

# A new probabilistic tool for the determination and optimization of multiphase equation of state parameters: Application to tin

E. Fraizier<sup>1,a</sup>, S. Delattre<sup>2</sup>, M. Mougeot<sup>2</sup>, Ph. Faure<sup>1</sup>, G. Roy<sup>1</sup>, and F. Poggi<sup>3</sup>

<sup>1</sup>CEA, DAM, Valduc, 21120 Is-sur-Tille, France

<sup>2</sup>LPMA, UFR de Mathématiques Paris 7, 75013 Paris, France

<sup>3</sup>CEA, DAM, Bruyères-le-Château, 91297 Arpajon, France

**Abstract.** A thermodynamically consistent Equation Of State (EOS) was developed to predict and analyse the behaviour of multiphase metals under shock wave loading. Assuming the Mie-Grüneisen hypothesis together with the Birch (for example) formulation, the EOS gives the relation between pressure  $P$ , temperature  $T$  and atomic volume  $V$ . Experimental data  $(P,V,T)$  for each phase are provided mainly by X-ray diffraction measurements with diamond anvil cells. In this work, mathematical tools are designed to optimize the determination of the EOS parameters and evaluate uncertainty. The general EOS form is  $y = f_{\theta}(x)$  where  $y = P$ ,  $x = (V,T)$  and  $\theta$  is the parameter vector to calibrate. Using experimental data  $(x_i, y_i)$ , the least square (non-linear) regression provides an optimal value  $\theta^*$  for the fit parameters. The measurement errors on  $y$  and  $x$  give biased estimation of  $\theta^*$  with the standard method. Assuming centered and known variance laws for the errors, a statistical procedure is proposed to estimate  $\theta^*$  and determine confidence intervals. Thanks to a Bayesian approach it is possible to introduce physical interval knowledge of the parameters in this procedure. Moreover, various EOS  $f_{\theta^*}$  formulations are evaluated with a chi-squared type statistical test. The present method is applied on experimental data for multi phase tin ( $\beta$  and  $\gamma$  phases and liquid state) in order to provide an optimized multi-phase model. Furthermore, the method is used to design further experimental campaign and to evaluate the gain of new experimental data with the corresponding estimated errors.

## 1 Introduction

In shock waves experiments on multiphase materials, the diagnostics generally measure only the free surface velocities. The pressure  $P$ , specific volume  $V$  and temperature  $T$  need being inferred from hypotheses and calculations. In order to reproduce these velocities and understand the behaviour of metals under these specific loading conditions, hydrocodes and models (in particular Equation of State - EOS) are developed. In previous works, thermodynamically based EOS (Fig. 1) were developed and validated for tin [1–3]. However, input EOS parameters gathered in these works are not always consistent as they originate from various experimental works and / or are determined from different EOS formalisms.

In situ x-ray diffraction experiments in Diamond Anvil Cells (DAC) provide Pressure-Volume-Temperature  $(P, V, T)$  data sets that can lead, through an adjustment process, to the determination of an equilibrium EOS [4–9]. This adjustment is classically obtained by a least square regression on the  $(P, V, T)$  data with a  $P = f(V, T)$  model. With the Mie-Grüneisen hypothesis for each phase the pressure model has the form:

$$P = P_k(V) + \frac{G_0}{V_0}(T - T_0) \quad (1)$$

The first term  $P_k(V)$  describes a reference isothermal (temperature  $T_0$ ) pressure, for instance with the Birch

formulation [10]:

$$P_k(V) = P_0 + \frac{3}{2}K_0 \left[ \left( \frac{V}{V_0} \right)^{-\frac{7}{3}} - \left( \frac{V}{V_0} \right)^{-\frac{5}{3}} \right] \\ \times \left[ 1 - \frac{3}{4}(4 - N_0) \left[ \left( \frac{V}{V_0} \right)^{-\frac{2}{3}} - 1 \right] \right] \quad (2)$$

where  $G_0 = \Gamma_0 C_{v_0}$  with  $C_{v_0}$  is the specific heat constant,  $\Gamma_0$  is the Grüneisen coefficient,  $K_0$  is the isothermal modulus and  $N_0$  is its first pressure derivative. Concerning the units,  $T$  is in Kelvin,  $P$  is in GPa,  $V$  is in  $\text{mm}^3 \text{kg}^{-1}$ , and  $G_0$  is in  $\text{Jkg}^{-1}\text{K}^{-1}$ .  $N_0$  does not have any unit.  $(T_0, P_0)$  defines the reference condition. It must be noted that experiments under extreme conditions of pressure and temperature are often difficult to perform and therefore data on the material of interest is often scarce, if not lacking. It is interesting, before planning a new experiment in order to increase the accuracy of a multiphase EOS, to evaluate the potential benefit of this experiment, taking into account the foreseen data accuracy and the amount of data to be recorded. For this purpose, some new mathematical tools are developed. These tools should give new information on Birch, Murnaghan [11], Vinet [12] and some others different [13–15] formulations for each phase. Furthermore, taking into account the phase transitions determined from experimental data and from the thermodynamically based EOS we plan to optimize each parameter (for each phase). This paper deals with the preliminary steps of this work. The methodology and the mathematical tools principles are detailed. The first results from literature data on tin [5, 6] present the determination of the EOS parameters for

<sup>a</sup> e-mail: emmanuel.fraizier@cea.fr

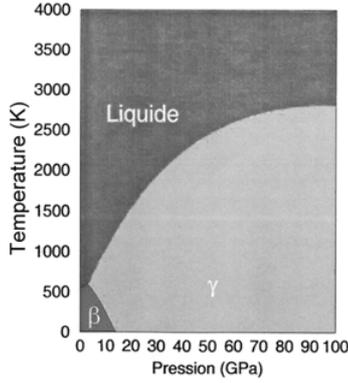


Fig. 1. Multiphase diagram for tin [3].

$\beta$  and  $\gamma$  phases. The maximum uncertainties are used for every (P,V,T) data [5,6] but, later, the uncertainties should be defined for each point.

## 2 Methodology description

This work presents mathematical tools designed for optimizing the determination of the EOS parameters and evaluate the associated errors. This paragraph gives a first overview of the mathematical methodology. The complete mathematical methodology is postponed at the end of the document (see part 5). The general EOS form is  $Y = f_{\vartheta}(X)$  where the output is  $y = P$ , the input vector is  $x = (V, T)$ , and  $\vartheta$  is the vector of parameters to be adjusted to best fit the experimental data  $(x_i, y_i)$ ,  $i = 1, \dots, n$ . This adjustment is classically obtained by a least square regression when the input vector  $x$  has no measurement error and the output quantity  $y$  may have measurement error. Indeed, one can prove that the parameters vector  $\vartheta_n^*$  which minimizes the quadratic misfit, which is the sum of the squared residuals  $(y_i - f_{\vartheta}(x_i))^2$ , is a consistent estimator of  $\vartheta$ , which means that  $\vartheta_n^*$  tends to the real  $\vartheta$  when  $n$  tends to infinity. The problem is more difficult when the input vector  $x$  has measurement error. In this paper, we propose an estimate of  $\vartheta$  that takes into account the measurement errors of both  $x$  and  $y$  and that has a reduced bias compared to the standard least square estimator. This estimator  $\vartheta_n$  is the  $\vartheta$  which minimizes  $K(\vartheta)$  (see part 5).

From this expression, we can study the estimation error distribution and we can deduce a confidence interval for  $\vartheta$ , which is a direct effect of the measurement errors on  $x$  and  $y$ . Finally, we study the possibility of taking an *a priori* knowledge into account. Then we propose a new estimator of  $\vartheta$ , whose expression is given in part 5.

Moreover, a goodness of fit test is proposed to answer to the question whether the EOS formulation is acceptable or not. A statistical  $P$ -value is computed. This quantity follows a uniform law on  $[0, 1]$  if the EOS formulation is right and is close to 0 if it is not the case.

These mathematical tools are applied to calibrate the studied EOS for tin from available experimental data [5,6].

Table 1.  $\beta$  phase, *a priori*.

	$V_0$	$K_0$	$N_0$	$G_0$
$\mu$	137230	54.73	5.75	476.7
$\sigma^2$	$10^5$	25	1	2885

Table 2.  $\beta$  phase, parameter estimation using raw data and then with *a priori* knowledge.

	$V_0$	$K_0$	$N_0$	$G_0$
raw data	137204.5	54.37	3.26	482.37
<i>a priori</i>	137218.3	53.98	3.38	482.70

Table 3.  $\beta$  phase, approximate correlation matrix.

	$V_0$	$K_0$	$N_0$	$G_0$
$V_0$	1.000	-0.78	0.49	-0.29
$K_0$	-0.78	1.000	-0.90	-0.02
$N_0$	0.49	-0.90	1.000	0.10
$G_0$	-0.29	-0.02	0.10	1.000

## 3 Application

We study here the EOS given by the Birch model assuming the Mie-Gruneisen hypothesis defined by (1) and (2). The parameters  $(K_0, N_0, G_0)$  should be estimated given some set of experimental data. A R program has been developed ([www.r-project.org](http://www.r-project.org)) to estimate the parameters using the proposed methodology. The experimental data correspond to non isotherm experiments for  $\beta$  [5], and  $\gamma$  phases [6].

### 3.1 $\beta$ phase

In [5], the  $\beta$  phase is characterized by  $n = 31$  observations where the temperature ranges from 298 to 529 K and the pressure ranges from 0 to 8.5 GPa. The reference state is defined by  $(P_0, T_0) = (0, 300)$ . The standard deviations for the measurement errors are  $\sigma_P = 0.13$ ,  $\sigma_V = 125$ , and  $\sigma_T = 2.5$ .

The *a priori* knowledge [1] on each parameter is introduced using a Gaussian law  $\mathcal{N}(\mu, \sigma^2)$  defined in Table 1.

The set of parameters is estimated using the raw data without introducing any *a priori* knowledge on the parameters and then using the *a priori*. The estimation results are presented in Table 2.

Based on the mathematical analysis, the error distribution on the parameters is computed with the *a priori* (cf. paragraph 5.4). Table 3 shows the approximate correlation matrix of the estimation errors, using (4) of section 5.5. We observe strong correlations for some parameters as for  $(K_0, N_0)$  and  $(K_0, V_0)$ .

The standard deviations are  $S_{V_0} = 157.31$ ,  $S_{K_0} = 1.80$ ,  $S_{N_0} = 0.45$ ,  $S_{G_0} = 36.40$ .

We have conducted a Monte Carlo study to estimate the correlation matrix in parallel (Table 4).

It should be stressed that both approaches bring similar results. The approximation given by 4 of section 5.5 seems

**Table 4.** Phase  $\beta$ , Monte-Carlo correlation matrix,  $K = 1000$  replications.

	$V_0$	$K_0$	$N_0$	$G_0$
$V_0$	1.00	-0.77	0.48	-0.27
$K_0$	-0.77	1.00	-0.90	-0.04
$N_0$	0.48	-0.90	1.00	0.11
$G_0$	-0.27	-0.04	0.11	1.00

**Table 5.**  $\gamma$  phase, *a priori*.

	$V_0$	$K_0$	$N_0$	$G_0$
$\mu$	118385	94	4.88	411.6
$\sigma^2$	$10^{10}$	25	1	2852

**Table 6.**  $\gamma$  phase, parameter estimation using raw data and with a priori knowledge.

	$V_0$	$K_0$	$N_0$	$G_0$
raw data	118248.3	93.23	5.53	564.32
a priori	118264.8	93.04	5.55	554.60

**Table 7.**  $\gamma$  phase, approximate correlation matrix.

	$V_0$	$K_0$	$N_0$	$G_0$
$V_0$	1.00	-0.28	-0.15	-0.66
$K_0$	-0.28	1.00	0.50	0.38
$N_0$	-0.15	0.50	1.00	-0.26
$G_0$	-0.66	0.38	-0.26	1.00

to be justified. The  $P$ -value to test the physical model equals 0.016. The goodness of the EOS formulation is not rejected at level 1%.

### 3.2 $\gamma$ phase

In [5], the  $\gamma$  phase is characterized by  $n = 38$  available observations where the temperature ranges from 298 K to 713 K and the pressure ranges from 4.49 to 16.15 GPa. The reference state is defined by  $(P_0, T_0) = (9.40, 300)$ .

The standard deviations for the measurement errors are  $\sigma_P = 0.24$ ,  $\sigma_V = 116$ , and  $\sigma_T = 4.2$ .

The a priori knowledge on each parameter is introduced using a Gaussian law  $\mathcal{N}(\mu, \sigma^2)$  defined in Table 5. The set of parameters is estimated using the raw data without introducing any a priori knowledge on the parameters and then using the a priori. The estimation results are presented in Table 6.

Table 7 shows the estimated correlation matrix.

The respective standard deviation are  $S_{V_0} = 93.39$ ,  $S_{K_0} = 1.61$ ,  $S_{N_0} = 0.54$ ,  $S_{G_0} = 37.24$ . As for the  $\beta$  phase, we have conducted a Monte Carlo study to estimate the correlation matrix in parallel (Table 8).

We observe similar correlation matrices computed with the mathematical approach or with Monte-Carlo. The  $P$ -value to test the physical model equals 0.027. The goodness of the EOS formulation is not rejected at level 1%.

**Table 8.**  $\gamma$  phase, Monte-Carlo correlation matrix,  $K = 1000$  replications.

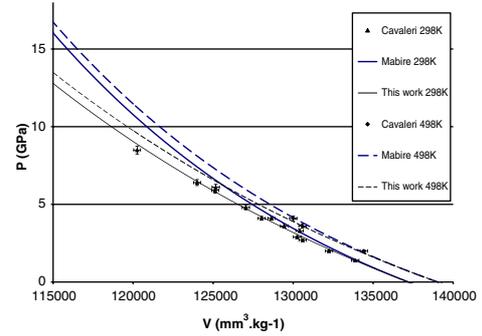
	$V_0$	$K_0$	$N_0$	$G_0$
$V_0$	1.00	-0.28	-0.14	-0.68
$K_0$	-0.28	1.00	0.50	0.34
$N_0$	-0.14	0.50	1.00	-0.25
$G_0$	-0.68	0.34	-0.25	1.00

**Table 9.** Literature and current work parameters for  $\beta$  tin.

$\beta$ phase	Literature	This work:	
		Value	Std
$V_0$ ( $\text{mm}^3 \cdot \text{kg}^{-1}$ )	137230	137218	157
$K_0$ (GPa)	54.73	53.98	1.8
$N_0$	5.75	3.38	0.45
$G_0$ ( $\text{J} \cdot \text{kg}^{-1} \cdot \text{K}^{-1}$ )	476.7	482.7	36.4

**Table 10.** Literature and current work parameters for  $\gamma$  tin.

$\gamma$ phase	Literature	This work:	
		Value	Std
$V_0$ ( $\text{mm}^3 \cdot \text{kg}^{-1}$ )	118385	118265	93.4
$K_0$ (GPa)	94	93.03	1.61
$N_0$	4.88	5.55	0.54
$G_0$ ( $\text{J} \cdot \text{kg}^{-1} \cdot \text{K}^{-1}$ )	411.6	554.3	37.2


**Fig. 2.** Isothermal Birch regression for  $\beta$  tin and experimental data [5].

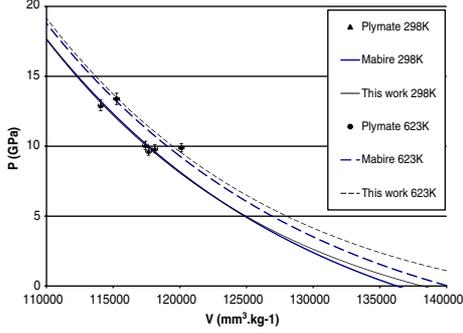
### 3.3 Isothermal Birch regressions for $\beta$ and $\gamma$ phases

The Table 9 summarizes the parameters from literature [1] previously used for  $\beta$  phase and the present results. Then, the isothermal Birch regressions for data tin at 298 K and 498 K are computed and represented on Fig. 2 with the corresponding experimental data from [1].

The Table 10 summarizes the parameters from literature [5] previously used for  $\gamma$  phase and the present results.

Then, the isothermal Birch regressions for  $\gamma$  phase at 298 K and 628 K are computed and represented on Fig. 3 with the corresponding experimental data from [6].

The present work regression fits the experimental data with a better accuracy than previously achieved with literature parameters, in particular for phase. It may be explained by our dedicated optimization for these particular



**Fig. 3.** Isothermal Birch regression for  $\gamma$  tin and experimental data [6].

data [5,6] and Birch formulation whereas the literature parameters integrate several experimental data.

## 4 Conclusion, benefits and perspectives

A first step of a new tool development to determine the multiphase EOS parameters is reached. The resulting parameters enable to compute isothermal Birch regression in good agreement with experimental data. We can now integrate more experimental data and test different  $P_k(V)$  formulations. Furthermore, the literature data can be used as an a priori on data. This tool will be used to determine the benefits of new experimental campaign or new diagnostics including the uncertainties. The matrix correlation of the  $V_0$ ,  $K_0$ ,  $N_0$  and  $G_0$  parameters is estimated and gives further information. A Monte Carlo study, conducted in parallel, shows similar results for this matrix and validates our approach. Furthermore, a P-value quantifies the goodness of physical model. Then, this new tool gives the opportunity to test different  $P_k(V)$  formulations. The last perspective is to optimize every parameters of each phase, taking into account the experimental transitions and our thermodynamically based EOS model.

## 5 Annex: Mathematical results

### 5.1 Statistical model

Consider two physical variables  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$  and  $y \in \mathbb{R}$  linked by the following equation  $y = \phi(\mathbf{x})$  where  $\phi$  is a function from  $\mathbb{R}^d$  to  $\mathbb{R}$ . An equation of state formulation is defined by a family of functions on  $\mathbb{R}^d$ :

$$f_\vartheta : \mathbb{R}^d \rightarrow \mathbb{R}, \quad \vartheta \in \Theta \subset \mathbb{R}^k$$

where  $\vartheta = (\vartheta_1, \vartheta_2, \dots, \vartheta_k)$  corresponds to the physical parameters. The EOS formulation is relevant if  $\varphi$  is mostly equal to  $f_{\vartheta^*}$  for a given value  $\vartheta^* \in \Theta$ . Considering  $n$  physical points  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$ , an approximation of  $\vartheta^*$  is given by  $\vartheta_n^*$  satisfying  $\kappa(\vartheta_n^*) = \min_{\vartheta \in \Theta} \kappa(\vartheta)$  where  $\kappa(\vartheta) = n^{-1} \sum_{i=1}^n (y_i - f_\vartheta(\mathbf{x}_i))^2$ . If the EOS formulation is relevant one can expect  $\vartheta_n^* = \vartheta^*$  for  $n \geq k$ , as in the Birch formulation for instance.

The physical points  $(\mathbf{x}_i, y_i)$  are only known through experimental measurements  $(\mathbf{X}_i, Y_i)$ . The measurement errors

are modeled as follows:

- $\mathbf{X}_i \sim \mathcal{N}(\mathbf{x}_i, \sigma_{\mathbf{X}}^2)$  is a Gaussian random vector with mean  $\mathbf{x}_i$  and covariance matrix  $\sigma_{\mathbf{X}}^2$  (typically a diagonal matrix).
- $Y_i \sim \mathcal{N}(y_i, \sigma_Y^2)$ .
- The random vectors  $\mathbf{X}_i, Y_i, 1 \leq i \leq n$  are supposed to be independent.

### 5.2 Estimation

The least square estimation  $\hat{\vartheta}$  of  $\vartheta^*$  is defined by  $K(\hat{\vartheta}) = \min_{\vartheta \in \Theta} K(\vartheta)$  where  $K(\vartheta) = n^{-1} \sum_{i=1}^n (Y_i - f_\vartheta(\mathbf{X}_i))^2$ . **A<sub>1</sub>**: The function  $(\vartheta, x) \mapsto f_\vartheta(x)$  is supposed to be sufficiently smooth.  $\Theta$  is a compact set. There exists a unique  $\vartheta_n^* \in \Theta$  such that  $\kappa(\vartheta_n^*) = \min_{\vartheta \in \Theta} \kappa(\vartheta)$ .

**Theorem 1** Assume **A<sub>1</sub>**. As  $\sigma_{\mathbf{X}}^2 \rightarrow 0$  and  $\sigma_Y^2 \rightarrow 0$ , the distance between  $\hat{\vartheta}$  and  $\vartheta_n^*$  tends to 0 in probability:

$$\text{for all } \varepsilon > 0, \quad P(\|\hat{\vartheta} - \vartheta_n^*\| > \varepsilon) \rightarrow 0.$$

### 5.3 Estimation error distribution

**Notation:** For  $\vartheta = (\vartheta_1, \dots, \vartheta_k)$  and  $\mathbf{x} = (x_1, \dots, x_d)$ , define the column vectors

$$\dot{f}_\vartheta(\mathbf{x}) = \left( \frac{\partial}{\partial \vartheta_i} f_\vartheta(\mathbf{x}) \right)_{1 \leq i \leq k}, \quad f'_\vartheta(\mathbf{x}) = \left( \frac{\partial}{\partial x_i} f_\vartheta(\mathbf{x}) \right)_{1 \leq i \leq d},$$

the  $k \times k$  matrix

$$\dot{f}_\vartheta(\mathbf{x}) = \left( \frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} f_\vartheta(\mathbf{x}) \right)_{1 \leq i, j \leq k}$$

and the  $d \times k$  matrix

$$\dot{f}'_\vartheta(\mathbf{x}) = \left( \frac{\partial^2}{\partial x_i \partial \vartheta_j} f_\vartheta(\mathbf{x}) \right)_{1 \leq i \leq d, 1 \leq j \leq k}.$$

If  $\mathbf{M}$  is a matrix,  $\mathbf{M}^\top$  denotes its transpose.

We introduce the assumption:

**A<sub>2</sub>**: **A<sub>1</sub>** holds true and  $\vartheta_n^*$  belongs to the interior of  $\Theta$ . Moreover the Hessian matrix of  $\kappa$  at  $\vartheta_n^*$  is positive definite. Note that the Hessian matrix of  $\kappa$  at  $\vartheta_n^*$  is equal to

$$\Lambda = n^{-1} \sum_{i=1}^n \dot{f}_{\vartheta_n^*}(\mathbf{x}_i) \dot{f}_{\vartheta_n^*}(\mathbf{x}_i)^\top - n^{-1} \sum_{i=1}^n (y_i - f_{\vartheta_n^*}(\mathbf{x}_i)) \dot{f}'_{\vartheta_n^*}(\mathbf{x}_i).$$

**Theorem 2** Assume **A<sub>2</sub>**. As  $\sigma_{\mathbf{X}}^2 \rightarrow 0$  and  $\sigma_Y^2 \rightarrow 0$ ,

$$\hat{\vartheta} - \vartheta_n^* = n^{-1/2} \Lambda^{-1} \Gamma^{1/2} Z + O(\|\sigma_{\mathbf{X}}^2\|) \quad (3)$$

where:

- $\|\sigma_{\mathbf{X}}^2\|$  denotes the largest eigenvalue of  $\sigma_{\mathbf{X}}^2$
- $Z$  is a  $k$ -dimensional random vector whose distribution converges to  $\mathcal{N}(0, \mathbf{I}_k)$  where  $\mathbf{I}_k$  denotes the  $k \times k$  identity matrix.

–  $\Gamma$  is the  $k \times k$  matrix

$$\frac{1}{n} \sum_{i=1}^n \left[ \sigma_Y^2 \dot{f}_{\vartheta_n^*}(\mathbf{x}_i) \dot{f}_{\vartheta_n^*}(\mathbf{x}_i)^\top + \mathbf{M}(y_i, \mathbf{x}_i, \vartheta_n^*) \sigma_X^2 \right. \\ \left. \times \mathbf{M}(y_i, \mathbf{x}_i, \vartheta_n^*)^\top \right]$$

$$\text{with } \mathbf{M}(y_i, \mathbf{x}_i, \vartheta_n^*) = (y_i - f_{\vartheta_n^*}(\mathbf{x}_i)) \dot{f}_{\vartheta_n^*}(\mathbf{x}_i)^\top - \dot{f}_{\vartheta_n^*}(\mathbf{x}_i) f'_{\vartheta_n^*}(\mathbf{x}_i)^\top.$$

When the second term of (3) is negligible, the first term provides the probability distribution of the estimation error, namely  $\mathcal{N}(0, \Sigma/n)$  with  $\Sigma = \Lambda^{-1} \Gamma \Lambda^{-1}$ , which order of magnitude is less than  $n^{-1/2}(\|\sigma_X^2\|^{1/2} + \sigma_Y)$ .

*Remark 1* For a large number  $n$  of observations and when  $\sigma_X^2$  is known, the standard least square contrast  $K(\vartheta)$  can be successfully replaced by

$$\frac{1}{n} \sum_{i=1}^n \left[ (Y_i - f_\vartheta(\mathbf{X}_i))^2 - f'_\vartheta(\mathbf{X}_i)^\top \sigma_X^2 f'_\vartheta(\mathbf{X}_i) \right]$$

which provides an estimation of  $\vartheta_n^*$  with a reduced bias, still of order  $O(\|\sigma_X^2\|)$ , when  $f_\vartheta(\mathbf{x})$  is a bilinear function of  $\vartheta$  and  $\mathbf{x}$ .

#### 5.4 Confidence interval

We construct confidence intervals for the component  $\vartheta_{n,i}^*$  of  $\vartheta_n^*$  and more generally for  $u^\top \vartheta_n^*$  where  $u$  is a vector of  $\mathbb{R}^k$ . From Theorem 2, if  $u^\top \Sigma u \ll \|\sigma_X^2\|$  then

$$n^{1/2}(u^\top \hat{\vartheta}_n - u^\top \vartheta_n^*) \sigma \mathcal{N}(0, u^\top \Sigma u).$$

Moreover  $u^\top \Sigma u$  can be approximated by  $u^\top \hat{\Sigma} u$  where  $\hat{\Sigma}$  is the empirical version of  $\Sigma$ , replacing  $(x_i, y_i, \vartheta_n^*)$  by  $(X_i, Y_i, \hat{\vartheta})$ . Finally we obtain that

$$n^{1/2}(u^\top \hat{\vartheta}_n - u^\top \vartheta_n^*) / (u^\top \hat{\Sigma} u)^{1/2} \sim \mathcal{N}(0, 1).$$

This gives directly confidence intervals for  $u^\top \hat{\vartheta}_n$ .

#### 5.5 Use of a priori knowledge

We propose here a method which takes benefit of a previous estimation of  $\vartheta_n^*$ . Assume that we have some information on the first  $p$  components of  $\vartheta_{n,1}^*, \dots, \vartheta_{n,p}^*$  of  $\vartheta_n^* = (\vartheta_{n,1}^*, \vartheta_{n,2}^*, \dots, \vartheta_{n,k}^*)$  ( $1 \leq p \leq k$ ), that is we observe a  $p$ -dimensional random vector  $\hat{\tau}$ , independent of  $\hat{\vartheta}$ , following the law  $\mathcal{N}(\tau_n^*, \tilde{\Sigma})$  where  $\tau_n^* = (\vartheta_{n,1}^*, \vartheta_{n,2}^*, \dots, \vartheta_{n,p}^*)$  and  $\tilde{\Sigma}$  is a given  $p \times p$  covariance matrix (positive definite).

$\hat{\vartheta}$  has been previously computed and satisfy

$$\hat{\vartheta} - \vartheta_n^* \sim \mathcal{N}(0, \hat{\Sigma}/n).$$

The new estimator  $\tilde{\vartheta}_n$  we can compute is the “best” linear combination of  $\hat{\vartheta}$  and  $\hat{\tau}$ :

$$\tilde{\vartheta} = (n\hat{\Sigma}^{-1} + \tilde{\Gamma})^{-1} \left[ n\hat{\Sigma}^{-1} \hat{\vartheta}_n + \begin{pmatrix} \tilde{\Sigma}^{-1} \hat{\tau} \\ 0 \end{pmatrix} \right]$$

where  $\tilde{\Gamma} = \begin{pmatrix} \tilde{\Sigma}^{-1} & 0 \\ 0 & 0 \end{pmatrix}$  is a  $k \times k$  matrix. The estimator  $\tilde{\vartheta}$  is approximatively Gaussian, with non bias and covariance matrix equal to

$$(n\hat{\Sigma}^{-1} + \tilde{\Gamma})^{-1}. \quad (4)$$

Note that  $\tilde{\vartheta}$  is always as good as  $\hat{\vartheta}$  since  $(n\hat{\Sigma}^{-1} + \tilde{\Gamma})^{-1} \leq \hat{\Sigma}/n$ .

#### 5.6 Feedback and test on the physical model

Can we detect whether the EOS formulation is good or not? In other words are the  $y_i - f_{\vartheta_n^*}(\mathbf{x}_i)$ ,  $1 \leq i \leq n$ , significantly not equal to zero?

Introduce the log-likelihood contrast: for  $\vartheta \in \Theta$  and  $n$  vectors  $\xi_1, \dots, \xi_n \in \mathbb{R}^d$  set

$$L(\vartheta, \xi_1, \dots, \xi_n) = \sum_{i=1}^n \left[ \frac{(Y_i - f_\vartheta(\xi_i))^2}{\sigma_Y^2} \right. \\ \left. + (\mathbf{X}_i - \xi_i)^\top (\sigma_X^2)^{-1} (\mathbf{X}_i - \xi_i) \right].$$

**A<sub>3</sub>:** **A<sub>1</sub>** holds.  $\vartheta_n^*$  belongs to the interior of  $\vartheta$ . Moreover  $y_i = f_{\vartheta_n^*}(\mathbf{x}_i)$  for  $i = 1, \dots, n$ .

**Theorem 3** Assume **A<sub>3</sub>**. As  $\sigma_Y^2 \rightarrow 0$  and  $\sigma_X^2 \rightarrow 0$ , the distribution of the random variable

$$D = \min_{\vartheta, \xi_1, \dots, \xi_n} L(\vartheta, \xi_1, \dots, \xi_n)$$

converges to the chi-square distribution with  $n - k$  degrees of freedom.

Therefore, if  $t$  denotes the computed value of  $D$ , the probability  $P(\chi_{n-k}^2 > t)$  is an approximate  $P$ -value for testing if the EOS formulation is relevant (the null hypothesis here).

#### References

1. C. Mabire, Ph.D. thesis, Poitiers Univ. France (1999)
2. C. Mabire, P. Hereil, Shock Compression of Condensed Matter (2000)
3. F. Buy, C. Voltz, F. Llorca, Shock Compression of Condensed Matter (2005)
4. M. Liu, L. Liu, High Temp. - High Pressures **18**, 78 (1986)
5. M. Cavaleri, T. Plymate, J. Stout, J. Phys. Chem. Solids **8**(49), 945 (1988)
6. T. Plymate, J. Stout, C. M.E., J. Phys. Chem. Solids **49**(11) (1988)
7. N. Rambert, P. Faure, B. Sitaud, 40th European High-Pressure Meeting, Edimbourg (2002)
8. R.J. Angel, in *High-pressure and high-temperature crystal chemistry* (2000), Vol. 41, pp. 35–60, r. hazen and r downs editors edn.
9. R.J. Angel, in *High-pressure crystallography* (NATO science series, Kluwer Academic Publishers, 2004), a. katrusiak and p. mcmillan edn.

10. F. Birch, J. Geophys. Res. **83**, 1257 (1978)
11. F.D. Murnaghan, *Finite Deformation of an Elastic Solid* (Wiley, New York, 1951)
12. P. Vinet et al, Phys. Rev. B **35** (1986)
13. P. Bose Roy, J. Bose Roy, J. Phys: Condens. Matter **17** (2005)
14. P. Bose Roy, J. Bose Roy, J. Phys: Condens. Matter **18** (2006)
15. F. Stacey, Reports of Progress in Physics **68** (2005)