Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 99 (2008) 2136–2153
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Regression with strongly correlated data

Christopher S. Jones[a,*], John M. Finn[a], Nicolas Hengartner[b]

[a] *T-15, Plasma Theory, Los Alamos National Laboratory, Los Alamos, NM 87545, United States*
[b] *CCS-3, Information Sciences, Los Alamos National Laboratory, Los Alamos, NM 87545, United States*

Received 15 June 2007
Available online 10 February 2008

Abstract

This paper discusses linear regression of strongly correlated data that arises, for example, in magnetohydrodynamic equilibrium reconstructions. We have proved that, generically, the covariance matrix of the estimated regression parameters for fixed sample size goes to zero as the correlations become unity. That is, in this limit the estimated parameters are known with perfect accuracy. Simple examples are shown to illustrate this effect and the nature of the exceptional cases in which the covariance of the estimate does not go to zero.

AMS subject classifications: 46N30; 62J02; 62J05; 62J10

Keywords: Regression; Least squares; Highly correlated errors; Peelle's pertinent puzzle; Infill asymptotics

1. Introduction

Magnetohydrodynamic (MHD) equilibrium reconstructions [1–4] play a vital role in the analysis of the states of plasmas in magnetic confinement devices such as tokamaks. Typically, such reconstructions are performed by least squares fitting of the nonlinear Grad–Shafranov equation to measurements of the magnetic field at spatially distinct points on the boundary of the device, complemented by measurements of the interior conditions of the plasma. Dynamical fluctuations associated with plasma turbulence are modeled as stochastic noise in the reconstructions, and these fluctuations are believed to exhibit strong spatial correlations [5], leading to strong correlations between measured signals.

* Corresponding author.
E-mail address: doc_plasmon@yahoo.com (C.S. Jones).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2008.02.008

Operationally, regression with correlated errors is understood [7,8,6]. However, the least squares equilibrium reconstruction studies of Ref. [9] exhibited unexpected properties. As a function of the degree of correlation (discussed further below), the variance of the fitted parameters was observed to have a maximum. Past this peak, the covariance matrix of the estimate converged to zero as the simulated measurements became fully correlated. The objective of the present paper is to explore this phenomenon and to show how generic it is. While our analysis is restricted to linear regression because the effect is most transparent there, our analysis and results are readily extended to nonlinear regression and were indeed first observed in a nonlinear context [9].

Specifically, consider the linear model

$$Y_i = X_i^t \beta + \eta_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where the covariates and the parameters are $X_i, \beta \in \mathbb{R}^m$ with $m \leq n$. The disturbances η_i have mean zero, variances $\mathbb{E}[\eta_i^2] = \sigma_i^2$ and covariances $\mathbb{E}[\eta_i \eta_j] = \Sigma_{ij} = \sigma_i \sigma_j \varrho_{ij}$. For our analysis and discussion, it is convenient to rewrite (1) in vector form

$$\mathsf{Y} = \mathsf{X}\beta + \eta, \tag{2}$$

where the vectors $\mathsf{Y}, \eta \in \mathbb{R}^n$, X is an $n \times m$ design matrix, $\mathbb{E}[\eta] = 0$ and $\mathbb{E}[\eta\eta^t] = \Sigma = \mathsf{SRS}$, where the deviations are $\mathsf{S} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$ and R is the matrix of correlation coefficients ϱ_{ij}. To simplify our exposition, we shall begin by assuming that the design matrix X is of rank m and, unless otherwise stated, further assume that the covariance matrix Σ is of full rank n.

It is well known that the best linear unbiased estimator (BLUE) $\hat{\beta}$ for β is the minimizer (with respect to β) of the quadratic

$$\chi^2(\beta) = (\mathsf{Y} - \mathsf{X}\beta)^t \Sigma^{-1} (\mathsf{Y} - \mathsf{X}\beta). \tag{3}$$

Under the stated assumptions, the BLUE $\hat{\beta}$ is then given by the unique solution of the normal equations

$$(\mathsf{X}^t \Sigma^{-1} \mathsf{X}) \hat{\beta} = \mathsf{X}^t \Sigma^{-1} \mathsf{Y}, \tag{4}$$

and the covariance matrix of the estimate $\hat{\beta}$ is given by

$$\mathsf{V} = (\mathsf{X}^t \Sigma^{-1} \mathsf{X})^{-1}. \tag{5}$$

Our main result deals with the approach to full correlation, defined as the limit

$$\min_{i,j} \frac{|\Sigma_{ij}|}{\sigma_i \sigma_j} = \min_{i,j} |\varrho_{ij}| \longrightarrow 1. \tag{6}$$

In this limit $\varrho_{ij} \to e_i e_j$ with $e_j = \pm 1$. This main result is that, if the vector of signed standard deviations $(e_1\sigma_1, \ldots, e_n\sigma_n)^t$ of the disturbances does not lie in the column space (the range space) of the design matrix X, then the sum of the variances, trace(V), of the estimated regression coefficients converges to zero as the noise becomes fully correlated.

Various authors have noted the unusual and unexpected effects of strong correlations in simple models. In the context of the optimal fit to a line, Canner noted in 1969 that strong positive correlations among measurements with unequal variances may lead to a regression line that lies entirely above or below the measurements, coinciding with a weighted average including negative weights [10]. More recently, several authors in the field of nuclear data

analysis have also remarked on this ostensible quandary, referring to it as Peelle's pertinent puzzle [11–13]. Sivia, again in the context of fitting data to a line, noted that strong correlations are beneficial with respect to the determination of the slope and detrimental with respect to the intercept [14]. In the context of forecast pooling, Winkler also remarked upon the appearance of negative weights [15]. Furthermore, Clemen and Winkler noted that in combining forecasts from strongly correlated sources of *equal* variance, the number of independent sources is effectively reduced [16]. Morrison and Schmittlein, applying a geometrical interpretation to the result of Clemen and Winkler, recognized that full correlation for sources of *unequal* variance can lead to vanishing variance of the estimate, but dismissed the effect as an aberration [17].

In the current work, we provide a thorough explication of the effect of strong correlations on multivariate regression. This paper is organized as follows. Section 2 sets the stage by analyzing several very simple models exhibiting our main result. With the aid of these models, the various observations of previous authors, such as negative weights and vanishing variance of the estimate, are connected and clarified. Our main result, on multivariate regression and how the relationship between the covariance and design matrices affects the variance of the estimates, is presented in Section 3. Section 4 deals with the issue of increasing the number of measurements in physical problems such as MHD equilibrium reconstruction, by adding measurements at necessarily more closely packed spatial positions. In Section 5 we deal with a generalization of the main result in which some, but not all, measurements become fully correlated. A summary and conclusions are presented in Section 6.

2. A simple example

This section presents simple examples which display some of the surprising effects associated with linear regression in the presence of strong correlations.

2.1. Estimation of a constant

We wish to estimate μ from a pair of observations

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \mu + \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \mathsf{X}\mu + \eta, \tag{7}$$

where the disturbances $\eta = (\eta_1, \eta_2)$ have mean zero and covariance and precision matrices given by

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \quad \text{and} \quad \Sigma^{-1} = \frac{1}{1-\varrho^2} \begin{pmatrix} \tau_1^2 & -\varrho\tau_1\tau_2 \\ -\varrho\tau_1\tau_2 & \tau_2^2 \end{pmatrix},$$

with $\tau_1 = 1/\sigma_1$, $\tau_2 = 1/\sigma_2$. Direct calculation shows that the weighted least squares estimate that satisfies the normal equation (4) is

$$\begin{aligned} \hat{\mu} &= \frac{\left(\tau_1^2 - \varrho\tau_1\tau_2\right) y_1 + \left(\tau_2^2 - \varrho\tau_1\tau_2\right) y_2}{\tau_1^2 - 2\varrho\tau_1\tau_2 + \tau_2^2} \\ &= \frac{\left(\tau_1^2 - \varrho\tau_1\tau_2\right) y_1 + \left(\tau_2^2 - \varrho\tau_1\tau_2\right) y_2}{(1-\varrho)\left(\tau_1^2 + \tau_2^2\right) + \varrho\left(\tau_1 - \tau_2\right)^2} \\ &= w_1(\sigma_1, \sigma_2, \varrho) y_1 + w_2(\sigma_1, \sigma_2, \varrho) y_2, \end{aligned} \tag{8}$$

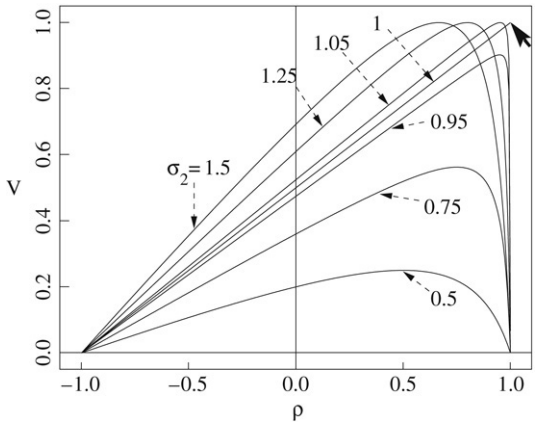

Fig. 1. Variance of the estimate V as a function of the correlation coefficient ϱ for $\sigma_1 = 1$ and seven values of σ_2 (0.5, 0.75, 0.95, 1, 1.05, 1.25, 1.5). For all cases, $V(\hat{\mu}) \to 0$ as $\varrho \to -1$. For $\sigma_1 \neq \sigma_2$, $V(\hat{\mu}) \to 0$ as $\varrho \to 1$. For $\sigma_2 = \sigma_1 = 1$, $1V(\hat{\mu}) = \sigma_1^2 = 1$ at $\varrho = 1$ (large arrow).

where we have written the estimate in terms of weights w_1 and w_2 on the last line. The variance of this estimate, computed from Eq. (5), is

$$V(\hat{\mu}) = \frac{1-\varrho^2}{\tau_1^2 - 2\varrho\tau_1\tau_2 + \tau_2^2} = \frac{1-\varrho^2}{(1-\varrho)\left(\tau_1^2+\tau_2^2\right) + \varrho\,(\tau_1-\tau_2)^2}. \tag{9}$$

The behavior of Eq. (9) as a function of ϱ for $\sigma_1 = 1$ and various σ_2 is shown in Fig. 1, showing clearly the limits $\varrho \to \pm 1$. In this simple example (and as displayed in Fig. 1), the variance of the estimate goes to zero as $\varrho \to \pm 1$, with the exception of $\varrho \to 1$ with $\sigma_1 = \sigma_2$. For this exceptional case, the eigenvector of Σ with the largest eigenvalue approaches (σ_1, σ_1), which is in the range space of the design matrix X in Eq. (7).

Regarding the exceptional case of equal variance, $\sigma_1 = \sigma_2$, the estimate for μ for all $\varrho \in [-1, 1]$ is the sample mean

$$\hat{\mu} = \frac{y_1 + y_2}{2}, \quad \text{which has variance } V(\hat{\mu}) = \frac{1+\varrho}{2}\sigma^2. \tag{10}$$

The variance vanishes for $\varrho \to -1$ while for $\varrho \to 1$, it is

$$\lim_{\varrho \to 1} V(\hat{\mu}) = \sigma^2. \tag{11}$$

The interpretation of Eqs. (10), (11) is the following. For a single measurement $y_1 = \mu + \eta_1$, the variance of $\hat{\mu} = y_1$ equals the variance of y_1, namely σ^2. For $\varrho \to 1$ and $\sigma_1 = \sigma_2$, the second measurement y_2 is equal to y_1 and provides no new information, leaving $V = \sigma^2$. This is the familiar interpretation of positive correlations, which are often assumed to imply redundancy in measurement (e.g. the Clemen and Winkler result [16]). For $\rho \to -1$, $y_1 = \mu + \eta_1$ and $y_2 = \mu - \eta_1$, implying $\mu = (y_1 + y_2)/2$ with zero error. We shall see below, however, that in the generic case ($\sigma_1 \neq \sigma_2$) positive correlations may also provide leverage with which to determine the estimator more accurately.

In the generic case of unequal variance, $\sigma_1 \neq \sigma_2$, the possibility arises of *negative weighting*, in which one of the two weights, w_1 or w_2, becomes negative. In particular, from Eq. (8) and assuming without loss of generality $\sigma_1 > \sigma_2$, y_1 is weighted negatively ($w_1 < 0$) if $\varrho > \sigma_2/\sigma_1$. Also, from Eq. (9) we conclude that $V(\hat{\mu})$ has a maximum with respect to ϱ at $\varrho = \sigma_2/\sigma_1$. At the maximum variance, note that the estimate (8) becomes $\hat{\mu} = y_2$ and Eq. (9) gives $V(\hat{\mu}) = \sigma_2^2$, i.e. $\hat{\mu}$ becomes independent of the measurement y_1 (recall $\sigma_1 > \sigma_2$). For $\varrho > \sigma_2/\sigma_1$, $V(\hat{\mu})$ decreases with respect to ϱ. The possibility of negative weighting, leading to an estimate outside the range of the measurements, has been noted with surprise in Refs. [10–13,15], as discussed in the introduction. These investigations did not, however, remark on our observation that for fixed $\sigma_1 \neq \sigma_2$, the appearance of negative weighting (as ϱ is increased) coincides with the decrease of the variance $V(\hat{\mu})$. Sivia noted the possibility of decreasing variance, but simply proffered a caveat regarding the careless interpretation of correlations as mere loss of information [14]. Morrison and Schmittlein also observed the possibility of vanishing variance [17], but dismissed the result as a poor choice for the covariance matrix of forecasts, leading to "very misleading and overly optimistic results." In situations such as MHD equilibrium reconstruction, large amounts of data allow for the reliable determination of the covariance matrix, and physical considerations suggest that those correlations may be strong. Thus, the limiting behavior of the variance of the estimate as measurements become fully correlated is of significant interest.

In the limit of full correlation ($\varrho \to 1$), the estimator for μ in the generic case of $\sigma_1 \neq \sigma_2$ is

$$\hat{\mu} = (\tau_1 y_1 - \tau_2 y_2)/(\tau_1 - \tau_2), \tag{12}$$

with negative weighting on y_1 if $\tau_1 < \tau_2$ ($\sigma_1 > \sigma_2$). This estimate has the variance (in the same limit)

$$V(\hat{\mu}) \to \frac{2(1-\varrho)}{(\tau_1 - \tau_2)^2} \to 0. \tag{13}$$

To interpret these results, consider the geometric description displayed in Fig. 2(a). For normally distributed noise, the measurements are distributed according to a probability density function proportional to $\exp(-\chi^2/2)$. The level sets of this function (contours of χ^2) are ellipses which circumscribe regions within which the measurements may be found with a given probability. For ϱ close to unity, the measurements are expected to be found within a thin ellipse whose major axis has slope near σ_2/σ_1. For $\varrho \to 1$, the two eigenvalues of the covariance matrix Σ are $\lambda_1 = \sigma_1^2 + \sigma_2^2$ (trace) and $\lambda_2 = 0$, and the ellipses become infinitely thin, i.e. line segments. In this limit the estimate must be on the intersection of the line $y_2 - \mu = (\sigma_2/\sigma_1)(y_1 - \mu)$ and the axis $y_1 = y_2$, giving $\hat{\mu} = \mu$ with zero uncertainty.

An alternative interpretation is shown in Fig. 2(b). The noise contributions η_1 and η_2 will necessarily have the same sign if $\varrho \to 1$. If, for example, they are both positive and $\sigma_1 \neq \sigma_2$, then y_1 and y_2 will both be above μ (as in Canner's example [10]), and an unbiased estimate will be possible only with negative weighting as in Eq. (12). To be more specific, for $\varrho \to 1$ but $\sigma_1 > \sigma_2$, we will have $\eta_2 = \eta_1 \sigma_2/\sigma_1$ or

$$y_1 = \mu + \sigma_1 \alpha, \quad y_2 = \mu + \sigma_2 \alpha, \tag{14}$$

where α is a single random variable with zero mean and unit variance. The BLUE (12) chooses the correct (negative) weighting to give the exact result $\hat{\mu}$ using a *single* realization α of the noise.

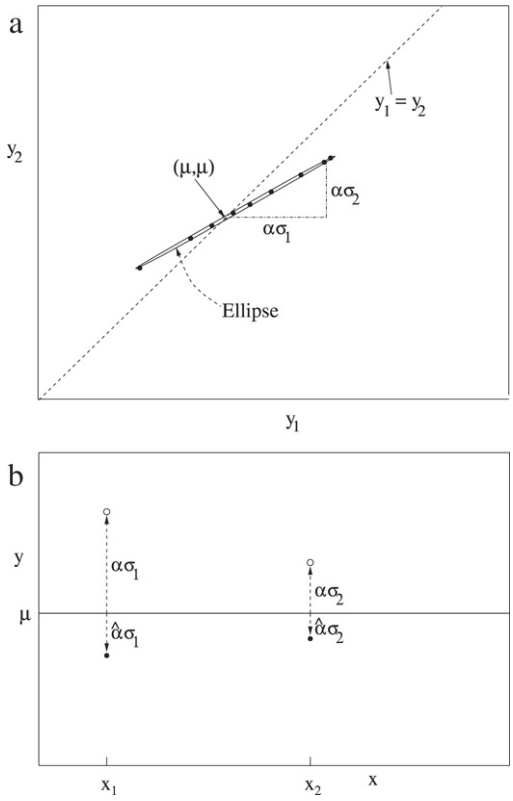

Fig. 2. Estimating a constant in the limit of large correlations $\varrho \to 1$. For $\sigma_1 > \sigma_2$ and $\varrho \to 1$, when the ellipse at constant χ^2 collapses (in the top figure), each measurement point (y_1, y_2) has $y_1 = \mu + \alpha\sigma_1$, $y_2 = \mu + \alpha\sigma_2$ for some α. Since the estimate must also be along the line $y_1 = y_2$, the estimate for $\varrho \to 1$ gives $\hat{\mu} = \mu$ with zero uncertainty. In (b), two noisy measurements of μ are made (a) at positions $x = x_1, x = x_2$ (open circles), corresponding to one value of α. The estimate, which has zero variance, has $(y_1 - \mu)/\sigma_1 = (y_2 - \mu)/\sigma_2$, exactly determining $\hat{\mu}$ if $\sigma_1 \neq \sigma_2$. A second set of correlated measurements (solid circles), with a second value $\widehat{\alpha}$, leads to the same estimate.

The noise term can be eliminated from Eq. (14), giving

$$\frac{y_1}{\sigma_1} - \frac{y_2}{\sigma_2} = \left(\frac{1}{\sigma_1} - \frac{1}{\sigma_2}\right)\mu, \tag{15}$$

in agreement with the estimate in Eq. (12). The variance of this estimate is zero because the noise has been eliminated. A different realization of the noise (different α) yields the same result. If, on the other hand, $\sigma_1 = \sigma_2$, then the measurements y_1 and y_2 are identical and the procedure leading to Eq. (15) cannot be followed, leaving μ subjected to the noise.

2.2. *Fitting a line*

The previous discussion may be easily reproduced for the case of the determination of a slope rather than a common mean from a pair of observations y_1 and y_2 located at distinct points x_1 and x_2, where the intercept is known to be the origin. In this case, the unit design matrix of Eq.

(7) is replaced by $\mathsf{X} = (x_1, x_2)^t$. For fully correlated noise ($\varrho \to 1$), the noise may again be eliminated as it was in Eq. (15), yielding

$$\frac{y_1}{\sigma_1} - \frac{y_2}{\sigma_2} = \left(\frac{x_1}{\sigma_1} - \frac{x_2}{\sigma_2}\right)\omega, \tag{16}$$

where ω is the slope. Note that in this case, the slope may be exactly determined from measurements of equal variance. In fact, μ is exactly determined for $\varrho \to 1$ as long as $x_1/\sigma_1 - x_2/\sigma_2 \neq 0$. When $x_1/\sigma_1 = x_2/\sigma_2$, α cannot be eliminated and ω remains subject to noise.

Now consider the fit of both the slope and intercept of a line from three measurements in the limit of full correlation. This example will display some of the more general aspects of the effects of full correlation addressed in detail in the following section. In particular, consider the determination of estimates for the intercept μ and slope ω from three observations y_1, y_2 and y_3 at distinct points x_1, x_2 and x_3:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \mathsf{X}\begin{pmatrix} \mu \\ \omega \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix}, \tag{17}$$

where the design matrix is

$$\mathsf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{pmatrix}. \tag{18}$$

In the limit in which the three disturbances are fully (positively) correlated, the disturbances are $\eta_1 = \sigma_1\alpha$, $\eta_2 = \sigma_2\alpha$, $\eta_3 = \sigma_3\alpha$, where as before, α is a single random variable with zero mean and unit variance. We may eliminate α, yielding [cf. Eqs. (15) and (16)]

$$\begin{pmatrix} y_1/\sigma_1 - y_2/\sigma_2 \\ y_1/\sigma_1 - y_3/\sigma_3 \end{pmatrix} = \begin{pmatrix} 1/\sigma_1 - 1/\sigma_2 & x_1/\sigma_1 - x_2/\sigma_2 \\ 1/\sigma_1 - 1/\sigma_3 & x_1/\sigma_1 - x_3/\sigma_3 \end{pmatrix}\begin{pmatrix} \mu \\ \omega \end{pmatrix}. \tag{19}$$

For the case in which all variances are equal, $\sigma_1 = \sigma_2 = \sigma_3$, the intercept μ disappears from these equations, and the slope is precisely determined while the intercept is subject to noise. This was the observation of Sivia [14]. Alternatively, if the variances are related to the measurement locations via $x_1/\sigma_1 = x_2/\sigma_2 = x_3/\sigma_3$, then the slope is not present in Eq. (19), and only the intercept is precisely determined. In each of these cases, the vector $(\sigma_1, \sigma_2, \sigma_3)^t$ is proportional to a column of the design matrix (18). Indeed, if this vector, the eigenvector associated with the largest eigenvalue of the fully correlated covariance matrix, is proportional to any linear combination of the columns of X, then there will exist a linear combination of μ and ω which is precisely determined and another which is subject to noise. Thus the covariance matrix of the estimate (μ, ω) has rank one. If the vector $(\sigma_1, \sigma_2, \sigma_3)^t$ is not a linear combination of the columns of X, as is generically the case, then both the intercept μ and slope ω are determined exactly. Indeed, this condition is exactly the condition that the determinant of the matrix in Eq. (19) is nonzero. These phenomena and their connection to the emergence of a *noise-free subspace* are generalized in the following section.

3. Regression in the limit of full correlation

This section generalizes the examples of Section 2 and shows under what conditions the variance of the weighted least squares regression estimator for the linear model (2) converges to zero in the limit of full correlation. For any covariance matrix, we define the parameter

$$\kappa = \max_{i,j}\left(1 - \frac{|\Sigma_{ij}|}{\sigma_i \sigma_j}\right) = \max_{i,j}\left(1 - |\varrho_{ij}|\right).$$

Any covariance matrix can be written as $\Sigma = \mathsf{SRS}$, where $\mathsf{S} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$ and R is the matrix of correlation coefficients ϱ_{ij}. In the limit of $\kappa = 0$, the entries of R are all ± 1 and satisfy $\varrho_{ik} = \varrho_{ij}\varrho_{jk} \forall i, j, k$, and therefore the correlation matrix is the rank one matrix $\mathsf{R} = \boldsymbol{e}\mathrm{e}^t$ or $\varrho_{ij} = e_i e_j$, where $e_j = \pm 1$. A simple example of such a class of covariance matrices is the autocorrelation model. In this model, the disturbances η consist of a zero mean random vector with covariance matrix $\Sigma = \mathbb{E}[\eta\eta^t] = \mathsf{SRS}$, where the correlation matrix is

$$\mathsf{R} = \begin{pmatrix} 1 & \varrho & \ldots & \varrho^{n-1} \\ \varrho & 1 & \ldots & \varrho^{n-2} \\ & & \ddots & \\ \ldots & \ldots & & \ldots \\ \varrho^{n-1} & \varrho^{n-2} & \ldots & 1 \end{pmatrix}, \tag{20}$$

or $\varrho_{ij} = \varrho^{|i-j|}$. For this model, we find $\kappa \equiv 1 - |\varrho|^{n-1}$. We are interested in the behavior of the variance $(\mathsf{X}^t\Sigma^{-1}\mathsf{X})^{-1}$ of the BLUE $\hat{\beta}_\kappa = (\mathsf{X}^t\Sigma^{-1}\mathsf{X})^{-1}\mathsf{X}^t\Sigma^{-1}\mathsf{Y}$ as $\kappa \to 0$. In this section, we emphasize the dependence of the estimated regression parameter on the parameter κ by subscripting the estimate $\hat{\beta}_\kappa$. Note that in the limit $\kappa = 0$, the correlation matrix of the autocorrelation model (20) is $\mathsf{R} = \boldsymbol{e}\mathrm{e}^t$, where $\boldsymbol{e}$ has entries $e_j = 1$ for $\varrho \to 1$, or $e_j = (-1)^j$ for $\varrho \to -1$.

We begin by presenting a heuristic that shows how the vanishing of the estimator variance is related to the emergence of a noise-free subspace in the limit of large correlations. A more rigorous proof is then presented at the end of the section.

3.1. Heuristic treatment

Fix the sample size to $n > m$, and let $\mathsf{Q} = [v_1, v_2, \ldots, v_n]$ denote the matrix of normalized column eigenvectors of the covariance matrix Σ, corresponding to the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$. That is, we are assuming that the covariance matrix Σ remains of full rank for $\kappa > 0$ (we discuss more general covariance matrices in Section 5). As $\kappa \longrightarrow 0$, we have $\Sigma_{ij} \to \sigma_i e_i \sigma_j e_j$, so that

$$\lambda_1 = \sum_{j=1}^{n} \sigma_j^2 + o(1), \quad \lambda_k = o(1), \quad k \geq 2, \quad \text{and} \quad v_{1j} = e_j \frac{\sigma_j}{\sqrt{\lambda_1}} + o(1). \tag{21}$$

Consider the transformation

$$\mathsf{Z} = \mathsf{Q}^t\mathsf{Y} = (\mathsf{Q}^t\mathsf{X})\beta + \mathsf{Q}^t\eta = (\mathsf{Q}^t\mathsf{X})\beta + \Lambda^{1/2}\xi, \tag{22}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and ξ is a vector of uncorrelated disturbances with mean zero and unit variance.

In light of (21), $\lambda_j = 0$, $j \geq 2$ in the limit of $\kappa = 0$, so that the transformed variables Z_j ($j \geq 2$) are noise-free. That is, the subset of Eq. (22) with $j \geq 2$ is not subjected to noise in the limit $\kappa = 0$:

$$\tilde{Z} = \tilde{\mathsf{X}}\beta. \tag{23}$$

Here, we define the vector of noise-free data $\tilde{Z} = (v_2^t\mathsf{Y}, \ldots, v_n^t\mathsf{Y})$ and the reduced design matrix $\tilde{\mathsf{X}} = [v_2, \ldots, v_n]^t\,\mathsf{X}$. Recall that $n > m$, so there exists at least one solution to Eq. (23). Further, if v_1 does not lie in the column space (the range space) of X, denoted $\mathcal{R}(\mathsf{X})$, then the system of Eq. (23) has a unique solution $\beta^\star$. Indeed, assume for contradiction that another solution β' exists, and denote the difference $\beta_0 = \beta^\star - \beta' \neq 0$. Then $v_j^t\mathsf{X}\beta_0 = 0$ for $2 \leq j \leq n$, and therefore $\mathsf{X}\beta_0$ is in the one-dimensional subspace spanned by v_1, i.e. v_1 is in the range of X, contrary to our original assumption.

We call the space spanned by the eigenvectors v_j for $j \geq 2$ the *noise-free subspace*. Since the weighted least squares estimator $\hat{\beta}_\kappa$ has the smallest variance among all linear estimators for β, we conclude that, assuming that the BLUE exists in the limit,

$$\operatorname{trace}(\mathsf{V}(\hat{\beta}_{\kappa=0})) \leq \operatorname{trace}(\mathsf{V}(\beta^\star)) = 0,$$

suggesting that $\hat{\beta}_{\kappa=0} = \beta^\star$.

Conversely, suppose that v_1 lies in $\mathcal{R}(\mathsf{X})$. We wish to show that under this condition the estimate $\hat{\beta}$ cannot be determined exactly, generalizing the observations at the end of Section 2 regarding the fitting of a line. Since $v_1 \in \mathcal{R}(\mathsf{X})$, there exists a vector of parameters, say $w_1 = (W_{11}, \ldots, W_{m1})^t$, such that $\mathsf{X}w_1 = v_1$. Let us choose this as the first column of an $m \times m$ nonsingular matrix W. We choose the remaining $m - 1$ columns $w_2, \ldots, w_m$ such that $\operatorname{span}\{\mathsf{X}w_2, \ldots, \mathsf{X}w_m\} = \mathcal{R}(\mathsf{X}) \cap \operatorname{span}\{v_2, \ldots, v_n\}$, ensuring that W is indeed of rank m and nonsingular. In this way we identify the linear combination of elements of β which is subject to noise. We obtain

$$\mathsf{Y} = \mathsf{X}\mathsf{W}\mathsf{W}^{-1}\beta + \eta = (\mathsf{X}\mathsf{W})\gamma + \eta,$$

with $\gamma = \mathsf{W}^{-1}\beta$. Then (21) and (22) imply that in the limit of $\kappa = 0$,

$$Z_1 = v_1^t\mathsf{Y} = \gamma_1 + \left(\sum_{j=1}^{n}\sigma_j^2\right)^{1/2}\xi_1, \tag{24}$$

$$Z_j = v_j^t\mathsf{Y} = v_j^t(\mathsf{X}\mathsf{W})\gamma, \quad j = 2, \ldots, n. \tag{25}$$

Because the first column of $(\mathsf{X}\mathsf{W})$ is v_1, Eq. (25) are independent of γ_1 and determine exactly the parameters $\gamma_2, \ldots, \gamma_m$. The estimator of γ_1, a particular linear combination of the elements of β, is subject to the total noise of the system (i.e. the variance of the estimate $\hat{\gamma}_1$ is the sum of the variances of the original measurements). Hence in the limit as $\kappa \longrightarrow 0$, the total variance for the BLUE for γ is

$$\operatorname{trace}(\mathsf{V}(\hat{\gamma}_\kappa)) = \sum_{j=1}^{n}\lambda_j = \sum_{j=1}^{n}\sigma_j^2 > 0,$$

which implies that $\operatorname{trace}(\mathsf{V}(\hat{\beta}_\kappa)) > 0$. Incidentally, in this case the rank of the covariance matrix of the estimate is necessarily unity.

In conclusion, we note that the preceding discussion implies that the condition $v_1 \notin \mathcal{R}(\mathsf{X})$ is equivalent to the condition that the reduced design matrix $\tilde{\mathsf{X}}$ be of full rank m. Thus, the full rank of the reduced design matrix $\tilde{\mathsf{X}}$ allows for the determination of the estimate $\hat{\beta}$ with vanishing variance in the limit $\kappa = 0$.

3.2. Rigorous treatment

The heuristic of Section 3.1 identifies, in the limit of $\kappa = 0$, the subspace orthogonal to the vector of signed deviations $v_1 = (e_1\sigma_1, \ldots, e_n\sigma_n)^t$ as a noise-free subspace that enables perfect estimation of β if v_1 is not in the range $\mathcal{R}(\mathsf{X})$. If $v_1 \in \mathcal{R}(\mathsf{X})$, a single linear combination of the parameters $\beta_1, \ldots, \beta_m$ is subject to the noise, and the covariance matrix of the estimate has rank one. This argument however does not prove that the variance of the BLUE for β converges to zero with $\kappa \longrightarrow 0$, because our argument lets κ converge to zero first before estimating β and showing that it resulted in an estimate that had zero variance. Theorem 1 below gives a rigorous proof of our claim.

Theorem 1. *For fixed sample size $n > m$, suppose that the design matrix X is of rank m and that the covariance matrix Σ is of rank n for $\kappa > 0$. If the eigenvector v_1 associated with the largest eigenvalue of the limiting covariance matrix Σ (when κ goes to zero) does not lie in the column space of the design matrix X, then the total variance of the least squares estimate approaches zero in the limit as $\kappa \longrightarrow 0$.*

Proof. Denote by $\mathsf{Q} = [v_1, \ldots, v_n]$ the matrix of column eigenvectors associated with the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$ of the covariance matrix Σ. As before, the continuity of the eigenvectors and eigenvalues as a function of Σ implies that as $\kappa \longrightarrow 0$, the eigenvalues and first eigenvector behave as in (21). We again consider the transformation defined in Eq. (22). Suppose that v_1 is not in the range of the design matrix X. Then

$$Z_1 = v_1^t \mathsf{X}\beta + \sqrt{\lambda_1}\xi_1 = \sqrt{\left(\sigma_1^2 + \cdots + \sigma_n^2\right)}\xi_1 + o(1),$$

so that Z_1 is a linear combination of the regression parameters β_i for which the signal becomes overwhelmed by the noise in the limit $\kappa \to 0$. Let us then consider the estimate $\tilde{\beta}_\kappa$ for $\kappa > 0$ and disregarding Z_1 [cf. Eq. (23)]

$$\tilde{\beta}_\kappa = (\tilde{\mathsf{X}}^t \tilde{\Lambda}^{-1} \tilde{\mathsf{X}})^{-1} \tilde{\mathsf{X}}^t \tilde{\Lambda}^{-1} \tilde{Z},$$

where the vector $\tilde{Z}$ and the reduced design matrix $\tilde{\mathsf{X}}$ are defined below Eq. (23), and $\tilde{\Lambda} = \operatorname{diag}(\lambda_2, \ldots, \lambda_n)$. Recall that $\tilde{\mathsf{X}}$ is of full rank m since $v_1 \notin \mathcal{R}(\mathsf{X})$. The covariance matrix of $\tilde{\beta}_\kappa$ is

$$\mathsf{V}(\tilde{\beta}_\kappa) = (\tilde{\mathsf{X}}^t \tilde{\Lambda}^{-1} \tilde{\mathsf{X}})^{-1}.$$

The estimate $\tilde{\beta}_\kappa$ is not the BLUE $\hat{\beta}_\kappa$ for $\kappa > 0$ since the equation for Z_1 has been discarded. However, the largest eigenvalue for the covariance matrix of the BLUE $\hat{\beta}_\kappa$ is bounded from above by the largest eigenvalue of the covariance matrix of the estimate $\tilde{\beta}_\kappa$:

$$\sup_{\|a\|=1} a^t \mathsf{V}(\hat{\beta}_\kappa) a \leq \sup_{\|a\|=1} a^t \mathsf{V}(\tilde{\beta}_\kappa) a.$$

To see this, consider the eigenvector a_0 for the largest eigenvalue of $\mathsf{V}(\hat{\beta}_\kappa)$, which satisfies $a_0^t \mathsf{V}(\hat{\beta}_\kappa) a_0 = \sup_{\|a\|=1} a^t \mathsf{V}(\hat{\beta}_\kappa) a$. According to the Gauss–Markov theorem [18], $a_0^t \mathsf{V}(\hat{\beta}_\kappa) a_0 \leq$

$a_0^t \mathsf{V}(\tilde{\beta}_\kappa) a_0$. Since $a_0^t \mathsf{V}(\tilde{\beta}_\kappa) a_0 \leq \sup_{\|a\|=1} a^t \mathsf{V}(\tilde{\beta}_\kappa) a$, the previous relation follows. Thus,

$$\begin{aligned}
\sup_{\|a\|=1} a^t \mathsf{V}(\hat{\beta}_\kappa) a &\leq \sup_{\|a\|=1} a^t (\tilde{\mathsf{X}}^t \tilde{\Lambda}^{-1} \tilde{\mathsf{X}})^{-1} a \\
&\leq \left[\inf_{\|a\|=1} a^t \tilde{\mathsf{X}}^t \tilde{\Lambda}^{-1} \tilde{\mathsf{X}} a \right]^{-1} \\
&\leq \left[\inf_{\forall a} \frac{\left(\tilde{\mathsf{X}} a\right)^t \tilde{\Lambda}^{-1} \left(\tilde{\mathsf{X}} a\right)}{\|\tilde{\mathsf{X}} a\|^2} \frac{\|\tilde{\mathsf{X}} a\|^2}{\|a\|^2} \right]^{-1}.
\end{aligned}$$

Noting that, for strictly positive quantities, the infimum of a product is greater than or equal to the product of infima, we have

$$\sup_{\|a\|=1} a^t \mathsf{V}(\hat{\beta}_\kappa) a \leq \lambda_2 / \inf_{\|a\|=1} \|\tilde{\mathsf{X}} a\|^2.$$

In light of (21) and the full rank of $\tilde{\mathsf{X}}$, the latter converges to zero with $\kappa \longrightarrow 0$. □

Remark 1. Our heuristic argument can be used to show the converse of the theorem, namely, if the column space of the design matrix X contains the eigenvector associated with the largest eigenvalue of the limiting covariance matrix, then the limiting variance of the BLUE $\hat{\beta}_\kappa$ is strictly positive. Indeed, note that the error distribution of η_κ converges in distribution to the limiting distribution of η_0. In light of Fatou's lemma (see Ref. [19]), we have for all vectors a,

$$\liminf_{\kappa \longrightarrow 0} a^t \mathsf{V}(\hat{\beta}_\kappa) a \geq a^t \mathsf{V}(\hat{\beta}_0) a. \tag{26}$$

We may then use the heuristic to show that the right side of (26) strictly positive.

Remark 2. In the discussion of the heuristic and the preceding proof of Theorem 1, it was assumed that the design matrix X is of full rank m. If instead, the rank of X is $s < m$, then there exist $m - s$ linear combinations of the parameters that cannot be determined by the model, regardless of the noise. The remaining s linear combinations are affected by the presence of fully correlated noise ($\kappa \to 0$) as in the previous discussion, but with m replaced by s.

4. Implications for sampling locations in experiments

The autocorrelation error model (20) of Section 3 provides a useful and simple framework in which to analyze parameter estimation from a large number of closely spaced measurements. In the context of magnetically confined plasmas, difficulty of access to the plasma typically implies that an increase in the number of measurements will be associated with a decrease in the spacing between measurements. As mentioned in the introduction, the noise in these devices arises in part from plasma turbulence, which may exhibit long-range characteristics. One may then be led to believe, in light of Section 3, that the increased correlations due to closer spacing could improve parameter estimation. We show below that this is not the case.

This section provides a detailed analysis of the variance of the BLUE for a single regression coefficient in the following setting: Suppose we observe the magnetic field at $n + 1$ locations in the interval [0, 1]. For the purpose of our discussion, we take these points to be equidistant, that is $x_{n,i} = i/n$, $i = 0, \ldots, n$, (spacing $\Delta x = 1/n$). At each location $x_{n,i}$, we observe

$$Y_{n,i} = \mu + \eta_{n,i}, \quad i = 0, 1, \ldots, n, \tag{27}$$

where the disturbances have mean zero, variance $\mathbb{E}[\eta_{n,i}^2] = \sigma^2(x_{n,i}) = \tau^{-2}(x_{n,i})$ and (strictly positive) correlation

$$\begin{aligned}\varrho_{ij} &= \tau(x_{n,i})\tau(x_{n,j})\mathbb{E}[\eta_{n,i}\eta_{n,j}] = \exp\left(-|x_{n,i} - x_{n,j}|/\delta\right) \\ &= \exp\left(-\frac{|i-j|}{\delta n}\right) \equiv \varrho^{|i-j|}.\end{aligned} \tag{28}$$

This is the autocorrelation model, Eq. (20) of Section 3, with $\varrho \equiv \exp(-(\delta n)^{-1})$. The parameter δ is interpreted as the correlation length, with $\delta \longrightarrow 0$ and $\delta \longrightarrow \infty$ corresponding to the uncorrelated and fully correlated error models, respectively. Note that $\delta n = \delta/\Delta x$ is the ratio of the correlation length to the spacing between measurements. We shall further assume that τ, a measure of the signal-to-noise ratio, is a smooth function of the sampling location.

In Section 3, we studied the limit of the variance $\mathcal{V}(n, \delta) \equiv V(\hat{\mu})$ as $\varrho \to 1$ (i.e. as the correlation length δ approached infinity). In this section, we fix δ and study the behavior of the estimated mean $\hat{\mu}$ and its variance as the sample size n goes to infinity. In this setting, as we increase the number of sampling locations within the unit interval, the correlation between neighboring measurements increases. This is the framework referred to as infill asymptotics [20,21], and the results of this analysis can provide guidelines for the usefulness of acquiring additional data by increasing the number of measurements done for MHD equilibrium reconstructions, or other estimation problems where acquiring more data necessitates packing the measurements more closely in space or time.

Theorem 2. *The inverse variance of the estimated mean for the autocorrelated model is, for* $\delta n \gg 1$ *and* $n \gg 1$,

$$\begin{aligned}\mathcal{V}^{-1}(n, \delta) = &\left[\frac{\delta}{2}\int_0^1 \tau'(s)^2 \mathrm{d}s + \frac{1}{2\delta}\int_0^1 \tau(s)^2 \mathrm{d}s\right]\left(1 + O((\delta n)^{-2})\right) \\ &+ \frac{1}{2}(\tau(0)^2 + \tau(1)^2) + \delta O(n^{-2}) + \delta^{-1} O(n^{-2}).\end{aligned} \tag{29}$$

Remark. This expression for the variance of the estimate provides insight into the estimated mean. In the limit of very large correlations ($\delta \to \infty$). The variance of the estimated mean approaches zero if $\int \tau'(s)^2 \mathrm{d}s \neq 0$, i.e. if the signal-to-noise ratio varies over the measurement region. On the other hand, if $\int \tau'(s)^2 \mathrm{d}s = 0$, the variance converges to $\sigma^2(0)$ for δ large. Indeed, vanishing of the integral of $\tau'(s)^2$ implies that $\tau(s) = \tau(0)$, in which case the variance converges with many measurements to

$$\lim_{n \longrightarrow \infty} \mathcal{V}(n, \delta) = \frac{2\delta}{2\delta + 1}\sigma^2(0) \xrightarrow{\delta \to \infty} \sigma^2(0).$$

Proof. Let us denote $\tau_i \equiv \tau(x_{n,i})$, suppressing the n-dependence. The inverse of the covariance matrix of the autocorrelation model is

$$\Sigma^{-1} = \mathsf{S}^{-1}\mathsf{R}^{-1}\mathsf{S}^{-1},$$

with R as in Eq. (20),

$$\mathsf{R}^{-1} = \frac{1}{1-\varrho^2}\begin{bmatrix} 1 & -\varrho & 0 & \cdots & 0 \\ -\varrho & 1+\varrho^2 & -\varrho & \cdots & 0 \\ 0 & -\varrho & 1+\varrho^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix},$$

and S the diagonal matrix with entries $\mathsf{S}_{ii} = \sigma_i = \tau_i^{-1}$. Using $\mathsf{X} = (1, 1, \ldots, 1)^t$ and Eq. (5), we find that the inverse of the variance $\mathcal{V}(n, \delta)$ of the estimated mean is

$$\begin{aligned} \mathcal{V}^{-1}(n,\delta) &= \sum_{i,j=0}^{n} (\mathsf{R}^{-1})_{ij}\tau_i\tau_j \\ &= \frac{1}{1-\varrho^2}\left[\sum_{i=0}^{n}\tau_i^2 + \varrho^2\sum_{i=1}^{n-1}\tau_i^2 - 2\varrho\sum_{i=0}^{n-1}\tau_i\tau_{i+1}\right] \\ &= \frac{1}{1-\varrho^2}\left[(1-\varrho)^2\sum_{i=0}^{n}\tau_i^2 + \varrho\sum_{i=0}^{n-1}(\tau_{i+1}-\tau_i)^2\right. \\ &\quad \left. + \varrho(1-\varrho)(\tau(0)^2+\tau(1)^2)\right]. \end{aligned} \tag{30}$$

Since $x_{n,i} = i/n$, we can use the trapezoidal rule for numerical integration to approximate the sums (see Ref. [22])

$$\frac{1}{n}\sum_{i=0}^{n}\tau_i^2 = \int_0^1 \tau(s)^2 \mathrm{d}s + \frac{1}{2n}\left(\tau(0)^2+\tau(1)^2\right) + O(n^{-2})$$

and, using the relation $(\tau_{i+1}-\tau_i) = (1/n)\tau_i' + (1/2n^2)\tau_i'' + O(n^{-3})$,

$$\begin{aligned} n\sum_{i=0}^{n-1}(\tau_{i+1}-\tau_i)^2 &= \frac{1}{n}\sum_{i=0}^{n}(\tau_i')^2 - \frac{1}{n}\tau'(1)^2 + \frac{1}{n^2}\sum_{i=0}^{n}\tau_i'\tau_i'' + O(n^{-2}) \\ &= \int_0^1 \tau'(s)^2\mathrm{d}s - \frac{1}{2n}\left(\tau'(1)^2 - \tau'(0)^2\right) \\ &\quad + \frac{1}{2n}\int_0^1 \frac{\mathrm{d}}{\mathrm{d}s}\left(\tau'(s)^2\right) + O(n^{-2}) \\ &= \int_0^1 \tau'(s)^2\mathrm{d}s + O(n^{-2}). \end{aligned}$$

Noting that $\varrho = \exp(-(\delta n)^{-1})$, it follows that

$$\begin{aligned} \mathcal{V}^{-1}(n,\delta) &= \frac{1}{\delta}f\left((\delta n)^{-1}\right)\left(\int_0^1 \tau(s)^2\mathrm{d}s + O(n^{-2})\right) + \frac{1}{2}(\tau(0)^2+\tau(1)^2) \\ &\quad + \delta g\left((\delta n)^{-1}\right)\left(\int_0^1 \tau'(s)^2\mathrm{d}s + O(n^{-2})\right), \end{aligned}$$

where $f(x) = \left(1 - e^{-x}\right)/x\left(1 + e^{-x}\right)$ and $g(x) = xe^{-x}/\left(1 - e^{-2x}\right)$. For $\delta n \gg 1$, these functions both behave as

$$\lim_{\delta n \to \infty} f\left((\delta n)^{-1}\right) = \frac{1}{2} + O\left((\delta n)^{-2}\right)$$
$$\lim_{\delta n \to \infty} g\left((\delta n)^{-1}\right) = \frac{1}{2} + O\left((\delta n)^{-2}\right)$$

which produces the desired result, Eq. (29). □

Note that the usual result for uncorrelated errors ($\delta = 0$) may be recovered from (30) by simply setting $\varrho = 0$. In this case, the inverse variance is simply the sum of n positive terms, implying the familiar relationship $\mathcal{V} \sim n^{-1}$ for n uncorrelated measurements. For a large number n of highly correlated measurements ($\delta n \gg 1$), we may ignore the higher order terms in Eq. (29) of Theorem 2, and the variance of the estimate converges to

$$\mathcal{V} = \frac{2\delta}{\int_0^1 \tau(s)^2 ds + \delta\left(\tau(0)^2 + \tau(1)^2\right) + \delta^2 \int_0^1 \tau'(s)^2 ds}. \tag{31}$$

Examples with $\delta > 0$ and linear inverse variance $\tau(s) = 1 + \alpha s$, with $\alpha = 1$ are shown in Fig. 3, with the results for finite n summed numerically and the limit $n \to \infty$ from Eq. (31). The value of $\mathcal{V}$ converges rapidly as $n \to \infty$ except near $\delta = 0$, where $\mathcal{V} \sim n^{-1}$. The form of Eq. (31), including the behavior $\mathcal{V}(n, \delta) \sim 1/\delta$ for large δ, is evident. The maximum of $\mathcal{V}(n, \delta)$ with respect to δ occurs at

$$\delta^2 = \frac{\int_0^1 \tau(s)^2 ds}{\int_0^1 \tau'(s)^2 ds}.$$

That is, $\mathcal{V}$ decreases if the correlation length δ is larger than the typical scale for change of $\tau(x)$.

In Fig. 4 we show $\mathcal{V}(n, \delta)$ as a function of n for $\alpha = 1$ and three values of δ. These results, similar to those of Ref. [9], show that $\mathcal{V}(n, \delta)$ converges to a positive value as $n \to \infty$, unless $\delta = 0$. There is an initial decrease, when $n \lesssim 1/\delta$; to the right of this region $\mathcal{V}(n, \delta)$ is nearly constant.

In Fig. 5 we show results for a constant inverse variance $\tau(s) = 1$ (i.e. $\alpha = 0$) both numerically for finite n and the asymptotic result [Eq. (31)] for $n \to \infty$. Again, the results converge rapidly with n except near $\delta = 0$, with $\mathcal{V} \to \tau^{-2} = 1$ as $\delta \to \infty$.

Notice that these results show that $\mathcal{V}(n, \delta)$ approaches a limiting curve as $n \to \infty$. Except for small δ, the convergence is quite rapid due to the absence of corrections of order n^{-1}, as mentioned above. The results for $\alpha = 1$ show the generic situation of $\mathcal{V}(n, \delta) \to 0$ as $\delta \to \infty$; those with $\alpha = 0$ show the special situation in which $\mathcal{V}(n, \delta)$ approaches a positive constant in that limit. As long as $\delta > 0$, the variance of the estimate $\mathcal{V}(n, \delta)$ becomes constant with respect to n for $n \gtrsim 1/\delta$ and has a finite limit as $n \to \infty$, showing that there is no advantage to be gained by increasing the number of measurements at points $x_{n,i}$ past $n \sim 1/\delta$.

5. Strong correlations among some measurements

In Sections 2 through 4, we have studied in depth the case in which the rank of the covariance matrix Σ becomes unity for $\kappa = 0$. To generalize, we suppose the rank becomes r' in some limit, with $1 \leq r' < n$. This can occur, for example, if two distinct and uncorrelated types of measurements are made. (This situation was present in the plasma reconstruction

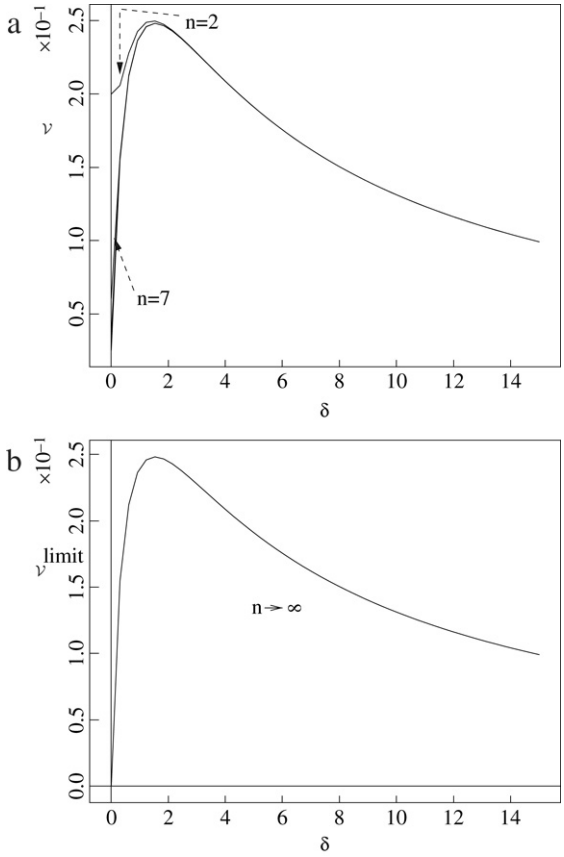

Fig. 3. Variance $\mathcal{V}$ as a function of δ for $\tau(1) = 1$ and $\alpha = 1$. The cases for $n = 2, 7$ are summed numerically and the case $\mathcal{V}^{\text{limit}}$ for $n \to \infty$ is from the analytic limit (31) in Theorem 2.

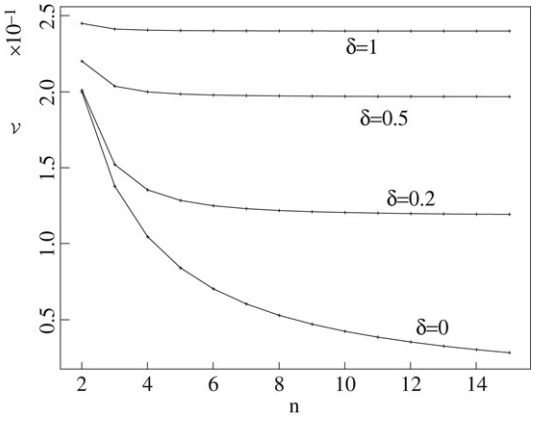

Fig. 4. Variance of the estimate $\mathcal{V}$ as a function of n for $\delta = 0, 0.2, 0.5, 1$, with $\alpha = 1$.

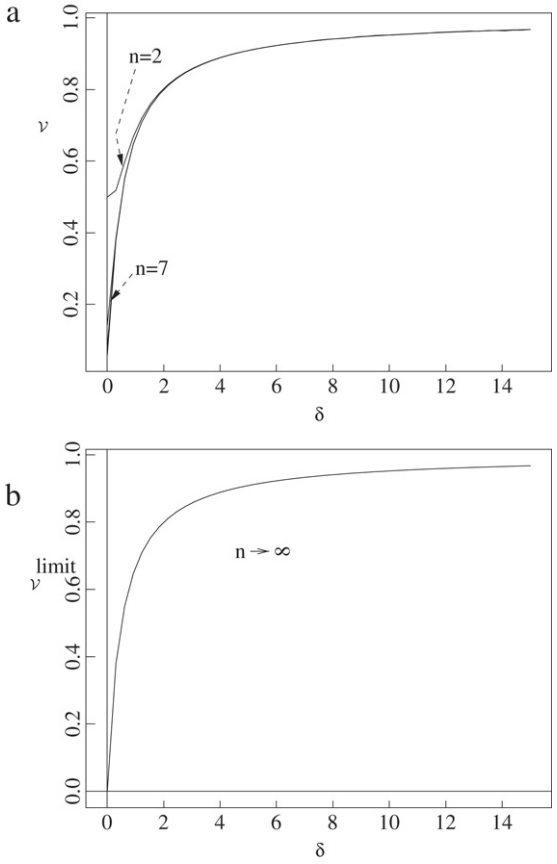

Fig. 5. Variance of the estimate $\mathcal{V}$ as a function of δ for $\tau(1) = 1$, $\alpha = 0$ and (a) $n = 2, 7$, summed numerically, and (b) the limiting value $\mathcal{V}^{\text{limit}}$ for $n \to \infty$ from the analytic limit (31) in Theorem 2.

studies of Ref. [9], where magnetic field measurements external to the plasma and pressure measurements internal to the plasma were used.) For example, suppose one type of measurement, for $i = 1, \ldots, r' - 1$ has a correlation matrix of the form $\mathsf{R}^{(1)} = \varrho_1^{|i-j|}$ and a second type, for $i = r', \ldots, n$ has $\mathsf{R}^{(2)} = \varrho_2^{|i-j|}$. We then have $\Sigma = \mathsf{SRS}$, with

$$\mathsf{R} = \begin{bmatrix} \mathsf{R}^{(1)} & 0 \\ 0 & \mathsf{R}^{(2)} \end{bmatrix}.$$

Then for $\varrho_2 = 1$, but $|\varrho_1| < 1$, the rank of R, and therefore the rank of Σ, equals r'.

In this case, the heuristic procedure described in Section 3 leads, in the limit $\varrho_2 \to 1$, to a linear system of equations subjected to noise

$$Z_j = v_j^t \mathsf{X}\beta + \lambda_j^{1/2}\xi_j \quad j = 1, \ldots, r',$$

and a noise-free subspace

$$Z_j = v_j^t \mathsf{X}\beta \quad j = r' + 1, \ldots, n.$$

If $n-r' \geq m$, the second set of equations determines $\hat{\beta}$ exactly and the variance of the estimate is zero. If, on the other hand, $n-r' < m$, the second set of equations has a null space of dimension $m-n+r'$. That is, there are $n-r'$ linearly independent linear combinations of the β_i that are determined exactly. In other words, the covariance matrix of the estimate has rank $n-r'$, i.e. $n-r'$ zero eigenvalues and $m-n+r'$ nonzero eigenvalues.

Note that the previous arguments allow for the relaxation of the assumption in Theorem 1 of Section 3.2 that the covariance matrix Σ remains of full rank n for $\kappa > 0$. A proof of the most general case includes the possibility of multiple rank changes $n \to r' \to r'' \to \cdots \to 1$ as κ decreases. At each rank change the discussion above applies and additional linear combinations of the parameters are determined until, in the limit, all are determined.

6. Summary and discussion

In its fundamental form, the main result of this paper, given in Section 3, is the following: in the limit of full correlation $\kappa \to 0$ [see Eq. (6)], the covariance matrix V of the estimate $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_m)$ generically vanishes. That is, its trace, the total variance of V, vanishes. The exceptions to this rule occur when the vector of signed deviations $(e_1\sigma_1, \ldots, e_n\sigma_n)^t$ of the measurements is in the range space of the design matrix X. We explained the decrease and eventual vanishing of trace(V) by means of simple examples in Section 2, and also showed the relationship between this phenomenon and negative weighting, in which the estimate is a weighted average of the measurements, with some weights negative.

This result is so surprising that it suggests a "free lunch" possibility. The idea that stronger correlations can be obtained simply by packing in closer measurements has been studied in Section 4. It is found that the covariance of the estimate does indeed decrease as the number n of measurements increases, but this decrease flattens when $\delta \gtrsim \Delta x$, where δ is the correlation length scale and Δx is the spacing between measurements. The interpretation of this result is that for the variance to decrease with increasing number of measurements, the measurement spacing must not be much smaller than the correlation length. Further, from (31) we have concluded that the variance decreases with correlation length δ if δ is greater than the typical length scale for variations in the signal-to-noise ratio.

We have addressed more general covariance structures in Section 5. We considered a generalization of the condition that, as $\kappa \to 0$, the rank of the covariance matrix Σ goes from n to unity. This generalization deals with cases in which this rank decreases from n to r' as κ decreases, i.e. cases in which some measurements, but not all, become fully correlated. In this case, the result is unchanged if $r' \geq m$; if $r' < m$, however, the estimate is not completely determined in the limit of large correlations, but a linearly independent set of $n-r'$ combinations are determined exactly.

References

[1] L.L. Lao, H.St. John, R.D. Stambaugh, A.G. Kellman, W. Pfeiffer, Nucl. Fusion 25 (1985) 1611.
[2] J.K. Anderson, C.B. Forest, T.M. Biewer, J.S. Sarff, J.C. Wright, Nucl. Fusion 44 (2004) 162.
[3] J.D. Hanson, S.F. Knowlton, P. Hirschman, E.A. Lazarus, B. Amer. Phys. Soc. 49 (2004) 307.
[4] L.L. Lao, H.St. John, R.D. Stambaugh, W. Pfeiffer, Nucl. Fusion 25 (1985) 1421.
[5] I. Furno, Private communication, 2007.
[6] Y.B. Lim, J. Sacks, W.J. Studden, W.J. Welch, Canad. J. Statist. 30 (2002) 109.
[7] R. Christensen, Plane Answers to Complex Questions: The Theory of Linear Models, Springer-Verlag, 2001.
[8] R. Christensen, Advanced Linear Modeling, Springer-Verlag, 2001.
[9] C.S. Jones, J.M. Finn, Nucl. Fusion 46 (2006) 335.

[10] P.L. Canner, Amer. Statist. 23 (1969) 239.
[11] G. D'Agostini, Nuclear Instrum. Methods Phys. Res. 346 (1994) 306.
[12] G. D'Agostini, Bayesian Reasoning in Data Analysis, World Scientific, 2003, p. 202.
[13] K.M. Hanson, T. Kawano, P. Talou, AIP Conf. Proc. 769 (2004) 304.
[14] D.S. Sivia, AIP Conf. Proc. 707 (2004) 303.
[15] R.L. Winkler, Manage. Sci. 27 (1981) 479.
[16] R.T. Clemen, R.L. Winkler, Oper. Res. 33 (1985) 427.
[17] D.G. Morrison, D.C. Schmittlein, Oper. Res. 39 (1991) 519.
[18] C.R. Rao, Linear Statistical Inference and Its Applications, John Wiley and Sons, 1965.
[19] W. Feller, An Introduction to Probability Theory and Its Applications, vol. II, John Wiley and Sons, 1971, p. 110.
[20] N.A.C. Cressie, Statistics for Spatial Data, John Wiley, 1993.
[21] M.L. Stein, J. Amer. Statist. Assoc. 90 (1995) 1277.
[22] M. Abramowitz, I.A. Stegun (Eds.), Handbook of Mathematical Functions, Dover Publications, 1970 (chapter 25).