e},\quad E(Z)=\mathbf{1_n}\mu,\quad Var(Z)=\theta_1 D+\theta_2 I\quad .$$

This represents the case of stationary, where $E(Z)=\mathbf{1_n}\mu$ is constant for all observations and corresponds to $X\beta=\mathbf{1_n}\mu$ in model (2.2.3), in which $X=\mathbf{1_n}$ is the vector of ones. The covariance matrix D was defined in Eq.(3.2.2) and I is an identity matrix or Kronecker delta matrix as it has defined in Eq. (2.2.7). By using a projection matrix which takes in this case the form

$$M = I - 1_n(1_n'1_n)^{-1}1_n' = I - \frac{1}{n}(1_n 1_n'),$$

M is a symmetric idempotent matrix and satisfies $MX = M1_n = 0$. Therefore the transformed model is given by

$$Y = MZ = Me, \quad E(Y) = 0, \quad Var(Y) = V_1\theta_1 + V_2\theta_2 = V(\theta),$$

where $V_1 = MDM$ and $V_2 = MIM = M$. Concerning the prior information, one could assume according to Jeffreys's prior (see Koch (1990) or Box and Tiao (1973)) that the parameters have the uniform distribution with probability density function such that

$$\begin{cases} p_1(\theta_1) = \frac{1}{\varphi_2 - \varphi_1} & ; \quad \varphi_1 \leq \theta_1 \leq \varphi_2 \\ p_2(\theta_2) = \frac{1}{\varphi_4 - \varphi_3} & ; \quad \varphi_3 \leq \theta_2 \leq \varphi_4 \end{cases}$$

where $\varphi_1,\ \varphi_2,\ \varphi_3,$ and φ_4 are constants. Let the parameters θ_1 and θ_2 are independent, so that,

$$p(\theta_1, \theta_2) = p(\theta_1) \cdot p(\theta_2), \quad Cov(\Theta_1, \Theta_2) = 0.$$

One needs only the first and second moments of the parameters, so the second order moment matrix C which represent the prior information matrix is given by

$$\begin{aligned} C &= E(\Theta)E(\Theta)' + Var(\Theta) \\ &= \begin{pmatrix} \frac{\varphi_1+\varphi_2}{2} \\ \frac{\varphi_3+\varphi_4}{2} \end{pmatrix} \begin{pmatrix} \frac{\varphi_1+\varphi_2}{2} & \frac{\varphi_3+\varphi_4}{2} \end{pmatrix} + \begin{pmatrix} \frac{(\varphi_2-\varphi_1)^2}{12} & 0 \\ 0 & \frac{(\varphi_4-\varphi_3)^2}{12} \end{pmatrix} \\ &= \begin{pmatrix} \frac{(\varphi_1+\varphi_2)^2}{4} & \frac{(\varphi_1+\varphi_2)(\varphi_3+\varphi_4)}{4} \\ \frac{(\varphi_1+\varphi_2)(\varphi_3+\varphi_4)}{4} & \frac{(\varphi_3+\varphi_4)^2}{4} \end{pmatrix} + \begin{pmatrix} \frac{(\varphi_1-\varphi_2)^2}{12} & 0 \\ 0 & \frac{(\varphi_3-\varphi_4)^2}{12} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\varphi_1^2+\varphi_1\varphi_2+\varphi_2^2}{3} & \frac{(\varphi_1+\varphi_2)(\varphi_3+\varphi_4)}{4} \\ \frac{(\varphi_1+\varphi_2)(\varphi_3+\varphi_4)}{4} & \frac{\varphi_3^2+\varphi_3\varphi_4+\varphi_4^2}{3} \end{pmatrix}. \end{aligned} \tag{3.2.12}$$

□

By taking the square root of this matrix as in Eq. (3.1.11). By doing that the partitioned equations system (3.2.11) can be formulate.

Example 3.7. Assume any design of two spatial points (u, v) and $E(Z) = \mathbf{1}\mu$. Hence the projection matrix M where $\mathbf{x} = (1, 1)'$ given by

$$\begin{aligned} M &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \mathbf{x}(\mathbf{x}'\mathbf{x})^{-\mathbf{1}}\mathbf{x} \\ &= \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}, \end{aligned}$$

while the distance matrix D between these two sites is

$$|D| = \begin{pmatrix} 0 & |u-v| \\ |u-v| & 0 \end{pmatrix} = \begin{pmatrix} 0 & \gamma_1 \\ \gamma_1 & 0 \end{pmatrix},$$

For purpose of facilitating to deal with the form of the partitioned system matrix, let $\zeta = 1/2$ and $\gamma_1 = |u - v|$, hence

$$M = \begin{pmatrix} \zeta & -\zeta \\ -\zeta & \zeta \end{pmatrix} \quad ; \quad D = \begin{pmatrix} 0 & \gamma_1 \\ \gamma_1 & 0 \end{pmatrix}.$$

Therefore the covariance matrix V_1 takes the form

$$\begin{aligned} V_1 &= \begin{pmatrix} \zeta & -\zeta \\ -\zeta & \zeta \end{pmatrix} \begin{pmatrix} 0 & \gamma_1 \\ \gamma_1 & 0 \end{pmatrix} \begin{pmatrix} \zeta & -\zeta \\ -\zeta & \zeta \end{pmatrix} \\ &= \begin{pmatrix} -2\gamma_1\zeta^2 & 2\gamma_1\zeta^2 \\ 2\gamma_1\zeta^2 & -2\gamma_1\zeta^2 \end{pmatrix}, \end{aligned}$$

put $\gamma = 2\gamma_1\zeta^2$, while $V_2 = M$. Consider the prior knowledge is the identity matrix , i.e. $R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, since $T_k = \sum_{i=1}^{2} r_{ik}V_i$ implies

$$T_1 = r_{11}V_1 = V_1 \ \textit{and} \ T_2 = r_{22}V_2 = V_2.$$

Hence one can obtain

$$\begin{aligned} T &= \begin{pmatrix} -\gamma & \gamma \\ \gamma & -\gamma \end{pmatrix} \bigotimes \begin{pmatrix} -\gamma & \gamma \\ \gamma & -\gamma \end{pmatrix} + \begin{pmatrix} \zeta & -\zeta \\ -\zeta & \zeta \end{pmatrix} \bigotimes \begin{pmatrix} \zeta & -\zeta \\ -\zeta & \zeta \end{pmatrix} \\ &= \begin{pmatrix} \alpha & -\alpha & -\alpha & \alpha \\ -\alpha & \alpha & \alpha & -\alpha \\ -\alpha & \alpha & \alpha & -\alpha \\ \alpha & -\alpha & -\alpha & \alpha \end{pmatrix}, \end{aligned}$$

where $\alpha = \gamma^2 + \zeta^2$. By substituting $vecV_1$, $vecV_2$ and T into matrix G in Eq.(3.2.11), the following matrix is obtained

$$G = \begin{pmatrix} -\gamma & \zeta & \alpha & -\alpha & -\alpha & \alpha \\ \gamma & -\zeta & -\alpha & \alpha & \alpha & -\alpha \\ \gamma & -\zeta & -\alpha & \alpha & \alpha & -\alpha \\ -\gamma & \zeta & \alpha & -\alpha & -\alpha & \alpha \\ 0 & 0 & -\gamma & \gamma & \gamma & -\gamma \\ 0 & 0 & \zeta & -\zeta & -\zeta & \zeta \end{pmatrix}.$$

It is clear this matrix is singular since their columns and rows are linearly dependent. This partitioned matrix can be represented in other form as follows

$$a = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}, \quad b = \begin{pmatrix} -\gamma \\ \zeta \end{pmatrix},$$

thus the matrix G is given by $\quad G = \begin{pmatrix} ab' & \alpha aa' \\ 0_{2\times 2} & ba' \end{pmatrix}$.
The least square solution $O = (G'G)^{-1}G'W$ here fails since $G'G$ is singular. That can be noted as follows

$$\begin{aligned} G'G &= \begin{pmatrix} ab' & \alpha aa' \\ 0_{2\times 2} & ba' \end{pmatrix}' \begin{pmatrix} ab' & \alpha aa' \\ 0_{2\times 2} & ba' \end{pmatrix} \\ &= \begin{pmatrix} ba'(ab') & ba'(\alpha aa') + 0_{2\times 2}(ba') \\ \alpha a'a(ab') + ab'(0_{2\times 2}) & \alpha a'a(\alpha aa') + ab'(ba') \end{pmatrix} \\ &= \begin{pmatrix} \zeta_1 & -\zeta_2 & -\zeta_3 & \zeta_3 & \zeta_3 & -\zeta_3 \\ -\zeta_2 & \zeta_4 & \zeta_5 & -\zeta_5 & -\zeta_5 & \zeta_5 \\ -\zeta_3 & \zeta_5 & \zeta_6 & -\zeta_6 & -\zeta_6 & \zeta_6 \\ \zeta_3 & -\zeta_5 & -\zeta_6 & \zeta_6 & \zeta_6 & -\zeta_6 \\ \zeta_3 & -\zeta_5 & -\zeta_6 & \zeta_6 & \zeta_6 & -\zeta_6 \\ -\zeta_3 & \zeta_5 & \zeta_6 & -\zeta_6 & -\zeta_6 & \zeta_6 \end{pmatrix}, \end{aligned}$$

where $\quad 4\gamma^2 = \zeta_1$, $-4\gamma\zeta = \zeta_2$, $-4\gamma\alpha = \zeta_3$, $4\zeta^2 = \zeta_4$, $4\zeta\alpha = \zeta_5$, and $4\alpha^2 + \gamma^2 + \zeta^2 = \zeta_6$.
It is easy to recognize that this matrix is also singular. □

It is useful to give numerical examples to realize more how this system does work and to find out the matrix of the BAIQUE.

Example 3.8. Suppose a one dimensional design of observations at $1, 3, 5$ with the design matrix $X = \mathbf{1}$, assume that the covariance model is

$$Var(Z) = \theta_1(\exp(-|D|) + \theta_2 I_{3\times 3}$$

where matrix $D = (h_{ij})_{i,j=1,2,3}$ represents the matrix of Euclidean distances and exp stands for the exponential function, so one can find

$$D = (h_{ij})_{i,j=1,2,3} = \begin{pmatrix} 0 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 0 \end{pmatrix},$$

and

$$exp(|-D|) = \begin{pmatrix} 1.0000 & 0.1353 & 0.0183 \\ 0.1353 & 1.0000 & 0.1353 \\ 0.0183 & 0.1353 & 1.0000 \end{pmatrix},$$

while

$$M = I - X(X'X)^{-1}X' = \begin{pmatrix} 0.6667 & -0.3333 & -0.3333 \\ -0.3333 & 0.6667 & -0.3333 \\ -0.3333 & -0.3333 & 0.6667 \end{pmatrix}.$$

Doing that matrices V_1 and V_2 corresponding to the transformed covariance model are

$$V_1 = \begin{pmatrix} 0.6285 & -0.2752 & -0.3532 \\ -0.2752 & 0.5504 & -0.2752 \\ -0.3532 & -0.2752 & 0.6285 \end{pmatrix}$$

and matrix $V_2 = MIM = M$. Suppose the uniform prior information on the parameters is given by

$$\begin{cases} p_1(\theta_1) = 0.25 & ; \quad 1 \leq \theta_1 \leq 5 \\ p_2(\theta_2) = 0.33 & ; \quad 0 \leq \theta_2 \leq 3 \end{cases},$$

Θ_1 and Θ_2 are independent. Thus, the prior information matrix C and its square root matrix are

$$C = \begin{pmatrix} 127.0271 & 59.9998 \\ 59.9998 & 29.2500 \end{pmatrix}, \quad R = \begin{pmatrix} 10.3333 & 4.5000 \\ 4.5000 & 3.0000 \end{pmatrix}.$$

respectively. The eigenvalues of matrix C are 0.7430 and 155.5341 while for matrix R are 0.8620 and 12.4713. In this context, Matlab program has used to compute the square root of the symmetric matrix.

So one can find the solution $O_{11\times 1} = G^{-}_{11\times 11} W_{11\times 1}$ where G^{-} stands for one generalized inverse of matrix G. Consider the case when $\mathbf{b} = (1,1)'$, then the solution of O in Eq.(3.2.11) gives

$$O = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ vecA \end{pmatrix} = \begin{pmatrix} -15.3284 \\ 12.8534 \\ 0.6217 \\ 0.2478 \\ -0.8696 \\ 0.1217 \\ 0.1304 \\ -0.2522 \\ -0.7435 \\ 0 \\ 0 \end{pmatrix}.$$

After deleting Lagrange multipliers λ_1 and λ_2 from the solution, matrix A is obtained as

$$A = \begin{pmatrix} 0.6217 & 0.2478 & -0.8696 \\ 0.1217 & 0.1304 & -0.2522 \\ -0.7435 & 0 & 0 \end{pmatrix}.$$

□

One can note that the matrix A is not symmetric whereas the quadratic estimator $Y'AY$ requires that the matrix A should be symmetric. It is clear that the symmetric property of this matrix is very important to get the Bayes invariant quadratic estimate by the quadratic form $Y'AY$. For this reason. The interest in this work to present the necessary assumptions for the model to achieve the symmetric matrix A. Consider now other numerical example which deals with Moore-penrose inverse to find the solution of the model.

Example 3.9. Consider the two dimensional design with observations at $(1,1), (7,3)$, $(5,6), (2,9)$ and take matrices U_1 and U_2 as in example (3.6). Hence one can obtain

$$D = (\parallel x_i - x_j \parallel)_{i,j=1,\ldots,4} = \begin{pmatrix} 0 & 6.3246 & 6.4031 & 8.0623 \\ 6.3246 & 0 & 3.6056 & 7.8102 \\ 6.4031 & 3.6056 & 0 & 4.2426 \\ 8.0623 & 7.8102 & 4.2426 & 0 \end{pmatrix}.$$

In this design $f(\mathbf{x_1}, \mathbf{x_2}) = (\mathbf{1}, \mathbf{x_1}, \mathbf{x_2})'$, therefore

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 7 & 3 \\ 1 & 5 & 6 \\ 1 & 2 & 9 \end{pmatrix},$$

$$M = \begin{pmatrix} 0.0027 & -0.0243 & 0.0413 & -0.0198 \\ -0.0243 & 0.2184 & -0.3721 & 0.1780 \\ 0.0413 & -0.3721 & 0.6339 & -0.3032 \\ -0.0198 & 0.1780 & -0.3032 & 0.1450 \end{pmatrix}.$$

The covariance matrices of the filtered model are

$$V_1 = MDM = \begin{pmatrix} -0.0069 & 0.0624 & -0.1063 & 0.0509 \\ 0.0624 & -0.5618 & 0.9571 & -0.4577 \\ -0.1063 & 0.9571 & -1.6306 & 0.7799 \\ 0.0509 & -0.4577 & 0.7799 & -0.3730 \end{pmatrix}$$

and $V_2 = M$, assume the prior information matrix is given by the identity matrix $I_{2\times 2}$, so matrices $T_{16\times 16}$ and $G_{18\times 18}$ can be found. By using the Moore-penerose inverse of matrix G and by taking the case $b = (1, 0)'$, the symmetric A matrix can be obtained

$$A = \begin{pmatrix} -0.0009 & 0.0082 & -0.0140 & 0.0067 \\ 0.0082 & -0.0738 & 0.1257 & -0.0601 \\ -0.0140 & 0.1257 & -0.2141 & 0.1024 \\ 0.0067 & -0.0601 & 0.1024 & -0.0490 \end{pmatrix}.$$

□

3.3 The symmetry of the matrix in BAIQUE

Now the matrix of the quadratic estimator $Y'AY$ has to be regarded. This matrix has an essential effect on the estimation achieved by Bayesian invariant quadratic unbiased estimators of spatial covariance components which is the interested object in this work. Therefore, we are trying to show that this matrix takes the symmetric form. This property is very important to obtain Bayes estimator of quadratic form which requires that this matrix has symmetric form. As we have seen in the last subsection the linear equations system (3.2.11) has the singularity property, therefore, the least square solution of this model which is $O = (G'G)^{-1}G'W$ fails to find out a solution since $G'G$ is singular. For this reason we consider Moore-Penrose inverse of matrix $G'G$ which appears here in the sense of the ordinary inverse form

to obtain the unique solution, such a solution has the minimal norm and enables to achieve the symmetric matrix of the quadratic estimator. Hence, in order to reach this aim we shall introduce some lemmas and conditions which help to realize the necessary assumptions for this purpose.

Lemma 3.10. *Let $Z = KK' \in R^{p\times p}$ of rank r, where $K'K \in R^{r\times r}$ is nonsingular, and $B \in R^{p\times p}$ nonsingular symmetric matrix. Then the Moore-Penrose inverse of ZBZ is given by*

$$(ZBZ)^- = K(K'K)^{-1}(K'BK)^{-1}(K'K)^{-1}K'.$$

Proof. In order to prove that

$$(ZBZ)^- = K(K'K)^{-1}(K'BK)^{-1}(K'K)^{-1}K'$$

is the Moore-Penrose inverse of matrix ZBZ, the conditions of Moore-Penrose inverse have to be satisfied, thus one can obtain

$$\begin{aligned}(ZBZ)(ZBZ)^-(ZBZ) &= (KK'BKK')(K(K'K)^{-1}(K'BK)^{-1}(K'K)^{-1}K')(KK'BKK')\\ &= KK'BK(K'BK)^{-1}K'BKK'\\ &= KK'BKK'\\ &= ZBZ.\end{aligned}$$

Thus $K(K'K)^{-1}(K'BK)^{-1}(K'K)^{-1}K'$ is g-inverse of (ZBZ) matrix. To satisfy the other conditions of Penrose inverse, one can find out

$$\begin{aligned}(ZBZ)^-(ZBZ)(ZBZ)^- &= (K(K'K)^{-1}(K'BK)^{-1}(K'K)^{-1}K')(KK'BKK')(K(K'K)^{-1}(K'BK)^{-1}\\ &\quad (K'K)^{-1}K')\\ &= K(K'K)^{-1}(K'BK)^{-1}(K'BK)(K'BK)^{-1}(K'K)^{-1}K'\\ &= K(K'K)^{-1}(K'BK)^{-1}(K'K)^{-1}K'\\ &= (ZBZ)^-.\end{aligned}$$

For the third condition, we have

$$\begin{aligned}((ZBZ)(ZBZ)^{-})' &= ((KK'BKK')(K(K'K)^{-1}(K'BK)^{-1}(K'K)^{-1}K'))' \\ &= (K(K'K)^{-1}(K'BK)^{-1}(K'K)^{-1}K')(KK'BKK') \\ &= K(KK')^{-1}(K'BK)^{-1}K'BKK' \\ &= K(K'K)^{-1}K' \\ &= K(K'BK)(K'BK)^{-1}(K'K)^{-1}K' \\ &= K(K'BK)(K'K)(K'K)^{-1}(K'BK)^{-1}(K'K)^{-1}K' \\ &= (KK'BKK')(K(K'K)^{-1}(K'BK)^{-1}(K'K)^{-1}K') \\ &= (ZBZ)(ZBZ)^{-}.\end{aligned}$$

Therefore

$$((ZBZ)(ZBZ)^{-})' = (ZBZ)(ZBZ)^{-}$$

And similarity one can get $((ZBZ)^{-}(ZBZ))' = (ZBZ)^{-}(ZBZ)$. Thus

$$(ZBZ)^{-} = K(K'K)^{-1}(K'BK)^{-1}(KK')^{-1}K'$$

is Penrose inverse of ZBZ.

□

Lemma 3.11. *Let $H = Z\mathcal{C}Z$, where $Z = M \bigotimes M$, M is a symmetric idempotent matrix satisfying the conditions of Lemma(3.10) and $u = Za$, where $a = vec(\mathcal{J})$ is the vec operation for any matrix $\mathcal{J}$, then*

$$(H + uu')^{-} = H^{-} - \frac{H^{-}uu'H^{-}}{1 + u'H^{-}u},$$

where $u'H^{-}u \neq -1$.

Proof. From Sherman-Morrison-Woodbury Corollary (5.7), see Appendix (5.2), in general if U^{-} is a generalized inverse of matrix U and for a vector u, we set that $u = b = c$ is a vector of $p \times 1$ satisfying $HH^{-}u = u$ and $u'HH^{-} = u'$ and $\delta = 1$, then one gets that

$$(H + uu')^{-} = H^{-} - \frac{H^{-}uu'H^{-}}{1 + u'H^{-}u}.$$

Thus the lemma is proved.

In this context, one can show that $HH^{-}u = u$ and $u'HH^{-} = u'$. Since that matrix M is an idempotent thus matrix $M \bigotimes M$ is also idempotent, this can be shown as follows

$$\begin{aligned}(M \bigotimes M)^2 &= (M \bigotimes M)(M \bigotimes M) \\ &= (M^2 \bigotimes M^2) \\ &= M \bigotimes M.\end{aligned}$$

If H^{-} is Moore-Penrose inverse of matrix H and matrix $\mathcal{C}$ is a nonsingular symmetric then one gets by using Lemma (3.10) that

$$\begin{aligned}HH^{-} &= Z\mathcal{C}Z(Z\mathcal{C}Z)^{-} \\ &= KK'\mathcal{C}KK'(K(K'K)^{-1}(K'\mathcal{C}K)^{-1}(K'K)^{-1}K') \\ &= KK'\mathcal{C}K(K'\mathcal{C}K)^{-1}(K'K)^{-1}K' \\ &= K(K'K)^{-1}K' \\ &= KK' = Z,\end{aligned}$$

where $K'K = I$ since that Z is a symmetric idempotent matrix, see Lemma (5.8) in Appendix (5.2), also one can find

$$\begin{aligned}H^{-}H &= (Z\mathcal{C}Z)^{-}Z\mathcal{C}Z \\ &= (K(K'K)^{-1}(K'\mathcal{C}K)^{-1}(K'K)^{-1}K')KK'\mathcal{C}KK' \\ &= K(K'K)^{-1}(K'\mathcal{C}K)^{-1}K'\mathcal{C}KK' \\ &= K(K'K)^{-1}K' \\ &= KK' = Z.\end{aligned}$$

Therefore one obtains

$$\begin{aligned} HH^{-}u &= KK'u \\ &= Zu \\ &= (M \bigotimes M)vec(M\mathcal{J}M) \\ &= vec(M^2\mathcal{J}M^2) \\ &= vec(M\mathcal{J}M) \\ &= u, \end{aligned}$$

and also

$$\begin{aligned} u'HH^{-} &= vec(M\mathcal{J}M)'Z \\ &= vec(M\mathcal{J}M)'M \bigotimes M \\ &= vec(M^2\mathcal{J}M^2)' \\ &= vec(M\mathcal{J}M)' \\ &= u'. \end{aligned}$$

□

Lemma 3.12. *If $H = Z\mathcal{C}Z$, $u = Za$, and $v = Zb$, where Z satisfies the conditions of Lemma (3.10), then*

$$(H + uu' + vv')^{-} = \frac{cH^{-} - H^{-}uu'H^{-}}{c} - \frac{(cH^{-} - H^{-}uu'H^{-})vv'(cH^{-} - H^{-}uu'H^{-})}{c^2 d}$$

Where $c = 1 + u'H^{-}u$ and $d = 1 + v'(H + uu')^{-}v$ are constants not equal zero.

Proof. Let $A = H + uu'$ then from Lemma (3.11) one gets

$$(A + vv')^{-} = A^{-} - \frac{A^{-}vv'A^{-}}{1 + v'A^{-}v}.$$

Now by making substitution of matrix A, one can find out,

$$
\begin{aligned}
(A+vv')^- &= A^- - \frac{A^-vv'A^-}{1+v'A^-v} \\
&= (H+uu')^- - \frac{(H+uu')^-vv'(H+uu')^-}{1+v'(H+uu')^-v} \\
&= (H^- - \frac{H^-uu'H^-}{1+u'H^-u}) - \frac{(H^- - \frac{H^-uu'H^-}{1+u'H^-u})vv'(H^- - \frac{H^-uu'H^-}{1+u'H^-u})}{1+v'(H+uu')^-v} \\
&= \frac{cH^- - H^-uu'H^-}{c} - \frac{(cH^- - H^-uu'H^-)vv'(cH^- - H^-uu'H^-)}{c^2d}
\end{aligned}
$$

where c and d are constants not equal zero such that $c = 1 + u'H^-u$ and $d = 1 + v'(H+uu')^-v$. □

Since the interested aim here is obtaining the solution of matrix A from the equations system of BAIQUE, it is useful to describe the general solution of this system by means of partitioned generalized inverse. This representation helps to find the form of the matrix A of the quadratic estimation and further it helps to show the symmetric of this matrix. Therefore, we give the following lemma to constitute the general solution through the partitioned generalized inverse.

Lemma 3.13. *Let*

$$
G = \begin{pmatrix} u & v & \vdots & T \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \vdots & u' \\ 0 & 0 & \vdots & v' \end{pmatrix}, \quad W = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ .... \\ b_1 \\ b_2 \end{pmatrix}
$$

where $u = vecU$, $v = vecV$ and matrices U, $V \in R^{p\times p}$, $T = T_1 \bigotimes T_1 + T_2 \bigotimes T_2 \in R^{p^2\times p^2}$ are symmetric, moreover b_1, b_2 are constants. The solution of the equations system $GO = W$ can be given in form of partitioned generalized inverse by

$$
O = (G'G)^-G'W = \begin{pmatrix} -(S^-BC^-)L \\ (C^- + C^-B'S^-BC^-)L \end{pmatrix}, \tag{3.3.1}
$$

where $S_{2\times 2}$ is the Schur complement of matrix C, i.e.

$$
S_{2\times 2} = A - BC^-B', \quad A = \begin{pmatrix} u'u & u'v \\ v'u & v'v \end{pmatrix}
$$

$$B_{2\times p^2}=\begin{pmatrix}u'T\\ v'T\end{pmatrix}, C_{p^2\times p^2}=T^2+uu'+vv' \;\; and \;\; L_{p^2\times 1}=b_1u+b_2v.$$

Proof. Firstly, one can find the matrix $G'G$ which expressible in the partitioned form as follows

$$G'G=\begin{pmatrix} u' & \vdots & 0 & 0\\ v' & \vdots & 0 & 0\\ \cdots & \cdots & \cdots & \cdots\\ T' & \vdots & u & v\end{pmatrix}\begin{pmatrix} u & v & \vdots & T\\ \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & \vdots & u'\\ 0 & 0 & \vdots & v'\end{pmatrix}$$
$$=\begin{pmatrix} u'u & u'v & \vdots & u'T\\ v'u & v'v & \vdots & v'T\\ \cdots & \cdots & \cdots & \cdots\cdots\\ Tu & Tv & \vdots & T^2+uu'+vv'\end{pmatrix}.$$

Let $H=T^2$, suppose the partitioned matrices as following
$A=\begin{pmatrix}u'u & u'v\\ v'u & v'v\end{pmatrix}_{2\times 2}$, $B=\begin{pmatrix}u'T\\ v'T\end{pmatrix}_{2\times p^2}$ and $C=\left(H+uu'+vv'\right)_{p^2\times p^2}$.
This implies to

$$G'G=\begin{pmatrix} A_{2\times 2} & \vdots & B_{2\times p^2}\\ \cdots & \cdots & \cdots\\ B'_{p^2\times 2} & \vdots & C_{p^2\times p^2}\end{pmatrix}.$$

Doing that the general case of the partitioned generalized inverse $(G'G)^-$ is given by

$$(G'G)^-=\begin{pmatrix} S^- & \vdots & -S^-BC^-\\ \cdots & \cdots & \cdots\\ -C^-B'S^- & \vdots & C^-+C^-B'S^-BC^-\end{pmatrix},$$

where $S=A-BC^-B'$ is Schur complement of matrix C, see the Appendix (5.2) . Since the partitioned matrix $G'G$ is symmetric then its generalized inverse $(G'G)^-$ is also has the symmetric form. Since that

$$W = \begin{pmatrix} 0_{p^2} \\ \cdots \\ b_1 \\ b_2 \end{pmatrix},$$

where b_1 and b_2 are constants. So one can find

$$\begin{aligned} G'W &= \begin{pmatrix} u' & \vdots & 0 & 0 \\ v' & \vdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ T & \vdots & u & v \end{pmatrix} \begin{pmatrix} O_{p^2} \\ \cdots \\ b_1 \\ b_2 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \\ \cdots \\ (b_1 u + b_2 v) \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \\ \cdots \\ L_{p^2 \times 1} \end{pmatrix}, \end{aligned}$$

where $L = b_1 u + b_2 v$. Therefore, the solution of the equations system through partitioned generalized inverse is given by

$$\begin{aligned} O = (G'G)^- G'W &= \begin{pmatrix} S^- & \vdots & -S^- BC^- \\ \cdots & \cdots & \cdots \\ -C^- B' S^- & \vdots & C^- + C^- B' S^- BC^- \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \cdots \\ L_{p^2 \times 1} \end{pmatrix} \\ &= \begin{pmatrix} -(S^- BC^-)L \\ (C^- + C^- B' S^- BC^-)L \end{pmatrix}. \end{aligned}$$

□

Lemma 3.14. *If $T = M \bigotimes M\mathcal{B}M \bigotimes M$ and $u = vecU$, $v = vecV$, where $\mathcal{B}$ is a symmetric nonsingular matrix and M is a symmetric idempotent matrix, then the solution of Lemma (3.13) has the form*

$$O = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ vec(A) \end{pmatrix},$$

where

$$vec(A) = H^{-}(\tilde{\zeta}_1 u + \tilde{\zeta}_2 v + \lambda_1 Tu + \lambda_2 Tv), \tag{3.3.2}$$

in which $H = T^2$, $\tilde{\zeta}_1$, $\tilde{\zeta}_2$, λ_1 *and* λ_2 *are constants.*

Proof. By using the assumptions of Lemma (3.12), put that

$$T = Z\mathcal{C}Z = M \bigotimes M\mathcal{B}M \bigotimes M,$$

where $Z = M \bigotimes M$ and $\mathcal{C} = \mathcal{B}$. Then one can obtain

$$(H+uu'+vv')^{-} = (\frac{cH^{-} - H^{-}uu'H^{-}}{c}) - \frac{(cH^{-} - H^{-}uu'H^{-})vv'(cH^{-} - H^{-}uu'H^{-})}{c^2 d},$$

where $c = 1 + u'H^{-}u$, $d = 1 + v'(H + uu')^{-}v$ are constants. Hence to find out the solution of the equations system, one has

$$\begin{aligned} -(S^{-}BC^{-})L &= -S^{-}\begin{pmatrix} u'TC^{-} \\ v'TC^{-} \end{pmatrix}(b_1 u + b_2 v) \\ &= -S^{-}_{2\times 2}\begin{pmatrix} (b_1 u'TC^{-}u + b_2 u'TC^{-}v) \\ (b_1 v'TC^{-}u + b_2 v'TC^{-}v) \end{pmatrix}_{2\times 1} \\ &= -\begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix}\begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} \\ &= -\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \end{aligned}$$

where $\delta_1 = (b_1 u'TC^{-}u + b_2 u'TC^{-}v)$, $\delta_2 = (b_1 v'TC^{-}u + b_2 v'TC^{-}v)$, $\lambda_1 = (s_{11}\delta_1 + s_{12}\delta_2)$ and $\lambda_2 = (s_{21}\delta_1 + s_{22}\delta_2)$ are represent Lagrange multipliers in this system. Now Lemma (3.12) has to be used in order to handle with the second part of the least squares solution in Eq. (3.3.1) of Lemma (3.13) which supplies the solution of matrix A. Since matrix $C = (H + uu' + vv')$ in the equations system (3.3.1), thus

one can obtain

$$\begin{aligned}
C^{-}L &= (\frac{cH^{-}-H^{-}uu'H^{-}}{c})L-(\frac{(cH^{-}-H^{-}uu'H^{-})vv'(cH^{-}-H^{-}uu'H^{-})}{c^2d})L\\
&=(\frac{cH^{-}-H^{-}uu'H^{-}}{c})L\\
&-(\frac{c^2H^{-}vv'H^{-}-cH^{-}vv'H^{-}uu'H^{-}-cH^{-}uu'H^{-}vv'H^{-}+H^{-}uu'H^{-}vv'H^{-}uu'H^{-}}{c^2d})L\\
&=\frac{(cH^{-})L}{c}-\frac{(H^{-}uu'H^{-})L}{c}-\frac{(c^2H^{-}vv'H^{-})L}{c^2d}+\frac{(cH^{-}vv'H^{-}uu'H^{-})L}{c^2d}\\
&+\frac{(cH^{-}uu'H^{-}vv'H^{-})L}{c^2d}-\frac{(H^{-}uuH^{-}vv'H^{-}uu'H^{-})L}{c^2d}\\
&=H^{-}(L-\frac{c_1}{c}u-\frac{c_2}{d}v+\frac{kc_1}{cd}v+\frac{kc_2}{cd}u-\frac{k^2c_1}{c^2d}u)\\
&=H^{-}(b_1u+b_2v-\frac{c_1}{c}u-\frac{c_2}{d}v+\frac{kc_1}{cd}v+\frac{kc_2}{cd}u-\frac{k^2c_1}{c^2d}u)\\
&=H^{-}(b_1-\frac{c_1}{c}+\frac{kc_2}{cd}-\frac{k^2c_1}{c^2d})u+(b_2-\frac{c_2}{d}+\frac{kc_1}{cd})v\\
&=H^{-}(\zeta_1u+\zeta_2v)
\end{aligned}$$

where $c_1 = u'H^{-}L$, $c_2 = v'H^{-}L$, $k = v'H^{-}u$, $\zeta_1 = (b_1 - \frac{c_1}{c} + \frac{kc_2}{cd} - \frac{k^2c_1}{c^2d})$ and $\zeta_2 = (b_2 - \frac{c_2}{d} + \frac{kc_1}{cd})$. Moreover, by using Lemma (3.12) one can get

$$\begin{aligned}
(C^{-}B'S^{-1}BC^{-})L &= C^{-}B'\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix}\\
&=C^{-}\begin{pmatrix}Tu & Tv\end{pmatrix}\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix}\\
&=C^{-}\left(\lambda_1Tu+\lambda_2Tv\right)\\
&=C^{-}(\lambda_1\tilde{u}+\lambda_2\tilde{v})\\
&=C^{-}\tilde{L}\\
&=\frac{(cH^{-})\tilde{L}}{c}-\frac{(H^{-}uu'H^{-})\tilde{L}}{c}-\frac{(c^2H^{-}vv'H^{-})\tilde{L}}{c^2d}+\frac{(cH^{-}vv'H^{-}uu'H^{-})\tilde{L}}{c^2d}\\
&+\frac{(cH^{-}uu'H^{-}vv'H^{-})\tilde{L}}{c^2d}-\frac{(H^{-}uuH^{-}vv'H^{-}uu'H^{-})\tilde{L}}{c^2d}\\
&=H^{-}(\tilde{L}-\frac{d_1}{c}u-\frac{d_2}{d}v+\frac{kd_1}{cd}v+\frac{kd_2}{cd}u-\frac{k^2d_1}{c^2d}u)\\
&=H^{-}(\lambda_1\tilde{u}+\lambda_2\tilde{v}-\frac{d_1}{c}u-\frac{d_2}{d}v+\frac{kd_1}{cd}v+\frac{kd_2}{cd}u-\frac{k^2d_1}{c^2d}u)\\
&=H^{-}(\lambda_1\hat{u}+\lambda_2\hat{v}+(-\frac{d_1}{c}+\frac{kd_2}{cd}-\frac{k^2d_1}{c^2d})u+(-\frac{d_2}{d}+\frac{kd_1}{cd})v)\\
&=H^{-}(\lambda_1\tilde{u}+\lambda_2\tilde{v}+\hat{\zeta}_1u+\hat{\zeta}_2v)
\end{aligned}$$

where
$\tilde{L} = \lambda_1\tilde{u} + \lambda_2\tilde{v}$, $\tilde{u} = Tu$, $\tilde{v} = Tv$, $d_1 = u'H^-\tilde{L}$, $d_2 = v'H^-\tilde{L}$,
$k = v'H^-u$, $\hat{\zeta}_1 = (-\frac{d_1}{c} + \frac{kd_2}{cd} - \frac{k^2d_1}{c^2d})$ and $\hat{\zeta}_2 = (-\frac{d_2}{d} + \frac{kd_1}{cd})$.
Therefore, one can find out

$$\begin{aligned}(C^- + C^-B'S^{-1}BC^-)L &= C^-L + (C^-B'S^{-1}BC^-)L \\ &= H^-(\zeta_1u + \zeta_2v) + H^-(\lambda_1\tilde{u} + \lambda_2\tilde{v} + \hat{\zeta}_1u + \hat{\zeta}_2v) \\ &= H^-((\zeta_1 + \hat{\zeta}_1)u + (\zeta_2 + \hat{\zeta}_2)v + \lambda_1\tilde{u} + \lambda_2\tilde{v}) \\ &= H^-(\tilde{\zeta}_1u + \tilde{\zeta}_2v + \lambda_1\tilde{u} + \lambda_2\tilde{v}),\end{aligned}$$

where $\tilde{\zeta}_1 = (\zeta_1 + \hat{\zeta}_1)$ and $\tilde{\zeta}_2 = (\zeta_2 + \hat{\zeta}_2)$. Since that $H_{p^2\times p^2} = T^2$, $T = T_1 \bigotimes T_1 + T_2 \bigotimes T_2$ and $(\tilde{\zeta}_1u+\tilde{\zeta}_2v+\lambda_1\tilde{u}+\lambda_2\tilde{v})_{p^2\times 1}$ is a vector which equivalent to vec operation, this implies

$$\begin{aligned}vec(A) &= H^-(\tilde{\zeta}_1u + \tilde{\zeta}_2v + \lambda_1\tilde{u} + \lambda_2\tilde{v}) \\ &= (T^2)^-(\tilde{\zeta}_1u + \tilde{\zeta}_2v + \lambda_1\tilde{u} + \lambda_2\tilde{v}) \\ &= ((T_1 \bigotimes T_1 + T_2 \bigotimes T_2)^2)^-(\tilde{\zeta}_1u + \tilde{\zeta}_2v + \lambda_1Tu + \lambda_2Tv).\end{aligned}$$

□

Lemma 3.15. *Let $M = KK'$ with $K'K = I_r$, $T_1 = MIM$, where $I = I_n$ is an identity matrix, $T_2 = MJM$, $I \bigotimes I + J \bigotimes J$ nonsingular, $\lim_{r\to\infty}(K'JK)^k = 0$. Then*

$$(T_1 \bigotimes T_1 + T_2 \bigotimes T_2)^- = \sum_{k=0}^{\infty}(-1)^kK(K'JK)^kK' \bigotimes K(K'JK)^kK'.$$

Proof. We have from Lemma (3.10) that

$$(ZBZ)^- = K(K'K)^{-1}(K'BK)^{-1}(K'K)^{-1}K'.$$

Since matrix M can be written as full rank factorization $M = KK'$ and $K'K = I$ where K is a matrix of full column rank. That implies

$$M \bigotimes M = KK' \bigotimes KK' = K \bigotimes K \cdot K' \bigotimes K',$$

and

$$K' \bigotimes K' \cdot K \bigotimes K = K'K \bigotimes K'K = I \bigotimes I.$$

Moreover, since the matrices $T_1 = MIM$ and $T_2 = MJM$, hence one can get

$$\begin{aligned} T_1 \bigotimes T_1 + T_2 \bigotimes T_2 &= MIM \bigotimes MIM + MJM \bigotimes MJM \\ &= (M \bigotimes M)(I \bigotimes I)(M \bigotimes M) + (M \bigotimes M)(J \bigotimes J)(M \bigotimes M) \\ &= M \bigotimes M(I \bigotimes I + J \bigotimes J)M \bigotimes M. \end{aligned}$$

Since that $\lim_{k\to\infty}(K'JK)^k = 0$ by using Lemma (3.10) and the Theorem (5.11) of the infinite series see Appendix (5.3), one can find out

$$\begin{aligned} &(T_1 \bigotimes T_1 + T_2 \bigotimes T_2)^- \\ &= ((M \bigotimes M)(I \bigotimes I + J \bigotimes J)(M \bigotimes M))^- \\ &= K \bigotimes K(K' \bigotimes K' \cdot K \bigotimes K)^{-1}[K' \bigotimes K'(I \bigotimes I + J \bigotimes J)K \bigotimes K]^{-1}(K' \bigotimes K' \cdot K \bigotimes K)^{-1} \\ &K' \bigotimes K' \\ &= K \bigotimes K(I \bigotimes I)^{-1}[K' \bigotimes K'(I \bigotimes I + J \bigotimes J)K \bigotimes K]^{-1}(I \bigotimes I)^{-1}K' \bigotimes K' \\ &= (K \bigotimes K)(K'K \bigotimes K'K + K'JK \bigotimes K'JK)^{-1}(K' \bigotimes K') \\ &= (K \bigotimes K)(I \bigotimes I + K'JK \bigotimes K'JK)^{-1}(K' \bigotimes K') \\ &= (K \bigotimes K)\sum_{k=0}^{\infty}(-1)^k((K'JK \bigotimes K'JK)^k)(K' \bigotimes K') \\ &= \sum_{k=0}^{\infty}(-1)^k K(K'JK)^k K' \bigotimes K(K'JK)^k K'. \end{aligned}$$

□

Lemma 3.16. *Let $Z = KK' \in R^{p\times p}$ a symmetric idempotent matrix of rank r, where $K'K = I \in R^{r\times r}$ is identity matrix, and $B \in R^{p\times p}$ nonsingular symmetric matrix. Then the Moore-Penrose inverse of $(ZBZ)^2$ is given by*

$$((ZBZ)^2)^- = K(K'BK)^{-2}K' = K[(K'BK)^{-1}]^2K'.$$

Proof. Since that matrix $Z = KK'$ and $K'K = I_r$, so one can obtain

$$\begin{aligned}(ZBZ)^2 &= (KK'BKK')(KK'BKK') \\ &= KK'BKK'BKK' \\ &= K(K'BK)^2K'.\end{aligned}$$

Now to prove that $((ZBZ)^2)^- = K(K'BK)^{-2}K'$, one needs to show that the Moore-Penrose inverse conditions here are satisfied, that can be done as follows

$$\begin{aligned}(ZBZ)^2((ZBZ)^2)^-(ZBZ)^2 &= (K(K'BK)^2K')(K(K'BK)^{-2}K')(K(K'BK)^2K') \\ &= K(K'BK)^2(K'BK)^{-2}(K'BK)^2K' \\ &= K(K'BK)^2K' \\ &= (ZBZ)^2.\end{aligned}$$

$$\begin{aligned}((ZBZ)^2)^-(ZBZ)^2((ZBZ)^2)^- &= (K(K'BK)^{-2}K')(K(K'BK)^2K')(K(K'BK)^{-2}K') \\ &= K(K'BK)^{-2}(K'BK)^2(K'BK)^{-2}K' \\ &= K(K'BK)^{-2}K' \\ &= ((ZBZ)^2)^-.\end{aligned}$$

Also one can get

$$\begin{aligned}((ZBZ)^2((ZBZ)^2)^-)' &= ((K(K'BK)^2K')(K(K'BK)^{-2}K'))' \\ &= K(K'BK)^{-2}K'K(K'BK)^2K' \\ &= K(K'BK)^{-2}(K'BK)^2K' \\ &= KK' \\ &= K(K'BK)^2(K'BK)^{-2}K' \\ &= K(K'BK)^2K'K(K'BK)^{-2}K' \\ &= (ZBZ)^2((ZBZ)^2)^-.\end{aligned}$$

similarly

$$\begin{aligned}(((ZBZ)^2)^-(ZBZ)^2)' &= ((K(K'BK)^{-2}K')(K(K'BK)^2K'))' \\ &= ((K(K'BK)^2K')(K(K'BK)^{-2}K')) \\ &= K(K'BK)^2(K'BK)^{-2}K' \\ &= KK' \\ &= K(K'BK)^{-2}(K'BK)^2K' \\ &= K(K'BK)^{-2}K'K(K'BK)^2K' \\ &= ((ZBZ)^2)^-(ZBZ)^2.\end{aligned}$$

In this case one can note that $(K'BK)$ is regular matrix in which

$$\begin{aligned}((K'BK)^2)^{-1} &= (K'BKK'BK)^{-1} \\ &= (K'BK)^{-1}(K'BK)^{-1} \\ &= ((K'BK)^{-1})^2.\end{aligned}$$

Hence

$$((ZBZ)^2)^- = K(K'BK)^{-2}K'$$

which represents Moore-Penrose inverse. □

Lemma 3.17. *Let $M = KK'$ with $K'K = I_r$, $T_1 = MIM$, where $I = I_n$ is an identity matrix, $T_2 = MJM$, $I \bigotimes I + J \bigotimes J$ nonsingular, $\lim_{k\to\infty}(K'JK)^k = 0$, and*

$$\begin{aligned}&\sum_{k=0}^{\infty}(-1)^k(K'JK)^k \bigotimes (K'JK)^k \sum_{l=0}^{\infty}(-1)^l(K'JK)^l \bigotimes (K'JK)^l \\ &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(-1)^{k+l}(K'JK)^{k+l} \bigotimes (K'JK)^{k+l}.\end{aligned}$$

Then

$$[(T_1 \bigotimes T_1 + T_2 \bigotimes T_2)^2]^- = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(-1)^{k+l}(K(K'JK)^{k+l}K' \bigotimes K(K'JK)^{k+l}K').$$

Proof. Since that M is a singular idempotent matrix which can be expressed as full rank decomposition $M = KK'$, thus the matrices $T_1 = MIM$ and $T_2 = MJM$ are also singular. Therefore, one can find the generalized inverse of the expression

$(T_1 \bigotimes T_1 + T_2 \bigotimes T_2)^2$ through using the Lemma (3.16) by putting $Z = M \bigotimes M$, $B = I \bigotimes I + J \bigotimes J$, thus as it has been shown in Lemma (3.15), one can write

$$Z = M \bigotimes M = KK' \bigotimes KK' = K \bigotimes K \cdot K' \bigotimes K',$$

$$T_1 \bigotimes T_1 + T_2 \bigotimes T_2 = M \bigotimes M(I \bigotimes I + J \bigotimes J)M \bigotimes M.$$

Hence by using Lemma (3.16) and Theorems (5.10) or (5.11), see the Appendix (5.3), one can find out

$$\begin{aligned}
&[(T_1 \bigotimes T_1 + T_2 \bigotimes T_2)^2]^- \\
&= [(M \bigotimes M(I \bigotimes I + J \bigotimes J)M \bigotimes M)^2]^- \\
&= K \bigotimes K(K' \bigotimes K'(I \bigotimes I + J \bigotimes J)K \bigotimes K)^{-2}K' \bigotimes K' \\
&= K \bigotimes K(K'K \bigotimes K'K + K'JK \bigotimes K'JK)^{-2}K' \bigotimes K' \\
&= K \bigotimes K((I \bigotimes I + K'JK \bigotimes K'JK)^{-1})^2K' \bigotimes K' \\
&= K \bigotimes K(\sum_{k=0}^{\infty}(-1)^k(K'JK \bigotimes K'JK)^k)^2K' \bigotimes K' \\
&= K \bigotimes K(\sum_{k=0}^{\infty}(-1)^k(K'JK)^k \bigotimes (K'JK)^k)(\sum_{k=0}^{\infty}(-1)^l(K'JK)^l \bigotimes (K'JK)^l)K' \bigotimes K' \\
&= (\sum_{k=0}^{\infty}(-1)^k K(K'JK)^k \bigotimes K(K'JK)^k)(\sum_{k=0}^{\infty}(-1)^l(K'JK)^l K' \bigotimes (K'JK)^l K') \\
&= (\sum_{k=0}^{\infty}(-1)^k(\sum_{l=0}^{\infty}(-1)^l K(K'JK)^k(K'JK)^l K' \bigotimes K(K'JK)^k(K'JK)^l K')) \\
&= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(-1)^{k+l}(K(K'JK)^{k+l}K' \bigotimes K(K'JK)^{k+l}K').
\end{aligned}$$

□

Theorem 3.18. *Under the assumptions of Lemma (3.17) with* $u = vec(MJM) = vec(T_2)$ *and* $v = vec(MIM) = vec(T_1)$*, the solution of matrix A which appears in equations system of Lemma (3.13) is given by*

$$vec(A) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(-1)^{k+l}(K(K'JK)^{k+l}K' \bigotimes K(K'JK)^{k+l}K')(\tilde{\zeta}_1 u + \tilde{\zeta}_2 v + \lambda_1 Tu + \lambda_2 Tv),$$

where $\tilde{\zeta}_1, \tilde{\zeta}_2, \lambda_1$ *and* λ_2 *are constants.*

Proof. From Lemma (3.14), it has been shown

$$vec(A) = H^{-}(\tilde{\zeta}_1 u + \tilde{\zeta}_2 v + \lambda_1 Tu + \lambda_2 Tv),$$

where $H = T^2 = (T_1 \bigotimes T_1 + T_2 \bigotimes T_2)^2$ and $\tilde{\zeta}_1, \tilde{\zeta}_2, \lambda_1$ and λ_2 are constants. Now from Lemma (3.17) and Theorem (5.10) see Appendix (5.3), one can get

$$\begin{aligned} vec(A) &= H^{-}(\tilde{\zeta}_1 u + \tilde{\zeta}_2 v + \lambda_1 Tu + \lambda_2 Tv) \\ &= ((T_1 \bigotimes T_1 + T_2 \bigotimes T_2)^2)^{-}(\tilde{\zeta}_1 u + \tilde{\zeta}_2 v + \lambda_1 Tu + \lambda_2 Tv) \\ &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(-1)^{k+l}(K(K'JK)^{k+l}K' \bigotimes K(K'JK)^{k+l}K')(\tilde{\zeta}_1 u + \tilde{\zeta}_2 v + \lambda_1 Tu + \lambda_2 Tv). \end{aligned}$$

□

Theorem 3.19. *The solution of matrix A given by Theorem (3.18) is symmetric.*

Proof. Since $U = MJM$, $V = MIM = M$, $u = vecU$, $v = vecV$ and matrices T_1 and T_2 are as given in Lemma (3.17), by using Lemmas (5.3) or (5.4), see Appendix (5.1), one can find

$$\begin{aligned} (T_1 \bigotimes T_1) \cdot u &= (MIM \bigotimes MIM)vec(MJM) \\ &= vec(MIMMJMMIM) \\ &= vec(MMJMM) \\ &= vec(MJM) \\ &= vec(U), \end{aligned}$$

$$\begin{aligned} (T_1 \bigotimes T_1) \cdot v &= (MIM \bigotimes MIM)vec(MIM) \\ &= vec(MIMMIMMIM) \\ &= vec(MMM) \\ &= vec(MIM) \\ &= vec(M) \\ &= vec(V). \end{aligned}$$

$$
\begin{aligned}
(T_2 \bigotimes T_2)\cdot u &= (MJM \bigotimes MJM)vec(MJM)\\
&= vec(MJMMJMMJM)\\
&= vec(MJMJMJM)\\
&= vec(UJU),
\end{aligned}
$$

$$
\begin{aligned}
(T_2 \bigotimes T_2)\cdot v &= (MJM \bigotimes MJM)vec(MIM)\\
&= vec(MJMMIMMJM)\\
&= vec(MJMJM)\\
&= vec(MJMMJM)\\
&= vec(U^2),
\end{aligned}
$$

Thus, from the solution of matrix A in Eq. (3.3.2) of Lemma (3.14), one has

$$
\begin{aligned}
&\tilde{\zeta_1}u + \tilde{\zeta_2}v + \lambda_1 Tu + \lambda_2 Tv\\
&= \tilde{\zeta_1}u + \tilde{\zeta_2}v + \lambda_1(T_1 \bigotimes T_1 + T_2 \bigotimes T_2)u + \lambda_2(T_1 \bigotimes T_1 + T_2 \bigotimes T_2)v\\
&= \tilde{\zeta_1}vec(U) + \tilde{\zeta_2}vec(V) + \lambda_1 vec(T_1UT_1) + \lambda_1 vec(T_2UT_2) + \lambda_2 vec(T_1VT_1) + \lambda_2 vec(T_2VT_2)\\
&= \tilde{\zeta_1}vec(U) + \tilde{\zeta_2}vec(V) + \lambda_1 vec(U) + \lambda_1 vec(UJU) + \lambda_2 vec(V) + \lambda_2 vec(U^2)\\
&= (\tilde{\zeta_1} + \lambda_1)vec(U) + (\tilde{\zeta_2} + \lambda_2)vec(V) + \lambda_1 vec(UJU) + \lambda_2 vec(U^2).
\end{aligned}
$$

Let $\tilde{\lambda}_1 = \tilde{\zeta_1} + \lambda_1$ and $\tilde{\lambda}_2 = \tilde{\zeta_2} + \lambda_2$ which are constants and suppose $\hat{K} = K'JK$ which is regular symmetric matrix. By using Theorem (3.18) and Lemma (5.3) in Appendix (5.1), one can find out

$$
\begin{aligned}
vec(A) &= H^{-}(\tilde{\zeta_1}u + \tilde{\zeta_2}v + \lambda_1 Tu + \lambda_2 Tv)\\
&= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(-1)^{k+l}(K(K'JK)^{k+l}K' \bigotimes K(K'JK)^{k+l}K')[\tilde{\lambda}_1 vec(U) + (\tilde{\lambda}_2)vec(V)\\
&+ \lambda_1 vec(UJU) + \lambda_2 vec(U^2)]\\
&= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(-1)^{k+l}(K(\hat{K})^{k+l}K' \bigotimes K(\hat{K})^{k+l}K')[\tilde{\lambda}_1 vec(U) + (\tilde{\lambda}_2)vec(V)\\
&+ \lambda_1 vec(UJU) + \lambda_2 vec(U^2)]\\
&= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(-1)^{k+l}[\tilde{\lambda}_1 vec(K(\hat{K})^{k+l}K'UK(\hat{K})^{k+l}K') + \tilde{\lambda}_2 vec(K(\hat{K})^{k+l}K'VK(\hat{K})^{k+l}K')\\
&+ \lambda_1 vec(K(\hat{K})^{k+l}K'(UJU)K(\hat{K})^{k+l}K') + \lambda_2 vec(K(\hat{K})^{k+l}K'(U^2)K(\hat{K})^{k+l}K')],
\end{aligned}
$$

where $\hat{K} = K'JK$. Since that matrix $M = KK'$ where $K'K = I_r$, one can show that

$$\begin{aligned}
K'UK &= K'(MJM)K = K'KK'JKK'K = K'JK,\\
K'VK &= K'(M)K = K'KK'K = I_r,\\
K'(UJU)K &= K'(MJMJMJM)K\\
&= K'KK'JKK'JKK'JKK'K\\
&= K'JKK'JKK'JK = (K'JK)^3,\\
K'(U)^2K &= K'(MJM)^2K\\
&= K'(MJM)(MJM)K\\
&= K'(KK'JKK')(KK'JKK')K\\
&= K'JKK'JK = (K'JK)^2.
\end{aligned}$$

Therefore, the solution in Eq. (3.3.2) can be written as follows

$$\begin{aligned}
vec(A) = &\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(-1)^{k+l}[\tilde{\lambda}_1 vec(K(\hat{K})^{k+l}K'UK(\hat{K})^{k+l}K') + \tilde{\lambda}_2 vec(K(\hat{K})^{k+l}K'VK(\hat{K})^{k+l}K')\\
&+ \lambda_1 vec(K(\hat{K})^{k+l}K'(UJU)K(\hat{K})^{k+l}K') + \lambda_2 vec(K(\hat{K})^{k+l}K'(U)^2K(\hat{K})^{k+l}K')]\\
= &\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(-1)^{k+l}[\tilde{\lambda}_1 vec(K(\hat{K})^{k+l}(K'JK)(\hat{K})^{k+l}K') + \tilde{\lambda}_2 vec(K(\hat{K})^{k+l}I_r(\hat{K})^{k+l}K')\\
&+ \lambda_1 vec(K(\hat{K})^{k+l}(K'JK)^3(\hat{K})^{k+l}K') + \lambda_2 vec(K(\hat{K})^{k+l}(K'JK)^2(\hat{K})^{k+l}K')]\\
= &\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(-1)^{k+l}[\tilde{\lambda}_1 vec(K(K'JK)^{(2(k+l)+1)}K') + \tilde{\lambda}_2 vec(K(K'JK)^{(2(k+l))}K')\\
&+ \lambda_1 vec(K(K'JK)^{(2(k+l)+3)}K') + \lambda_2 vec(K(K'JK)^{(2(k+l)+2)}K')].
\end{aligned}$$

Hence

$$\begin{aligned}
A = &\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(-1)^{k+l}[\tilde{\lambda}_1(K(K'JK)^{(2(k+l)+1)}K') + \tilde{\lambda}_2(K(K'JK)^{(2(k+l))}K') + \lambda_1(K(K'JK)^{(2(k+l)+3)}K')\\
&+ \lambda_2(K(K'JK)^{(2(k+l)+2)}K')]
\end{aligned}$$

which represents the summation of symmetric matrices, therefore matrix A is symmetric. □

3.4 Infinite matrix series and the eigenvalues of $K^{'}JK$

In the last paragraph the solution of the matrix in BAIQUE has been obtained and also the symmetric property of the solution matrix has been found. This solution has dealt with the Neumann series of an infinite sum of matrices. It has shown that

$$\begin{aligned}
&[(T_1 \bigotimes T_1 + T_2 \bigotimes T_2)^2]^- \\
&= [(M \bigotimes M(I \bigotimes I + J \bigotimes J)M \bigotimes M)^2]^- \\
&= K \bigotimes K((I \bigotimes I + K'JK \bigotimes K'JK)^{-1})^2 K' \bigotimes K' \\
&= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(-1)^{k+l}(K(K'JK)^{k+l}K' \bigotimes K(K'JK)^{k+l}K'). \qquad (3.4.1)
\end{aligned}$$

This form contains the term $(I \bigotimes I + K'JK \bigotimes K'JK)^{-1}$ which is represented by the sum of the infinite series of matrices, so the purpose now is to investigate when this form is converging. Clearly, the central part in this formula is the matrix $K'JK$. As it has been mentioned before, K is a full column rank matrix which has found out by decomposing the projection matrix M into $M = KK'$, while matrix J is a covariance structure matrix. Concerning the matrix norm which we used here, we consider the Euclidean norm (largest singular values) or called spectral radius (see Appendix (5.3) of the matrix to deal with the convergence of matrix series. In order to describe the eigenvalues of the term $K'JK$ and $K'JK \bigotimes K'JK$, Gerschgorin Discs and Poincare's separation theorem shall be used. Firstly, let us introduce the following results.

Theorem 3.20. *See Lancaster and Tismenetsky (1985). If $\|\cdot\|$ denotes any matrix norm for which $\|I\| = 1$ and if $\|M\| < 1$, then $(I + M)^{-1}$ exists,*

$$(I + M)^{-1} = I - M + M^2 - \dots$$

and

$$\|(I + M)^{-1}\| \leq \frac{1}{1 - \|M\|}$$

Corollary 3.21. *See Graybill(1983). Let A be an $n \times n$ matrix. If $|\lambda| < 1$ for every characteristic root λ of A, then the following holds.*
1) $\sum_{i=0}^{\infty} A^i$ converges to $(I - A)^{-1}$.
2) There exists a positive integer K such that for all positive integers $k \geq K$ the matrix B^k is nonsingular where $B^k = I + A + A^2 + \ldots + A^k$.

Gerschgorin Discs is one of the methods used to estimate the eigenvalues of a matrix. This method is useful for providing bounds to the magnitudes of the eigenvalues of a matrix. Gerschgorin Discs are defined as follows

Definition 3.22. (Gerschgorin Discs) Let $A \in \mathcal{C}$ matrix have entries a_{ij}. For each $i = 1, \ldots, n$ we define the Gerschgorin disc D_i associated to the $i - th$ row of A by

$$D_i = \{z \in \mathcal{C} :\mid z - a_{ii} \mid \leq \sum_{i \neq j; j=1}^{n} \mid a_{ij} \mid\}. \tag{3.4.2}$$

These discs $D_1, D_2, \ldots, D_n$ all lie in the complex plane. In other words, the eigenvalues are trapped in the collection of circles centered at a_{ii} with a radius given by the sum of absolute values A with a_{ii} deleted. Furthermore, if a union $\bigcup$ of k Gerschgorin circles does not touch any of the other $n - k$ circles, then there are exactly k eigenvalues in the circles of the union $\bigcup$. See Lewis (1991).

Example 3.23. Assume the one dimensional design points $(0.1, 0.4, 0.5, 0.8, 1)$, so the distance matrix takes the form

$$D = \begin{pmatrix} 0 & 0.3000 & 0.4000 & 0.7000 & 0.9000 \\ 0.3000 & 0 & 0.1000 & 0.4000 & 0.6000 \\ 0.4000 & 0.1000 & 0 & 0.3000 & 0.5000 \\ 0.7000 & 0.4000 & 0.3000 & 0 & 0.2000 \\ 0.9000 & 0.6000 & 0.5000 & 0.2000 & 0 \end{pmatrix}$$

The Gerschgorin Discs for the rows of the distance matrix D are as follows;

$$\mid z \mid \leq 2.3, \mid z \mid \leq 1.4, \mid z \mid \leq 1.3, \mid z \mid \leq 1.6, \mid z \mid \leq 2.2$$

The union of these five discs is the single disc $\mid z \mid \leq 2.3$. Suppose the covariance structure matrix is in the form $J = \exp(-|D|)$, hence J for our example is

$$J = \begin{pmatrix} 1.0000 & 0.7408 & 0.6703 & 0.4966 & 0.4066 \\ 0.7408 & 1.0000 & 0.9048 & 0.6703 & 0.5488 \\ 0.6703 & 0.9048 & 1.0000 & 0.7408 & 0.6065 \\ 0.4966 & 0.6703 & 0.7408 & 1.0000 & 0.8187 \\ 0.4066 & 0.5488 & 0.6065 & 0.8187 & 1.0000 \end{pmatrix}.$$

Its Gerschgorin Discs are

$$| z-1 | \leq 2.3143, | z-1 | \leq 2.8648, | z-1 | \leq 2.9225, | z-1 | \leq 2.7265, | z-1 | \leq 2.3806.$$

The union $\bigcup$ is the third Disc which stands for the third row (column). Since distance matrix D is symmetric, the Gerschgorin Discs for the columns of this matrix are the same for the rows, hence the information obtained from the columns is the same as the information from the rows. □

Theorem 3.24. *(Gerschgorin's theorem) See Lewis (1991). Let $A \in \mathcal{C}$ matrix, the eigenvalues of A lie in the union of the n Gerschgorin discs associated to the rows of A.*

Example 3.25. Consider the covariance structure matrix J as it is defined above, then the five Gerschgorin Discs of the matrix J have $(centre, radius) = (a_{ii}, \sum_{i \neq j; j=1}^{3} |a_{ij}|)$ given by (1,2.3143), (1,2.8648), (1,2.9225), (1,2.7265) and (1,2.3806). The relation (3.4.2) guarantees that the eigenvalues are in (or on) the five circles centered at 1, or all the eigenvalues lie in (or on) the largest circle $(1, 2.3806)$. While the actual eigenvalues are 0.0830, 0.1593, 0.3259, 0.7713 and 3.6605 lie well within the union of these Discs, the maximum eigenvalue lies in the largest circle, i.e. $3.6605 < 3.3806$. That can be seen in the Figure 1. □

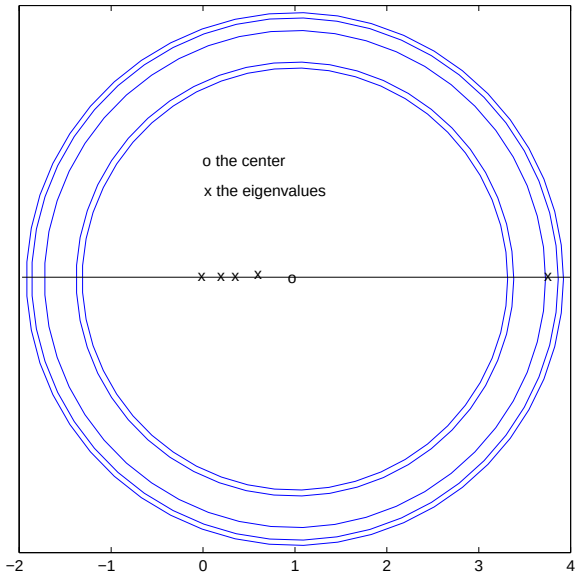

Figure 1: Gerschgorin Discs and the eigenvalues of matrix J.

Clearly, the Euclidean norm (largest eigenvalue) of the matrix $K'JK \bigotimes K'JK$ plays an important role in order to apply the Theorem (3.20), and to get the infinite series sum. Therefore, it is necessary to deal with the eigenvalues of the matrix $K'JK$. For this purpose, consider again that the covariance structure matrix is $J = \exp(-|D|)$, whereas D is the distance matrix $D = \parallel x^i - x^j \parallel i, j = 1, \ldots, n$ as it has been mentioned before, while K is a full column rank matrix and satisfies $KK' = M$ and $K'K = I$. To get a relation between the eigenvalues of the matrices $K'JK$ and J, one can handle here with Poincare's separation theorem.

Lemma 3.26. *For any real symmetric $n \times n$ matrix A and vector $\mathbf{x}$,*

$$\lambda_1 \leq \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} \leq \lambda_n \ ,$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of matrix A in increasing order, so that $\lambda_1 \leq \lambda_2 \ldots \leq \lambda_n$.

Theorem 3.27. *(Poincare's separation theorem) See Magnus and Neudecker (1999). Let A be a real symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, and let G be a semi-orthogonal $n \times r$ matrix $(1 \leq r \leq n)$, so that $G'G = I_r$. Then the eigenvalues $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_r$ of $G'AG$ satisfy*

$$\lambda_i \leq \mu_i \leq \lambda_{n-r+i} \qquad (i = 1, 2, \ldots, r)$$

Note: For $r = 1$, Theorem (3.27) reduces to Lemma (3.26). For $r = n$, we obtain the well-known result that the symmetric matrices A and $G'AG$ have the same set of eigenvalues, if G is orthogonal.

The interesting thing here is to know more about the Euclidean norm or the largest singular value of the matrix $K'JK \bigotimes K'JK$. If the covariance structures matrix J has the form $\exp(-|D|)$, then matrix J represents here a correlation matrix with all elements on the diagonal equal one while the off diagonal elements are symmetric and lie between zero and one. Since that matrix K is semi-orthogonal $K'K = I_r$, one gets from Poincare theorem (3.27) that

$$\lambda_i \leq \mu_i \leq \lambda_{n-r+i} \qquad (i = 1, 2, \ldots, r)$$

where $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_r$ are the eigenvalues of the matrix $K'JK$ and $\lambda_1, \ldots, \lambda_n$ the eigenvalues of matrix J.

Since matrix J is a symmetric matrix, from the spectral theorem (Theorem (5.20) see Appendix 5.3), there exists an orthogonal matrix Q and diagonal matrix Δ such that $J = Q\Delta Q'$. Let $\mathbf{x}$ be a vector of order n such as

$$\mathbf{x} = c_1\mathbf{q_1} + \ldots + c_n\mathbf{q_n}, \tag{3.4.3}$$

whereas $q_1, \ldots, q_n$ are the eigenvectors of the symmetric matrix Q and $c_1, \ldots, c_n$ are constants. This equation may be expressed in the form

$$\mathbf{x} = Qc,$$

where $c = (c_1, \ldots, c_n)$. According to the spectral theorem (Theorem (5.20) see Appendix 5.3) or Corollary (5.30) in Appendix 5.5, consider any eigenvalue δ of the matrix J and corresponding the eigenvector $\mathbf{x}$. Premultiplying the eigenvalue equation by $\mathbf{x}'$ gives

$$\mathbf{x}'J\mathbf{x} = \delta\mathbf{x}'\mathbf{x}. \tag{3.4.4}$$

Expressing the vectors $\mathbf{x}$ as a linear combination of the eigenvectors of matrix J according to Eq.(3.4.3) gives

$$\mathbf{x}'J\mathbf{x} - \delta\mathbf{x}'\mathbf{x} = c'Q'JQc - \delta c'Q'Qc = 0.$$

Since $Q'Q = I$ and $JQ = Q\Delta$, thus

$$c'\Delta c - \delta c'c = \sum_{i=1}^{n} c_i^2(\lambda_i - \delta) = 0$$

since all the terms $(\lambda_i - \delta)$ cannot have the same sign, then

$$\lambda_1 \leq \delta \leq \lambda_n. \tag{3.4.5}$$

Now we focus on applying the Gerschgorin theorem on matrix J, considering the eigenvalues equation $Jq = \lambda q$ in which the eigenvector has been normalized so that the largest element $q_k = 1$. This equation may be expressed as

$$\begin{pmatrix} * & * & \dots & * & \dots & * \\ * & * & \dots & * & \dots & * \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ j_{k1} & j_{k2} & \dots & j_{kk} & \dots & j_{kn} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ * & * & \dots & * & \dots & * \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ \cdot \\ 1 \\ \cdot \\ q_n \end{pmatrix} = \lambda \begin{pmatrix} q_1 \\ q_2 \\ \cdot \\ 1 \\ \cdot \\ q_n \end{pmatrix} \tag{3.4.6}$$

the k^{th} elemental equation of which is

$$\lambda - j_{kk} = \sum_{j \neq k} j_{kj} q_j.$$

Since $|q_j| \leq 1$, it follows that

$$|\lambda - j_{kk}| \leq \sum_{j \neq k} |j_{kj}|. \tag{3.4.7}$$

This can be interpreted on a diagram where λ must lie within a circle with its center j_{kk} and its radius $\sum_{j \neq k} |j_{kj}|$. Since the position on the largest element in an eigenvector is normally unknown, it is only possible to say that every eigenvalue must lie within the union of the discs constructed from n rows of the matrix according to Eq.(3.4.7). It can be shown from Eq.(3.4.5) that

$$\lambda_1 \leq j_{ii} \leq \lambda_n$$

By adopting a vector $\mathbf{x}$ which is null except for a unit term in position i in the Eq.(3.4.4), one gets for a real symmetric matrix that

$$\begin{cases} (j_{ii})_{max}) \leq \lambda_n \leq (j_{kk} + \sum_{j \neq k} |j_{jk}|)_{max} \\ (j_{kk} + \sum_{j \neq k} |j_{jk}|)_{min} \leq \lambda_1 \leq (j_{ii})_{min}. \end{cases} \tag{3.4.8}$$

See Jennings and Mckeown (1993). Hence, the relation (3.4.8) gives the constraints of the maximum eigenvalue of matrix J. From that one can deal with $\parallel K'JK \parallel$ and examine when it could satisfy the condition $\parallel K'JK \bigotimes K'JK \parallel < 1$.

Example 3.28. Continue with the example above, the projection matrix M can be found where in this case the design matrix is

$$X = \begin{pmatrix} 1.0000 & 0.1000 \\ 1.0000 & 0.4000 \\ 1.0000 & 0.5000 \\ 1.0000 & 0.8000 \\ 1.0000 & 1.0000 \end{pmatrix}.$$

Thus

$$M = I - X(X'X)^{-1}X' = \begin{pmatrix} 0.3699 & -0.3496 & -0.2561 & 0.0244 & 0.2114 \\ -0.3496 & 0.7480 & -0.2195 & -0.1220 & -0.0569 \\ -0.2561 & -0.2195 & 0.7927 & -0.1707 & -0.1463 \\ 0.0244 & -0.1220 & -0.1707 & 0.6829 & -0.4146 \\ 0.2114 & -0.0569 & -0.1463 & -0.4146 & 0.4065 \end{pmatrix}.$$

This matrix has a rank equal to three ($rank(K) = 3$), so one can find matrix K which satisfies that $M = KK'$ and $K'K = I$, this matrix is given by

$$K = \begin{pmatrix} -0.6082 & -0.0000 & -0.0000 \\ 0.5748 & -0.6171 & 0.1916 \\ 0.4211 & 0.7783 & 0.0980 \\ -0.0401 & -0.0944 & -0.8200 \\ -0.3475 & -0.0668 & 0.5303 \end{pmatrix}.$$

Therefore, in order to examine Theorem (3.27) which tells us that the maximum eigenvalue of $K'JK$ is less than or equal to the maximum eigenvalue of matrix J. As mentioned before this is $\lambda_i \leq \mu_i \leq \lambda_{n-r+i} \qquad (i = 1, 2, \ldots, r)$. The eigenvalues of matrix $K'JK$ in this example are 0.0830, 0.1673 and 0.3518 while matrix J has the eigenvalues 0.0830, 0.1593, 0.3259, 0.7713 and 3.6605. It is clear from that, that the condition of Poincare's separation theorem is satisfied since $\mu_r = 0.3518 < \lambda_n = 3.6605$, where μ_r is the maximum eigenvalue (norm) of $K'JK$ and λ_n stands for the maximum eigenvalue of J. Furthermore, one finds out that, $\| K'JK \bigotimes K'JK \| = 0.1238 < \| K'JK \| = 0.3518 < 1$.

In this context, one can investigate whether the order of the summation in Eq. (3.4.1) is exchanged as Theorem (5.19) in the Appendix (5.3) asserts. Also one can check the assumptions of Lemma (3.17) and Theorem (3.18). For this purpose, one has to determine the radius of convergence R in Theorem (5.19) in Appendix (5.3), where the sums of power series convergent for $|z| < R$. So for our case, we can

choose $R = 1$. In this case all the characteristics roots of the power series matrix $K'JK$ lie in the interior of the circle of $|R| < 1$. One can note that in Figure 2.

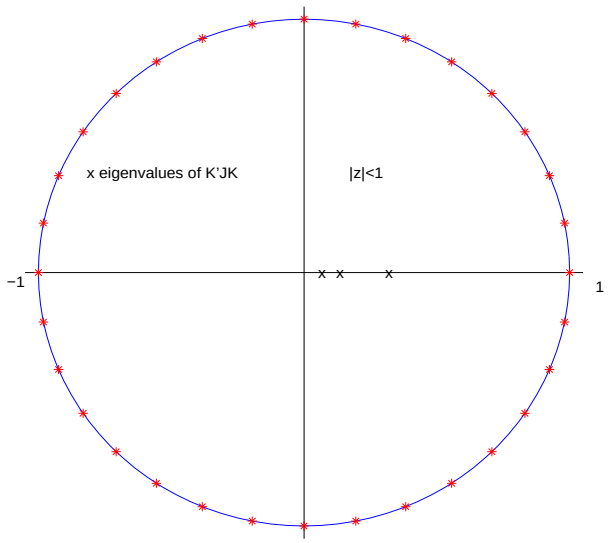

Figure 2: All the eigenvalues of matrix $K'JK$ lie in the interior of the circle $|R| < 1$.

Therefore, under this assumption one can deal with the exchangeability of the order of the power of series matrices $(K'JK)^{k+l}$ as Theorem (5.19) is postulated. In this way, the assumptions of Lemma (3.17) and Theorem (3.18) can be shown. For $\lim_{k\to\infty}(K'JK)^k = 0$, consider $k = 5$ and 10 and find the corresponding results as follows

$$(K'JK)^5 = \begin{pmatrix} 0.0052 & 0.0010 & -0.0002 \\ 0.0010 & 0.0002 & -0.0000 \\ -0.0002 & -0.0000 & 0.0001 \end{pmatrix},$$

$$(K'JK)^{10} = 1.0e^{-004} * \begin{pmatrix} 0.2808 & 0.0519 & -0.0087 \\ 0.0519 & 0.0096 & -0.0016 \\ -0.0087 & -0.0016 & 0.0004 \end{pmatrix}.$$

The largest singular values (spectral radius) of these matrices are 0.0054 and $2.9072e^{-005}$ respectively, while for the matrix $K'JK$ was 0.3518. So one can note that the power series of matrices $(K'JK)^k$ converges to zero when k converges to infinity. On the other hand, concerning computing the both sides of the following relation

$$\sum_{k=0}^{\infty}(-1)^k(K'JK)^k\bigotimes(K'JK)^k\sum_{l=0}^{\infty}(-1)^l(K'JK)^l\bigotimes(K'JK)^l$$
$$=\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(-1)^{k+l}(K'JK)^{k+l}\bigotimes(K'JK)^{k+l}. \tag{3.4.9}$$

Since in our example the matrix $(K'JK)\bigotimes(K'JK)$ will be of size (9×9). For purpose of dealing with matrices of small-size, consider the same example above but with four points design such as $\mathbf{x}=(0.1,0.4,0.5,1)$, then matrix $K'JK$ is given by

$$(K'JK)=\begin{pmatrix}0.0955 & 0.0486\\ 0.0486 & 0.3424\end{pmatrix}.$$

Assume that $k=3$ and $l=4$, by using Matlab program, the following results are obtained

$$\sum_{k=0}^{3}(-1)^k(K'JK)^k\bigotimes(K'JK)^k=\begin{pmatrix}-0.0000 & -0.0000 & -0.0000 & -0.0001\\ -0.0000 & -0.0001 & -0.0001 & -0.0003\\ -0.0000 & -0.0001 & -0.0001 & -0.0003\\ -0.0001 & -0.0003 & -0.0003 & -0.0018\end{pmatrix},$$
$$\sum_{l=0}^{4}(-1)^l(K'JK)^l\bigotimes(K'JK)^l=1.0e^{-003}*\begin{pmatrix}0.0003 & 0.0016 & 0.0016 & 0.0078\\ 0.0016 & 0.0086 & 0.0078 & 0.0411\\ 0.0016 & 0.0078 & 0.0086 & 0.0411\\ 0.0078 & 0.0411 & 0.0411 & 0.2175\end{pmatrix},$$

$$(\sum_{k=0}^{3}(-1)^k(K'JK)^k\bigotimes(K'JK)^k)(\sum_{l=0}^{4}(-1)^l(K'JK)^l\bigotimes(K'JK)^l)$$
$$=1.0e^{-006}*\begin{pmatrix}-0.0005 & -0.0028 & -0.0028 & -0.0148\\ -0.0028 & -0.0148 & -0.0148 & -0.0779\\ -0.0028 & -0.0148 & -0.0148 & -0.0779\\ -0.0148 & -0.0779 & -0.0779 & -0.4106\end{pmatrix},$$

$$\sum_{k=0}^{3}\sum_{l=0}^{4}(-1)^{k+l}(K'JK)^{k+l}\bigotimes(K'JK)^{k+l}$$

$$= 1.0e^{-006} * \begin{pmatrix} -0.0005 & -0.0028 & -0.0028 & -0.0148 \\ -0.0028 & -0.0148 & -0.0148 & -0.0779 \\ -0.0028 & -0.0148 & -0.0148 & -0.0779 \\ -0.0148 & -0.0779 & -0.0779 & -0.4106 \end{pmatrix}.$$

It is clearly that the both side of relation (3.4.9) have the same results and thus the assumptions of Lemma (3.17) have examined. □

4 Comparison of designs and a practical example

4.1 The influence of the position of the design points in a fixed interval

The interesting thing here is to study the effect of the design points on the results obtained by handling the system such as the eigenvalues or the norms of the matrices D, J, $K'JK$ and $K'JK \bigotimes K'JK$ and Bayes risk. It is necessary to mention here that Bayes risk has been obtained by taking always $\mathbf{b} = (1, 1)'$ in the equations system (3.2.11) to find the matrix A. In this context, one can consider some regressor functions in one and two dimensional designs and investigate the relations between them with corresponding results for different covariance structures. For this purpose, one can consider different sites of the design points on the fixed interval $[0, 1]$. Let us assume the following four designs of four points.

1- $(a, a + 0.5c, 1 - 0.5c, 1)$, (border sites) the points lie at both ends of the interval.
2- $(a, a + c, a + 2c, 1)$, (regular sites) the points have the same distances between them.
3- $(a, a+1.3c, a+1.7c, 1)$, (middle sites) some points near to each other in the center and the design has both interval edges.
4- $(c+0.1, c+0.2, c+0.3, 1)$, (neighbor sites) some sites near to each other and the design has one of the interval edges. Where a is the first point, 1 the last point, and $c = (1 - a)/3$. For example, let $a = 0.1$ then $c = (1 - 0.1)/3 = 0.3$. Now, one can get for the above designs the following results.

1- $(0.1, 0.25, 0.85, 1)$,
2- $(0.10, 0.40, 0.70, 1)$,
3- $(0.1, 0.49, 0.61, 1)$,
4- $(0.4, 0.5, 0.6, 1)$.

Figure 3 shows these designs.

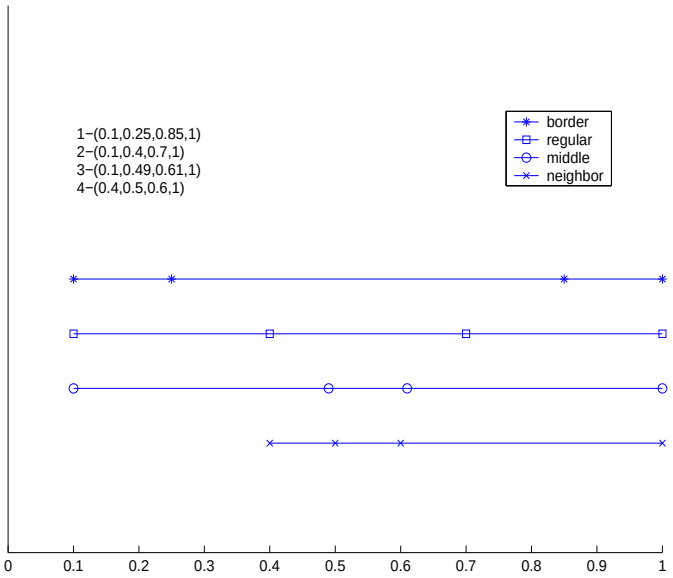

Figure 3: Designs of four points in one dimension.

Firstly one can consider the case where the model has regression function of the stationary case i.e. $X = \mathbf{1}$ and also when $X = (\mathbf{1}\ \ \mathbf{x})$, where $\mathbf{x}$ refers to the design points. Here, let us suppose a covariance structure such as $J = I - \exp(-|D|)$ where $D = |d_{ij}|$ is the distance between the points i and j in this case in one dimension, while I stands for an identity matrix. The function $I - \exp(-|D|)$ has been used by Stein(1987). Schuenemeyer and Power (2000) have used the form of the exponential model $1 - \exp(-3d/a)$ where a is constant stands for the estimated range. For the first design one can find out matrices D and J as follows

$$D = \begin{pmatrix} 0 & 0.1500 & 0.7500 & 0.9000 \\ 0.1500 & 0 & 0.6000 & 0.7500 \\ 0.7500 & 0.6000 & 0 & 0.1500 \\ 0.9000 & 0.7500 & 0.1500 & 0 \end{pmatrix}; J = \begin{pmatrix} 0 & -0.8607 & -0.4724 & -0.4066 \\ -0.8607 & 0 & -0.5488 & -0.4724 \\ -0.4724 & -0.5488 & 0 & -0.8607 \\ -0.4066 & -0.4724 & -0.8607 & 0 \end{pmatrix}.$$

Matrix K, as it has been explained before, is a full column rank matrix. This matrix satisfies that $M = KK'$ and $K'K = I$, where M is the projection matrix in Eq. (3.1.5). Matrix K for the cases when the design matrix is $X = \mathbf{1}$ and $X = (\mathbf{1}\ \mathbf{x})$ takes the following forms

$$\begin{pmatrix} -0.0000 & -0.8660 & -0.0000 \\ 0.7715 & 0.2887 & 0.2673 \\ -0.1543 & 0.2887 & -0.8018 \\ -0.6172 & 0.2887 & 0.5345 \end{pmatrix}, \begin{pmatrix} -0.6355 & -0.0000 \\ 0.7565 & 0.1543 \\ 0.0303 & -0.7715 \\ -0.1513 & 0.6172 \end{pmatrix}$$

respectively. Bayes risk can be computed from the relation in Eq. (3.1.12) which is

$$\begin{aligned} r(\hat{\alpha}) &= 2\sum_{k=1}^{2} trA(\sum_{i=1}^{r} r_{ik}V_i)A(\sum_{j=1}^{r} r_{kj}V_j) \\ &= 2\sum_{k=1}^{r} trAT_kAT_k \end{aligned} \tag{4.1.1}$$

where $T_k = \sum_{i=1}^{r} r_{ik}V_i$ and r_{ik} are the components of the quare root matrix R of the prior information matrix C as it has been given in Eq. (3.1.4). Assume that the prior information matrix C is the identity matrix $I_{2\times 2}$. In this case Bayes risk is given by

$$r(\hat{\alpha}) = 2\sum_{k=1}^{2} trAV_kAV_k, \tag{4.1.2}$$

where $V_1 = MJM$ and $V_2 = MIM = M$. Moreover, the same vector $\mathbf{b} = (1,1)'$ has been used in the equations system (3.2.11) to get the matix A which is included in Bayes risk, as one notes that in Eq. (4.1.2). Table 1 shows the corresponding outcomes of matrices D, J, $K'JK$, $K'JK \bigotimes K'JK$ and Bayes risk of these designs.

Designs	Matrix X	Eig.(D)	$\parallel D \parallel$	Eig.(J)	$\parallel J \parallel$	$\parallel K'JK \parallel$	Bayes risk
1	$X = \mathbf{1}$	-1.3685,-0.1624, -0.1315,1.6624	1.6624	-1.8127,0.0829, 0.8573,0.8725	1.8127	0.8725	0.2609
2		-1.0243,-0.3487, -0.1757,1.5487	1.5487	-1.8741,0.3191, 0.7267,0.8282	1.8741	0.8282	4.2497
3		-0.9180,-0.4709, -0.1020,1.4909	1.4909	-1.9467,0.3947, 0.6532,0.8988	1.9467	0.8988	4.7920
4		-0.7648,-0.1870, -0.0660,1.0177	1.0177	-2.2476,0.4760, 0.8367,0.9349	2.2476	0.9348	2.9788
1	$X = (\mathbf{1}\ \mathbf{x})$					0.8635	4.6090e+003
2						0.8226	43.2778
3		=	=	=	=	0.8986	11.7031
4						0.9340	32.2821

Table 1: Eigenvalues, norms and Bayes risk of some matrices of four point designs with two regressor functions.

The norm used here is the Euclidean norm which represents the maximum eigenvalue of the matrix. One can note that the eigenvalues of the distance matrix D are all negative except one of them as Lemma (5.34, see Appendix 5.5) asserts, while the eigenvalues of the covariance structure matrix J are all positive except one of them is negative. For these four designs, the last case has the minimum norm for matrix

D. On the other hand, the first design has the minimum norm for the covariance structure matrix J. One may explain this by saying that the matrix J has the exponential function $\exp(-|D|)$ which entries lie between zero and one. That means, the small entries in matrix D would have big values near to one in the corresponding position in matrix J and this would reflect on the norm of both matrices. For the norm of the decomposed matrix $K'JK$, the four designs achieve similar norms and all less than one for both cases $X = \mathbf{1}$ and $X = (\mathbf{1}\ \ \mathbf{x})$. Anyway, the second design has a little less norm than the other. Regarding Bayes risk, one can note that for the first case of the regression function, the first design has the minimum Bayes risk while in the second case of regression function the third design has the minimum Bayes risk.

For the case of two dimensional designs of regression function $X = (\mathbf{1}\ \ \mathbf{x_1}\ \ \mathbf{x_2})$, three types of Minkowski distance in a metric space are used here. Minkowski distance is a generalization of Euclidean distance in Eq. (3.2.2) which is defined of order r between two objects i and j by

$$d_{ij} = (\sum_{k=1}^{n}(x_{ik} - x_{jk})^r)^{\frac{1}{r}},$$

For $r = 2$ it gives Euclidean distance measure and, when $r = 1$, it gives first norm distance or called city-block distance or rectangular distance measure. The third type of the distance measure is the L_{max} or L_∞ which is the maximum difference between any attribute of the objects. It can be defined by

$$d_{ij} = \lim_{r\to\infty}(\sum_{k=1}^{n}(x_{ik} - x_{jk})^r)^{\frac{1}{r}}.$$

Since one needs here two dimensional design points, the other vector points for these four designs are given here by using a permutation with the third component of the design points as follows
1-$(0.8667, 0.2, 0.3333, 1)$
2-$(0.7333, 0.2, 0.4667, 1)$
3-$(0.6533, 0.2, 0.5467, 1)$
4-$(0.5667, 0.3667, 0.4667, 1)$
where $a = 0.2$, $c = 0.2667$. One can note from the Figure 4 below the points of these four designs in the interval $[0, 1]^2$.

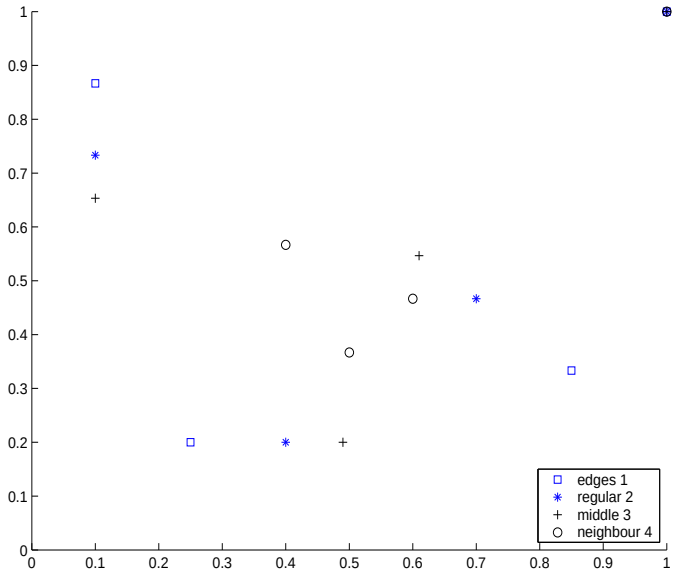

Figure 4: Designs of four points in two dimensions.

Doing that, the corresponding distance matrices of these three types of distance measures can be found. Consider the first design points, then one can get the following distance matrices respectively

$$\begin{pmatrix} 0 & 0.6834 & 0.9203 & 0.9098 \\ 0.6834 & 0 & 0.6146 & 1.0966 \\ 0.9203 & 0.6146 & 0 & 0.6834 \\ 0.9098 & 1.0966 & 0.6834 & 0 \end{pmatrix}; \begin{pmatrix} 0 & 0.8167 & 1.2834 & 1.0333 \\ 0.8167 & 0 & 0.7333 & 1.5500 \\ 1.2834 & 0.7333 & 0 & 0.8167 \\ 1.0333 & 1.5500 & 0.8167 & 0 \end{pmatrix}.$$

$$\begin{pmatrix} 0 & 0.6667 & 0.7500 & 0.9000 \\ 0.6667 & 0 & 0.6000 & 0.8000 \\ 0.7500 & 0.6000 & 0 & 0.6667 \\ 0.9000 & 0.8000 & 0.6667 & 0 \end{pmatrix}.$$

In this case matrix K of the full column rank is given by

$$\begin{pmatrix} -0.3723 \\ 0.5683 \\ -0.6075 \\ 0.4115 \end{pmatrix}.$$

For every case of these distance types one can compute the eigenvalues of matrices $D = \| x_i - x_j \|_{i,j=1,\ldots,n}$, $J = I - \exp(-|D|)$ as it has been defined above, matrix

$K'JK$ and Bayes risk. The results corresponding to the first design are shown in Table 2 below.

Designs	D type	Eig.(D)	$\parallel D \parallel$	Eig.(J)	$\parallel J \parallel$	$\parallel K'JK \parallel$	Bayes risk
1	L_2	-1.1540,-0.9018, -0.4074,2.4632	2.4632	-1.3461,0.3067, 0.4119,0.6275	1.3461	0.6265	0.1072
2		-1.1017,-0.7178, -0.3271,2.1467	2.1467	-1.5283,0.3288, 0.4901,0.7094	1.5283	0.7037	0.1569
3		-1.0930,-0.6337, -0.3232,2.0500	2.0500	-1.5941,0.3339, 0.5324,0.7279	1.5941	0.7155	0.1664
4		-1.1033,-0.2510, -0.1380,1.4923	1.4923	-1.9910,0.3418, 0.7774,0.8719	1.9910	0.8672	0.4841
1	L_1	-1.5802,-1.2769, -0.2699,3.1270	3.1270	-1.1081,0.1976, 0.2849,0.6256	1.1081	0.6240	0.1058
2		-1.4768,-0.9505, -0.4497,2.8771	2.8771	-1.2225,0.2257, 0.3837,0.6130	1.2225	0.6060	0.0960
3		-1.4694,-0.8625, -0.4087,2.7406	2.7406	-1.3198,0.2291, 0.4203,0.6704	1.3198	0.6461	0.1185
4		-1.5521,-0.3332, -0.1951,2.0804	2.0804	-1.7525,0.2143, 0.7156,0.8226	1.7525	0.8160	0.3048
1	L_{max}	-0.9372,-0.6981, -0.5644,2.1997	2.1997	-1.4546,0.3884, 0.4996,0.5666	1.4546	0.5536	0.0699
2		-0.9717,-0.6454, -0.2567,1.8738	1.8738	-1.6733,0.3743, 0.5324,0.7666	1.6733	0.7576	0.2078
3		-1.0050,-0.4923, -0.2783,1.7756	1.7756	-1.7293,0.3543, 0.6189,0.7561	1.7293	0.7433	0.1920
4		-0.8712,-0.2342, -0.0978,1.2031	1.2031	-2.1301,0.4316, 0.7908,0.9077	2.1301	0.9033	0.7430

Table 2: Eigenvalues, norms and Bayes risk of some matrices of four point designs with three distance measures.

From this table, one can note some similar properties as it is the case with one dimensional design and the stationary case above. The minimum norm of that distance matrix D obtained in the fourth design while the first design achieves the minimum norm of the covariance structure matrix J. One can also note that all the norms of matrix $K'JK$ are less than one. Moreover, the results obtained from using the Euclidean and L_{max} distances are more similar than using the distance (L_1) in the second case. In the cases L_2 and L_{max}, the minimum norm of matrix $K'JK$ and minimum Bayes risk are received in the first design while for the L_1 distance type, the second design has the minimum norm of matrix $K'JK$ and also for Bayes risk.

Consider again the same four design above but with five points, suppose $a = 0.1$, then $c = 0.2250$ for the first axis $\mathbf{x_1}$. As for the second axis $\mathbf{x_2}$, let $a = 0.17$, so $c = 0.2075$, for these five points the designs can be given by

1-(a, a+0.5c ,a+c, 1-0.5c, 1).
2-(a, a+c, a+2c, a+3c, a+4c=1).
3-(a, a+1.8c, a+2c, a+2.2c, 1).
4-(c+0.1, c+0.2, c+0.3, c+0.4, 1).
Doing that, one can find out the five sites of the points in two dimensional design as follows

$1 - \mathbf{x_1} = (0.1, 0.2125, 0.3250, 0.8875, 1)$, $\mathbf{x_2} = (0.3775, 0.8962, 0.17, 0.2737, 1)$.
$2 - \mathbf{x_1} = (0.1, 0.3250, 0.5500, 0.7750, 1)$, $\mathbf{x_2} = (0.5850, 0.7925, 0.17, 0.3775, 1)$.
$3 - \mathbf{x_1} = (0.1, 0.5050, 0.5500, 0.5950, 1)$, $\mathbf{x_2} = (0.5850, 0.6265, 0.17, 0.5435, 1)$.
$4 - \mathbf{x_1} = (0.3250, 0.4250, 0.5250, 0.6250, 1)$, $\mathbf{x_2} = (0.5075, 0.6075, 0.17, 0.3075, 0.4075)$.
Figure 5 shows these pairs of points.

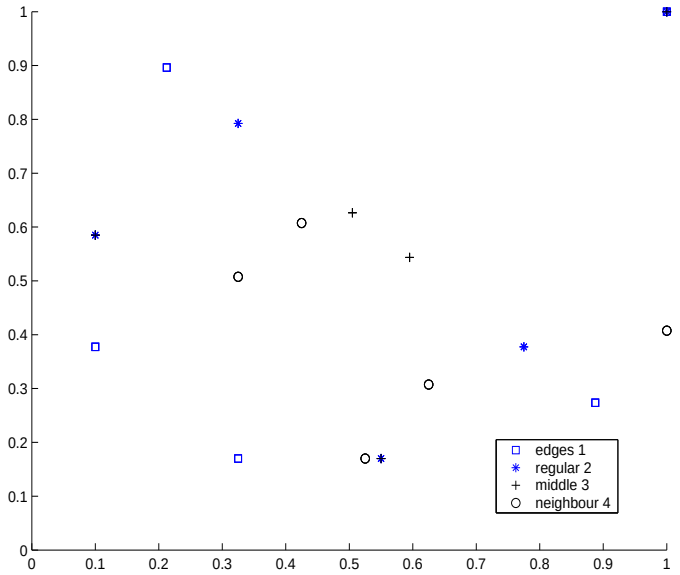

Figure 5: Designs of five points in two dimensions.

The intention here is to see if the designs have similar qualities as if the number of designs points were four. Table 3 shows the corresponding outcomes of this case.

Designs	Matrix X	$\parallel D \parallel$	$\parallel J \parallel$	$\parallel K'JK \parallel$	Bayes risk
1	$X = \mathbf{1}$	2.0160	2.5674	0.9264	2.6785
2		1.8648	2.6270	0.8765	11.2874
3		1.6793	2.8740	0.9704	9.5247
4		1.3451	3.0183	0.9420	8.2532
1	$X = (\mathbf{1}\ \mathbf{x})$			0.9259	4.8422
2				0.8455	7.1271
3		=	=	0.9704	2.1617
4				0.9418	3.2463
1	$X = (\mathbf{1}\ \mathbf{x_1}\ \mathbf{x_2})$	3.0499	1.9478	0.7690	19.9829
2	with	2.6446	2.1486	0.7740	137.7259
3	Euclidean(L_2)	2.3741	2.3839	0.8931	3.8205
4	distance	1.6771	2.7067	0.8768	38.2191
1	$X = (\mathbf{1}\ \mathbf{x_1}\ \mathbf{x_2})$	3.8491	1.6587	0.6723	1.2382e+003
2	with	3.5123	1.7656	0.6824	550.7681
3	L_1 distance	3.1001	2.1265	0.8457	6.2297
4		2.1419	2.4224	0.8063	238.2867
1	$X = (\mathbf{1}\ \mathbf{x_1}\ \mathbf{x_2})$	2.7217	2.0913	0.8377	5.0844
2	with	2.3198	2.3406	0.9133	3.8542
3	L_{max} distance	2.0745	2.5224	0.9250	2.6537
4		1.5378	2.8137	0.9276	5.5254

Table 3: Eigenvalues, norms and Bayes risk of some matrices of five point designs with some regressor functions.

Furthermore, one can note in this case that the minimum norm of the distance matrix D is obtained in the fourth design while the least norm for matrix J is given in the first design as in the case with four points. Regarding the norm of the decomposed matrix $K'JK$, there is also a similarity with the case of four point designs. Here, one finds that for the design matrix when $X = \mathbf{1}$ and $X = (1\ \mathbf{x})$, the second design has the least norm and the first design has the minimum norm for the case of two dimensional design. Finally, concerning Bayes risk, one notes in both of four and five points, the first and third design achieve the minimum Bayes risk for $X = \mathbf{1}$ and $X = (\mathbf{1}\ \mathbf{x})$ respectively. But for the two dimensional design case, the third design for five points has the least risk while in the previous case of four pairs points it was in general the first design has the minimum risk.

4.2 The influence of scaling on the design points

We investigate here the effect of the scaling on the design points. By doing so, one can note the changing of the eigenvalues and norms of the interested matrices in our system. These are the distance matrix D, covariance structure matrix J and the decomposed matrix $K'JK$. Let one considers the covariance structure matrices have the forms $J = I - \exp(-|D|)$ and $J = \exp(-|D|)$. For this purpose, assume the second design (regular case) with four points in the interval $[0, 1]$ as it has mentioned before. This design sites shall be studied with one and two dimensional designs, i.e. $X = (\mathbf{1}\ \mathbf{x})$ and $X = (\mathbf{1}\ \mathbf{x_1}\ \mathbf{x_2})$ respectively. Here only the Euclidean distance is used for the two dimensional design. By multiplying the points vector with scalers

such as 10 and 100. One can move the sites of these design through three intervals, $(0, 1)$, $(1, 10)$ and $(10, 100)$. For these three cases the norms and the eigenvalues of matrices D, J and $K'JK$ are given in Table 4.

Designs	Eig.(D) ; Cov. function Eig.(J) ; $J = I - exp(-\|D\|)$	$\parallel D \parallel$	$\parallel J \parallel$	$\parallel K'JK \parallel$
1-(0.1,0.4,0.7,1)	-1.0243,-0.3487,-0.1757,1.5487 -1.8741, 0.3191,0.7267,0.8282	1.5487	1.8741	0.8226
2-(1,4,7,10)	-10.2426, -3.4868,-1.7574,15.4868 -0.0828,-0.0285,0.0329,0.0784	15.4868	0.0828	0.0732
3-(10,40,70,100)	-102.4264,-34.8683,-17.5736,154.8683 -0.0000,-0.0000 ,0.0000,0.0000	154.8683	1.5141e-013	1.4036e-013
1-(0.1,0.4,0.7,1) (0.7333,0.2,0.4667,1)	-1.1017,-0.7178,-0.3271,2.1467 -1.5283, 0.3288,0.4901,0.7094	2.1467	1.5283	0.7037
2-(1,4,7,10) (7.333,2,4.667,10)	-11.0172,-7.1784,-3.2711,21.4668 -0.0185,-0.0001,0.0004,0.0182	21.4668	0.0185	0.0162
3-(10,40,70,100) (73.33,20,46.67,100)	-110.1719,-71.7845,-32.7112,214.6675 -0.0000,-0.0000,0.0000,0.0000	214.6675	3.6902e-018	3.0750e-018
Designs	eig.(D) ; Cov. function eig.(J) ; $J = exp(-\|D\|)$	$\parallel D \parallel$	$\parallel J \parallel$	$\parallel K'JK \parallel$
1-(0.1,0.4,0.7,1)	-1.0243,-0.3487,-0.1757,1.5487 0.1718,0.2733,0.6809,2.8741	1.5487	2.8741	0.2841
2-(1,4,7,10)	-10.2426,-3.4868,-1.7574,15.4868 0.9216,0.9671,1.0285,1.0828	15.4868	1.0828	0.9727
3-(10,40,70,100)	-102.4264,-34.8683,-17.5736,154.8683 1.0000,1.0000,1.0000,1.0000	154.8683	1.0000	1.0000
1-(0.1,0.4,0.7,1) (0.7333,0.2,0.4667,1)	-1.1017,-0.7178,-0.3271,2.1467 0.2906,0.5099,0.6712,2.5283	2.1467	2.5283	0.2963
2-(1,4,7,10) (7.333,2,4.667,10)	-11.0172,-7.1784,-3.2711,21.4668 0.9818,0.9996,1.0001,1.0185	21.4668	1.0185	0.9838
3-(10,40,70,100) (73.33,20,46.67,100)	-110.1719,-71.7845,-32.7112,214.6675 1.0000,1.0000,1.0000,1.0000	214.6675	1.0000	1.0000

Table 4: Eigenvalues and norms of some matrices show the influence of the scaling on the design points.

One can note that the eigenvalues of matrix J are all positive while for matrix D all negative except one is positive as it has been mentioned before. Moreover, one can see how the norm of the covariance structure matrix J is convergent to one when the distance between the sites becomes larger while this norm has its maximum in the first interval case when the distances between the sites lie in the interval $[0, 1]$. This is one of the important aspect of the exponential function effecting on the distances between the positions of the points. It can be seen that $\parallel K'JK \parallel$ is always less than $\parallel J \parallel$, too, which is what Poincare's separation theorem has asserted. Furthermore, the difference between the outcomes of $\parallel K'JK \parallel$ for the one and two dimensional designs are very near to each other compared to the outcomes of $\parallel J \parallel$. For example, for the covariance structure $\exp(-|D|)$ in the first case, the difference is $0.2963 - 0.2841 = 0.0122$ while for norm of matrix J is $2.8741 - 2.5283 = 0.3458$, the same thing for the first covariance model. Anyway, $\parallel K'JK \parallel$ is also convergent to one when the design points scaled by a high number such 100, as it is shown in

the third case. But in such case, the rounding errors play a rule to realize when exactly the norm of matrix J would be equal one or equal the size of this matrix. For this reason, one might expect for the third case of the model $\exp(-|D|)$ above that $\parallel J \parallel > 1$ and $\parallel K'JK \parallel < 1$ or very near to one. This can be noted for the first covariance model, where $(1.5141e^{-013})$ is a little bigger than $(1.4036e^{-013})$ but both near to zero.

It is clear that the norm of the covariance structure matrix J lies between the size of this matrix and one. This happens when the distance matrix D is multiplied by scalers which are small and big enough. In the first case, the spectral norm (largest singular value) of matrix J would be convergent to the size of the this matrix while in the second case when the scaler is big enough, the norm would be convergent to one. Let four sites of the observations ($n = 4$) in one dimensional design and consider two scalers such as z_{11} (small enough) and z_{22} (big enough), therefore one can write that

$$J = exp(-|z_{11}D|) \xrightarrow{z_{11} \to 0} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$J = exp(-|z_{22}D|) \xrightarrow{z_{22} \to \infty} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

For that, suppose the design of sites $\mathbf{x} = (0.2; 0.3; 0.6; 0.8)$ and let $z_{11} = 0.01$, $z_{22} = 10$, hence the distance matrix D and $J = \exp(-|D|)$ of these sites are given by

$$D = \begin{pmatrix} 0 & 0.1000 & 0.4000 & 0.6000 \\ 0.1000 & 0 & 0.3000 & 0.5000 \\ 0.4000 & 0.3000 & 0 & 0.2000 \\ 0.6000 & 0.5000 & 0.2000 & 0 \end{pmatrix} ; J = \begin{pmatrix} 1.0000 & 0.9048 & 0.6703 & 0.5488 \\ 0.9048 & 1.0000 & 0.7408 & 0.6065 \\ 0.6703 & 0.7408 & 1.0000 & 0.8187 \\ 0.5488 & 0.6065 & 0.8187 & 1.0000 \end{pmatrix}$$

Hence one finds that $\parallel J \parallel = 3.1493$. By multiplying the distance matrix D mentioned above by $z_{22} = 10$ and $z_{11} = 0.01$, new matrices arise

$$z_{22}D = \begin{pmatrix} 0 & 1.0000 & 4.0000 & 6.0000 \\ 1.0000 & 0 & 3.0000 & 5.0000 \\ 4.0000 & 3.0000 & 0 & 2.0000 \\ 6.0000 & 5.0000 & 2.0000 & 0 \end{pmatrix} ; z_{11}D = \begin{pmatrix} 0 & 0.0100 & 0.0400 & 0.0600 \\ 0.0100 & 0 & 0.0300 & 0.0500 \\ 0.0400 & 0.0300 & 0 & 0.0200 \\ 0.0600 & 0.0500 & 0.0200 & 0 \end{pmatrix}.$$

Assume that $J_1 = \exp(-|z_{22} * D|)$ and $J_2 = \exp(-|z_{11} * D|)$. Now, one gets

$$J_1 = \begin{pmatrix} 1.0000 & 0.3679 & 0.0183 & 0.0025 \\ 0.3679 & 1.0000 & 0.0498 & 0.0067 \\ 0.0183 & 0.0498 & 1.0000 & 0.1353 \\ 0.0025 & 0.0067 & 0.1353 & 1.0000 \end{pmatrix} ; J_2 = \begin{pmatrix} 1.0000 & 0.9900 & 0.9608 & 0.9418 \\ 0.9900 & 1.0000 & 0.9704 & 0.9512 \\ 0.9608 & 0.9704 & 1.0000 & 0.9802 \\ 0.9418 & 0.9512 & 0.9802 & 1.0000 \end{pmatrix}.$$

From that one gets $\parallel J_1 \parallel = 1.3758$ which is smaller than the norm of matrix J and near to one. On the other hand, $\parallel J_2 \parallel = 3.8973$, is clearly bigger than $\parallel J \parallel$ and in the same time near to the size of matrix J which is in this case equal four ($n = 4$).

4.3 The influence of the covariance structure on the best design

The aim here is to study the four designs which has been concentrated on in the first subsection (4.1) but with other covariance structure functions, through that one can note the influence of covariance structure on the design or on the best design which has supplied the minimum bayes risk. For this purpose, some covariance structures are presented here which have appeared in some references such $\exp(-|D|)$ which has been used by Putter and Young (2001). They used the form $\theta_1 \exp(-\theta_2|D|)$, where we fix $\theta_2 = 1$ in our system. One can use also the function $\exp(-|D^2|)$ which has been handled by Schmidt and O'Hagan (2003) and Lee et al. (2004). They used $\exp(-\beta D^2)$, where β can be fixed as a constant such $1/2a$, where a is stands for the range as they have mentioned. Also the function $J = (D)^{-1}$ used here which has been introduced in Davis (1986) as a distance weighting functions in contouring programs. Last type of covariance structure handled here is $(I+D^2)^{-1}$, this function is a special case of the model $(I + D^{\alpha})^{-\beta}$, $0 < \alpha < 2$, $\beta > 0$, which was presented by Gneiting and Schlather (2004). Tables 5 and 6 show the corresponding results of using these covariance structures with the four designs of four points considered before in cases of stationary, one and two dimensional designs.

Designs	Matrix X	Cov. function	Eig.(D)	Eig.(J)	$\parallel J \parallel$	$\parallel K'JK \parallel$	$\parallel K'JK \otimes K'JK \parallel$	Bayes risk
1-x=(0.1,0.25,0.85,1)	$X = \mathbf{1}$	$\exp(-\lvert D \rvert)$	-1.3685,-0.1624,-0.1315,1.6624	0.1275,0.1427,0.9171,2.8127	2.8127	0.9171	0.8411	0.8856
2-x=(0.1,0.4,0.7,1)			-1.0243,-0.3487,-0.1757,1.5487	0.1718,0.2733,0.6809,2.8741	2.8741	0.6809	0.4636	3.9366
3-x=(0.1,0.49,0.61,1)			-0.9180,-0.4709,-0.1020,1.4909	0.1012,0.3468,0.6053,2.9467	2.9467	0.6053	0.3664	4.6978
4-x=(0.4,0.5,0.6,1)			-0.7648,-0.1870,-0.0660,1.0177	0.0651,0.1633,0.5240,3.2476	3.2476	0.5417	0.2934	2.9779
1	$X = \mathbf{1}$	$\exp(-\lvert D^2 \rvert)$	=	-0.9593,-0.0086,-0.0012,2.2421	2.2421	0.9593	0.9203	28.1069
2			=	-0.5230,-0.0630,-0.0059,2.3247	2.3247	0.5230	0.2736	35.4538
3			=	-0.3932,-0.1333,-0.0031,2.4242	2.4242	0.3932	0.1546	56.1581
4			=	-0.4092,-0.0316,-0.0025,3.1440	3.1440	0.4091	0.1674	39.4348
1	$X = \mathbf{1}$	$(D)^{-1}$	=	0.6015,-0.7307,-7.6026,-6.1571	7.6026	7.6026	57.7993	5.7173e+147
2			=	-5.6904,-2.8679,0.6457,-0.9763	5.6904	5.6904	32.3802	2.3142e+127
3			=	-9.8082,0.6708,-2.1237,-1.0893	9.8082	9.8082	96.2001	1.8372e+167
4			=	-15.1603,-5.3481,0.9826,-1.3075	15.1603	15.1556	229.6931	5.5683e+198
1	$X = \mathbf{1}$	$(I + D^2)^{-1}$	=	0.2657,0.3481,0.9743,0.9830	0.9830	0.9830	0.9663	1.1458
2			=	0.2943,0.4880,0.8916,0.9700	0.9700	0.9700	0.9410	2.4391
3			=	0.3103,0.5427,0.8185,0.9897	0.9897	0.9897	0.9795	4.3494
4			=	0.4912,0.6310,0.9662,0.9957	0.9957	0.9955	0.9911	1.9236
1	$X = \mathbf{1}$	$I - \exp(-\lvert D \rvert)$	=	-1.8127,0.0829,0.8573,0.8725	1.8127	0.8725	0.7613	0.2609
2			=	-1.8741,0.3191,0.7267,0.8282	1.8741	0.8282	0.6859	4.2497
3			=	-1.9467,0.3947,0.6532,0.8988	1.9467	0.8988	0.8078	4.7920
4			=	-2.2476,0.4760,0.8367,0.9349	2.2476	0.9348	0.8739	2.9788
1	$X = (\mathbf{1}\ \ \mathbf{x})$	$\exp(-\lvert D \rvert)$	=		2.8127	0.1446	0.0209	4.5577e+004
2			=		2.8741	0.2841	0.0807	213.5108
3			=		2.9467	0.3692	0.1363	29.4799
4			=		3.2476	0.1710	0.0292	279.0482
1	$X = (\mathbf{1}\ \ \mathbf{x})$	$\exp(-\lvert D^2 \rvert)$	=		2.2421	0.0158	2.4842e-004	4.6622e+004
2			=		2.3247	0.0536	0.0029	3.4488e+003
3			=		2.4242	0.1207	0.0146	332.5452
4			=		3.1440	0.0380	0.0014	3.4688e+003
1	$X = (\mathbf{1}\ \ \mathbf{x})$	D^{-1}	=		7.6026	7.5641	57.2156	3.0580e+038
2			=		5.6904	5.6667	32.1111	8.3490e+031
3			=		9.8082	9.8057	96.1524	5.5909e+040
4			=		15.1603	15.1367	229.1189	1.3728e+049
1	$X = (\mathbf{1}\ \ \mathbf{x})$	$(I + D^2)^{-1}$	=		0.9830	0.9794	0.9593	509.2495
2			=		0.9700	0.9676	0.9363	31.8256
3			=		0.9897	0.9896	0.9793	18.4657
4			=		0.9957	0.9952	0.9904	32.2143
1	$X = (\mathbf{1}\ \ \mathbf{x})$	$I - \exp(-\lvert D \rvert)$	=		1.8127	0.8635	0.7456	4.6090e+003
2			=		1.8741	0.8226	0.6767	43.2778
3			=		1.9467	0.8986	0.8075	11.7031
4			=		2.2476	0.9340	0.8724	32.2821

Table 5: Some covariance structures with four point designs in two cases of regressor functions.

Designs	D type	Cov. function	$\parallel D \parallel$	Eig.(J)	$\parallel J \parallel$	$\parallel K'JK \parallel$	Bayes risk
1		$\exp(-\lvert D\rvert)$	2.4632	0.3725,0.5881,0.6933,2.3461	2.3461	0.3735	0.0112
2	L_2		2.1467	0.2906,0.5099,0.6712,2.5283	2.5283	0.2963	0.0028
3			2.0500	0.2721,0.4676,0.6661,2.5941	2.5941	0.2845	0.0021
4			1.4923	0.1281,0.2226,0.6582,2.9910	2.9910	0.1328	2.4593e-006
1		$\exp(-\lvert D^2\rvert)$		-0.2706,-0.1823,0.0164,0.9495	0.9495	0.0144	1.5743e-013
2				-0.3066,-0.1602,-0.0118,1.3642	1.3642	0.0147	1.7966e-013
3			=	-0.3165,-0.1249,-0.0401,1.5185	1.5185	0.0371	2.7740e-010
4				-0.5694,-0.0347,-0.0104,2.4151	2.4151	0.0131	1.4042e-006
1		D^{-1}		-2.4546,0.4060,-1.1089,-0.8666	2.4546	2.4509	3.2857e+014
2				-3.0571,0.4658,-1.3931,-0.9077	3.0571	3.0199	2.1806e+018
3			=	0.4878,-0.9149,-1.5779,-3.0938	3.0938	3.0103	1.9096e+018
4				-7.2466,-3.9838,0.6701,-0.9064	7.2466	7.1744	5.7358e+033
1		$(I+D^2)^{-1}$		0.8577,0.1415,0.5515,0.4289	0.8577	0.8567	0.4348
2				0.1783,0.9033,0.6599,0.4517	0.9033	0.8960	0.6761
3			=	0.1922,0.4557,0.7135,0.9054	0.9054	0.8904	0.6300
4				0.3099,0.4510,0.9407,0.9813	0.9813	0.9752	2.3658
1		$I-\exp(-\lvert D\rvert)$		-1.3461,0.3067,0.4119,0.6275	1.3461	0.6265	0.1072
2				-1.5283,0.3288,0.4901,0.7094	1.5283	0.7037	0.1569
3			=	-1.5941,0.3339,0.5324,0.7279	1.5941	0.7155	0.1664
4				-1.9910,0.3418,0.7774,0.8719	1.9910	0.8672	0.4841
1		$\exp(-\lvert D\rvert)$	3.1270	0.3744,0.7151,0.8024,2.1081	2.1081	0.3760	0.0117
2	L_1		2.8771	0.3870,0.6163,0.7743,2.2225	2.2225	0.3940	0.0151
3			2.7406	0.3296,0.5797,0.7709,2.3198	2.3198	0.3539	0.0083
4			2.0804	0.1774,0.2844,0.7857,2.7525	2.7525	0.1840	0.0001
1		$\exp(-\lvert D^2\rvert)$		-0.2236,-0.1819,0.0890,0.4514	0.4514	0.0804	1.5841e-007
2				-0.2065,-0.1250,0.0205,0.6115	0.6115	0.0077	1.0958e-015
3			=	-0.2111,-0.0894,-0.0284,0.7596	0.7596	0.0192	1.4917e-012
4				-0.6126,-0.0352,-0.0067,1.5920	1.5920	0.0095	2.2719e-006
1		D^{-1}		-3.7051,0.3198,-0.7832,-0.6328	3.7051	3.6802	8.1453e+021
2				-2.2237,0.3476,-1.0521,-0.6771	2.2237	2.1951	2.9628e+012
3			=	0.3649,-2.4467,-1.1594,-0.6805	2.4467	2.3660	7.3249e+013
4				0.4807,-0.6443,-3.0012,-5.1248	5.1248	5.0733	4.3544e+027
1		$(I+D^2)^{-1}$		0.9321,0.0928,0.3802,0.2860	0.9321	0.9271	1.0399
2				0.8318,0.1078,0.5254,0.3144	0.8318	0.8238	0.3240
3			=	0.8569,0.5734,0.1175,0.3165	0.8569	0.8342	0.3532
4				0.1877,0.2933,0.9001,0.9633	0.9633	0.9559	1.6622
1		$I-\exp(-\lvert D\rvert)$		-1.1081,0.1976,0.2849,0.6256	1.1081	0.6240	0.1058
2				-1.2225,0.2257,0.3837,0.6130	1.2225	0.6060	0.0960
3			=	-1.3198,0.2291,0.4203,0.6704	1.3198	0.6461	0.1185
4				-1.7525,0.2143,0.7156,0.8226	1.7525	0.8160	0.3048
1		$\exp(-\lvert D\rvert)$	2.1997	0.4334,0.5004,0.6116,2.4546	2.4546	0.4464	0.0282
2	L_{max}		1.8738	0.2334,0.4676,0.6257,2.6733	2.6733	0.2424	0.0007
3			1.7756	0.2439,0.3811,0.6457,2.7293	2.7293	0.2567	0.0011
4			1.2031	0.0923,0.2092,0.5684,3.1301	3.1301	0.0967	5.2944e-007
1		$\exp(-\lvert D^2\rvert)$		-0.2263,-0.1248,-0.0840,1.2388	1.2388	0.0869	2.0629e-007
2				-0.3182,-0.1657,-0.0187,1.7550	1.7550	0.0207	2.7479e-012
3			=	-0.3871,-0.0995,-0.0322,1.9265	1.9265	0.0418	7.0438e-010
4				-0.4616,-0.0373,-0.0073,2.8544	2.8544	0.0107	1.0208e-005
1		D^{-1}		0.4546,-1.0670,-1.4325,-1.7718	1.7718	1.7182	6.9800e+007
2				-3.8959,0.5337,-1.5494,-1.0291	3.8959	3.8201	3.8068e+022
3			=	0.5632,-0.9950,-2.0314,-3.5933	3.5933	3.4539	5.9818e+020
4				-10.2276,-4.2706,0.8312,-1.1478	10.2276	10.1211	6.2631e+039
1		$(I+D^2)^{-1}$		0.1713,0.5324,0.6724,0.7584	0.7584	0.7413	0.1900
2				0.2217,0.5144,0.7059,0.9382	0.9382	0.9276	0.8992
3			=	0.2408,0.4975,0.8049,0.9281	0.9281	0.9122	0.8374
4				0.4086,0.5685,0.9480,0.9905	0.9905	0.9853	2.8798
1		$I-\exp(-\lvert D\rvert)$		-1.4546,0.3884,0.4996,0.5666	1.4546	0.5536	0.0699
2				-1.6733,0.3743,0.5324,0.7666	1.6733	0.7576	0.2078
3			=	-1.7293,0.3543,0.6189,0.7561	1.7293	0.7433	0.1920
4				-2.1301,0.4316,0.7908,0.9077	2.1301	0.9033	0.7430

Table 6: Some covariance structures with four point designs of three types of distance measures in two dimensions.

For these cases one can see that the first design has also the minimum norm of matrix J in case the covariance structures contain exponential function such as $\exp(-|D|)$ and $\exp(-|D^2|)$. Therefore, these models with $I - \exp(-|D|)$ assert that the design which has maximum norm of matrix D would has minimum norm for covariance structure matrix J. The models $\exp(-|D|)$ and $(I + D^2)^{-1}$ have all positive eigenvalues while the model D^{-1} has all negative eigenvalues except one. It is clear from this that if the $\parallel K'JK \parallel$ is less than one then $\parallel K'JK \bigotimes K'JK \parallel$ is also less than one and if the first is greater than one, the second one also would be so. By the way, the minimum norm of the decomposed matrix $K'JK$ is obtained from the case of the covariance structure $\exp(-|D^2|)$, the model D^{-1} gives maximum norm of matrix J, $K'JK$ and $K'JK \bigotimes K'JK$ where it is greater than one most of time. Furthermore, for this model the design which has the minimum norm for the matrices J and $K'JK$ is the same design which has minimum Bayes risk. On the other hand, the model $(I+D^2)^{-1}$ achieves always $\parallel J \parallel$ and $\parallel K'JK \parallel$ less than one, in other word this model satisfies always the convergence condition of the infinite series matrices.

Regarding Bayes risk, the models D^{-1} and $\exp(-|D^2|)$ gives the maximum values for Bayes risk while the covariance structures $I - \exp(-|D|)$, $\exp(-|D|)$ and $(I + D^2)^{-1}$ achieve minimum values and their outcomes of Bayes risk are near each other. In case of the design matrix $X = \mathbf{1}$, all the covariance structures except the model D^{-1} support that the first design has the minimum Bayes risk, therefore this design is the best one in this case and one can see that this design has a stable behavior with different covariance structures. The same thing is repeated with the case $X = (\mathbf{1}\ \mathbf{x})$ but this time the third design gives the minimum Bayes risk. Concern the case $X = (\mathbf{1}\ \mathbf{x_1}\ \mathbf{x_2})$, the Euclidean distance L_2 gives for most of the covariance models the first design with minimum Bayes risk which was the same also with the model $I - \exp(-|D|)$ in the first subsection. While for the city-block (L_1) and L_{max} distance measures, most the models achieve the second and first design to be the best with minimum Bayes risk respectively.

Now one can consider again these four designs which have presented in first subsection but with five points. Since the covariance structure models $I - \exp(-|D|)$, $exp(-|D|)$ and $(I + D^2)^{-1}$ have supplied similar results regarding minimum Bayes risk, these three models shall be now considered. Tables 7 and 8 give the results of these covariance models for the designs with five points.

Designs	Matrix X	Cov. function	$\parallel D \parallel$	$\parallel J \parallel$	$\parallel K'JK \parallel$	Bayes risk
1-x=(0.1,0.2125,0.3250,0.8875,1) 2-x=(0.1,0.3250,0.5500,0.7750,1) 3-x=(0.1,0.5050,0.5500,0.5950,1) 4-x=(0.3250,0.4250,0.5250,0.6250,1)	$X = \mathbf{1}$	$\exp(-\lvert D \rvert)$	2.0160 1.8648 1.6793 1.3451	3.5674 3.6270 3.8740 4.0183	1.0747 0.7895 0.5998 0.6057	5.5442e+005 6.6348 11.3032 4.3306
1 2 3 4	$X = \mathbf{1}$	$I - \exp(-\lvert D \rvert)$	2.0160 1.8648 1.6793 1.3451	2.5674 2.6270 2.8740 3.0183	0.9264 0.8765 0.9704 0.9420	2.6785 11.2874 9.5247 8.2532
1 2 3 4	$X = \mathbf{1}$	$(I + D^2)^{-1}$	2.0160 1.8648 1.6793 1.3451	0.9945 0.9845 0.9991 0.9966	0.9944 0.9842 0.9991 0.9966	2.5516 4.7575 5.9034 3.5895
1 2 3 4	$X = (\mathbf{1}\ \ \mathbf{x})$	$\exp(-\lvert D \rvert)$	=	=	0.2026 0.3062 0.4666 0.2545	161.0905 42.4556 4.2905 58.2566
1 2 3 4	$X = (\mathbf{1}\ \ \mathbf{x})$	$I - \exp(-\lvert D \rvert)$	=	=	0.9259 0.8455 0.9704 0.9418	4.8422 7.1271 2.1617 3.2463
1 2 3 4	$X = (\mathbf{1}\ \ \mathbf{x})$	$(I + D^2)^{-1}$	=	=	0.9942 0.9743 0.9991 0.9965	0.1133 0.9785 1.4274 0.3877

Table 7: Three covariance structures with five point designs in two cases of regressor functions.

Designs ; L_2 for D	Cov. function	$\parallel D \parallel$	$\parallel J \parallel$	$\parallel K'JK \parallel$	Bayes risk
$1 - \mathbf{x_1} = (0.1, 0.2125, 0.3250, 0.8875, 1)$ $\mathbf{x_2} = (0.3775, 0.8962, 0.17, 0.2737, 1)$	$\exp(-\lvert D \rvert)$	3.0499	2.9478	0.3993	69.3032
$2 - \mathbf{x_1} = (0.1, 0.3250, 0.5500, 0.7750, 1)$ $\mathbf{x_2} = (0.5850, 0.7925, 0.17, 0.3775, 1)$		2.6446	3.1486	0.2762	941.7829
$3 - \mathbf{x_1} = (0.1, 0.5050, 0.5500, 0.5950, 1)$ $\mathbf{x_2} = (0.5850, 0.6265, 0.17, 0.5435, 1)$		2.3741	3.3839	0.4087	21.4041
$4 - \mathbf{x_1} = (0.3250, 0.4250, 0.5250, 0.6250, 1)$ $\mathbf{x_2} = (0.5075, 0.6075, 0.17, 0.3075, 0.4075)$		1.6771	3.7067	0.1679	1.4884e+003
1	$I - \exp(-\lvert D \rvert)$	3.0499	1.9478	0.7690	19.9829
2		2.6446	2.1486	0.7740	137.7259
3		2.3741	2.3839	0.8931	3.8205
4		1.6771	2.7067	0.8768	38.2191
1	$(I + D^2)^{-1}$	3.0499	0.9459	0.9457	2.7255
2		2.6446	0.9481	0.9481	3.1666
3		2.3741	0.9885	0.9883	1.3632
4		1.6771	0.9848	0.9842	0.1459
Designs ; L_1 for D	Cov.	$\parallel D \parallel$	$\parallel J \parallel$	$\parallel K'JK \parallel$	Bayes risk
1	$\exp(-\lvert D \rvert$	3.8491	2.6587	0.3505	3.7486e+003
2		3.5123	2.7656	0.3513	1.7276e+003
3		3.1001	3.1265	0.4306	24.6477
4		2.1419	3.4224	0.2242	2.8029e+003
1	$I - \exp(-\lvert D \rvert)$	3.8491	1.6587	0.6723	1.2382e+003
2		3.5123	1.7656	0.6824	550.7681
3		3.1001	2.1265	0.8457	6.2297
4		2.1419	2.4224	0.8063	238.2867
1	$(I + D^2)^{-1}$	3.8491	0.9690	0.9675	1.3350
2		3.5123	0.9105	0.9002	15.1345
3		3.1001	0.9743	0.9720	1.7296
4		2.1419	0.9615	0.9521	3.4363
Designs ; L_{max} for D	Cov.	$\parallel D \parallel$	$\parallel J \parallel$	$\parallel K'JK \parallel$	Bayes risk
1	$\exp(-\lvert D \rvert$	2.7217	3.0913	0.5137	13.0178
2		2.3198	3.3406	0.3019	52.3366
3		2.0745	3.5224	0.4067	17.6324
4		1.5378	3.8137	0.1444	612.9243
1	$I - \exp(-\lvert D \rvert)$	2.7217	2.0913	0.8377	5.0844
2		2.3198	2.3406	0.9133	3.8542
3		2.0745	2.5224	0.9250	2.6537
4		1.5378	2.8137	0.9276	5.5254
1	$(I + D^2)^{-1}$	2.7217	0.9814	0.9799	1.6390
2		2.3198	0.9993	0.9992	0.5932
3		2.0745	0.9951	0.9948	1.2186
4		1.5378	0.9962	0.9961	0.0305

Table 8: Three covariance structures and some distance measure with five point designs in two dimensions.

For these outcomes, one can note the same properties which were achieved for the case with four points. For example, the models $\exp(-|D|)$ and $I - \exp(-|D|)$ give also the first design as the minimum norm of matrix J and in the same time the last design is the minimum norm of matrix D. The model $(I + D^2)^{-1}$ gives always the norm of matrix J less than one, so the matrices $K'JK$ and $K'JK \bigotimes K'JK$ have always norm or maximum eigenvalues less than one according to Poincare's separation theorem. The covariance structure $\exp(-|D|)$ supplies large values for Bayes risk comparing with the other two models. Anyway, the model $I - \exp(-|D|)$ in the cases $X = \mathbf{1}$ and $X = (\mathbf{1}\ \mathbf{x})$ achieves that the first and third designs are the best with respect to minimum Bayes risk, like in the case of four points. But for the case of two dimensions, it has shown for the case of four points that the first design

in two types of distance measures has minimum Bayes risk, while here all the three distance measures give the third design with least Bayes risk. Hence, for the cases $X = \mathbf{1}$ and $X = (\mathbf{1}\ \mathbf{x})$, two from these covariance models achieve that the first and third design have minimum Bayes risk respectively. While for the two dimensional case, also two from these covariance structures support that the third design has the minimum Bayes risk. So one might consider for the case of $X = \mathbf{1}$ that the first design as the best while for the case of $X = (\mathbf{1}\ \mathbf{x})$ and two dimensional designs the third design as the best concerning achieving the minimum Bayes risk.

Example 4.1. Consider two dimensional designs with four points in the interval $[1, 15]^2$, these points compose some geometrical forms (rectangular, square, triangular, kite, and trapezoid) as it is shown in the Figure 6 below. The summation of all the axes coordinates of these designs are equal to (40), for instance, $\mathbf{x_1} = (1, 12, 1, 12), \mathbf{x_2} = (1, 1, 6, 6)$, which is $1 + 12 + 1 + 12 + 1 + 1 + 6 + 6 = 40$ and so on. Suppose the covariance structure matrix is $J = \exp(-|D|)$, where D is the distance matrix in the metric space and for the prior information let $C = I_{2\times 2}$. Here three types of distance measures (Euclidean, city block L_1 norm and L_{max} norm) are considered for the purpose of comparing the outcomes of these designs. The outcomes are shown in Table 9.

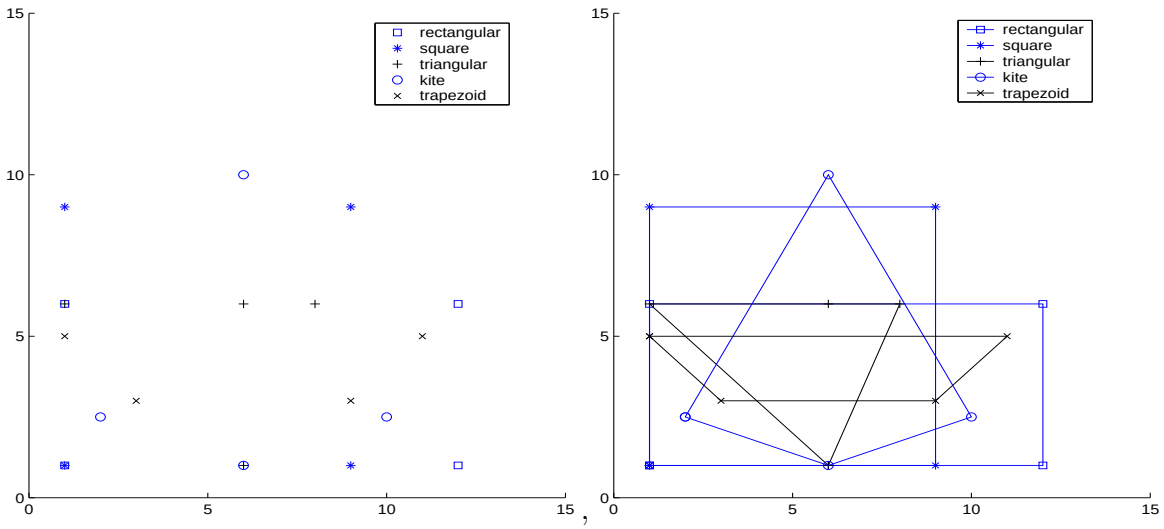

Figure 6: Some geometrical shapes are formed by four points.

D and design shape	Designs	$\parallel D \parallel$	$\parallel J \parallel$	$\parallel K'JK \parallel$	$\parallel K'JK \bigotimes K'JK \parallel$	Bayes risk
L_2: rectangular	1-x1=(1, 12, 1, 12) x2=(1 , 1,6, 6)	28.0830	1.0068	0.9933	0.9865	3.3805
square	2-x1=(1, 9, 1, 9) x2=(1, 1, 9, 9)	27.3137	1.0007	0.9993	0.9987	3.8362
triangular	3-x1=(6, 1, 6, 8) x2=(1, 6, 6, 6)	16.0727	1.1360	0.8764	0.7680	0.5341
kite	4-x1=(2, 6, 10, 6) x2=(2.5, 10, 2.5, 1)	21.5661	1.0199	0.9811	0.9625	2.6369
trapezoid	5-x1=(1, 3, 9, 11) x2=(5, 3, 3, 5)	19.2538	1.0606	0.9462	0.8954	1.4195
L_1: =	1	32.0000	1.0068	0.9932	0.9865	3.3801
=	2	32.0000	1.0007	0.9993	0.9987	3.2325
=	3	18.7761	1.1358	0.8764	0.7680	0.5364
=	4	26.0141	1.0060	0.9946	0.9892	3.4884
=	5	22.1421	1.0197	0.9820	0.9644	2.7287
L_∞: =	1	27	1.0068	0.9933	0.9866	3.3812
=	2	24	1.0010	0.9997	0.9993	3.3503
=	3	14.6611	1.1363	0.8764	0.7680	0.5352
=	4	20.2133	1.0261	0.9750	0.9507	2.3821
=	5	18.1980	1.1369	0.8790	0.7727	0.5505

Table 9: Outcomes of some geometrical shapes of four point designs in two dimensions with some type of distance measures.

From these outcomes clearly the third design which represents the triangular case has the minimum Bayes risk and minimum norm of matrices D, $K'JK$ and $K'JK \bigotimes K'JK$. The minimum norm for the covariance matrix J is obtained in the second case (square shape) and very near to the first one (rectangular shape). These results are valid for all three types of the distance measures. Bayes risk achieved by using these distance types are very near to each other and the least one is given in case of Euclidean distance (L_2 norm). For the norm of the distance matrix D, one can note that the L_∞ distance achieves the least one since this kind of distance takes only the maximum difference of the distance values between the pairs of the designs points. On the other hand, the norms of matrices J, $K'JK$ and $K'JK \bigotimes K'JK$ for all three metric distances are very near to each other. □

4.4 Prior knowledge about the parameters

The parameter of interest in the survey is the value of the probability p. Conceivable, this could be any value in the interval $(0, 1)$. From Bayesian viewpoint a person expresses one's belief about the location of population probability, before observing any data, by means of probability density which called a *prior density*, since it reflects a person's beliefs prior to observing survey data. To illustrate such a prior density in inference, consider a person who has little knowledge about this parameter. For this individual all values of the probability p between 0 and 1 may be equally approved a priori. To deal with this information about p, this person's prior beliefs may be summarized by a uniform density on the unit interval; that is

$$g(p) = 1, \quad 0 < p < 1,$$

as in Figure 7 (left). This density might be summarized by computing probability or area under the density. This prior is often called *vague* or *noninformative* since it reflects a lack of prior information about the value of the parameter.

A second person may have more precise beliefs about the location of the parameter's value. Suppose this person believes or has some information about the location of the parameter. For this person, the prior density for p might be concentrated on small values in the area. One density that might represent this prior information is displayed in Figure 7 (right). One notes that most of the mass of this density lies between 0 and 0.7. Such a prior is called an *informative prior*. See Johnson and Albert (1999).

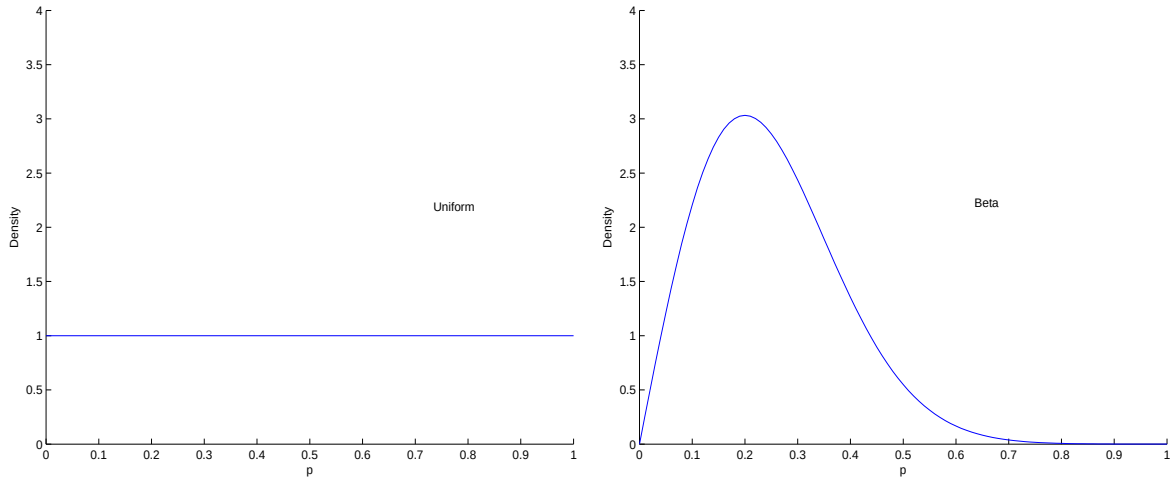

Figure 7: Two prior densities,(left) uniform reflects vague prior beliefs and(right) informative density reflects prior some beliefs.

For purpose of dealing with prior knowledge, let us consider here some prior information about the parameters. Jeffrey (1961) has given a famous rule based on the invariance of a transformation of the parameters. According to this approach, Koch (1990) has given the noninformative prior density function for the variance σ^2 as follows

$$p(\sigma^2) \propto \frac{1}{\sigma^2} \quad with\ 0 < \sigma^2 < \infty.$$

In this context, we have given a prior information in Section 2 (Eq.(3.2.12)) according to noninformative Jeffrey's prior. The uniform distribution which represents

noninformative prior has considered and has obtained the matrix

$$C = \begin{pmatrix} \frac{\varphi_1^2+\varphi_1\varphi_2+\varphi_2^2}{3} & \frac{(\varphi_1+\varphi_2)(\varphi_3+\varphi_4)}{4} \\ \frac{(\varphi_1+\varphi_2)(\varphi_3+\varphi_4)}{4} & \frac{\varphi_3^2+\varphi_3\varphi_4+\varphi_4^2}{3} \end{pmatrix} \tag{4.4.1}$$

where $\varphi_1 \leq \theta_1 \leq \varphi_2$ and $\varphi_3 \leq \theta_2 \leq \varphi_4$. The square root of this matrix represents the prior information as it has mentioned before. By the way, the square root of the symmetric matrix has been calculated by using Matlab program. Assume that $0 \leq \theta_1 \leq 2.5$ and $0 \leq \theta_2 \leq 0.5$, one gets the following prior matrix C with its square root matrix R

$$C = \begin{pmatrix} 2.0833 & 0.3125 \\ 0.3125 & 0.0833 \end{pmatrix} \ , R = \begin{pmatrix} 1.4309 & 0.1896 \\ 0.1896 & 0.2177 \end{pmatrix}.$$

One can also consider the diagonal matrix of C as a prior knowledge matrix, i.e.

$$C = \begin{pmatrix} \frac{\varphi_1^2+\varphi_1\varphi_2+\varphi_2^2}{3} & 0 \\ 0 & \frac{\varphi_3^2+\varphi_3\varphi_4+\varphi_4^2}{3} \end{pmatrix}. \tag{4.4.2}$$

So for the last prior information matrix C, one gets the follows diagonal square root matrix

$$R = \begin{pmatrix} 1.4434 & 0 \\ 0 & 0.2887 \end{pmatrix}.$$

In this context, Berger and Bernardo(1989) and Datta and Ghosh(1996) have considered two independent normal random variables X_1 and X_2 with unit variances and means $\mu_1(>0)$ and $\mu_2(>0)$, respectively. They have given the prior matrix of the parameterization $\theta_1 = \mu_1\mu_2$ and $\theta_2 = \sqrt{(}\mu_2/\mu_1)$ as follows

$$R = \begin{pmatrix} \frac{\theta_2^2+\theta_2^{-2}}{4\theta_1} & \frac{\theta_2(1-\theta_2^{-4})}{2} \\ \frac{\theta_2(1-\theta_2^{-4})}{2} & \theta_1(1+\theta_2^{-4}) \end{pmatrix}. \tag{4.4.3}$$

Assume for that, $\mu_1 = 2.5$ and $\mu_2 = 0.5$ then from Eq. (4.4.3) above one obtains that

$$R = \begin{pmatrix} 1.6250 & -5.3666 \\ -5.3666 & 32.5000 \end{pmatrix}.$$

Another kind of the prior knowledge which has presented by Stein(1987). He has supposed gaussian process which has semivariogram and has shown that

$$var(\hat{\theta}_1) \sim (2\theta_2\theta_1^3/r)^{1/2}4N^{-1/2}, var(\hat{\theta}_2) \sim 2\theta_2^2N^{-1} \ \ and \ \ cov(\hat{\theta}_1, \hat{\theta}_2) \sim -2\theta_1\theta_2N^{-1}, \tag{4.4.4}$$

where N is the number of the observations and r stands for the rank of design matrix X. Hence, when one considers a design with four sites i.e. $N = 4$ and regression

functions in cases $r = 1$, $r = 2$ and $r = 3$, these correspond to $\mathbf{X} = \mathbf{1}$, $\mathbf{X} = (\mathbf{1}\ \ \mathbf{x})$ and $\mathbf{X} = (\mathbf{1}\ \mathbf{x_1}\ \mathbf{x_2})$ respectively. Suppose that $\theta_1 = 2.5$ and $\theta_2 = 0.5$, so one gets from Eq. (4.4.4) for $r = 1$, $r = 2$ and $r = 3$ the following matrices,

$$\begin{pmatrix} 7.9057 & -0.6250 \\ -0.6250 & 0.1250 \end{pmatrix}, \begin{pmatrix} 5.5902 & -0.6250 \\ -0.6250 & 0.1250 \end{pmatrix}, \begin{pmatrix} 4.5644 & -0.6250 \\ -0.6250 & 0.1250 \end{pmatrix}$$

respectively. Let us denote for the square root matrix R obtained from Eq. (4.4.1) and Eq. (4.4.2) by $pr(1)$ and $pr(2)$ respectively. Also $pr(3)$ for the prior matrix corresponds to Eq. (4.4.3) and $pr(4)$ for the prior matrices achieved from Eq. (4.4.4). Table 10 shows the results corresponding to the first design of four points case, which is $\mathbf{x_1} = (0.1, 0.25, 0.85, 1)$ with covariance structure $\exp(-|D|)$ for these four types of the prior information.

Designs	Matrix X	Prior	Distance kind	Bayes risk
1-x=(0.1,0.25,0.85,1)	$\mathbf{X} = \mathbf{1}$	pr(1) pr(2) pr(3) pr(4)	one dimension	1.0966e+005 2.9635e+006 7.9335e+013 2.7134e+007
	$\mathbf{X} = (\mathbf{1}\ \mathbf{x})$	pr(1) pr(2) pr(3) pr(4)	one dimension	1.1441e+004 2.9895e+005 4.8211e+019 9.6211e+008
1-$\mathbf{x_1}$=(0.1,0.25,0.85,1) $\mathbf{x_2}$=(0.8667,0.2,0.3333,1)	$\mathbf{X} = (\mathbf{1}\ \mathbf{x_1}\ \mathbf{x_2})$	pr(1) pr(2) pr(3) pr(4)	Euclidean distance	2.0333 44.7514 1.3492e+015 0.0324
		pr(1) pr(2) pr(3) pr(4)	L_1 distance	1.9453 42.4271 1.3349e+015 0.0523
		pr(1) pr(2) pr(3) pr(4)	L_{max} distance	0.5399 9.6796 9.6586e+014 9.2411

Table 10: Some prior information with the first design of four points.

One notes that the first prior type achieves the minimum Bayes risk compared to the other prior knowledge in case the regressor functions are in the one dimensional designs but its results are very near to the second prior type, both of them based on Jeffrey' prior. On the other hand, for the case of two dimensional designs, one can note that the first and fourth type of these priors give the least Bayes risk. The third type gives in all these cases bigger results compared to the other. Furthermore, these prior information achieve Bayes risk for case of two dimensional designs less than the other cases of one dimensional designs except the third prior type.

4.5 A practical example

It is nice to find applications where one could handle with the theoretical formulas of any mathematical model. Hence, we turn now to apply our procedure of estimating spatial covariance components through using a real data, these data describe the pH measurements at 38 sites in Lower Saxony precipitation network in Germany for April 1987. Berke(1998) has used this data with the exponential variogram model, the model components are used to interpret the spatial variation of the process. He has dealt with restricted maximum likelihood (REML) to find the estimations of the spatial parameters.

It is interesting to find such estimations of these parameters by using our system (BAIQUE), furthermore, to compare the results with REML results which appeared in Berke's article. In our model the parameter θ_1 corresponding to the sill and θ_2 to the nugget effect which are famous largely in spatial statistics, moreover the variance is $\sigma^2 = \theta_1 + \theta_2$. According to the covariance function which Berke has used in his model, we consider the covariance structure function in the form $J = \exp(-|0.4 * D/0.1|)$, where 0.1 stands for the range of the spatial pH phenomena which it regards here as a constant. REML estimates which have found in his work are $(\hat{\theta}_1, \hat{\theta}_2) = (0.4, 0.07)$, hence $\sigma^2 = \widehat{(\theta_1 + \theta_2)} = 0.47$. Therefore, one can use some prior information about the parameters θ_1 and θ_2, for this purpose, one can consider Jeffrey's prior when the parameters are independent. Hence, corroding to the outcomes of REML, let us suppose the prior information matrix as a diagonal, such as

$$\begin{pmatrix} 0.3 & 0 \\ 0 & 0.05 \end{pmatrix}.$$

Doing that, since the number of observations in our data here is $n = 38$, one has to solve the equations system of size 1446×1446. For this equations system BAIQUE has given the following outcomes for this data

$$(\widehat{\theta_1 + \theta_2}, \hat{\theta}_1, \hat{\theta}_2) = (0.5014, 0.4618, 0.0395).$$

While their Bayes risk are

$$(0.0114, 0.0172, 0.0013),$$

respectively. Table 11 shows the results of our procedure BAIQUE and REML which has found by Berke. Moreover, we have computed some other methods to find their estimators corresponding to the same data, through that one can compare the results

of these methods with our procedure. For this aim, we have computed maximum likelihood estimation of spatial covariance parameter which has been dealt by Eulogio Pardo-Iguzquiza (1998) and Pardo-Iguzquiza and Dowd (1998). Also MINQUE (minimum norm quadratic unbiased estimator) has computed for this data. This approach has been used by Sultan and Quassim (1997) as well. Another procedure deals with minimum mean square which has found by LaMotte (1973). This procedure considers the translation invariant unbiased estimator of variance components. We have also calculated here the corresponding results of this procedure for this data. The other methods which have calculated here were introduced by Volaufova and Witkovsky (1991). These methods are LBIMSE which stands for locally best-MSE and invariant estimator, OLS which represents ordinary least squares estimator and finally, MLSE denotes modified least squares estimator.

For that, the outcomes of these methods are shown in the Figures 8 and 9. These figures show three estimated values $\hat{\theta}_1$, $\hat{\theta}_2$ and $(\widehat{\theta_1 + \theta_2})$ of each method. In each figure the outcomes obtained by BAIQUE and also the empirical variance (which is 0.4982) of the data observations are given. The number of data observations are $(n = 38)$ as mentioned before. One can note from that, that BAIQUE gives very near estimates to the empirical variance of the data, see figure (8), also both of BAIQUE and REML achieve very similar estimates compared to the other procedures. In the second class, one can diagnose that MLSE, OLS and MINQUE have similar results. While the estimates of LBIMSE ML and LaMotte(1973) are farther from the real variance of the data and thus also farther from the estimates obtained by BAIQUE and REML. Finally, we mention again that we have used Matlab program to calculate the outcomes.

Method type	$\hat{\theta}_1$	$\hat{\theta}_2$	$\widehat{\theta_1 + \theta_2}$	Empirical variance
BAIQUE	0.4618	0.0395	0.5014	0.4982
REML	0.40	0.07	0.47	0.4982
MINQUE	0.6798	-0.0299	0.6499	0.4982
ML	0.6323	0.3340	0.8900	0.4982
LBIMSE	0.2305	0.1153	0.3458	0.4982
MLSE	0.4393	-0.0272	0.4121	0.4982
OLS	0.5136	0.0944	0.6080	0.4982
LaMotte(1973)	0.6056	0.3340	0.9396	0.4982

Table 11: Estimates are obtained by applying BAIQUE and other approaches for pH data of Lower Saxony.

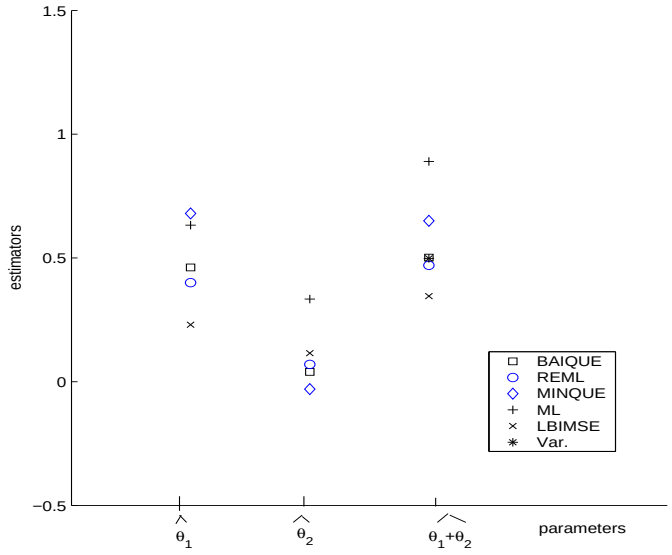

Figure 8: Comparing BAIQUE and REML results with the estimators of other methods.

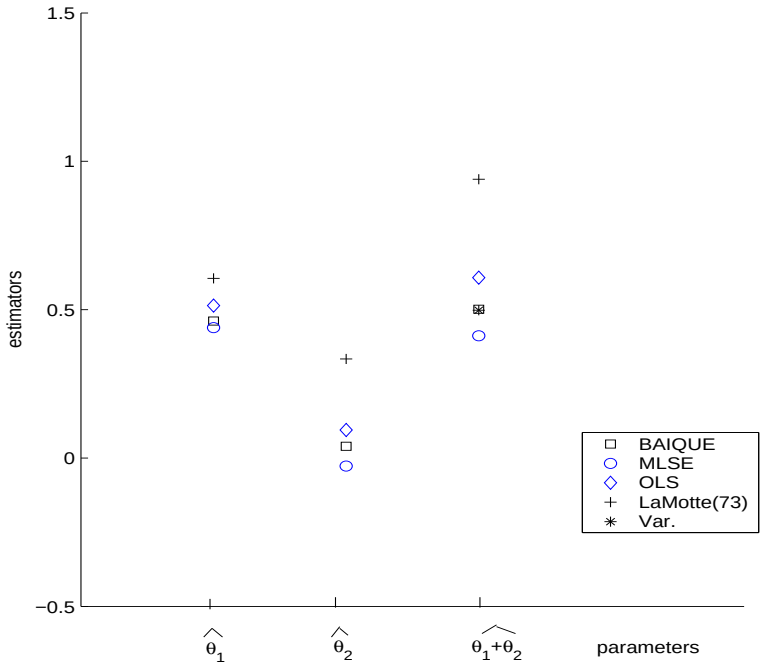

Figure 9: Comparing BAIQUE results with the estimators of other methods.

5 Appendices

5.1 Some properties of Kronecker product and vec operation

Definition 5.1. Let $A = (a_1|\dots|a_p) \in \mathbb{R}^{p\times p}$ with $a_i \in \mathbb{R}^p$. The vec operation is defined as

$$vec(A) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}.$$

And the Mat_p operation is defined as

$$Mat_p\left(\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}\right) = (a_1|,\dots,|a_p).$$

Definition 5.2. See Harville(1997). The *Kronecker product* of two matrices, say an $m \times n$ matrix $A = a_{ij}$ and a $p \times q$ matrix $B = b_{ij}$, is denoted by the symbol $A \bigotimes B$ and is defined to be the $mp \times nq$ matrix expressible in the partitioned form as

$$A \bigotimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & \dots & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{pmatrix}.$$

Here some of properties of Kronecker product which are used in this work.

1) For any $m \times n$ matrix A, $p \times q$ matrix B, $n \times u$ matrix C and $q \times v$ matrix D. Then

$$(A \bigotimes B)(C \bigotimes D) = (AC) \bigotimes (BD).$$

2) For any m-dimensional column vector a and n-dimensional vector b. Then

$$vec(ba') = a \bigotimes b \quad and \quad a \bigotimes b' = b' \bigotimes a = ab'.$$

3) For any k scalers $c_1, c_2, \dots, c_k$ and for any k matrices $A_1, A_2, \dots, A_k$ of the same

size,

$$vec(\sum_{i=1}^{k} c_i A_i) = \sum_{i=1}^{k} c_i vec(A_i).$$

4) If there are a matrix such $A = (a_1, a_2, \ldots, a_n)$ and identity matrix

$$I_n = \begin{pmatrix} u_1' \\ u_2' \\ \vdots \\ u_n' \end{pmatrix}$$

where a_j denote the j^{th} column of A and u_j' is the j^{th} row of the matrix I_n (j=1,. . . ,n), then the matrix A can be written matrix as

$$A = \sum_{j=1}^{n} a_j u_j' = a_1 u_1' + a_2 u_2' + \ldots + a_n u_n'.$$

Lemma 5.3. *See Harville (1997) p.341. For any $m \times n$ matrix A, $n \times p$ matrix B, and $p \times q$ matrix C,*

$$vec(ABC) = (C' \bigotimes A) vec(B).$$

Proof. Making use of the properties of Kronecker product, matrix B is expressible as,

$$B = \sum_{j=1}^{p} v_j u_j'$$

where v_j is the j^{th} column of B and u_j' is the j^{th} row of I_p. Thus one can find out

$$\begin{aligned} vec(ABC) &= vec[A(\sum_{j}^{p} v_j u_j')C] \\ &= \sum_{j}^{p} vec(Av_j u_j' C) \\ &= \sum_{j}^{p} [(C'u_j) \bigotimes (Av_j)] \\ &= \sum_{j}^{p} (C' \bigotimes A)(u_j \bigotimes v_j) \\ &= \sum_{j}^{p} (C' \bigotimes A) vec(v_j u_j') \\ &= (C' \bigotimes A) vec(\sum_{j=1}^{p} v_j u_j') \\ &= (C' \bigotimes A) vec(B). \end{aligned}$$

By using Mat_p operation which has defined before, one gets

$$Mat_p(vec(ABC) = Mat_p(C' \bigotimes A \cdot vec(B)) = ABC.$$

□

Lemma 5.4. *Let V,$T \in R^{p \times p}$ are symmetric matrices. Then it holds*

$$Mat_p(T \bigotimes T \cdot vec(V)) = TVT,$$

where TVT is of dimension $p \times p$ symmetric matrix.

5) There are alternative expressions for the vec of the product AB of an $m \times n$ matrix A and an $n \times p$ matrix B can be obtained as a special cases of Lemma (5.3) which are as follows.
a)

$$vec(AB) = (I_p \bigotimes A) vec(B).$$

b)

$$vec(AB) = (B' \bigotimes I_m) vec(A).$$

c)

$$vec(AB) = (B' \bigotimes A)vec(I_n).$$

6) For any $m \times n$ matrices A and B,

$$tr(A'B) = (vecA)'vecB.$$

5.2 Some aspects of generalized inverse matrices

A matrix has an inverse only if it is square, and even then only if it is nonsingular or, in other word , if its columns (or row) are linearly independent. In recent years, some kind of partial inverse of a matrix that is singular or even rectangular have been needed in some numerous areas of applied mathematics. Let $A_{m\times n}$ matrix then a matrix $A^{-}_{n\times m}$ is said to be a *generalized inverse* (or a $g-inverse$) of A if $AA^{-}A = A$. If A is square and nonsingular, then $A^{-1} = A^{-}$ is unique, otherwise, A has infinitely many $g-inverse$.

In 1955 Penrose showed that, for every finite matrix A (square or rectangular) of real or complex elements, there is a unique matrix X satisfying the four equations (that we call the Penrose equations)

$$AXA = A, \;\; XAX = X, \;\; (AX)^* = AX, \;\; (XA)^* = XA,$$

where A^* denotes the conjugate transpose of A. It is commonly known as the $Moore-Penrose \;\; inverse$, and is often denoted by A^{+}. If A is nonsingular, then $X = A^{-1}$ trivially satisfies the four equations and it follows that the Moore-Penrose inverse of a nonsingular matrix is the same as the ordinary inverse. See Ben-Isreal and Greville (2003).

Definition 5.5. (Schur complement), see (Searle(1982), p.261).
The matrix of the expression

$$S = D - CA^{-1}B$$

which occurring in

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} -A^{-1}B \\ I \end{pmatrix} (D - CA^{-1}B)^{-1} \begin{pmatrix} -CA^{-1} & I \end{pmatrix}.$$

is called the *Schur complement* of matrix A in

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

when A is nonsingular. And for singular A, the matrix $D - CA^{-}B$ is said to the *generalized Schur complement* relative to A^{-}.

Lemma 5.6. *See Ehrbar (2003) or Henderson and Searle (1981).*
Given a $m \times n$ matrix A, a $m \times h$ matrix B, a $k \times n$ matrix C and $k \times h$ matrix D satisfying $AA^{-}B = BD^{-}D$ and $DD^{-}C = CA^{-}A$ where D^{-} is the Penrose inverse of D. Then

$$(A + BD^{-}C)^{-} = A^{-} - A^{-}B(D + CA^{-}B)^{-}CA^{-}.$$

Proof. Note: This lemma has proved by Ehrbar (2003) but in a short style, we give here a detailed proof of this lemma.

Define $E = D + CA^{-}B$, then it follows from the assumptions that

$$\begin{aligned}(A + BD^{-}C)(A^{-} - A^{-}BE^{-}CA^{-}) &= AA^{-} - AA^{-}BE^{-}CA^{-} + BD^{-}CA^{-} - BD^{-}CA^{-}BE^{-}CA^{-} \\ &= AA^{-} - BD^{-}DE^{-}CA^{-} + BD^{-}CA^{-} - BD^{-}CA^{-}BE^{-}CA^{-} \\ &= AA^{-} + BD^{-}(I - DE^{-} - CA^{-}BE^{-})CA^{-} \\ &= AA^{-} + BD^{-}(I - (D + CA^{-}B)E^{-})CA^{-} \\ &= AA^{-} + BD^{-}(I - EE^{-})CA^{-}.\end{aligned}$$

Since that $AA^{-}(A + BD^{-}C) = A + BD^{-}C$, we have to show that the second term on the right hand side annuls $(A + BD^{-}C)$, indeed,

$$\begin{aligned}&BD^{-}(I - EE^{-})CA^{-}(A + BD^{-}C) \\ &= BD^{-}CA^{-}A + BD^{-}CA^{-}BD^{-}C - BD^{-}EE^{-}CA^{-}A - BD^{-}EE^{-}CA^{-}BD^{-}C \\ &= BD^{-}DD^{-}C + BD^{-}CA^{-}BD^{-}C - BD^{-}EE^{-}DD^{-}C - BD^{-}EE^{-}CA^{-}BD^{-}C \\ &= BD^{-}(D + CA^{-}B - EE^{-}D - EE^{-}CA^{-}B)D^{-}C \\ &= BD^{-}(D + CA^{-}B - EE^{-}(D + CA^{-}B))D^{-}C \\ &= BD^{-}(E - EE^{-}E)D^{-}C \\ &= BD^{-}(E - E)D^{-}C = 0.\end{aligned}$$

Then one can get ,

$$\begin{aligned}&(A+BD^{-}C)(A^{-}-A^{-}BE^{-}CA^{-})(A+BD^{-}C)\\&=AA^{-}(A+BD^{-}C)+BD^{-}(I-EE^{-})CA^{-}(A+BD^{-}C)\\&=A+BD^{-}C+0.\end{aligned}$$

Therefore,

$$(A+BD^{-}C)^{-}=A^{-}-A^{-}B(D+CA^{-}B)^{-}CA^{-}.$$

□

Corollary 5.7. *(Sherman-Morrison-Woodbury), see Ehrbar (2003) and Henderson and Searle (1981).*
Given a $m\times n$ matrix A, a $m\times 1$ vector b satisfying $AA^{-}b=b$, a $n\times 1$ vector c satisfying $c'AA^{-}=c'$, and a scaler δ. If A^{-} is a g-inverse of A, then

$$A^{-}-\frac{A^{-}bc'A^{-}}{c'A^{-}b+\delta}\quad is\ \ a\ \ g-inverse\ \ of\quad A+\frac{bc'}{\delta}$$

Lemma 5.8. *Let $Z\in R^{p\times p}$ of rank $r<p$ and idempotent matrix. Then there exists $K\in R^{p\times r}$ such that $Z=KK'$ and $K'K$ is nonsingular and $K'K=I$. See searle (1982), Lee et al. (1977) and Bapat (2000).*

Proof. Note: We give here a detailed proof which has been combined from the above references.

Since matrix Z is symmetric , there exist nonsingular orthogonal matrix P such that

$$P'ZP=\begin{pmatrix}I_r & 0\\0 & 0\end{pmatrix}.$$

Which is equivalent to

$$Z=P\begin{pmatrix}I_r & 0\\0 & 0\end{pmatrix}P',$$

where $P'=P^{-1}$. So matrix P can be partitioned as

$$P=[K_{p\times r}\quad W_{p\times(p-r)}].$$

Since P is nonsingular orthogonal so its columns are linear independent. In particular, the r columns of K are linear independent and so rank(K)=r, (of course, with Z being $p \times p$ of rank r, it is clear that $r \leq p$, the number of row in K). So one has

$$\begin{aligned} Z &= P \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} P' \\ &= \begin{pmatrix} K_{p\times r} & W_{p\times(p-r)} \end{pmatrix} \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} K'_{r\times p} \\ W'_{(p-r)\times p} \end{pmatrix} \\ &= \begin{pmatrix} K_{p\times r} & 0 \end{pmatrix} \begin{pmatrix} K'_{r\times p} \\ W'_{(p-r)\times p} \end{pmatrix} \\ &= K_{p\times r} K'_{r\times p} \\ &= Z'. \end{aligned}$$

Now since the columns of matrix K are linear independent, let $K = [x_1, x_2, \ldots, x_r]$ where $x_1, x_2, \ldots, x_r$ are the eigenvectors of matrix Z corresponding to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_r$ respectively. Since that $x_i' x_i = 1$ and $x_i' x_j = 0 \quad (i \neq j)$, therefore one gets $K'K = I$. □

5.3 Infinite series representation of the matrix inverse

Let $A_1, A_2, \ldots$ represent a sequence of $m \times n$ matrices, and for $(i = 1, \ldots, m, j = 1, \ldots, n$ and $k = 1, 2 \ldots)$ let $a_{ij}^{(k)}$ represent the ij^{th} element of A_k. If for every i and j there exists a scaler a_{ij} such that a_{ij} is the limit of the sequence $a_{ij}^{(1)}, a_{ij}^{(2)}, \ldots$, we say that the $m \times n$ matrix A, whose ij^{th} element is a_{ij}, is the limit of the sequence $A_1, A_2, \ldots$(or that the sequence $A_1, A_2, \ldots$ converges to A) and write $\lim_{k\to\infty} A_k = A$ (or $A_k \to A$). If the sequence $A_1, A_2, \ldots$ has a limit, it is said to be *convergent*. If the sequence does not have a limit for some i and j, it is said to be *divergent*. Some basic properties of the sequence of matrices are presented in the following.

1) For any $m \times n$ matrix A, the sequence A, A,... (each of whose members equal A) converges to A.
2) For any scaler c (including $c = -1$) and for any sequence $A_1, A_2, \ldots$ of $m \times n$ matrices that converge to an $(m \times n)$ matrix A, $\lim_{k\to\infty}(cA_k) = cA$.
3) If $A_1, A_2, \ldots$ and $B_1, B_2, \ldots$ are sequence of $m \times n$ that converge to $(m \times n)$ matrices A and B, respectively, then

$$\lim_{k\to\infty} (A_k + B_k) = A + B.$$

4) For a square matrix A, if it has multiplied n times with own, gives

$$A.A \ldots A = A^n.$$

5) If $A_1, A_2, \ldots$ is a sequence of $m \times n$ matrices that converges to an $(m \times n)$ matrix A and $B_1, B_2, \ldots$ a sequence of $n \times p$ matrices that converges to an $(n \times p)$ matrix B, then $\lim_{k\to\infty} A_k B_k = AB$. See Harville(1997).

Lemma 5.9. *Let $A_0, A_1, A_2, \ldots$ represent a sequence of $m \times n$ matrices. If the infinite series $\sum_{k=0}^{\infty} A_k$ converges, then $\lim_{k\to\infty} A_k = 0$.*

Proof. Suppose that $\sum_{k=0}^{\infty} A_k$ converges. And, for $k = 0, 1, 2, \ldots$, let $S_k = \sum_{i=0}^{k} A_i$. Then in light of our supposition the sequence $S_0, S_1, S_2, \ldots$ has a limit, say S. Thus, observing that (for $k = 1, 2, \ldots$) $A_k = S_k - S_{k-1}$ and that $\lim_{k\to\infty} S_{k-1} = S$, it follows that

$$\lim_{k\to\infty} A_k = \lim_{k\to\infty} (S_k - S_{k-1}) = \lim_{k\to\infty} S_k - \lim_{k\to\infty} S_{k-1} = S - S = 0.$$

□

Theorem 5.10. *(Neumann series), see Harville(1997),p.429-430.*
Let A represent an $n \times n$ matrix. Then, the infinite series $I + A + A^2 + A^3 + \ldots$ converges if and only if $\lim_{k\to\infty} A^k = 0$, in which case $I - A$ is nonsingular and

$$(I - A)^{-1} = \sum_{k=0}^{\infty} A^k = I + A + A^2 + A^3 + \ldots \tag{5.3.1}$$

where $A^0 = I$.

Proof. If the infinite series $I + A + A^2 + A^3 + \ldots$ converges, then it is clear from Lemma (5.9) that $\lim_{k\to\infty} A^k = 0$.
Conversely, suppose that $\lim_{k\to\infty} A^k = 0$. And observe that

$$(I + A + A^2 + \ldots + A^k)(I - A) = I - A^{k+1} \tag{5.3.2}$$

also

$$\lim_{k\to\infty} (I - A^{k+1}) = I - \lim_{k\to\infty} A^{k+1} = I - 0 = I. \tag{5.3.3}$$

Then, for any $n \times 1$ vector $\mathbf{x}$ such that $(I - A)\mathbf{x} = 0$, one finds

$$(I - A^{k+1})\mathbf{x} = (I + A + A^2 + \ldots + A^k)(I - A)\mathbf{x} = 0$$

$(k = 0, 1, 2, \ldots)$ and consequently

$$\mathbf{x} = I\mathbf{x} = [\lim_{k\to\infty}(I - A^{k+1})]\mathbf{x} = \lim_{k\to\infty}(I - A^{k+1})\mathbf{x} = \lim_{k\to\infty} 0 = 0.$$

Thus, $I - A$ is nonsingular. Further, postmultiplying both sides of equality (5.3.2) by $(I - A)^{-1}$, one obtains the equality

$$(I + A + A^2 + \ldots + A^k) = (I - A^{k+1})(I - A)^{-1}.$$

And, in light of result (5.3.3), one concludes that the infinite series $\sum_{k=0}^{\infty} A^k$ converges and that

$$\begin{aligned}\sum_{k=0}^{\infty} A^k &= \lim_{k\to\infty}(I - A^{k+1})(I - A)^{-1} \\ &= [\lim_{k\to\infty}(I - A^{k+1})](I - A)^{-1} = I(I - A)^{-1} = (I - A)^{-1}.\end{aligned}$$

□

This theorem can be generalized as follows.

Theorem 5.11. *See Harville(1997), p.430-431. Let A and B represent $n \times n$ matrices. Suppose that B is nonsingular, and define $F = B^{-1}A$. Then the infinite series $B^{-1} + FB^{-1} + F^2B^{-1} + F^3B^{-1} + \ldots$ converges if and only if $\lim_{k\to\infty} F^k = 0$, in which case $B - A$ is nonsingular and*

$$(B - A)^{-1} = \sum_{k=0}^{\infty} F^k B^{-1} = B^{-1} + FB^{-1} + F^2B^{-1} + F^3B^{-1} + \ldots \qquad (5.3.4)$$

where $F^0 = I$.

Formula (5.3.1) is applicable if and only if $\lim_{k\to\infty} A^k = 0$, and, more generally, formula (5.3.4) is applicable if and only if $\lim_{k\to\infty} F^k = 0$. Depending on the nature of A (or F), it may be difficult to determine whether the condition $\lim_{k\to\infty} A^k = 0$ (or the condition $\lim_{k\to\infty} F^k = 0$) is satisfied. A condition that is more stringent, but that is typically easier to check, can be obtained from the Theorem (5.14).

Definition 5.12. The spectral norm of a matrix $A \in C^{m\times n}$ is given by

$$\| A \|_2 = max\{\sqrt{\lambda} : \lambda \; an \; eigenvalue \; of \; A^*A\} \tag{5.3.5}$$

which is the Euclidean norm of the $mn-dimensional$ vector obtained by listing all components of A. Where $(*)$ stands for the conjugate transpose.

Definition 5.13. The spectral radius $\rho(A)$ of a square matrix $A \in C^{n\times n}$ is the maximal value among the n moduli of the eigenvalues of A,

$$\rho(A) = max\{|\lambda| : \lambda \in \lambda(A)\}. \tag{5.3.6}$$

Here two relations between the spectral radius and norms of the matrix.
1) Let $\| \, . \, \|$ be any multiplicative norm on $C^{n\times n}$. Then, for any $A \in C^{n\times n}$,

$$\rho(A) \leq \| A \| \, .$$

2) For any $A \in C^{n\times n}$, the spectral norm $\| \, . \, \|_2$ of (5.3.5) equals

$$\| A \|_2 = \rho^{1/2}(A^*A) = \rho^{1/2}(AA^*).$$

In particular, if A is Hermitian, then $\| A \|_2 = \rho(A)$. See Ben-Israel and Greville (2003).

Theorem 5.14. *See Harville (1997), p.431. Let A represent $n \times n$ matrix, If $||A|| < 1$, then $\lim_{k\to\infty} A^k = 0$. Where the norm is the usual norm.*

Corollary 5.15. *See Kincaid and Cheney(1991), p.172-173.*
If A is an $n \times n$ matrix such that $||A|| < 1$, then $I - A$ is invertible, and

$$(I - A)^{-1} = \sum_{k=0}^{\infty} A^k.$$

Corollary 5.16. *See Mirsky (1990).*
The series $\sum_{m=0}^{\infty} A_m$ is absolutely convergent if and only if $\sum_{m=0}^{\infty} ||A_m||$ is convergent.

Theorem 5.17. *See Mirsky (1990)*
i) If all characteristic roots of A lie in the interior of the circle of convergence of the power series

$$\phi(z) = \sum_{m=0}^{\infty} c_m z^m \tag{5.3.7}$$

then the matrix power series

$$\sum_{m=0}^{\infty} c_m A^m \tag{5.3.8}$$

converges absolutely.
ii) If at least on characteristic root of A lies outside the circle of convergence of (5.3.7), then (5.3.8) diverges

Definition 5.18. If all characteristic roots of A lie in the interior of the circle of convergence of the power series (5.3.7), the $\phi(A)$ is defined as the sum of the series (5.3.8).

Theorem 5.19. *See Mirsky (1990), p.337-338. Let*

$$\phi(z) = \sum_{m=0}^{\infty} a_m z^m, \qquad \psi(z) = \sum_{m=0}^{\infty} b_m z^m, \qquad \chi(z) = \sum_{m=0}^{\infty} c_m z^m,$$

so that

$$c_m = a_0 b_m + a_1 b_{m-1} + \ldots + a_m b_0 \qquad (m = 0, 1, 2, \ldots).$$

be sums of power series convergent for $|z| < R$ and suppose that

$$\phi(z)\psi(z) = \chi(z) \qquad (|z| < R).$$

If all characteristic roots of A are less than R in modulus, then

$$\phi(A)\psi(A) = \chi(A).$$

Proof. Let

$$d_m = \sum_{v=0}^{m} |a_v||b_{m-v}|.$$

Then $\sum_{m=0}^{\infty} d_m z^m$ is convergent for $|z| < R$ and therefore by Theorem (5.17) (i), $\sum_{m=0}^{\infty} d_m A^m$ is absolutely convergent. Hence, by Corollary (5.16),

$$\sum_{m=0}^{\infty}(\sum_{v=0}^{m} |a_v||b_{m-v}|) \parallel A^m \parallel = \sum_{m=0}^{\infty} \parallel d_m A^m \parallel$$

is convergent. Since this is a series of non-negative terms, we see that

$$\sum_{m=0}^{\infty}\sum_{v=0}^{m} |a_v||b_{m-v}| \parallel A^m \parallel = \sum_{m=0}^{\infty}\sum_{v=0}^{m} \parallel a_v b_{m-v} A^m \parallel$$

is convergent. Hence, again by Corollary (5.16), the matrix series

$$\sum_{m=0}^{\infty}\sum_{v=0}^{m} a_v b_{m-v} A^m$$

is absolutely convergent and may therefore be rearranged. Accordingly, we have

$$\begin{aligned}
\chi(A) &= \sum_{m=0}^{\infty} c_m A^m = \sum_{m=0}^{\infty}(\sum_{v=0}^{m} a_v b_{m-v}) A^m = \sum_{m=0}^{\infty}\sum_{v=0}^{m} a_v b_{m-v} A^m \\
&= \sum_{v=0}^{\infty} a_v \sum_{m=v}^{\infty} b_{m-v} A^m = \sum_{v=0}^{\infty} a_v \sum_{\mu=0}^{\infty} b_\mu A^{v+\mu} = \sum_{v=0}^{\infty} a_v A^v \sum_{\mu=0}^{\infty} b_\mu A^\mu \\
&= \phi(A)\psi(A).
\end{aligned}$$

□

Theorem 5.20. *(The spectral theorem), see Bapat (2000), p.20.*

Let A be a symmetric $n \times n$ matrix. Then there exists an orthogonal matrix P such that

$$P'AP = diag(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

5.4 Distribution of quadratic forms

When the random variables in $\mathbf{x}'=[x_1 \quad x_2 \ldots x_n]$ have a multivariate normal distribution with vector of means μ and variance-covariance matrix V, we write "$\mathbf{x}$ is $N(\mu, V)$" or "$\mathbf{x}\sim N(\mu, \sigma^2)$". When $E(x_i = \mu)$ for all i then $\mu = \mu 1$; and if the $x_i's$ are mutually independent, all with the same variance σ^2, then $V = \sigma^2 I$ and we write "$\mathbf{x}$ is $N(\mu 1, \sigma^2 I)$". This is equivalent to the more usual notation $NID(\mu, \sigma^2)$, but by retaining the matrix notation of $N(\mu 1, \sigma^2 I)$ we emphasize that this is just special case of the general multivariate normal $N(\mu, V)$. Where $V > 0$ positive definite, then the joint probability density function is

$$f(x_1, x_2, \ldots, x_n) = \frac{\exp(-\frac{1}{2}(\mathbf{x}-\mu)'\, V^{-1}(\mathbf{x}-\mu))}{(2\pi)^{\frac{1}{2}n}|V|^{\frac{1}{2}}}.$$

If $V = \sigma^2 I_n$ then the probability density function simplifies to

$$f(\mathbf{x}) = \frac{\exp(-\frac{1}{2\sigma^2}(\mathbf{x}-\mu)'(\mathbf{x}-\mu))}{(2\pi\sigma^2)^{\frac{1}{2}n}}.$$

Consider that $\mathbf{X}$ is a random variable whose moment generating function is $M(t) = E(\exp(t\mathbf{x}))$ for all t. Since $M(0) = 1 \neq 0$ and that the composition of two functions which have power series expansion itself has a power series expansion, we may write

$$logM(t) = \sum_{n=0}^{\infty} \frac{K_n t^n}{n!},$$

for small t where the numbers $K_1, K_2, \ldots$in this expansion are called the cumulants. Notice that

$$K_n = [logM(t)]^{(n)}(0)$$

for all n, in particular $K_0 = log(M(0)) = log(1) = 0$. Since the function $logM$ generate the cumulants it is called cumulants generating function. For examples let $\mathbf{x} \sim \mathcal{N}(\mu,\, \sigma^2)$then

$$M(t) = \exp(t\mu + \frac{\sigma^2 t^2}{2}) \;\; and \;\; \log M(t) = t\mu + \frac{\sigma^2 t^2}{2},$$

for all t near zero and it gives $K_1 = \mu$, $K_2 = \sigma^2$, $K_3 = K_4 = \ldots = 0$. Since the first and second cumulants of any random variable are its mean and variance, the normal distribution has the simplest possible cumulants.

Lemma 5.21. *See Searle(1971).*
For any vector h and any positive definite symmetric matrix N then

$$(2\pi)^{\frac{1}{2}n}|N|^{\frac{1}{2}}\exp(\frac{1}{2}h'Nh) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp(-\frac{1}{2}\boldsymbol{x}'N^{-1}\boldsymbol{x}+h'\boldsymbol{x})dx_1,\ldots,dx_n \quad (5.4.1)$$

Proof. From the integral of a multivariate normal density $\mathcal{N}(\mu,\, N)$. We have

$$\begin{aligned}(2\pi)^{\frac{1}{2}n}|N|^{\frac{1}{2}} &= \textstyle\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}(\mathbf{x}-\mu)'N^{-1}(\mathbf{x}-\mu)\right]dx_1,\ldots dx_n\\ &= \textstyle\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\mathbf{x}'N^{-1}\mathbf{x}+\frac{1}{2}\mu'N^{-1}\mathbf{x}+\frac{1}{2}\mathbf{x}'N^{-1}\mu-\frac{1}{2}\mu'N^{-1}\mu\right]dx_1,\ldots,dx_n\\ &= \textstyle\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}exp\left[-\frac{1}{2}\mathbf{x}'N^{-1}\mathbf{x}+\mu'N^{-1}\mathbf{x}'-\frac{1}{2}\mu'N^{-1}\mu\right]dx_1,\ldots,dx_n\end{aligned}$$

Where $\frac{1}{2}\mu'N^{-1}\mathbf{x}$ and $\frac{1}{2}\mathbf{x}'N^{-1}\mu$ are scalers and have the same dimensions. Now writing h' for $\mu'N^{-1}$ this gives (5.4.1).

$$(2\pi)^{\frac{1}{2}n}|N|^{\frac{1}{2}}\exp(\frac{1}{2}h'Nh) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\mathbf{x}'N^{-1}\mathbf{x}+\mu'N^{-1}\mathbf{x}\right]dx_1,\ldots,dx_n$$

where

$$\frac{1}{2}h'Nh = \frac{1}{2}(\mu'N^{-1})N(N^{-1}\mu) = \frac{1}{2}\mu'N^{-1}NN^{-1}\mu = \frac{1}{2}\mu'N^{-1}\mu.$$

□

Theorem 5.22. *See Searle(1971).*
When $\mathbf{x}\sim N(\mu,V)$ and A then
1)

$$E(\mathbf{x}'A\mathbf{x}) = tr(AV)+\mu'A\mu;$$

(true also when $\mathbf{x}$ is non-normal);
2) the r^{th} cumulant of $\mathbf{x}'A\mathbf{x}$ is

$$K_r(\mathbf{x}'A\mathbf{x}) = 2^{r-1}(r-1)![tr(AV)^r + r\mu'A(VA)^{r-1}\mu];$$

where $r=1$ it gives $E(\mathbf{x}'A\mathbf{x})$ and where $r=2$ gives the variance of $\mathbf{x}'A\mathbf{x}$

$$Var(\mathbf{x}'A\mathbf{x}) = 2tr(AVAV)+4\mu'AVA\mu;$$

Proof. Note: This theorem has proved by Searle (1971) but in a short style, we give here a detailed proof of this theorem.

1) With $E(\mathbf{x})=\mu$ and Var($\mathbf{x}$)$=V$ we have

$$E(\mathbf{x}\mathbf{x}') = V+\mu\mu'.$$

Hence

$$\begin{aligned} E(\mathbf{x}'A\mathbf{x}) &= Etr(A\mathbf{x}\mathbf{x}') \\ &= tr[AE(\mathbf{x}\mathbf{x}')] \\ &= tr(AV + A\mu\mu') \\ &= tr(AV) + \mu' A\mu. \end{aligned}$$

It is clear from the proof that this part of the theorem holds whether $\mathbf{x}$ is normal or not.

2) The moment generating function (m.g.f.) of the quadratic form $\mathbf{x}'A\mathbf{x}$ is

$$\begin{aligned} M_{\mathbf{x}'A\mathbf{x}}(t) &= E(\exp(t\mathbf{x}'A\mathbf{x})) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(t\mathbf{x}'A\mathbf{x}) f(x_1, \ldots, x_n) dx_1, \ldots, dx_n \\ &= (2\pi)^{-\frac{1}{2}n} |V|^{-\frac{1}{2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[t\mathbf{x}'A\mathbf{x} - \frac{1}{2}(\mathbf{x}-\mu)'V^{-1}(\mathbf{x}-\mu) \right] dx_1, \ldots, dx_n \\ &= (2\pi)^{-\frac{1}{2}n} |V|^{-\frac{1}{2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[t\mathbf{x}'A\mathbf{x} - \frac{1}{2}\mathbf{x}'V^{-1}\mathbf{x} + \frac{1}{2}\mu'V^{-1}\mathbf{x} + \frac{1}{2}\mathbf{x}'V^{-1}\mu - \frac{1}{2}\mu'V^{-1}\mu \right] \\ & dx_1, \ldots, dx_n \\ &= \frac{\exp(-\frac{1}{2}\mu'V^{-1}\mu)}{(2\pi)^{\frac{1}{2}n}|V|^{\frac{1}{2}}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[t\mathbf{x}'A\mathbf{x} - \frac{1}{2}\mathbf{x}'V^{-1}\mathbf{x} + \mu'V^{-1}\mathbf{x} \right] dx_1, \ldots, dx_n, \end{aligned}$$

and on rearranging the exponent this becomes

$$M_{\mathbf{x}'A\mathbf{x}}(t) = \frac{\exp(-\frac{1}{2}\mu'V^{-1}\mu)}{(2\pi)^{\frac{1}{2}n}|V|^{\frac{1}{2}}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\mathbf{x}'(I - 2tAV)V^{-1}\mathbf{x} + \mu'V^{-1}\mathbf{x} \right] dx_1, \ldots, dx_n. \tag{5.4.2}$$

Now in Lemma (5.21) put $h' = \mu'V^{-1}$ and $N = [(I-2tAV)V^{-1}]^{-1} = V(I-2tAV)^{-1}$. So the right-hand side in (5.4.1) then equals the multiple integral in (5.4.2) which

becomes

$$\begin{aligned}
M_{\mathbf{x}'A\mathbf{x}}(t) &= \exp(-\frac{1}{2}\mu' V^{-1}\mu)|V|^{-\frac{1}{2}}|V(I-2tAV)^{-1}|^{\frac{1}{2}}\exp\left[\frac{1}{2}\mu' V^{-1}V(I-2tAV)^{-1}V^{-1}\mu\right] \\
&= \exp(-\frac{1}{2}\mu' V^{-1}\mu)|V^{-1}|^{\frac{1}{2}}|V(I-2tAV)^{-1}|^{\frac{1}{2}}\exp\left[\frac{1}{2}\mu'(I-2tAV)^{-1}V^{-1}\mu\right] \\
&= \exp(-\frac{1}{2}\mu' V^{-1}\mu)(|V^{-1}V(I-2tAV)^{-1}|)^{\frac{1}{2}}\exp\left[\frac{1}{2}\mu'(I-2tAV)^{-1}V^{-1}\mu\right] \\
&= \exp(-\frac{1}{2}\mu' V^{-1}\mu)(|(I-2tAV)|)^{-\frac{1}{2}}\exp\left[\frac{1}{2}\mu'(I-2tAV)^{-1}V^{-1}\mu\right] \\
&= |(I-2tAV)|^{-\frac{1}{2}} exp\left[-\frac{1}{2}\mu'(I-(I-2tAV)^{-1})V^{-1}\mu\right]. \qquad (5.4.3)
\end{aligned}$$

The cumulant generating function is the logarithm of the m.g.f. hence

$$\begin{aligned}
\sum_{r=1}^{\infty} K_r t^r/r! &= log[M_{\mathbf{x}'A\mathbf{x}}(t)] \\
&= -\frac{1}{2}log|I-2tAV| - \frac{1}{2}\mu'(I-(I-2tAV)^{-1})V^{-1}\mu \qquad (5.4.4)
\end{aligned}$$

The two parts of this are evaluated as follows. Use "λ_i of X" to denote the "ith latent root of X". Then for sufficiently small $|t|$

$$\begin{aligned}
-\frac{1}{2}log|I-2tAV| &= -\frac{1}{2}\sum_{i=1}^{n} log[\lambda_i \quad of \quad (I-2tAV)] \\
&= -\frac{1}{2}\sum_{i=1}^{n} log[1-2t(\lambda_i \ \ of \ \ AV)] \\
&= -\frac{1}{2}\sum_{i=1}^{n}\sum_{r=1}^{\infty} -[2t(\lambda_i \ \ of \ \ AV)]^r/r \\
&= \sum_{r=1}^{\infty} 2^{r-1}t^r/r \sum_{i=1}^{n}[(\lambda_i of AV)]^r \\
&= \sum_{r=1}^{\infty}(2^{r-1}t^r/r)tr(AV)^r.
\end{aligned}$$

And, by direct binomial expansion, for sufficiently small $|t|$, one obtains

$$\begin{aligned} I-(I-2tAV)^{-1} &= I-(I+2t(AV)+2^2t^2(AV)^2+\ldots) \\ &= -(2t(AV)+2^2t^2(AV)^2+\ldots) \\ &= -\sum_{r=1}^{\infty} 2^r t^r (AV)^r. \end{aligned}$$

Making these substitutions in (5.4.4) and equating the coefficients of t^r gives

$$\begin{aligned} \sum_{r=1}^{\infty} K_r t^r/r! &= \sum_{r=1}^{\infty}(2^{r-1}t^r/r)tr(AV)^r - \frac{1}{2}\mu'(-\sum_{r=1}^{\infty} 2^r t^r (AV)^r)V^{-1}\mu \\ &= \sum_{r=1}^{\infty} t^r[(2^{r-1}/r)tr(AV)^r + \frac{1}{2}\mu' 2^r (AV)^r V^{-1}\mu]. \end{aligned} \tag{5.4.5}$$

Since that $(AV)^r = A(VA)^{r-1}V$ therefore one gets

$$\begin{aligned} K_r(\mathbf{x}'A\mathbf{x}) &= (r-1)!\ 2^{r-1}tr(AV)^r + \frac{r!}{2}\mu' 2^r A(VA)^{r-1}VV^{-1}\mu \\ &= (r-1)!2^{r-1}tr(AV)^r + r!2^{r-1}\mu' A(VA)^{r-1}\mu \\ &= 2^{r-1}(r-1)![tr(AV)^r + r\mu' A(VA)^{r-1}\mu]. \end{aligned} \tag{5.4.6}$$

Equation (5.4.6) represent the r^{th} cumulant of $\mathbf{x}^{'}A\mathbf{x}$. An important application of this part of the theorem when $r = 1$ it gives $E(\mathbf{x}'A\mathbf{x})$ and when $r = 2$ it gives the variance of $\mathbf{x}^{'}A\mathbf{x}$

$$\begin{aligned} K_2(\mathbf{x}'A\mathbf{x}) &= Var(\mathbf{x}'A\mathbf{x}) \\ &= 2tr(AV)^2 + 4\mu' A(VA)\mu \\ &= 2tr(AVAV) + 4\mu' AVA\mu. \end{aligned}$$

So the theorem is proved. □

Corollary 5.23. *(Square root of a symmetric matrix) If A is a non-singular symmetric matrix, then for any integer n,*

$$\Lambda^n = diag(\lambda_i^n) \text{ and } A^n = \Gamma\Lambda^n\Gamma'. \tag{5.4.7}$$

If all the eigenvalues of A are positive then one can define the rational powers

$$A^{r/s} = \Gamma\Lambda^{r/s}\Gamma' \text{ where } \Lambda^{r/s} = diag(\lambda_i^{r/s}), \tag{5.4.8}$$

for integers $s > 0$ and r. If some of the eigenvalues of A are zero, then (5.4.7) and (5.4.8) hold if the exponents are restricted to be nonnegative.

Important special case of (5.4.8) are

$$A^{1/2} = \Gamma\Lambda^{1/2}\Gamma' \text{ where } \Lambda^{1/2} = diag(\lambda_i^{1/2}), \tag{5.4.9}$$

when $\lambda_i \geq 0$ for all i, so if $A = R^2$ then

$$R = \Gamma diag(\sqrt{\lambda}_1, \ldots, \sqrt{\lambda}_n)\Gamma'. \tag{5.4.10}$$

Moreover,

$$A^{-1/2} = \Gamma\Lambda^{-1/2}\Gamma' \text{ where } \Lambda^{-1/2} = diag(\lambda_i^{-1/2}), \tag{5.4.11}$$

when $\lambda_i > 0$ for all i. The decomposition in Eq. (5.4.9) or (5.4.10) is called the square root decomposition of A. This has the following properties

1) $(A^{1/2})' = A^{1/2}$ (that is, $A^{1/2}$ is symmetric).
2) $A^{1/2}A^{1/2} = A$.
3) $A^{1/2}A^{-1/2} = A^{-1/2}A^{1/2} = I$ and $A^{-1/2}A^{-1/2} = A^{-1}$ where $A^{-1/2} = (A^{1/2})^{-1}$. See Johnson and Wichern (1988) and Mardia et al. (1979).

Theorem 5.24. *See Koch (1999). Let $\boldsymbol{x}$ be an $n \times 1$ random vector with the covariance matrix $D(\boldsymbol{x})$ and A an $m \times n$ matrix and b an $m \times 1$ vector of constants. Then the covariance matrix of the $m \times 1$ random vector $\boldsymbol{y}$ resulting from the linear transformation $\boldsymbol{y} = A\boldsymbol{x} + b$ is given by*

$$D(\boldsymbol{y}) = D(A\boldsymbol{x} + b) = AD(\boldsymbol{x})A'.$$

Theorem 5.25. *Let $f(\boldsymbol{x})$ be a real-valued differentiable function of the vector $\boldsymbol{x} \in \mathbf{E}^n$. Furthermore, let $m \leq n$ and*

$$g_i(\boldsymbol{x}) = 0 \quad for \ \ i \in \{1, \ldots, m\},$$

where the functions $g_i(\boldsymbol{x})$ are real-valued and differentiable and the $n \times n$ matrix $B = (\partial g_i/\partial x_j)$ has full row rank m. Let $f(\boldsymbol{x})$ have a local extremum at the point $\boldsymbol{x}_0$ subject to the constrains $g_i(\boldsymbol{x}) = 0$. Then there exists the $m \times 1$ vector $\boldsymbol{k}= (k_i)$ of Lagrange multipliers such that the Lagrange function $w(\boldsymbol{x})$ with

$$w(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} k_i g_i(\boldsymbol{x})$$

has the stationary point $\boldsymbol{x}_0$, that is $\partial w(\boldsymbol{x})/\partial \boldsymbol{x}|\boldsymbol{x} = \boldsymbol{x}_0$.

Theorem 5.26. *a) Let A be an $m \times n$ and B an $n \times m$ matrix, then*

$$\partial tr(AB)/\partial A = B'.$$

b) Let A and B be quadratic and A in addition symmetrical, then

$$\partial tr(AB)/\partial A = B + B' - diag(B).$$

c) Let A be an $m \times n$ matrix and B and C two $n \times m$ matrices, then

$$\partial tr(ABAC)/\partial A = (BAC + CAB)'.$$

d) Let A, B and C be quadratic matrices, then

$$\partial tr(ABA'C)/\partial A = (BA'C)' + CAB.$$

5.5 Some qualities of non-negative matrices

Definition 5.27. 1) *Positive semidefinite matrix.* An $n \times n$ matrix A is defined to be positive semi-definite if and only if $A = A'$ and $\mathbf{y}'A\mathbf{y} \geq 0$ for each and every vector $\mathbf{y}$ in R^n and the equality holds for at least one vector $\mathbf{y}$ such that $\mathbf{y} \neq 0$.
2) *Positive definite matrix.* An $n \times n$ matrix A is defined to be positive definite if and only if $A = A'$ and $\mathbf{y}'A\mathbf{y} > 0$ for each and every vector $\mathbf{y}$ in R^n such that $\mathbf{y} \neq 0$.
(3) *Non − negative matrix.* A matrix is defined to be non-negative if and only if it is either positive definite or positive semidefinite.

Lemma 5.28. *See Graybill (1983). 1) If A is an $n \times n$ positive semidefinite matrix then,*
a) $a_{ii} \geq 0$ for all $i = 1, \ldots, n$; if $a_{tt} = 0$, then each element in the t^{th} row and t^{th} column of A is equal to zero.
b) $P'AP$ is a non-negative matrix for any $n \times n$ matrix P.
2) If A is an $n \times n$ positive definite matrix then,
a) $a_{ii} > 0$ for all $i = 1, \ldots, n$.
b) $P'AP$ is positive definite matrix for any nonsingular $n \times n$ matrix P, in particular, A^{-1} is positive definite.

Theorem 5.29. *See Harville(1997).*
Let A represent an $n \times n$ matrix, and P an $n \times m$ matrix.
1) If A is nonnegative definite, then $P'AP$ is nonnegative definite.
2) If A is nonnegative definite and $rank(P) < m$, then $P'AP$ is positive semidefinite.
3) If A is positive definite and $rank(P) = m$, then $P'AP$ is positive definite.

Corollary 5.30. *A symmetric nonnegative definite matrix is positive definite if and only if it is nonsingular (or equivalently, is positive semidefinite if and only if it is singular).*

Corollary 5.31. *An $n \times n$ matrix A is a symmetric positive definite matrix if and only if there exists a nonsingular matrix P such that $A = P'P$.*

Lemma 5.32. *See Bapat and Raghavan (1997). 1) If A is positive semidefinite then the matrix $\exp(a_{ij})$ is positive semidefinite.*
2) For any $\beta > 0$, the matrix $\exp(\beta a_{ij})$ is positive semidefinite.

Definition 5.33. *Conditional positive definite* (*conditional negative definite*). A real, symmetric $n \times n$ matrix A is said to be conditional positive definite (conditional negative definite) if $x'Ax \geq 0$ (≤ 0) for any $x \in H^n$, where

$$H^n = \{x \in R^n : \sum_{i=1}^{n} x_i = 0\}$$

Lemma 5.34. *1) If A is a conditional negative definite matrix, then A has at most one positive eigenvalue.*
2) Let A be a nonnegative, nonzero matrix that is conditional negative definite. Then A has exactly one positive eigenvalue.

If $x^1, \ldots, x^n$ are points in an Euclidean space, their squared distance matrix may be defined as the matrix of inter-point squared distances ($\| x_i - x_j \|^2$). Note that such a matrix is symmetric and has zeros on the main diagonal.

Lemma 5.35. *Let $x^1, \ldots, x^n \in R$ for some s, and let $A = (a_{ij})$ be the $n \times n$ matrix defined as $a_{ij} = \| x^i - x^j \|^2$, $i, j = 1, 2, \ldots, n$. Then A is conditional negative definite.*

Lemma 5.36. *Let A be a symmetric $n \times n$ matrix with zero diagonal entries. Then there exist vectors $x^1, \ldots, x^n \in R^s$ for some s such that $a_{ij} = \| x^i - x^j \|^2$ for all i, j if and only if A is conditional negative definite.*

References

Armstrong, M. (1998). Basic linear geostatistics. Springer-Verlag, Berlin.

Bapat, R.B. (2000). Linear algebra and linear models. Springer-Verlag, New York.

Bapat, R.B. and Raghavan, T.E.S. (1997). Nonnegative matrices and applications. Cambridge university press, Cambridge.

Ben-Isreal, A. and Greville, T.N.E. (2003). Generalized inverse: Theory and applications. Springer-Verlag, New York.

Berger, J.O. (1985). Statistical decision theory and Bayesian analysis. Springer-Verlag, Now York.

Berger, J. and Bernardo, J.M. (1989). Estimating a product of means: Bayesian analysis with reference priors. Journal of the American statistical association, vol.84, pp.200-207.

Berke, O. (1999). Estimation and prediction in the spatial linear model. Water, air and soil pollution 110, pp.215-237.

Besag, J. and Higdon, D. (1999). Bayesian analysis of agricultural field experiments. Journal of the royal statistical society B, vol.61, pp.691-746.

Box, B.E. and Tiao, G.C. (1973). Bayesian inference in statistical analysis. Addison-Wesley, London.

Christensen, R. (1991). Linear models for multivariate, time series, and spatial data. Springer-Verlang, New York.

Christensen, R., Johnson, W. and Pearson, L.M. (1993). Covariance function diagnostics for spatial linear models. Mathematical geology, vol.25, pp.145-160.

Cressie, N.A.C. (1993). Statistics for spatial data. John Wiley, New York.

Davis, J.C. (1986). Statistics and data analysis in geology. John Wiley.

Datta, G.S. and Ghosh, M. (1996). On the invariance of noninformative priors.

The annals of statistics, vol. 24, pp.141-159.

De Waal, D.J. and Groenewald, P.C.N. (1995). Bayesian estimation of ground water levels. South African statistical journal, vol.29, pp.65-78.

Ehrbar, H.G. (2003). Class note Economical statistics. Economics department, university of Utah.

Fathy, Y.M. and Qassim, M.N.I. (2002). Bayesian estimation of spatial covariance function with two and three parameters. Al-Rafedain journal of science, vol. 13, pp.43-50.

Fuentes, M., Guttorp, P. and Challenor, P. (2003). Statistical assessment of numerical methods. International statistical review, vol.71, pp.201-221.

Gneiting, T. and Schlather, M. (2004). Stochastic models that separate fractal dimension and the Hurt effect. SIAM review, vol.46, pp.269-282.

Gnot, S. and Kleffe, J. (1983). Quadratic estimation in mixed linear models with two variance components. Journal of statistical planning and inference. vol.8, pp.267-279.

Graybill, F.A. (1983). Matrices with applications in statistics. Wadsworth Inc., California.

Handcock, M.S. and Stein, M.L. (1993). A Bayesian analysis of Kriging. Technometrics, vol.35, PP.403-410.

Hartley, H.O., Rao, J.N.K and LaMotte, L.R. (1978). A simple 'synthesis'-based method of variance component estimation. Biometrics, vol.34, pp.233-242.

Hartley, H.O. and Rao, J.N.K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. Biometrika, vol.54, pp.93-108.

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

Helmert, F. R. (1924), Die Ausgleichungsrechnung nach der Methode der Kleinsten Quadratic, 3.AUFL., Leipzig.

Henderson, C.R. (1984). Applications of linear models in animal breeding. Uni-

versity of Guelph.

Henderson, H.V. and Searle, S.R. (1981). On deriving the inverse of a sum of matrices. Siam review. Vol.23, pp.53-60.

Hjort, N.L. and Omre, H. (1994). Topics in spatial statistics. Scandinavian journal of statistics, vol.21, pp.289-357.

Jeffreys, H. (1961). Theory of probability. Clarenden Press, London.

Jennings, A. and Mckeown, J.J. (1993). Matrix computation. John Wiley, Chichester.

Johnson, R. A. and Wichern, D.W. (1988). Applied multivariate statistical analysis. Prentice Hall, New Jersey.

Johnson, V.E. and Albert, J.H. (1999). Ordinal data modeling. Springer-verlag, Now York.

Journel, A.G. and Huijbregts, C.T. (1978). Mining geostatistics, Academic press, New York.

Kincaid, D. and Cheney, W. (1991). Numerical analysis. Wadsworth Inc., California.

Kitanidis, P.K. (1983). Statistical estimation of polynomial generalized covariance functions and hydrologic applications. Water resources research, vol.19, pp.909-921.

Kitanidis, P.K. and Lane, R. W. (1985). Maximum likelihood parameter estimation of hydrologic spatial processes by the gauss-newton method. Journal of hydrology, vol.79, pp.53-71.

Kitanidis, P.K. (1985) Minimum variance unbiased quadratic estimation of covariance of regionalized variable. Mathematical geology, vol.17, pp.195-208.

Kitanidis, P.K. (1987). Parametric estimation of covariance of ragionalized variables. Water resources bulletin, vol.23, pp.557-567.

Kleffe, J. and Pincus, R. (1974). Bayes and best quadratic estimators for parameters of the covariance matrix linear model. Mathematische Operationsforschung

und Statistik, vol.5, pp.43-67.

Koch, K. (1987). Bayesian inference for variance components. Manuscripta geodaetica, vol.12,pp.309-313.

Koch, K. (1990). Bayesian inference with geodetic applications. Springer-Verlag, Berlin.

Koch, K. (1999). Parameter estimation and hypothesis testing in linear models. Springer-Verlang, Berlin.

Kubacek, L. (1985). Locally best quadratic estimators. Mathematica Slovaca, vol.35, pp.393-408.

Kuiper, H. (1986). Distribution of distances in pregeographical space. Gower company, England.

LaMotte, L. R. (1973). Quadratic estimation of variance components. Biometrics, vol.29, pp.311-333.

Lancaster, P. and Tismenetsky, M. (1985). The theory of matrices with applications. Academic press, San Diego.

Lark, R.M. and Cullis, B.R. (2004). Model-based analysis using REML for inference from systematically sampled data on soil. Europen journal of soil science, vol.55, pp.799-813.

Le, N.D. and Zidek, J.V. (1992). Interpolation with uncertain spatial covariances: A Bayesian alternative to Kriging. Journal of multvariate analysis, vol.43, pp.351-374.

Lee, P.M. (1997). Bayesian statistics: An introduction. John Wiley, New York.

Lee, T.C., Judge, G.G. and Zellner, A. (1977). Estimating the parameters of the Markov probability model from aggregate time series data. North-Holland publishing company, Amsterdam.

Lewis, D.W. (1991). Matrix theory. World scientific, Singapore.

Lee, H.K.H, Higdon, D.M., Calder, C.A. and Holloman, C.H. (2005). Efficient models for correlated data via convolutions of intrinsic processes. Statistical mod-

elling: An international journal, vol.5, pp.53-74.

Magnus, J.R. and Neudecker, H. (1999). Matrix differential calculus with applications in statistics and Econometrics. John Wiley, Ontario.

Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). Multivariate analysis. Academic Press, London.

Mardia, K.V. and Marshall, R.J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika, vol.71, pp.135-146.

Mardia, K.V. and Watkins, A.J. (1989). On multimodality of the likelihood in the spatial linear model. Biometrika, vol.76, pp.289-295.

Marshall, R.J. and Mardia, K.V. (1985). Minimum norm quadratic estimation of components of spatial covariance. Mathematical geology, vol.17, pp.517-525.

Matheron, G. (1963). Principles of geostatistics. Economic geology, vol.58, pp.1246-1266.

Mirsky, L. (1990). An introduction to linear algebra. Clarendon Press, Oxford.

Ou, Z.(1991). Aproximative bayes estimation for variance components. Manuscripta geodaetica, vol.16, pp.168-172.

Pardo-Iguzquiza, E. (1998). Maximum likelihood estimation of spatial covariance parameters. Mathematical geology, vol.30, pp.95-108.

Pardo-Iguzquiza, E. and Wowd, P. (1998). The second order stationary universial kriging model revisited. Mathematical geology, vol.30, pp.347-378.

Patterson, H.D. and Thompson, R. (1974). Maximum likelihood estimation of components of variance. Proceedings of the eighth international congress of Biomechanics, pp.197-209.

Pelletier, B., Dutilleul, P., Larocque, G. and Fyles, J. W. (2004). Fitting the linear model of coregionalization by generalized least squares. Mathematical geology, vol.36, pp.323-343.

Penrose, R. (1955). A generalized inverse for matrices. Mathematical proceedings of the Cambridge philosophical society, vol.51, pp.406-413.

Putter, H. and Young, G.A. (2001). On the effect of covariance function estimation on the accuracy of kriging predictors. Bernoulli, vol.7, pp.421-438.

Rao, C.R. (1971). Estimation of variance components-MINQUE theory. Journal of multivariate statistics, vol.1, pp.257-275.

Rao, C.R. (1972). Estimation of variance and covariance components in linear models. Journal of the American statistical association, vol.67, pp.112-115.

Rao, C.R. and Kleffe, J. (1988). Estimation of variance components and application. Elsevier science publishers, Amsterdam.

Salem, S.A.M. and Qassim, M.N.I. (1997). An estimation of components of spatial covariance. Journal of educational sciences, vol.27, pp.175-188.

Schmidt, A.M. and O'Hagan, A. (2003). Bayesian inference for non-stationary spatial covariance structure via spatial deformations. Journal of the royal statistical society B, vol.65, pp.743-758.

Schuenemeyer, J.H. and Power, H.C. (2000). Uncertainty estimation for resources assessment-an application to coal. Mathematical geology, vol.32, pp.521-541.

Searle, S.R. (1971). Linear models. John Wiley, New York.

Searle, S.R. (1982). Matrix algebra useful for statistics. John Wiley, New York.

Searle, S.R., Casella, G. and Mcculloch, C.E. (1992). Variance components. John Wiley, New York.

Seely, J. (1971). Quadratic subspaces and completeness. The annals of mathematical statistics, vol.42, pp.710-721.

Sjöberg, L. (1995). The best quadratic minimum biased non-negative definite estimator for an additive two variance component model. Manuscripta geodaetica, vol.20, pp.139-144.

Slanger, W.D. (1996). Least squares Lehmann-Scheffe estimation of variances and covariances with mixed linear models. Journal of animal sciences, vol.74, pp.2577-2585.

Steel, P. and Kammeyer-Mueller, J. (2003). Bayesian Variance Estimation: Avoiding Nonsensical Negative Estimates. Poster session presented at the annual meeting of the society of industrial organizational psychology, Orlando.

Stein, M.L. (1987). Minimum norm quadratic estimation of spatial variograms. Journal of the American statistical association, vol.82, pp.765-772.

Stein, M.L., Chi, Z., and Welty, L.J. (2004). Approximating likelihoods for large spatial data sets. Journal of the royal statistical society B, vol.66, pp.275-296.

Stuchly, H.J. (1989). Bayes unbiased estimation in a model with three variance components. Aplikace matematikly, vol.34, pp.375-386.

Townsend, E.C. and Searle, S.R. (1971). Best quadratic unbiased estimation of variance components from unbalanced data in the 1-way classification. Biometrics, vol.27, pp.643-657.

Verdooren, L. R. (1988). Least squares estimators and non-negative estimators of variance components. Communications in statistics/Theory and methods, vol.17, pp.1027-1051.

Volaufova, J. and Witkovsky, V. (1991). Least squares and minimum MSE estimators of variance components in linear models. Biometrical journal, vol.33, No.8, pp.923-936.

Wackernagel, H. (2003). Multivariate geostatistics: An introduction with applications. Springer-Verlag, Heidelberg.

Wang, S. and Chow, S. (1994). Advanced linear models. Marcel Dekker, New York.

Witkovsky, V. (1996). On variance-covariance components estimation in linear models with AR(1) disturbances. Acta mathematica Universitatis Comenianae, vol.1, pp.129-139.

Zimmermann, D.L. and Harville, D.A. (1991). A random field approach to the analysis of field-plot experiments and other spatial experiments. Biometrics, vol.47, pp.223-239.