

An Efficient Calculation of Fisher Information Matrix: Monte Carlo Approach using Prior Information

Sonjoy Das, James C. Spall and Roger Ghanem

Abstract—The Fisher information matrix (FIM) is a critical quantity in several aspects of system identification, including input selection and confidence region calculation. Analytical determination of the FIM in a general system identification setting may be difficult or almost impossible due to intractable modeling requirements and/or high-dimensional integration. A Monte Carlo (MC) simulation-based technique was introduced by the second author to address these difficulties [1]. This paper proposes an extension of the MC algorithm in order to enhance the statistical qualities of the estimator of the FIM. This modified MC algorithm is particularly useful in those cases where the FIM has a structure with some elements being analytically known from prior information and the others being unknown. The estimator of the FIM, obtained by using the proposed MC algorithm, simultaneously preserves the analytically known elements and reduces the variances of the estimators of the unknown elements by capitalizing on the information contained in the known elements.

Index Terms—Fisher information matrix, Monte Carlo simulation.

I. INTRODUCTION AND MOTIVATING FACTORS

The Fisher information matrix (FIM) plays a key role in estimation and identification, and information theory. A standard problem in the practical application and theory of statistical estimation and identification is to estimate the unobservable parameters, θ , of the probability distribution function from a set of observed data set drawn from that distribution [2]. The FIM is an indicator of the amount of information contained in this observed data set about θ . Some important areas of applications of FIM include, to name a few, confidence interval computation of model parameter, configuration of experimental design, and determination of noninformative prior distribution (Jeffreys' prior) for Bayesian analysis.

However, the analytical determination of the FIM may be a formidable undertaking in a general setting, specially in nonlinear models, due to intractable modeling requirements and/or high-dimensional integration. To avoid this difficulty, the resampling approach [3, Section 13.3.5], [1], a Monte Carlo (MC) based simulation technique, may be employed to estimate the FIM.

There may be also instances in practice when some elements of the FIM are analytically known from prior information while the other elements are unknown (and

need to be estimated) [2]. In such cases, the resampling approach, however, still yields the *full* FIM without taking any advantage of the prior information contained in the analytically known elements. The resampling based estimates of the known elements are also “wasted” because these estimates are simply replaced by the analytically known elements. The issue yet to be examined is whether there is a way of focusing the averaging process (required in the resampling algorithm) — on the elements of interest (unknown elements that need to be estimated) — that is more effective than simply extracting the estimates of those elements from the full FIM estimated by employing the existing resampling algorithm.

The current work, similar in some sense to the one for Jacobian/Hessian estimates presented earlier [4], presents a modified and improved (in the sense of variance reduction) version of the resampling approach for estimating the unknown elements of the FIM by “borrowing” the information contained in the analytically known elements.

II. FISHER INFORMATION MATRIX: DEFINITION AND NOTATION

Consider a set of n random data vector (to be treated as column vector), $\{\mathcal{Z}_1, \dots, \mathcal{Z}_n\}$, and form \mathbf{Z}_n by $\mathbf{Z}_n = [\mathcal{Z}_1^T, \dots, \mathcal{Z}_n^T]^T$. Here, the superscript, T , is transpose operator. Let the multivariate joint probability density or mass (or hybrid density/mass) function (pdf) of \mathbf{Z}_n be denoted by $p_{\mathbf{Z}_n}(\cdot|\theta)$ that is parameterized by θ . The likelihood function of θ is then given by $\ell(\theta|\mathbf{Z}_n) = p_{\mathbf{Z}_n}(\mathbf{Z}_n|\theta)$ and the associated log-likelihood function by $L(\theta|\mathbf{Z}_n) \equiv \ln \ell(\theta|\mathbf{Z}_n)$.

Let us define the $p \times 1$ gradient vector, \mathbf{g} , of L by $\mathbf{g}(\theta|\mathbf{Z}_n) = \partial L(\theta|\mathbf{Z}_n)/\partial \theta$ and the $p \times p$ Hessian matrix, \mathbf{H} , by $\mathbf{H}(\theta|\mathbf{Z}_n) = \partial^2 L(\theta|\mathbf{Z}_n)/\partial \theta \partial \theta^T$. Then, the $p \times p$ FIM, $\mathbf{F}_n(\theta)$, is defined [3, Section 13.3.2] as follows,

$$\mathbf{F}_n(\theta) \equiv E[\mathbf{g}(\theta|\mathbf{Z}_n) \cdot \mathbf{g}^T(\theta|\mathbf{Z}_n)|\theta] = -E[\mathbf{H}(\theta|\mathbf{Z}_n)|\theta], \quad (1)$$

provided that the derivatives and expectation (the expectation operator, E , is with respect to the probability measure of \mathbf{Z}_n) exist. The equality ‘=’ in (1) is followed [3, p. 352 – 353] by assuming that L is twice differentiable with respect to θ and the regularity conditions [5, Section 3.4.2] hold for ℓ . The Hessian-based form above is more amenable to the practical computation for FIM than the gradient-based form that is used for defining the FIM.

III. CURRENT RESAMPLING ALGORITHM — NO USE OF PRIOR INFORMATION

The current resampling approach is based on producing a set of large (say N) number of Hessian estimates from either

Sonjoy Das is with University of Southern California, Department of Civil and Environmental Engineering, Los Angeles, California 90089. Sonjoy.Das@usc.edu

James C. Spall is with The Johns Hopkins University, Applied Physics Laboratory, Laurel, Maryland 20723. james.spall@jhuapl.edu

Roger Ghanem is with University of Southern California, Los Angeles, California 90089. ghanem@usc.edu

the values of the log-likelihood function or (if available) its exact stochastic gradient both of which, in turn, are computed from a set of statistically independent pseudo data vector, $\{\mathbf{Z}_{\text{pseudo}}(1), \dots, \mathbf{Z}_{\text{pseudo}}(N)\}$, $\mathbf{Z}_{\text{pseudo}}(k) \sim p_{\mathbf{Z}_n}(\cdot|\boldsymbol{\theta})$, $k = 1, \dots, N$. This set of pseudo data vector acts as a proxy for the observed data set in the resampling algorithm. The average of the negative of these Hessian estimates is reported as an estimate of $\mathbf{F}_n(\boldsymbol{\theta})$.

Denote the k -th estimate of the Hessian matrix, $\mathbf{H}(\boldsymbol{\theta}|\mathbf{Z}_{\text{pseudo}}(k))$, in the resampling algorithm by $\hat{\mathbf{H}}_k$. Then, $\hat{\mathbf{H}}_k$, as per resampling scheme, is computed as [1],

$$\hat{\mathbf{H}}_k = \frac{1}{2} \left\{ \frac{\delta \mathbf{G}_k}{2c} [\Delta_{k1}^{-1}, \dots, \Delta_{kp}^{-1}] + \left(\frac{\delta \mathbf{G}_k}{2c} [\Delta_{k1}^{-1}, \dots, \Delta_{kp}^{-1}] \right)^T \right\}, \quad (2)$$

in which $c > 0$ is a small number, $\delta \mathbf{G}_k \equiv \mathbf{G}(\boldsymbol{\theta} + c\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(k)) - \mathbf{G}(\boldsymbol{\theta} - c\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(k))$ and the perturbation vector, $\boldsymbol{\Delta}_k = [\Delta_{k1}, \dots, \Delta_{kp}]^T$, is a user-generated random vector statistically independent of $\mathbf{Z}_{\text{pseudo}}(k)$. The random variables, $\Delta_{k1}, \dots, \Delta_{kp}$, are mean-zero and statistically independent and, also the inverse moments, $E[|1/\Delta_{km}|]$, $m = 1, \dots, p$, are finite.

The symmetrizing operation (the multiplier 1/2 and the indicated sum) as shown in (2) is useful in optimization problems to compute a symmetric estimate of the Hessian matrix with finite samples [1]. This also maintains a symmetric estimate of $\mathbf{F}_n(\boldsymbol{\theta})$, which itself is a symmetric matrix.

Depending on the setting, $\mathbf{G}(\cdot|\mathbf{Z}_{\text{pseudo}}(k))$, as required in $\delta \mathbf{G}_k$, represents the k -th direct measurement or approximation of the gradient vector, $\mathbf{g}(\cdot|\mathbf{Z}_{\text{pseudo}}(k))$. If the direct measurement or computation of \mathbf{g} is feasible, $\mathbf{G}(\boldsymbol{\theta} \pm c\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(k))$ represent the *direct* k -th measurements of $\mathbf{g}(\cdot|\mathbf{Z}_{\text{pseudo}}(k))$ at $\boldsymbol{\theta} \pm c\boldsymbol{\Delta}_k$. Otherwise, $\mathbf{G}(\boldsymbol{\theta} \pm c\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(k))$ represents the k -th *approximation* of $\mathbf{g}(\boldsymbol{\theta} \pm c\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(k))$ based on the values of $L(\cdot|\mathbf{Z}_{\text{pseudo}}(k))$.

If the direct measurements or computations of \mathbf{g} are not feasible, \mathbf{G} in (2) can be computed by using the classical finite-difference (FD) technique [3, Section 6.3] or the simultaneous perturbation (SP) gradient approximation technique [6], [3, Section 7.2] from the values of $L(\cdot|\mathbf{Z}_{\text{pseudo}}(k))$. For the computation of gradient approximation based on the values of L , there are advantages to using *one-sided* [3, p. 199] SP gradient approximation (relative to the standard two-sided SP gradient approximation) in order to reduce the total number of function measurements or evaluations for L . The SP technique for gradient approximation is quite useful when p is large and usually superior to FD technique when the objective is to estimate $\mathbf{F}_n(\boldsymbol{\theta})$ by employing the resampling algorithm. The formula for the one-sided gradient approximation using SP technique is given by,

$$\mathbf{G}^{(1)}(\boldsymbol{\theta} \pm c\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(k)) = (1/\tilde{c}) \left[L(\boldsymbol{\theta} + \tilde{c}\tilde{\boldsymbol{\Delta}}_k \pm c\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(k)) - L(\boldsymbol{\theta} \pm c\boldsymbol{\Delta}_k|\mathbf{Z}_{\text{pseudo}}(k)) \right] \begin{bmatrix} \tilde{\Delta}_{k1}^{-1} \\ \vdots \\ \tilde{\Delta}_{kp}^{-1} \end{bmatrix}, \quad (3)$$

in which superscript, (1), in $\mathbf{G}^{(1)}$ indicates that it is one-sided gradient approximation ($\mathbf{G} = \mathbf{G}^{(1)}$), $\tilde{c} > 0$ is a small number and $\tilde{\boldsymbol{\Delta}}_k = [\tilde{\Delta}_{k1}, \dots, \tilde{\Delta}_{kp}]^T$ is generated in the same statistical manner as $\boldsymbol{\Delta}_k$, but otherwise statistically independent of $\boldsymbol{\Delta}_k$ and $\mathbf{Z}_{\text{pseudo}}(k)$. It is usually recommended that $\tilde{c} > c$.

At this stage, let us also formally state that the perturbation vectors, $\boldsymbol{\Delta}_k$ and $\tilde{\boldsymbol{\Delta}}_k$, satisfy the following condition,[3, Chapter 7].

C.1: (Statistical properties of the perturbation vector) The random variables, Δ_{km} (and $\tilde{\Delta}_{km}$), $k = 1, \dots, N$, $m = 1, \dots, p$, are statistically independent and almost surely (a.s.) uniformly bounded for all k, m , and, are also mean-zero and symmetrically distributed satisfying $E[|1/\Delta_{km}|] < \infty$ (and $E[|1/\tilde{\Delta}_{km}|] < \infty$).

Let us also assume that the moments of Δ_{km} and $1/\Delta_{km}$ (and, of $\tilde{\Delta}_{km}$ and $1/\tilde{\Delta}_{km}$) up to fifth order exist (this condition will be used later in Section IV-C). Since Δ_{km} (and $\tilde{\Delta}_{km}$) is symmetrically distributed, $1/\Delta_{km}$ (and $1/\tilde{\Delta}_{km}$) is also symmetrically distributed implying that,

I: (Statistical properties implied by C.1) All the odd moments of Δ_{km} and $1/\Delta_{km}$ (and of $\tilde{\Delta}_{km}$ and $1/\tilde{\Delta}_{km}$) up to fifth order are zeros, $E[(\Delta_{km})^q] = 0$ and $E[(1/\Delta_{km})^q] = 0$ ($E[(\tilde{\Delta}_{km})^q] = 0$ and $E[(1/\tilde{\Delta}_{km})^q] = 0$), $q = 1, 3, 5$.

The random vectors, $\boldsymbol{\Delta}_k$ (and $\tilde{\boldsymbol{\Delta}}_k$), are also independent across k . The random variables, $\Delta_{k1}, \dots, \Delta_{kp}$ (and $\tilde{\Delta}_{k1}, \dots, \tilde{\Delta}_{kp}$), can also be chosen identically distributed. In fact, independent and identically distributed (i.i.d.) (across both k and m) mean-zero random variable satisfying C.1 is a perfectly valid choice for Δ_{km} (and $\tilde{\Delta}_{km}$). In particular, Bernoulli ± 1 random variable for Δ_{km} (and $\tilde{\Delta}_{km}$) is a valid — but not the necessary — choice among other probability distributions satisfying C.1.

Given the form of the Hessian estimate in (2), an estimate, $\hat{\mathbf{F}}_n$, of $\mathbf{F}_n(\boldsymbol{\theta})$ is computed by taking the (negative) average of the Hessian estimates, $\hat{\mathbf{H}}_k$, $\hat{\mathbf{F}}_n = -(1/N) \sum_{k=1}^N \hat{\mathbf{H}}_k$.

IV. IMPROVED RESAMPLING ALGORITHM — USING PRIOR INFORMATION

Let the k -th estimate of Hessian matrix, $\mathbf{H}(\boldsymbol{\theta}|\mathbf{Z}_{\text{pseudo}}(k))$, per proposed resampling algorithm be denoted by $\tilde{\mathbf{H}}_k$. In this section, the estimator, $\tilde{\mathbf{H}}_k$, is shown separately for two different cases: Case 1 :— when only the measurements of L are available and, Case 2 :— when the measurements of \mathbf{g} are available. To contrast the two cases, the superscript, (L), is used in $\tilde{\mathbf{H}}_k^{(L)}$ and $\hat{\mathbf{H}}_k^{(L)}$ to represent the dependence on L for Case 1 and, the superscript, (\mathbf{g}), in $\tilde{\mathbf{H}}_k^{(\mathbf{g})}$ and $\hat{\mathbf{H}}_k^{(\mathbf{g})}$ for Case 2.

A. Additional Notation

Denote the (i, j) -th element of $\mathbf{F}_n(\boldsymbol{\theta})$ by $F_{ij}(\boldsymbol{\theta})$. Let \mathbb{I}_i , $i = 1, \dots, p$, be the set of column indices of the known elements of the i -th row of $\mathbf{F}_n(\boldsymbol{\theta})$ and \mathbb{I}_i^c be the complement

of \mathbb{I}_i . Consider a $p \times p$ matrix, $\mathbf{F}_n^{(\text{given})}$, whose (i, j) -th element, $F_{ij}^{(\text{given})}$, is defined as follows,

$$F_{ij}^{(\text{given})} = \begin{cases} F_{ij}(\boldsymbol{\theta}), & \text{if } j \in \mathbb{I}_i \\ 0, & \text{if } j \in \mathbb{I}_i^c \end{cases}, \quad i = 1, \dots, p. \quad (4)$$

Consider another $p \times p$ matrix, \mathcal{D}_k , defined by, $\mathcal{D}_k = \Delta_k [\Delta_{k1}^{-1}, \dots, \Delta_{kp}^{-1}]$ together with the matrix, $\tilde{\mathcal{D}}_k$, obtained by replacing all Δ_{ki} in \mathcal{D}_k with the corresponding $\tilde{\Delta}_{ki}$ (note that \mathcal{D}_k is symmetric when the perturbations are i.i.d. Bernoulli distributed).

B. The Step-by-Step Description of the Proposed Resampling Algorithm

The new estimate, $\tilde{\mathbf{H}}_k$, is extracted from $\tilde{\mathbf{H}}_{k0}$ that is defined below separately for Case 1 and Case 2.

Case 1: only the measurements of L are available,

$$\tilde{\mathbf{H}}_{k0}^{(L)} = \hat{\mathbf{H}}_k^{(L)} - \frac{1}{2} \left[\tilde{\mathcal{D}}_k^T (-\mathbf{F}_n^{(\text{given})}) \mathcal{D}_k + (\tilde{\mathcal{D}}_k^T (-\mathbf{F}_n^{(\text{given})}) \mathcal{D}_k)^T \right]. \quad (5)$$

Case 2: measurements of \mathbf{g} are available,

$$\tilde{\mathbf{H}}_{k0}^{(\mathbf{g})} = \hat{\mathbf{H}}_k^{(\mathbf{g})} - \frac{1}{2} \left[(-\mathbf{F}_n^{(\text{given})}) \mathcal{D}_k + ((-\mathbf{F}_n^{(\text{given})}) \mathcal{D}_k)^T \right]. \quad (6)$$

The estimates, $\tilde{\mathbf{H}}_k^{(L)}$ and $\tilde{\mathbf{H}}_k^{(\mathbf{g})}$, are readily obtained from, respectively, $\tilde{\mathbf{H}}_{k0}^{(L)}$ in (5) and $\tilde{\mathbf{H}}_{k0}^{(\mathbf{g})}$ in (6) by replacing the (i, j) -th element of $\tilde{\mathbf{H}}_{k0}^{(L)}$ and $\tilde{\mathbf{H}}_{k0}^{(\mathbf{g})}$ with known values of $-F_{ij}(\boldsymbol{\theta})$, $j \in \mathbb{I}_i$, $i = 1, \dots, p$. The new estimate, $\tilde{\mathbf{F}}_n$, of $\mathbf{F}_n(\boldsymbol{\theta})$ is then computed by taking the (negative) average of the Hessian estimates, $\tilde{\mathbf{H}}_k$.

A summary of the salient steps, required to produce the estimate, $\tilde{\mathbf{F}}_n$ (i.e., $\tilde{\mathbf{F}}_n^{(L)}$ or $\tilde{\mathbf{F}}_n^{(\mathbf{g})}$ with the appropriate superscript) of $\mathbf{F}_n(\boldsymbol{\theta})$ per modified resampling algorithm as proposed here, is presented next.

Step 0. Initialization: Construct $\mathbf{F}_n^{(\text{given})}$ in (4) by using the analytically known elements of the FIM. Determine $\boldsymbol{\theta}$, the sample size (n) and the number (N) of pseudo data vectors that will be generated. Determine whether log-likelihood, $L(\cdot)$, or gradient vector, $\mathbf{g}(\cdot)$, will be used to compute the Hessian estimates, $\tilde{\mathbf{H}}_k$. Pick a small number, c , (perhaps $c = 0.0001$) to be used for Hessian estimation (see (2)) and, if required, another small number, \tilde{c} (perhaps $\tilde{c} = 0.00011$), for gradient approximation (see (3)). Set $k = 1$.

Step 1. At the k -th step perform the following tasks,

- a. **Generation of pseudo data:** Based on $\boldsymbol{\theta}$, generate the k -th pseudo data vector, $\mathbf{Z}_{\text{pseudo}}(k)$, by using MC simulation technique.
- b. **Computation of $\hat{\mathbf{H}}_k$:** Generate Δ_k (and also $\tilde{\Delta}_k$, if required, for gradient approximation) by satisfying C.1. Using $\mathbf{Z}_{\text{pseudo}}(k)$, Δ_k or/and $\tilde{\Delta}_k$, evaluate $\hat{\mathbf{H}}_k$ (i.e., $\hat{\mathbf{H}}_k^{(L)}$ or $\hat{\mathbf{H}}_k^{(\mathbf{g})}$) by using (2).

- c. **Computation of \mathcal{D}_k and $\tilde{\mathcal{D}}_k$:** Use Δ_k or/and $\tilde{\Delta}_k$, as generated in the above step, to construct \mathcal{D}_k or/and $\tilde{\mathcal{D}}_k$ as defined in section IV-A.

- d. **Computation of $\tilde{\mathbf{H}}_{k0}$:** Modify $\hat{\mathbf{H}}_k$ as produced in **Step 1b** by employing (5) or (6) as appropriate in order to generate $\tilde{\mathbf{H}}_{k0}$ (i.e., $\tilde{\mathbf{H}}_{k0}^{(L)}$ or $\tilde{\mathbf{H}}_{k0}^{(\mathbf{g})}$).

Step 2. Average of $\tilde{\mathbf{H}}_{k0}$: Repeat **Step 1** until N estimates, $\tilde{\mathbf{H}}_{k0}$, are produced. Compute the (negative) mean of these N estimates. (The standard recursive representation of sample mean can be used here to avoid the storage of N matrices, $\tilde{\mathbf{H}}_{k0}$). The resulting (negative) mean is $\tilde{\mathbf{F}}_{n0}$.

Step 3. Evaluation of $\tilde{\mathbf{F}}_n$: The new estimate, $\tilde{\mathbf{F}}_n$, of $\mathbf{F}_n(\boldsymbol{\theta})$ per modified resampling algorithm is simply obtained by replacing the (i, j) -th element of $\tilde{\mathbf{F}}_{n0}$ with the analytically known elements, $F_{ij}(\boldsymbol{\theta})$, $j \in \mathbb{I}_i$, $i = 1, \dots, p$, of $\mathbf{F}_n(\boldsymbol{\theta})$. To avoid the possibility of having a non-positive semi-definite estimate, it may be desirable to take the symmetric square root of the square of the estimate (the `sqrtm` function in MATLAB may be useful here).

The new estimator, $\tilde{\mathbf{F}}_n$, is better than $\hat{\mathbf{F}}_n$ in the sense that it would preserve exactly the analytically known elements of $\mathbf{F}_n(\boldsymbol{\theta})$ as well as reduce the variances of the estimators of the unknown elements of $\mathbf{F}_n(\boldsymbol{\theta})$.

C. Theoretical Basis for the Modified Resampling Algorithm

For notational simplification, the subscript ‘pseudo’ in $\mathbf{Z}_{\text{pseudo}}(k)$ and the dependence of $\mathbf{Z}(k)$ on k would be suppressed (note that $\mathbf{Z}_{\text{pseudo}}(k)$ is identically distributed across k). Since, Δ_k is usually assumed to be statistically independent across k and an identical condition for $\tilde{\Delta}_k$ is also assumed, their dependence on k would also be suppressed in the forthcoming discussion. Let also the (i, j) -th element of $\hat{\mathbf{H}}_k$ and $\tilde{\mathbf{H}}_k$ be, respectively, denoted by \hat{H}_{ij} and \tilde{H}_{ij} with the appropriate superscript. The two cases as described earlier by (5) and (6) are considered next.

1) *Case 1 — only the measurements of L are available:* The main objective here is to compare variance of $\hat{H}_{ij}^{(L)}$ and variance of $\tilde{H}_{ij}^{(L)}$ to show the superiority of $\tilde{\mathbf{H}}_k^{(L)}$, which leads to the superiority of $\tilde{\mathbf{F}}_n^{(L)}$.

It is assumed here that the gradient estimate is based on one-sided gradient approximation using SP technique given by (3). Based on a Taylor expansion, the i -th component of $\mathbf{G}^{(1)}(\boldsymbol{\theta}|\mathbf{Z})$, $i = 1, \dots, p$, that is an approximation of the i -th component, $g_i(\boldsymbol{\theta}|\mathbf{Z}) \equiv \partial L(\boldsymbol{\theta}|\mathbf{Z})/\partial \theta_i$, of $\mathbf{g}(\boldsymbol{\theta}|\mathbf{Z})$ based on the values of $L(\cdot|\mathbf{Z})$, can be shown to given by [7, Section 2.3.1],

$$G_i^{(1)}(\boldsymbol{\theta}|\mathbf{Z}) = \sum_l g_l(\boldsymbol{\theta}) \frac{\tilde{\Delta}_l}{\tilde{\Delta}_i} + \frac{1}{2} \tilde{c} \sum_{l,m} H_{lm}(\boldsymbol{\theta}) \frac{\tilde{\Delta}_m \tilde{\Delta}_l}{\tilde{\Delta}_i} + \frac{1}{6} \tilde{c}^2 \sum_{l,m,s} \frac{\partial H_{lm}(\boldsymbol{\theta})}{\partial \theta_s} \frac{\tilde{\Delta}_s \tilde{\Delta}_m \tilde{\Delta}_l}{\tilde{\Delta}_i}, \quad (7)$$

in which $H_{lm}(\boldsymbol{\theta}|\mathbf{Z}) \equiv \partial^2 L(\boldsymbol{\theta}|\mathbf{Z})/\partial\theta_l \partial\theta_m$ is the (l, m) -th element of $\mathbf{H}(\boldsymbol{\theta}|\mathbf{Z})$, $\tilde{\boldsymbol{\theta}} = \lambda(\boldsymbol{\theta} + \tilde{c}\tilde{\boldsymbol{\Delta}}) + (1 - \lambda)\boldsymbol{\theta} = \boldsymbol{\theta} + \tilde{c}\lambda\tilde{\boldsymbol{\Delta}}$ (with $\lambda \in [0, 1]$ being some real number) denotes a point on the line segment between $\boldsymbol{\theta}$ and $\boldsymbol{\theta} + \tilde{c}\tilde{\boldsymbol{\Delta}}$ and, in the right-hand-side of (7), the condition on \mathbf{Z} is suppressed for notational clarity and, also the summations are expressed in abbreviated format where the indices span their respective and appropriate ranges.

Given $G_i(\cdot|\mathbf{Z}) \equiv G_i^{(1)}(\cdot|\mathbf{Z})$ by (7), the (i, j) -th element of $\hat{\mathbf{H}}_k^{(L)}$ can be readily obtained from,

$$\hat{H}_{ij}^{(L)} = \frac{1}{2} \left[\hat{J}_{ij}^{(L)} + \hat{J}_{ji}^{(L)} \right], \quad (8)$$

in which the expression of $\hat{J}_{ij}^{(L)}$ based on a Taylor expansion can be shown to be given by [7, Section 2.3.1],

$$\begin{aligned} \hat{J}_{ij}^{(L)} &= \sum_{l,m} H_{lm}(\boldsymbol{\theta}|\mathbf{Z}) \frac{\Delta_m}{\Delta_j} \frac{\tilde{\Delta}_l}{\tilde{\Delta}_i} + O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(c^2) \\ &+ O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(\tilde{c}) + O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(\tilde{c}^2). \end{aligned} \quad (9)$$

The subscripts in the ‘big- O ’ terms, $O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(\cdot)$, explicitly indicate that they depend on $\tilde{\Delta}$, Δ and \mathbf{Z} . In these *random* ‘big- O ’ terms, the point of evaluation, $\boldsymbol{\theta}$, is suppressed for notational clarity. By the use of C.1 and further assumptions on the continuity and uniformly (in k) boundedness conditions on all the derivatives (up to fourth order) of L , it can be shown that $|O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(c^2)/c^2| < \infty$ almost surely (a.s.) (a.s. with respect to the joint probability measure of $\tilde{\Delta}$, Δ and \mathbf{Z}) as $c \rightarrow 0$ and, both $|O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(\tilde{c})/\tilde{c}| < \infty$ a.s. and $|O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(\tilde{c}^2)/\tilde{c}^2| < \infty$ a.s. as $\tilde{c} \rightarrow 0$. The effects of $O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(\tilde{c}^2)$ are not included in $O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(\tilde{c})$. The reason for showing $O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(\tilde{c})$ separately in (9) is that this term vanishes upon expectation because it involves either $E[\tilde{\Delta}_r]$ or $E[1/\tilde{\Delta}_r]$, $r = 1, \dots, p$, both of which are zero by implication I and rest of the terms associated with $O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(\tilde{c})$ do not depend on $\tilde{\Delta}$. The other terms, $O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(c^2)$ and $O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(\tilde{c}^2)$, do not vanish upon expectation. By using (8)-(9), it can then be shown that [7, Section 2.3.1],

$$E[\hat{H}_{ij}^{(L)}|\boldsymbol{\theta}] = -F_{ij}(\boldsymbol{\theta}) + O(c^2) + O(\tilde{c}^2). \quad (10)$$

Note that the ‘big- O ’ terms, $O(c^2)$ and $O(\tilde{c}^2)$, satisfying $|O(c^2)/c^2| < \infty$ as $c \rightarrow 0$ and $|O(\tilde{c}^2)/\tilde{c}^2| < \infty$ as $\tilde{c} \rightarrow 0$, are *deterministic* unlike the random ‘big- O ’ terms in (9).

Next, the variance of $\hat{H}_{ij}^{(L)}$ is given by,

$$\text{var}[\hat{H}_{ij}^{(L)}|\boldsymbol{\theta}] = \frac{1}{4} \left(\text{var}[\hat{J}_{ij}^{(L)}|\boldsymbol{\theta}] + \text{var}[\hat{J}_{ji}^{(L)}|\boldsymbol{\theta}] + 2\text{cov}[\hat{J}_{ij}^{(L)}, \hat{J}_{ji}^{(L)}|\boldsymbol{\theta}] \right). \quad (11)$$

The expression of the typical variance term, $\text{var}[\hat{J}_{ij}^{(L)}|\boldsymbol{\theta}]$, and the covariance term, $\text{cov}[\hat{J}_{ij}^{(L)}, \hat{J}_{ji}^{(L)}|\boldsymbol{\theta}]$, would now be considered.

Based on (9), it can also be shown that [7, Section 2.3.1],

$$\begin{aligned} \text{var}[\hat{J}_{ij}^{(L)}|\boldsymbol{\theta}] &= \sum_{l,m} a_{lm}(i, j) \text{var}[H_{lm}(\boldsymbol{\theta}|\mathbf{Z})|\boldsymbol{\theta}] \\ &+ \sum_{\substack{l,m \\ lm \neq ij}} a_{lm}(i, j) (E[H_{lm}(\boldsymbol{\theta}|\mathbf{Z})|\boldsymbol{\theta}])^2 \\ &+ O(c^2) + O(\tilde{c}^2) + O(c^2\tilde{c}^2), \end{aligned} \quad (12)$$

in which $a_{lm}(i, j) = E[\Delta_m^2/\Delta_j^2]E[\tilde{\Delta}_l^2/\tilde{\Delta}_i^2]$. It should be remarked here that, in deriving (12), it is assumed that all the combinations of covariance terms involving $H_{lm}(\boldsymbol{\theta}|\mathbf{Z})$, $H_{lm,s}(\boldsymbol{\theta}|\mathbf{Z})$ and $H_{lm,rs}(\boldsymbol{\theta}|\mathbf{Z})$, $l, m, s, r = 1, \dots, p$, exist *around* $\boldsymbol{\theta}$ that indicates the point of evaluation of these functions.

Similarly, the expression of $\text{cov}[\hat{J}_{ij}^{(L)}, \hat{J}_{ji}^{(L)}|\boldsymbol{\theta}]$, $j \neq i$, can also be deduced and it is given by [7, Section 2.3.1], $j \neq i$,

$$\begin{aligned} \text{cov}[\hat{J}_{ij}^{(L)}, \hat{J}_{ji}^{(L)}|\boldsymbol{\theta}] &= 2 \left\{ \text{var}[(H_{ij}(\boldsymbol{\theta}|\mathbf{Z}))|\boldsymbol{\theta}] + (E[(H_{ij}(\boldsymbol{\theta}|\mathbf{Z}))|\boldsymbol{\theta}])^2 \right\} \\ &+ 2E[H_{ii}(\boldsymbol{\theta}|\mathbf{Z})H_{jj}(\boldsymbol{\theta}|\mathbf{Z})|\boldsymbol{\theta}] - F_{ij}^2(\boldsymbol{\theta}) + O(c^2) + O(\tilde{c}^2) + O(c^2\tilde{c}^2). \end{aligned} \quad (13)$$

Now, the variance of $\hat{H}_{ij}^{(L)}$, $\text{var}[\hat{H}_{ij}^{(L)}|\boldsymbol{\theta}]$, for $j \neq i$, can be readily obtained from (11) by using (12) and (13). Note that $\text{var}[\hat{H}_{ii}^{(L)}|\boldsymbol{\theta}]$ is same as $\text{var}[\hat{J}_{ii}^{(L)}|\boldsymbol{\theta}]$ that can be directly obtained from (12) by replacing j with i . The contributions of the variance and covariance terms (as appeared in (11)) to $\text{var}[\hat{H}_{ij}^{(L)}|\boldsymbol{\theta}]$ are compared next with the contributions of the respective variance and covariance terms to $\text{var}[\hat{H}_{ij}^{(L)}|\boldsymbol{\theta}]$.

Consider the (i, j) -th element of $\tilde{\mathbf{H}}_k$ associated with (5) that is given by,

$$\tilde{H}_{ij}^{(L)} = \frac{1}{2} \left(\tilde{J}_{ij}^{(L)} + \tilde{J}_{ji}^{(L)} \right), \quad \forall j \in \mathbb{I}_i^c, \quad (14)$$

$$\tilde{H}_{ij}^{(L)} = -F_{ij}(\boldsymbol{\theta}), \quad \forall j \in \mathbb{I}_i. \quad (15)$$

In (14), $\tilde{J}_{ij}^{(L)}$ is defined as,

$$\tilde{J}_{ij}^{(L)} = \hat{J}_{ij}^{(L)} - \sum_l \sum_{m \in \mathbb{I}_i} (-F_{lm}(\boldsymbol{\theta})) \frac{\Delta_m}{\Delta_j} \frac{\tilde{\Delta}_l}{\tilde{\Delta}_i}, \quad \forall j \in \mathbb{I}_i^c. \quad (16)$$

Uses of (14)-(16) now yield [7, Section 2.3.1] $E[\tilde{H}_{ij}^{(L)}|\boldsymbol{\theta}] = E[\hat{H}_{ij}^{(L)}|\boldsymbol{\theta}]$, $\forall j \in \mathbb{I}_i^c$, and $E[\tilde{H}_{ij}^{(L)}|\boldsymbol{\theta}] = -F_{ij}(\boldsymbol{\theta})$, $\forall j \in \mathbb{I}_i$.

While $\text{var}[\tilde{H}_{ij}^{(L)}|\boldsymbol{\theta}] = 0$, $\forall j \in \mathbb{I}_i$, by (15) clearly implying that $\text{var}[\tilde{H}_{ij}^{(L)}|\boldsymbol{\theta}] < \text{var}[\hat{H}_{ij}^{(L)}|\boldsymbol{\theta}]$, $\forall j \in \mathbb{I}_i$, comparing $\text{var}[\tilde{H}_{ij}^{(L)}|\boldsymbol{\theta}]$ to $\text{var}[\hat{H}_{ij}^{(L)}|\boldsymbol{\theta}]$, $\forall j \in \mathbb{I}_i^c$, is the task that will be considered now. In fact, this is the main result associated with the variance reduction from prior information available in terms of the known elements of $\mathbf{F}_n(\boldsymbol{\theta})$.

The first step in determining $\text{var}[\tilde{H}_{ij}^{(L)}|\boldsymbol{\theta}]$ is to note that the expression of $\hat{J}_{ij}^{(L)}$ in (9) can be decomposed into two parts as shown below,

$$\begin{aligned} \hat{J}_{ij}^{(L)} &= \sum_l \left[\sum_{m \in \mathbb{I}_i} H_{lm}(\boldsymbol{\theta}|\mathbf{Z}) \frac{\Delta_m}{\Delta_j} \frac{\tilde{\Delta}_l}{\tilde{\Delta}_i} + \sum_{m \in \mathbb{I}_i^c} H_{lm}(\boldsymbol{\theta}|\mathbf{Z}) \frac{\Delta_m}{\Delta_j} \frac{\tilde{\Delta}_l}{\tilde{\Delta}_i} \right] \\ &+ O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(c^2) + O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(\tilde{c}) + O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(\tilde{c}^2). \end{aligned} \quad (17)$$

The elements, $H_{lm}(\boldsymbol{\theta}|\mathbf{Z})$, of $\mathbf{H}(\boldsymbol{\theta}|\mathbf{Z})$ in the right-hand-side of (17) are not known. However, since by (Hessian-based) definition $E[H_{lm}(\boldsymbol{\theta}|\mathbf{Z})|\boldsymbol{\theta}] = -F_{lm}(\boldsymbol{\theta})$, approximation of the unknown elements of $\mathbf{H}(\boldsymbol{\theta}|\mathbf{Z})$ in the right-hand-side of (17), particularly those elements that correspond to the elements of the FIM that are known *a priori*, by the negative of those elements of $\mathbf{F}_n(\boldsymbol{\theta})$ is the primary idea based on

which the modified resampling algorithm is developed. This approximation introduces an error term, $e_{lm}(\boldsymbol{\theta}|\mathbf{Z})$, that can be defined by, $\forall m \in \mathbb{I}_l, l = 1, \dots, p$,

$$H_{lm}(\boldsymbol{\theta}|\mathbf{Z}) = -F_{lm}(\boldsymbol{\theta}) + e_{lm}(\boldsymbol{\theta}|\mathbf{Z}), \quad (18)$$

and this error term satisfies the following two conditions that directly follow from (18), $\forall m \in \mathbb{I}_l, l = 1, \dots, p$,

$$E[e_{lm}(\boldsymbol{\theta}|\mathbf{Z})|\boldsymbol{\theta}] = 0, \quad (19)$$

$$\text{var}[e_{lm}(\boldsymbol{\theta}|\mathbf{Z})|\boldsymbol{\theta}] = \text{var}[H_{lm}(\boldsymbol{\theta}|\mathbf{Z})|\boldsymbol{\theta}]. \quad (20)$$

Also, introduce $X_{lm}, l = 1, \dots, p$, as defined below,

$$X_{lm}(\boldsymbol{\theta}|\mathbf{Z}) = \begin{cases} e_{lm}(\boldsymbol{\theta}|\mathbf{Z}), & \text{if } m \in \mathbb{I}_l, \\ H_{lm}(\boldsymbol{\theta}|\mathbf{Z}), & \text{if } m \in \mathbb{I}_l^c. \end{cases} \quad (21)$$

Now, substitution of (18) in (17) results in a known part in the right-hand-side of (17) involving the analytically known elements of FIM. This known part is transferred to the left-hand-side of (17) and, consequently, acts as a feedback to the current resampling algorithm yielding, in the process, the expression of $\tilde{J}_{ij}^{(L)}$ in (16). By making use of (21), it can be shown that,

$$\begin{aligned} \tilde{J}_{ij}^{(L)} &= \sum_{l,m} X_{lm}(\boldsymbol{\theta}|\mathbf{Z}) \frac{\Delta_m}{\Delta_j} \frac{\tilde{\Delta}_l}{\tilde{\Delta}_i} + O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(c^2) + O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(\tilde{c}) \\ &+ O_{\tilde{\Delta}, \Delta, \mathbf{Z}}(\tilde{c}^2), \quad \forall j \in \mathbb{I}_i^c. \end{aligned} \quad (22)$$

The variance of $\tilde{J}_{ij}^{(L)}, \forall j \in \mathbb{I}_i^c$, can be computed by using (19)-(22) and, subsequently, it can be shown [7, Section 2.3.1] that, $\forall j \in \mathbb{I}_i^c, i = 1, \dots, p$,

$$\begin{aligned} \text{var}[\hat{J}_{ij}^{(L)}|\boldsymbol{\theta}] - \text{var}[\tilde{J}_{ij}^{(L)}|\boldsymbol{\theta}] &= \sum_l \sum_{m \in \mathbb{I}_l} a_{lm}(i, j) (F_{lm}(\boldsymbol{\theta}))^2 \\ &+ O(c^2) + O(\tilde{c}^2) + O(c^2\tilde{c}^2) > 0. \end{aligned} \quad (23)$$

The inequality above follows from the fact that $a_{lm}(i, j) = (E[\Delta_m^2/\Delta_j^2] E[\tilde{\Delta}_l^2/\tilde{\Delta}_i^2]) > 0, l, m = 1, \dots, p$, for any given (i, j) and assuming that at least one of the known elements, $F_{lm}(\boldsymbol{\theta})$, in (23) is not equal to zero. It must be remarked that the bias terms, $O(c^2), O(\tilde{c}^2)$ and $O(c^2\tilde{c}^2)$, can be made negligibly small by selecting c and \tilde{c} small enough that are primarily controlled by users. Note that if $\Delta_1, \dots, \Delta_p$ and $\tilde{\Delta}_1, \dots, \tilde{\Delta}_p$ are both assumed to be Bernoulli ± 1 i.i.d. random variables, then $a_{lm}(i, j)$ turns out to be unity.

At this point it should be already clear that $\text{var}[\tilde{H}_{ii}^{(L)}|\boldsymbol{\theta}] < \text{var}[\hat{H}_{ii}^{(L)}|\boldsymbol{\theta}]$, if $j = i \in \mathbb{I}_i^c$, by (23).

Next, by using (22), the expression of $\text{cov}[\tilde{J}_{ij}^{(L)}, \tilde{J}_{ji}^{(L)}|\boldsymbol{\theta}]$ can be deduced and, subsequently, it can be concluded that [7, Section 2.3.1],

$$\begin{aligned} &\text{cov}[\hat{J}_{ij}^{(L)}, \hat{J}_{ji}^{(L)}|\boldsymbol{\theta}] - \text{cov}[\tilde{J}_{ij}^{(L)}, \tilde{J}_{ji}^{(L)}|\boldsymbol{\theta}] \\ &= 2(E[H_{ii}(\boldsymbol{\theta}|\mathbf{Z})H_{jj}(\boldsymbol{\theta}|\mathbf{Z})|\boldsymbol{\theta}] - E[X_{ii}(\boldsymbol{\theta}|\mathbf{Z})X_{jj}(\boldsymbol{\theta}|\mathbf{Z})|\boldsymbol{\theta}]) \\ &+ O(c^2) + O(\tilde{c}^2) + O(c^2\tilde{c}^2), \quad j \neq i, \forall j \in \mathbb{I}_i^c. \end{aligned} \quad (24)$$

Here, $E[H_{ii}(\boldsymbol{\theta}|\mathbf{Z})H_{jj}(\boldsymbol{\theta}|\mathbf{Z})|\boldsymbol{\theta}] - E[X_{ii}(\boldsymbol{\theta}|\mathbf{Z})X_{jj}(\boldsymbol{\theta}|\mathbf{Z})|\boldsymbol{\theta}] \geq 0$ that follows from (18) and by the positive-definiteness of $\mathbf{F}_n(\boldsymbol{\theta})$. Using this fact and (23), and also noting that the bias terms, $O(c^2), O(\tilde{c}^2)$ and $O(c^2\tilde{c}^2)$, above can be

made negligibly small by selecting c and \tilde{c} small enough, the following can be immediately concluded,

$$\text{var}[\tilde{H}_{ij}^{(L)}|\boldsymbol{\theta}] < \text{var}[\hat{H}_{ij}^{(L)}|\boldsymbol{\theta}], \quad i, j = 1, \dots, p.$$

In deducing the expressions of mean and variance, several assumptions related to the existences of derivatives of L with respect to $\boldsymbol{\theta}$ and also the existences of expectations of these derivatives are required as ‘hinted’ earlier sporadically. For a complete list of assumptions and rigorous derivation of these expressions, readers are referred to [7].

Since Case 2 is simpler than Case 1 that has already been considered in full possible detail within the limited space, the next section simply presents the final results for Case 2 highlighting $\text{var}[\tilde{H}_{ij}^{(\mathbf{g})}|\boldsymbol{\theta}] < \text{var}[\hat{H}_{ij}^{(\mathbf{g})}|\boldsymbol{\theta}]$, $i, j = 1, \dots, p$.

2) *Case 2 — measurements of \mathbf{g} are available:* As shown for Case 1, it can also be shown here that $\forall i = 1, \dots, p$, $E[\tilde{H}_{ij}^{(\mathbf{g})}|\boldsymbol{\theta}] = E[\hat{H}_{ij}^{(\mathbf{g})}|\boldsymbol{\theta}]$, $\forall j \in \mathbb{I}_i^c$, and $E[\tilde{H}_{ij}^{(\mathbf{g})}|\boldsymbol{\theta}] = -F_{ij}(\boldsymbol{\theta}), \forall j \in \mathbb{I}_i$ [7, Section 2.3.2].

While the difference between $\text{var}[\hat{H}_{ii}^{(\mathbf{g})}|\boldsymbol{\theta}]$ and $\text{var}[\tilde{H}_{ii}^{(\mathbf{g})}|\boldsymbol{\theta}]$ in this case is given by $\forall j \in \mathbb{I}_i^c$ [7, Section 2.3.2],

$$\text{var}[\hat{J}_{ij}^{(\mathbf{g})}|\boldsymbol{\theta}] - \text{var}[\tilde{J}_{ij}^{(\mathbf{g})}|\boldsymbol{\theta}] = \sum_{l \in \mathbb{I}_i} b_l(j) (F_{il}(\boldsymbol{\theta}))^2 + O(c^2) > 0, \quad (25)$$

the difference between the covariance terms is given by [7, Section 2.3.2],

$$\begin{aligned} &\text{cov}[\hat{J}_{ij}^{(\mathbf{g})}, \hat{J}_{ji}^{(\mathbf{g})}|\boldsymbol{\theta}] - \text{cov}[\tilde{J}_{ij}^{(\mathbf{g})}, \tilde{J}_{ji}^{(\mathbf{g})}|\boldsymbol{\theta}] = E[H_{ii}(\boldsymbol{\theta}|\mathbf{Z})H_{jj}(\boldsymbol{\theta}|\mathbf{Z})|\boldsymbol{\theta}] \\ &- E[X_{ii}(\boldsymbol{\theta}|\mathbf{Z})X_{jj}(\boldsymbol{\theta}|\mathbf{Z})|\boldsymbol{\theta}] + O(c^2), \quad j \neq i, \forall j \in \mathbb{I}_i^c. \end{aligned} \quad (26)$$

In (25), $b_l(j) = E[\Delta_l^2/\Delta_j^2] > 0, l = 1, \dots, p$, and it turns out to be unity if $\Delta_1, \dots, \Delta_p$ is assumed to be Bernoulli ± 1 i.i.d. random variables.

Therefore, as for Case 1, we also have the following conclusion for Case 2,

$$\text{var}[\tilde{H}_{ij}^{(\mathbf{g})}|\boldsymbol{\theta}] < \text{var}[\hat{H}_{ij}^{(\mathbf{g})}|\boldsymbol{\theta}], \quad i, j = 1, \dots, p.$$

Finally, since $\hat{\mathbf{H}}_k$ and $\tilde{\mathbf{H}}_k, k = 1, \dots, N$, both are statistically independent across k , it can be concluded straightway for any (i, j) -th element, $i, j = 1, \dots, p$, of $\hat{\mathbf{F}}_n$ and $\tilde{\mathbf{F}}_n$ that,

$$\text{var}[\tilde{F}_{ij}|\boldsymbol{\theta}] = \frac{\text{var}[\tilde{H}_{ij}|\boldsymbol{\theta}]}{N} < \text{var}[\hat{F}_{ij}|\boldsymbol{\theta}] = \frac{\text{var}[\hat{H}_{ij}|\boldsymbol{\theta}]}{N}. \quad (27)$$

V. NUMERICAL ILLUSTRATIONS AND DISCUSSIONS

Consider independently distributed scalar-valued random data \mathbf{z}_i with $\mathbf{z}_i \sim N(\mu, \sigma^2 + c_i\alpha), i = 1, \dots, n$, in which μ and $(\sigma^2 + c_i\alpha)$ are, respectively, mean and variance of \mathbf{z}_i with c_i being some known nonnegative constants and $\alpha > 0$. Here, $\boldsymbol{\theta}$ is considered as $\boldsymbol{\theta} = [\mu, \sigma^2, \alpha]^T$. This is a simple extension of an example problem already considered in literature [3, Example 13.7]. The analytical FIM, $\mathbf{F}_n(\boldsymbol{\theta})$, can be readily determined for this case so that the MC resampling-based estimates of $\mathbf{F}_n(\boldsymbol{\theta})$ can be verified with the analytical FIM. It can be shown that the analytical FIM is given by,

$$\mathbf{F}_n(\boldsymbol{\theta}) = \begin{bmatrix} F_{11} & 0 & 0 \\ 0 & F_{22} & F_{33} \\ 0 & F_{33} & F_{33} \end{bmatrix},$$

Cases	Error in FIM estimates		MSE (variance) reduction
	relMSE($\hat{\mathbf{F}}_n$) [MSE($\hat{\mathbf{F}}_n$)]	relMSE($\tilde{\mathbf{F}}_n$) [MSE($\tilde{\mathbf{F}}_n$)]	
Case 1	0.3815 % [1.9318]	0.0033 % [0.0169]	99.1239 % (97.7817 %)
Case 2	0.0533 % [0.2703]	0.0198 % [0.1005]	62.8420 % (97.5856 %)

TABLE I

MSE AND MSE REDUCTION OF FIM ESTIMATES ($N = 2000$).

in which $F_{11} = \sum_{i=1}^n (\sigma^2 + c_i \alpha)^{-1}$, $F_{22} = (1/2) \sum_{i=1}^n (\sigma^2 + c_i \alpha)^{-2}$ and $F_{33} = (1/2) \sum_{i=1}^n c_i (\sigma^2 + c_i \alpha)^{-2}$. Here, the value of θ , that is used to generate the pseudo data vector (as a proxy for $\mathbf{Z}_n = [\mathbf{z}_1, \dots, \mathbf{z}_n]^T$) and to evaluate $\mathbf{F}_n(\theta)$, is assumed to correspond to $\mu = 0$, $\sigma^2 = 1$ and $\alpha = 1$. The values of c_i across i are chosen between 0 and 1, which are generated by using MATLAB uniform random number generator, `rand`, with a given seed (`rand('state', 0)`). Based on $n = 30$ yields a positive definite $\mathbf{F}_n(\theta)$ whose eigenvalues are given by 0.5696, 8.6925 and 20.7496.

Let us assumed for the sake of illustration that only the upper-left 2×2 block of the analytical FIM is known *a priori*. Using this known information, both the existing [1] and the modified resampling algorithm as proposed in this work are employed to estimate the FIM. For Hessian estimation per (2), c is considered as 0.0001 and, for gradient-approximation per (3), \tilde{c} is considered as 0.00011. Bernoulli ± 1 random variable components are considered to generate Δ_k and $\hat{\Delta}_k$.

The results are summarized in Table I. The mean-squared error (MSE) of $\hat{\mathbf{F}}_n$ and $\tilde{\mathbf{F}}_n$ are first computed; for example, in the case of $\hat{\mathbf{F}}_n$, $\text{MSE}(\hat{\mathbf{F}}_n)$ is computed as $\text{MSE}(\hat{\mathbf{F}}_n) = \sum_{ij} (\hat{F}_{ij} - F_{ij}(\theta))^2$. The relative MSE are computed, for example, in the case of $\hat{\mathbf{F}}_n$, as $\text{relMSE}(\hat{\mathbf{F}}_n) = 100 \times \text{MSE}(\hat{\mathbf{F}}_n) / \sum_{ij} (F_{ij}(\theta))^2$. The effectiveness of the modified resampling algorithm can be clearly seen from the fourth column of the table that shows substantial MSE reduction. The relative MSE reduction in the table is computed as $100 \times (\text{MSE}(\hat{\mathbf{F}}_n) - \text{MSE}(\tilde{\mathbf{F}}_n)) / \text{MSE}(\hat{\mathbf{F}}_n)$. In this column also shown within parentheses are variance reduction. The relative variance reduction are computed as $100 \times (A - B) / A$, in which $A = \sum_{ij} \text{var}[\hat{F}_{ij} | \theta]$ and $B = \sum_{ij} \text{var}[\tilde{F}_{ij} | \theta]$.

It would also be interesting to investigate the effect of the modified resampling algorithm on the MSE reduction in the estimators of the unknown elements of the FIM in contrast to a rather 'naive approach' in which the estimates of the unknown elements are simply extracted from $\hat{\mathbf{F}}_n$. To see the improvement in terms of MSE reduction of the estimators of the *unknown* elements of the FIM, the elements corresponding to the upper-left 2×2 block of $\hat{\mathbf{F}}_n$ obtained from the current resampling algorithm are simply replaced by the corresponding *known* analytical elements of the FIM, $\mathbf{F}_n(\theta)$. Therefore, the results (shown in Table II) only display the contributions of the MSE from the estimators of the unknown elements of FIM. This table clearly reflects the superiority of the modified resampling algorithm as presented in this work over the current resampling algorithm.

Table I-II essentially highlight the substantial improvement of the results (in the sense of MSE reduction as well as

Cases	MSE($\hat{\mathbf{F}}_n$) (and A): naive approach	MSE($\tilde{\mathbf{F}}_n$) (and B)	MSE (variance) reduction
Case 1	0.1288 (0.0159)	0.1021 (0.0006)	20.7235 % (95.9179 %)
Case 2	0.0885 (0.0030)	0.0878 (0.0002)	0.7930 % (94.4222 %)

TABLE II

MSE COMPARISON FOR $\hat{\mathbf{F}}_n$ AND $\tilde{\mathbf{F}}_n$ ONLY FOR THE UNKNOWN ELEMENTS OF $\mathbf{F}_n(\theta)$ ACCORDING TO $\mathbf{F}_n^{\text{GIVEN}}$ ($N = 100000$)

(SIMILAR RESULTS ON VARIANCE ARE REPORTED WITHIN

PARENTHESES, $A = \sum_{i=1}^p \sum_{j \in \mathbb{I}_i^c} \text{var}[\hat{F}_{ij} | \theta]$ AND

$$B = \sum_{i=1}^p \sum_{j \in \mathbb{I}_i^c} \text{var}[\tilde{F}_{ij} | \theta].$$

variance reduction) of the modified MC based resampling algorithm over the results of the current MC based resampling algorithm. Of course, this degree of improvement is controlled by the values of the known elements of the analytical FIM; see (23) and (24) for Case 1 and (25) and (26) for Case 2.

VI. CONCLUSIONS

The present work re-visits the resampling algorithm and computes the variance of the estimator of an arbitrary element of the FIM. A modification in the existing resampling algorithm is proposed simultaneously preserving the known elements of the FIM and improving the statistical characteristics of the estimators of the unknown elements (in the sense of variance reduction) by utilizing the information available from the known elements. The numerical example showed significant improvement of the results (in the sense of MSE reduction as well as variance reduction) of the proposed resampling algorithm over that of the current resampling algorithm.

REFERENCES

- [1] J. C. Spall, "Monte Carlo computation of the Fisher information matrix in nonstandard settings," *J. Comput. Graph. Statist.*, vol. 14, no. 4, pp. 889–909, 2005.
- [2] S. Das, R. Ghanem, and J. C. Spall, "Asymptotic Sampling Distribution for Polynomial Chaos Representation of Data: A Maximum Entropy and Fisher information approach," in *Proc. of the 45th IEEE Conference on Decision and Control*, San Diego, CA, USA, Dec 13-15, 2006, CD rom.
- [3] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation and Control*. Wiley-Interscience, 2003.
- [4] —, "Feedback and weighting mechanisms for improving Jacobian (Hessian) estimates in the adaptive simultaneous perturbation algorithm," in *Proc. of the 2006 American Control Conference*, Minneapolis, Minnesota, USA, June 14-16, 2006, pp. 3086–3091.
- [5] P. Bickel and K. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics, Vol I*. New Jersey: Prentice Hall, 2001.
- [6] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Trans. Automat. Control*, vol. 37, no. 3, pp. 332–341, 1992.
- [7] S. Das, "Efficient calculation of Fisher information matrix: Monte Carlo approach using prior information," Master's thesis, Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, Maryland, USA, May 2007, <http://dspace.library.jhu.edu/handle/1774.2/32459>.