



## Comparison of methods for multivariate moment inversion—Introducing the independent component analysis<sup>☆</sup>

Jovani L. Favero<sup>a</sup>, Luiz Fernando L.R. Silva<sup>b</sup>, Paulo L.C. Lage<sup>a,\*</sup>

<sup>a</sup> Programa de Engenharia Química, COPPE, Universidade Federal do Rio de Janeiro, P.O. Box 68502, Rio de Janeiro, RJ 21941-972, Brazil

<sup>b</sup> Escola de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ 21941-909, Brazil



### ARTICLE INFO

#### Article history:

Received 29 April 2013

Received in revised form 23 July 2013

Accepted 10 August 2013

Available online 21 August 2013

#### Keywords:

Population balance

Numerical analysis

Optimization

Mathematical modeling

Multivariate moment inversion

Independent component analysis

### ABSTRACT

Despite the advantages of the moment based methods in solving multivariate population balances models, these methods still suffer with the so-called multivariate moment inversion problem. Although univariate moment inversion is achieved without major problems this is not true for multivariate cases, for which there is no well established methodology. This work presents a comparative analysis of the existing methods regarding their accuracy and robustness. A new moment inversion method based on the independent component analysis was proposed and analyzed. Improvements in accuracy and robustness were achieved by combination of different moment inversion methods.

© 2013 Elsevier Ltd. All rights reserved.

### 1. Introduction

The simulation of polydisperse multiphase flows depends on the development of robust and accurate methods to solve the population balance equation (PBE) coupled to a computational fluid dynamics simulation (PB-CFD) (Ramkrishna, 2000; Silva & Lage, 2011; Zucca, Marchisio, Vanni, & Barresi, 2007).

There are several different numerical approaches for solving the PBE, for example, Monte Carlo stochastic methods (Irizarry, 2008; Krallis, Meimarglou, & Kiparissides, 2008; Meimarglou & Kiparissides, 2007), weighted residuals methods (WRM) (Hulbert & Akiyama, 1969; Ramkrishna, 1971; Subramanianand & Ramkrishna, 1971), methods of classes (MoC) (Hill & Ng, 1996; Kumar & Ramkrishna, 1996; Lister, Smit, & Hounslow, 1995) and moment based methods (MoM) (Hulbert & Katz, 1964; Marchisio & Fox, 2005; McGraw, 1997). Also, it is not difficult to find studies comparing some of the existing methods (Attarakih, Bart, & Faqir, 2006; Lemanowicz & Gierczycki, 2010; Silva, Rodrigues, Mitre, & Lage, 2010), but in fact, there is still no robust method to

accurately solve the PBE including all possible phenomena that can be modeled in it.

Because of the complexity involved to solve the multivariate PBE, most of published works consider only approximate univariate models. However, this approximation is not adequate for all problems. There are cases where it is necessary to take into account more than one variable in order to have a good description of the particulate system behavior (Ramkrishna, 2000). The number of variables needed depends on the application. Usually, particle volume, surface area, temperature, concentration of different components, among others, could be necessary to model the dynamics of the particulate system.

Due to their easy of application, reasonable accuracy and moderate computational cost, the moment based methods are preferable to couple the PBE solution with CFD codes for polydisperse multiphase flow simulations. The moment based methods solve for a selected set of lower order moments of the particle distribution function. These methods depend on the solution of the finite moment inversion problem, that is, the ability of determining a Gauss–Christoffel quadrature rule from this set of lower order moments that can accurately evaluate them. For example, the QMoM (McGraw, 1997) applies the inversion to evolve the moment equations, the DQMoM (Marchisio & Fox, 2005) may use it to generate its initial condition and the DuQMoGeM (Favero & Lage, 2012; Lage, 2011) needs it for the PB-CFD coupling.

For univariate problems, moment inversion can be successfully performed by the modified Chebyshev algorithm (Gautschi, 1994)

<sup>☆</sup> This article is dedicated to Professor Alberto Luiz Coimbra, in the 50th anniversary of COPPE (1963–2013), the Graduate School of Engineering of the Federal University of Rio de Janeiro.

\* Corresponding author. Tel.: +55 21 2562 8346; fax: +55 21 2562 8300.

E-mail address: [paulo@peq.coppe.ufrj.br](mailto:paulo@peq.coppe.ufrj.br) (P.L.C. Lage).

URL: <http://www.peq.coppe.ufrj.br/pesquisa/tfd/> (P.L.C. Lage).

## Nomenclature

<b>D</b>	eigenvalue diagonal matrix
<b>e</b>	unit vector
<b>f</b>	distribution function
$F_{obj}$	objective function
$G$	moment generating function
$g$	Gaussian distribution
<b>H</b>	Hotelling transformation matrix
$h$	problem dimensionality
$I$	mutual information
$J$	negentropy
<b>K</b>	cumulant tensor
$K$	cumulant of a multivariate distribution
<b>M</b>	mixture matrix
<b>S</b>	separation matrix
<b>Q</b>	a possible joint diagonalization matrix
<b>R</b>	matrix that optimize the contrast function
$s$	total order of a multivariate moment
<b>T</b>	moment tensor
$T$	moment of a multivariate distribution
<b>W</b>	whitening matrix
$\mathbf{x}$	vector of variables in the original coordinate frame
$\tilde{\mathbf{x}}$	vector of zero mean variables
$\mathbf{y}$	vector of variables in the transformed coordinate frame

### Greek letters

$\delta$	Dirac delta function
$\epsilon$	error
$\kappa$	cumulant of a univariate distribution
$\lambda$	eigenvalue
$\mu$	mean vector
$\mu$	moment
$\tilde{\mu}$	centered moment
$\omega$	quadrature weight
$\Psi$	contrast function
$\Phi$	cumulant generating function
$\Sigma$	covariance matrix
$\Xi$	differential entropy

### Abbreviations

BSS	blind source separation
CLT	Central Limit Theorem
CQMoM	Conditional Quadrature Method of Moments
CuBICA	Cumulant Based ICA
DPCM	Direct Product Cartesian Method
DQMoM	Direct Quadrature Method of Moments
DuQMoGeM	Dual-Quadrature Method of Generalized Moments
ICA	Independent Component Analysis
JADE	Joint Approximate Diagonalization of Eigenmatrices
KLT	discrete Karhunen–Loève transform
MoC	Method of Classes
TPM	Tensor Product Method
MoM	Method of Moments
PBE	Population Balance Equation
PCA	Principal Component Analysis
PDA	Product-Difference Algorithm
PDF	Probability Density Function
POD	Proper Orthogonal Decomposition
QMoM	Quadrature Method of Moments
SHIBBS	Shifted Blocks for Blind Separation
SVD	Singular Value Decomposition
WRM	Weighted Residuals Method

or the product-difference algorithm (PDA) (Gordon, 1968). These procedures are direct and accurate, consisting basically on the solution of an eigenvalue problem. John and Thein (2012) compared these methods, recommending the modified Chebyshev algorithm due to its better robustness. For multivariate problems, these algorithms are no longer directly applicable and there is no direct approach to solve the finite multivariate moment inversion problem, which is still an open problem.

Wright, McGraw, and Rosner (2001) made the first attempt to extend the QMoM to a bivariate case. They used 3 and 12-point quadratures and performed the bivariate moment inversion using either an adapted univariate technique or direct optimization.

Later, Yoon and McGraw (2004a, 2004b) presented the mathematical and statistical foundation for the multivariate extension of the QMoM using the principal component analysis (PCA), creating the so-called PCA-QMoM.

Fox (2006) solved the same problem proposed by Wright et al. (2001), but he used DQMoM instead of QMoM. Fox (2006) bypassed the moment inversion problem defining the initial values for weights and abscissas with only one weight having non zero value. The abscissa values for the quadrature points with zero weight were generated in an arbitrary way. This procedure does not need moment inversion, but he reported problems related to stiffness at the beginning of the simulation until the values of abscissas and weights become compatible to the set of moments that are used in their determination.

Later, Fox (2009b) proposed a complex brute-force procedure to determine optimal multivariate moment sets, that is, sets of moments for which the matrix in the DQMoM linear system of equations does not become singular. This procedure considers Gaussian-like distributions and the author himself pointed out that further investigation is necessary for non-Gaussian distributions. For an  $h$ -dimensional problem, this procedure derived optimal multivariate moment sets for  $n^h$ -point quadratures with  $h = 1–3$  and  $n = 1–3$ .

Fox (2008) developed a quadrature-based third-order moment method for dilute gas-particle flows in two and three dimensions. The proposed method, called by the author as the tensor product method (TPM), is similar to that developed by Yoon and McGraw (2004a, 2004b) using the principal component analysis (PCA) for  $2^h$ -point quadratures. Afterwards, for three-dimensional problems, Fox (2009a) extended the method to arbitrary higher-order quadratures with  $n^3$  points, with  $n = 3$  and 4.

Cheng and Fox (2010) proposed the conditional method of moments to multivariate moment inversion and used this together with QMoM to simulate a nano-precipitation process. This same method was also used in a more recent work by Yuan and Fox (2011) to solve the kinetic equations for the velocity distribution function.

From the above, it is clear that a comparative evaluation of the existing methods for multivariate moment inversion is needed, which is the first aim of the present work. Our second goal is to introduce a method based on the independent component analysis, ICA (Comon, 1994). The methods cited above as well as their combinations were analyzed in terms of accuracy and robustness.

## 2. Multivariate moment inversion problem

Consider a generic multivariate distribution,  $f(\mathbf{x})$ , where  $\mathbf{x} = [x_1, x_2, \dots, x_h]^T$  is the  $h$ -dimensional vector of internal variables. The moments of this distribution give important statistical information, for example, the mean, the variance, the skewness and the kurtosis. The standard multivariate moment is defined as:

$$\mu_{\mathbf{k}} = \mu_{k_1, k_2, \dots, k_h} = \langle x_1^{k_1} x_2^{k_2} \dots x_h^{k_h} \rangle = \int \int \dots \int \left( \prod_{i=1}^h x_i^{k_i} \right) f(\mathbf{x}) d\mathbf{x} \quad (1)$$

where  $\mu_{\mathbf{k}}$  is a generic multivariate moment and  $\mathbf{k} = [k_1, k_2, \dots, k_h]^T$  is a vector whose  $i$  component represents the order of the moment relative to the  $x_i$  variable. The total order of the moment is given by  $s = \sum_i k_i$ .

The total number of moments up to the total order  $O$  is given by

$$N_{mom}^{(O)} = \sum_{s=0}^O \binom{h+s-1}{s} \quad (2)$$

For a bivariate problem ( $h=2$ ), this reduces to  $N_{mom}^{(O)} = (O+1)(O+2)/2$ .

The moment inversion problem consists of determining an  $N$ -point Gauss–Christoffel quadrature from a finite set of  $f$  moments. This quadrature can be interpreted as a discretization of the  $f$  distribution, which is given by its abscissas  $\mathbf{x}_j$  and weights  $\omega_j, j = 1, \dots, N$ .

The order of the quadrature is defined as the maximum value of the total order  $s$  for which the  $f$  moments are calculated exactly for all possible combinations of  $k_i, i = 1, \dots, h$ . Considering an  $N$ -point quadrature rule, these moments are exactly calculated by:

$$\mu_{k_1, k_2, \dots, k_h} = \sum_{j=1}^N x_{1j}^{k_1} x_{2j}^{k_2} \cdots x_{hj}^{k_h} \omega_j \quad (3)$$

It should be noted that a quadrature rule in the  $h$ -dimensional space has  $N(h+1)$  values that have to be determined. From Eq. (2), it is clear that it is not always possible to choose  $N$  for a given  $O$  in order that  $N_{mom}^{(O)} = N(h+1)$ .

### 3. Existing methods for multivariate moment inversion

The methods for multivariate moment inversion are briefly presented in the following. Most of them apply one of the existing algorithms (Gautschi, 2004; Gordon, 1968; Press, Teukolsky, Vetterling, & Flannery, 1992) for univariate moment inversion in a one variable at a time.

#### 3.1. Direct optimization

In principle, the multidimensional quadrature can be calculated by solving Eq. (3) for the abscissas and weights using  $N(h+1)$  multivariate moments. However, due to the strong non-linearity of the system of equations, it is usually better to obtain such a solution by optimizing a convenient objective function that includes the differences between the actual multivariate moments and their approximation given by Eq. (3).

The multivariate moment inversion using optimization was first reported by Wright et al. (2001) for a bivariate population balance problem. They used a conjugate-gradient minimization algorithm and reported that large computational times and good initial guesses were required.

Details of optimization algorithms can be found in the literature (Miller, 1999; Pierre, 1987; Weise, 2009). The relevance point here is the definition of the objective function. For the sake of simplicity, consider a bivariate case and an  $N$ -point quadrature with abscissas,  $x_{1i}$  and  $x_{2i}$ , and weights  $\omega_i$ . Then, the objective function to be minimized can be defined using the relative errors in the moments that, for a moment set up to the total order  $O$ , is given by:

$$F_{obj} = \sum_{k=0}^O \sum_{l=0}^{O-k} \left| \frac{\epsilon_{kl}}{\mu_{kl}} \right|, \quad \epsilon_{kl} = \sum_{i=1}^N x_{1i}^k x_{2i}^l \omega_i - \mu_{kl} \quad (4)$$

When the number of parameters is equal to the number of moments in  $F_{obj}$ , the optimization might obtain the solution of the non-linear system of equations given by Eq. (3).

#### 3.2. Direct Cartesian Product Method (DCPM)

The use of pure moments in population balance problems has already been reported in some works (Buffo, Vanni, & Marchisio, 2012; Marchisio, 2009). This method assumes that all variables are independent, which allows the following factorization of the PDF:

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_h) = f_1(x_1)f_2(x_2) \cdots f_h(x_h) \quad (5)$$

where  $f_i(x_i)$  are the marginal PDFs, defined by:

$$f_i(x_i) = \int_{\bar{\mathbf{x}}} f(\mathbf{x}) d\bar{\mathbf{x}} \quad (6)$$

where  $\bar{\mathbf{x}}$  is the  $\mathbf{x}$  vector without variable  $x_i$ . Therefore, the mixed moments can be written as the product of pure moments in each variable:

$$\mu_{k_1, k_2, \dots, k_h} = \mu_{k_1, 0, \dots, 0} \mu_{0, k_2, \dots, 0} \cdots \mu_{0, 0, \dots, k_h} \quad (7)$$

and the set of multivariate moments up to the total order  $O$  has just  $(Oh+1)$  degrees of freedom. Thus, an  $h$ -dimensional quadrature can be determined by the Cartesian product of  $h$  unidimensional quadratures obtained from a univariate moment inversion method using the first  $2N_i$  moments for the  $x_i$  variable, as shown in Appendix A.

However, if the variables are not independent, the multidimensional quadrature is only first order accurate, being this a strong drawback of this method.

#### 3.3. Principal component analysis, PCA

The principal component analysis (PCA) was first proposed by Pearson (1901) as a helpful tool for multidimensional data analysis. It is also known in literature as the discrete Karhunen–Loève transform (KLT), the Hotelling transform and proper orthogonal decomposition (POD) (Liang et al., 2002). The application of the PCA to multivariate population balance problems was first proposed by Yoon and McGraw (2004a).

The PCA diagonalizes the covariance matrix to find a rotated orthogonal coordinate frame in which the new variables are uncorrelated, whose first direction has the largest variance, the second orthogonal direction has the second largest variance and so on. This is the so-called principal coordinate frame and its orthogonal axes are known as the principal components or principal directions.

The rotated covariant matrix provides three moments in each principal direction that can be used to generate a one-dimensional 2-point quadrature with equal weights. However, any complete set of moments can be transformed to the principal coordinate frame, as shown in Appendix B. Therefore, the DCPM can then be applied using just the pure moments in the principal directions. Then, the abscissas are transformed back to the original coordinate frame. Since the transformed variables are uncorrelated, the PCA quadrature rules are second-order accurate (Yoon & McGraw, 2004a). When higher order moments are used in the multidimensional quadrature determination, they are correctly calculated by the quadrature only if the distribution is factorable in the principal coordinate frame, that is, the transformed variables are independent (Yoon & McGraw, 2004a).

Therefore, the PCA basically applies the DCPM using uncorrelated variables. However, variable uncorrelatedness does not guarantee variable independence, which is required by DCPM to achieve the  $(2N_i - 1)$ -order accuracy in each  $x_i$  direction.

### 3.4. The tensor product method, TPM

This method was developed by Fox (2008, 2009a) and it is similar to the PCA. Basically, new coordinate variables are determined in which the covariant matrix is diagonal, which implied that the new variables are uncorrelated. Then, a set of moments are transformed to the new coordinate frame and a univariate moment inversion method is applied to determine the abscissas. The multivariate abscissas are constructed from the univariate ones by a Cartesian product as in the DCPM, what was called by the author as a ‘tensor product’. The quadrature abscissas are then transformed back to the original coordinates.

There exist two main differences between the TPM and PCA. First, PCA solves an eigenvalue problem to find the transformation matrix that diagonalizes the covariant matrix, whereas the TPM uses the Cholesky decomposition, which also makes the covariance matrix unitary in the new coordinate system.

The second difference is that the quadrature weights are calculated by solving a linear system of equations for a multivariate moment set. Fox (2008) developed the method for  $2^h$  quadrature points. For  $h=2$ , the system is naturally closed by the fact that the new variables are uncorrelated. For  $h=3$ , the uncorrelatedness of the new variables is not enough to close the system and a mixed third-order moment has to be used. For three-dimensional problems, Fox (2009a) gave the moment sets for the 27 and 64-point quadratures. Nonetheless, it is still possible to generate negative weights depending on the complexity of the distribution function.

As pointed out by the author, an advantage of using the Cholesky decomposition method is that the transformation matrix varies smoothly with the components of the covariance matrix. On the other hand, it has the disadvantage of depending on the ordering of the variables in the covariance matrix. Therefore, the method gives a different result for each of the possible permutations of the coordinates. The TPM reproduces all the moments used to determine the quadrature weights.

### 3.5. The conditional quadrature method of moments, CQMoM

The conditional quadrature method of moments was proposed by Cheng and Fox (2010) and recently used by Yuan and Fox (2011). This method begins with the representation of the multivariate distribution function as the product of its marginal distribution and conditional distributions. For the sake of simplicity, consider a bivariate distribution written as:

$$f(x_1, x_2) = f(x_1)f(x_2|x_1) \quad (8)$$

where  $f(x_1)$  is the marginal distribution of the variable  $x_1$  and  $f(x_2|x_1)$  is the conditional distribution of  $x_2$  for a given value of  $x_1$ . Conditional moments can be defined from the conditional distribution function (Yuan & Fox, 2011):

$$\langle x_2^l | x_1 \rangle \equiv \int x_2^l f(x_2|x_1) dx_2 \quad (9)$$

The order in which the distribution variables are numbered is not established by CQMoM. Therefore, any of the possible permutations of the variables can be used. If one decides that a direction needs more discretization points than the others, than the corresponding distribution variable should be  $x_1$ .

The CQMoM applied a univariate moment inversion method to the moments of the marginal distribution. The resulting one-dimensional quadrature is used to express a set of mixed moments in terms of the moments of the conditional distribution. These expressions form a linear system of equations that can be solved for the conditional moments, which are then used together with a univariate moment inversion method to obtain the one-dimensional

quadrature in the other direction. Details are given in Appendix C, which shows that the CQMoM does not use a symmetric set of moments. This method also reproduces all the moments used in determining the multivariate quadrature.

A drawback of this method is the possibility of obtaining a set of non-realizable conditional moments. This was reported by Yuan and Fox (2011), who pointed out that a partial solution to this problem is to use all the possible permutations of the distribution variables in the search of a realizable moment set. If this does not solve the problem, they recommended the reduction of the number of quadrature points.

## 4. The independent component analysis, ICA

The independent component analysis (ICA) is a statistical based method used to transform a vector of random variables, measures or signals to a coordinate frame in which the independence of the variables is maximized. Its definition was presented by Comon (1994), who described the method using the information theory (Shannon, 1948, 1964). Considering its formulation, the ICA can also be seen as a variant of the projection pursuit (Friedman & Tukey, 1974; Jones & Sibson, 1987). A classical example of the application of the ICA is the blind source separation (BSS) problem (Haykin, 1994).

The ICA can be considered an extension of the PCA. As mentioned before, the PCA finds a coordinate frame in which the variables are uncorrelated, which does not imply that the variables are independent. The ICA uses high order statistics to maximize the independence of the transformed variables and, unlike the PCA, it is not restricted to search for coordinates obtained by orthogonal transformations. A more rigorous definition of the ICA can be found in Hyvärinen and Oja (2000) and Hyvärinen, Karhunen, and Oja (2001).

As commented in Section 3.2, variables  $x_i$ ,  $i = 1, \dots, h$ , are independent if and only if the joint PDF is represented by the product of its  $h$  marginal PDFs. This allows the DCPM to determine a highly accurate  $h$ -multidimensional quadrature by finding  $h$  unidimensional quadratures. Therefore, as the ICA searches for new independent variables, it seems more suitable for multivariate moment inversion than the PCA.

In this work, we considered only the linear ICA, which assumes that a vector of observable variables,  $\mathbf{x}$ , are formed by a linear combination of unknown independent components,  $\mathbf{y}$ , given by a mixture matrix,  $\mathbf{M}$ , also unknown. Without loss of generality, it can be assumed that the variables have zero mean,  $\tilde{\mathbf{x}} = \mathbf{x} - \boldsymbol{\mu}$ . The linear mixture model is written as  $\tilde{\mathbf{x}} = \mathbf{My}$  and, therefore,  $\mathbf{y} = \mathbf{S}\tilde{\mathbf{x}}$  where  $\mathbf{S} = \mathbf{M}^{-1}$  is the separation matrix. The goal of linear ICA is to find the separation matrix  $\mathbf{S}$  that makes the components of the vector  $\mathbf{y}$  as independent as possible.

Therefore, the ICA needs a measure of variable independence which is related to the nongaussianity of the distribution, that is, how a given distribution differs from the Gaussian distribution. This is supported by a classical result in probability theory known as the Central Limit Theorem, which states that the linear combination of two independent random variables has a distribution that is closer to a Gaussian distribution than any of the two original random variables (Hyvärinen & Oja, 2000). Several methods for measuring variable independence were proposed in the literature (Haykin, 1994; Hyvärinen et al., 2001). A review of all possible methods is not in the scope of this work, but a brief description of the underlying theory and main results is presented in Appendix D.

### 4.1. Contrast functions

Comon (1994) defined that a contrast function,  $\Omega$ , must be a maximum for the linear transformation that generates the

independent components  $\mathbf{y} = \mathbf{S}(\mathbf{x} - \boldsymbol{\mu})$ , that is  $\Omega(f_y) \geq \Omega(f_x)$ . Besides, it must be invariant to variable permutation and scaling. Therefore, in order to make the solution unique, the variables should be ordered accordingly to the characteristic values of their covariance matrix and then scaled to generate a unitary covariance matrix. This is accomplished by the whitening transformation,  $\mathbf{b} = \mathbf{W}(\mathbf{x} - \boldsymbol{\mu})$ , that does not change the value of the contrast function (see Appendix D).

Due to the properties of the mutual information,  $I(f_x)$ , Comon (1994) showed that a convenient contrast function can be defined by:

$$\Omega(f_x) = -I(f_z), \quad \mathbf{z} = \mathbf{Q}\mathbf{b} \quad (10)$$

and the maximization process consists of determining the correct orthogonal transformation,  $\mathbf{Q}$ . Using Eq. (D.19), we can write:

$$\Omega(f_x) = \sum_{i=1}^h J(f_{z_i}) - J(f_x) \quad (11)$$

It should be noted that  $J(f_x)$  is a constant and, therefore, it does not affect the maximization of the contrast function. As  $J(f_{z_i})$  can be approximated by an expansion around the correspond standard Gaussian distribution (Comon, 1994):

$$\Omega(f_x) \approx \frac{1}{48} \sum_{i=1}^h (4K_{z_{ii}}^2 + K_{z_{iiii}}^2 + 7K_{z_{iiii}}^4 - 6K_{z_{ii}}^2 K_{z_{iiii}}) - J(f_x). \quad (12)$$

This contrast function can be maximized if the diagonal terms of the third and forth order cumulants are maximized by the orthogonal transformation. However, different ICA methods have different contrast functions.

It should be noted that  $\Omega$  is a function of  $\mathbf{Q}$  because of the transformation of the cumulants. For instance, for the 4th-order cumulant, the transformation is:

$$K_{z_{mnop}} = \sum_{i,j,k,l=1}^h Q_{mi} Q_{nj} Q_{ok} Q_{pl} K_{ijkl} \quad (13)$$

At the point of maximum of  $\Omega$ ,  $\mathbf{Q} = \mathbf{R}$  and  $\mathbf{z} = \mathbf{y}$ . Thus, the final variable transformation is:

$$\mathbf{y} = \mathbf{RW}(\mathbf{x} - \boldsymbol{\mu}) \quad (14)$$

#### 4.2. ICA methods

Most of the works found in the literature apply the ICA directly on the data set instead of using the data statistics. For the moment inversion problem only the ICA algorithms that can be applied to the statistics are important and were considered.

The most used ICA algorithms based on statistics are the JADE (Joint Approximate Diagonalization of Eigen-matrices), the SHIBBS (SHIFTed Blocks for Blind Separation), which is a simplified version of JADE (Cardoso, 1999; Cardoso & Souloumiac, 1993), and the CuBICA (Cumulant Based ICA) (Blaschke & Wiskott, 2004).

The JADE uses the Cardoso and Souloumiac (1993) contrast function to find the independent components. This contrast function is based on the fourth order cumulants and is defined as:

$$\Psi_{JADE}(\mathbf{Q}) = \sum_{i,k,l=1}^h K_{z_{iilk}}^2 \quad (15)$$

The CuBICA considers both the third and fourth order statistics in its contrast function, which is given by (Blaschke & Wiskott, 2004):

$$\Psi_{CuBICA}(\mathbf{Q}) = \frac{1}{3!} \sum_{i=1}^h K_{z_{iii}}^2 + \frac{1}{4!} \sum_{i=1}^h K_{z_{iiii}}^2 \quad (16)$$

It is clear that it uses the first two terms of Eq. (12). More details about the JADE and CuBICA algorithms can be found in the works of Cardoso and Souloumiac (1993) and Blaschke and Wiskott (2004), respectively.

#### 4.3. Application of ICA to multidimensional quadrature calculation

The ICA application to the derivation of a multidimensional quadrature is similar to the procedure for the PCA. First, the moments are calculated and whitened. Then, the ICA is applied to determine the orthogonal transformation  $\mathbf{R}$ , which is used to transform all moment tensors up to a given order to the transformed coordinates,  $\mathbf{y}$ . Then, the DCPM is applied. Finally, the abscissas are transformed back to the original coordinate frame, by the inverse transformation:

$$\mathbf{x} = \mathbf{W}^{-1} \mathbf{R}^T \mathbf{y} + \boldsymbol{\mu} \quad (17)$$

There is no guarantee that the linear ICA model correctly fits the multidimensional data structure in order to find variables that are really independent. Thus, the ICA just provides a method to minimize the dependence among the variables, being, in this way, a more robust and elaborated statistical method compared with the PCA. The drawback of the ICA is to assume a linear model. However, the ICA has been receiving much attention, mainly in the electrical engineering research field, and non-linear versions of ICA, already have been reported (Almeida, 2000; Hyvärinen & Pajunen, 1999; Jutten & Karhunen, 2004; Schölkopf, Smola, & Müller, 1996). Basically, the idea behind this new version of ICA is to use a non-linear transformation between the  $\mathbf{y}$  and  $\mathbf{x}$  variables instead of  $\mathbf{y} = \mathbf{S}(\mathbf{x} - \boldsymbol{\mu})$ .

#### 5. Numerical procedure

All the methods presented above can be applied for the inversion of a multivariate moment set, but for the sake of simplicity, only bivariate cases were analyzed. The procedure consisted of the following steps:

- 1 A normalized bivariate PDF in variables  $x_1$  and  $x_2$  was chosen.
- 2 The necessary set of bivariate moments were calculated in MAPLE v.12 (Maplesoft Inc., 2008).
- 3 The bivariate quadrature points were obtained using one of the previously described methods.
- 4 The obtained quadrature points were used to reconstruct the moments up to fifth order and the errors were calculated.

The methods for multivariate moment inversion were mainly implemented in C. The ORTHOPOL package (Gautschi, 1994) was used for the univariate moment inversion. The SVD routine, which was used to find the Hotelling transformation, and the routine that applies the Cholesky decomposition were both obtained from Press et al. (1992). The solution of the linear system of equations was carried out by a Gaussian elimination routine from Pinto and Lage (2001). The application of the method JADE was carried out using a slightly adapted version of the code provided by Cardoso (2011) and the CuBICA algorithm from Blaschke and Wiskott (2004) was implemented in C. The tolerance used in the Jacobi rotation

**Table 1**

Distributions used to evaluate the methods for multivariate moment inversion.

$g(x_1, x_2)$	
$\exp[-(2.5\sqrt{2}(x_1 - x_2))^2 - (0.5\sqrt{2}(x_1 + x_2) - 5)^2]$	(20)
$\exp[-\exp(-x_1 + 5) - \exp(-x_2 + 5) + x_1 + y_2 - 10]$	(21)
$2\exp[-(2x_1 - 4)^2 - (x_2 - 6)^2] + 2\exp[-(x_1 - 3)^2 - (x_2 - 4)^2] + \exp[-(x_1 - 4)^2 - (x_2 - 7)^2]$	(22)
$2\exp[- x_1 + x_2 - 10  -  2x_2 - 9 ] + \exp[-(6x_1 - 4x_2 - 60)^2 - (x_2 - 8)^2]$	(23)
$\exp[-(x_1 - 3)^2 - (x_2 - 3)^2] + \exp[-(x_1 - 3)^2 - (x_2 - 5)^2] + \frac{1}{2}\exp[-(x_1 - 7)^2 - (x_2 - 5)^2] + \frac{1}{2}\exp[-(x_1 - 7)^2 - (x_2 - 7)^2]$	(24)
$\exp[-(x_1 - 6)^2 - (3x_2 - 7)^4] + \exp[-(3x_1 - 9)^2 - (x_2 - 6)^4]$	(25)

algorithm for the jointly diagonalization of the cumulant tensor needed by JADE or CuBICA was  $10^{-14}$ .

The NLOpt open-source library of Johnson (2012) was used for the non-linear global optimization. Among the several possible choices, the controlled random search (CRS) with local mutation (Kaelo & Ali, 2006) followed by the BOBYQA (Powell, 2009) for solution refinement was shown to be the best optimization method. The CRS algorithm starts with a random population of points, and randomly evolve these points by heuristic rules. An initial guess can enter as an additional point in the population. For all cases the global optimization was carried out using a population with  $150 \times (3N+1)$  points, where  $N$  is total number of quadrature points. The relative and absolute tolerance applied for both the variables and the objective function were set to  $10^{-15}$ , which is close to the machine precision. The maximum number of evaluations was set to  $5 \times 10^8$ . The search domains were chosen to be  $0 \leq \mathbf{x} \leq 10$  for the abscissas and  $0 \leq \boldsymbol{\omega} \leq 1$  for the weights.

All computations were performed in standard double precision using GNU GCC 4.6.3 in a Linux Intel i7-2600K platform. The CPU times for the optimization were obtained in a single run and those for the DCPM, PCA, TPM, ICA and CQMoM were obtained by averaging the CPU times of  $10^5$  sequential runs. These CPU times do not take into account the calculation of the moments of the PDF.

## 6. Results and discussion

**Table 1** shows the bivariate distribution functions,  $g(x_1, x_2)$ , used in this work. The distribution functions were normalized by:

$$f(x_1, x_2) = \frac{g(x_1, x_2)}{\int_0^\infty \int_0^\infty g(x_1, x_2) dx_1 dx_2} \quad (18)$$

The bivariate moments of  $f(x_1, x_2)$  were those used in the moment inversion methods. Fig. 1 shows these distribution functions.

The results were organized in three sections. First, the results for DCPM, PCA, ICA and CQMoM are compared for the quadratures with  $2 \times 2$ ,  $3 \times 2$ ,  $2 \times 3$  and  $3 \times 3$  points. The TPM was applied just for the  $3 \times 3$ -point quadrature because it is equivalent to the PCA for the  $2 \times 2$ -point quadrature. Both CQMoM and TPM were applied for the two possible permutation of variables. Then, in an attempt to improve the previous results, the results of combining the PCA and ICA methods with CQMoM are presented. In this case, the moments were first transformed to the PCA or ICA variables and the CQMoM was then applied in this new coordinate frame. Finally, the results using optimization are presented.

The cumulative mean quadratic error of the reconstructed moments was calculated by:

$$\epsilon_0 = \sqrt{\frac{1}{N_{mom}} \sum_{k=0}^O \sum_{l=0}^{O-k} \left( \frac{\mu_{kl} - \mu_{kl}^{rec}}{\mu_{kl}} \right)^2} \quad (19)$$

where  $\mu$  is the moment value and  $\mu^{rec}$  is the reconstructed moment obtained by the bidimensional quadrature rule,  $O$  is the total order of the mixed moment and  $N_{mom}$  is the total number of moments up

to the order  $O$ . It should be pointed out that if the CQMoM and TPM do not fail, they give multidimensional quadratures that reproduce exactly all moments that were used for their determination.

### 6.1. Comparison of the DCPM, PCA, ICA, TPM and CQMoM

The DCPM, PCA, ICA, TPM and CQMoM were applied to all distribution functions presented in **Table 1**. Both JADE and CuBICA were used for ICA, but the results obtained with the latter were a little better and, therefore, only CuBICA results are presented.

#### 6.1.1. Rotated Gaussian distribution

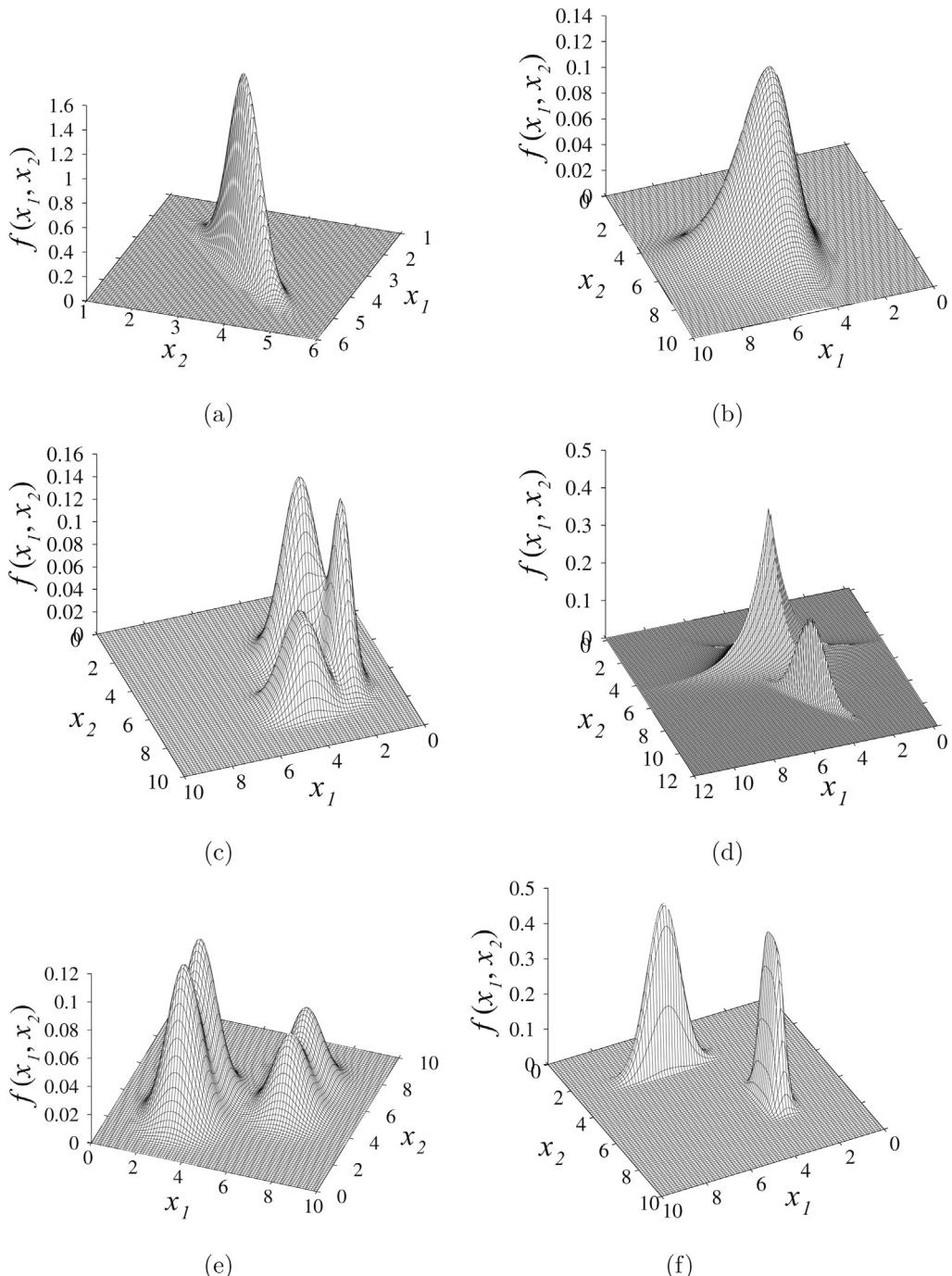
The unimodal rotated Gaussian distribution is shown in Fig. 1(a) and it is given by Eq. (20) in **Table 1**. This function was chosen because of the existing correlation between the variables. The resulting base-10 logarithm of the cumulative errors for each one of the methods are shown in **Table 2**. It can be seen that the DCPM did not provide good results because it does not consider the existing correlation between the variables. Increasing the number of quadrature points did not improve its results. This is a serious drawback of this method and this shows that it is necessary to consider the mixed-moment information.

On the other hand, the PCA and ICA obtained good results because both methods rotate the coordinate frame to obtain a new one where the variables are uncorrelated. These two methods obtained  $2 \times 2$ ,  $3 \times 2$  and  $2 \times 3$ -point quadratures that could accurately reconstruct moments up to 3rd order. For the  $3 \times 2$  and  $2 \times 3$ -point quadratures, the moments of 4th and 5th orders were reasonably well reconstructed. On the other hand, the  $3 \times 3$ -point quadratures are 5th order accurate. The quadrature points obtained

**Table 2**

Cumulative relative errors  $\log(\epsilon_0)$  in the recovered moments of the rotated Gaussian distribution (Eq. (20)) up to a given order,  $O$ , using PCA, ICA, TPM and CQMoM.

Method	$\log \epsilon_1$	$\log \epsilon_2$	$\log \epsilon_3$	$\log \epsilon_4$	$\log \epsilon_5$	CPU ( $\mu$ s)
DPCM $_{(2 \times 2)}$	-15.99	-2.11	-1.76	-1.53	-1.35	0.8
PCA $_{(2 \times 2)}$	-15.55	-15.51	-14.09	-3.38	-2.74	2.3
ICA $_{(2 \times 2)}$	-15.71	-15.73	-14.10	-3.39	-2.74	11.0
CQMoM $_{(2 \times 2)}^{X_2 X_1}$	Unrealizable conditional moment set					-
CQMoM $_{(2 \times 2)}^{X_1 X_2}$	Unrealizable conditional moment set					-
DPCM $_{(2 \times 2)}^{X_1 X_2}$	-15.60	-2.11	-1.76	-1.53	-1.35	1.4
DPCM $_{(2 \times 2)}^{X_2 X_1}$	-15.49	-2.11	-1.76	-1.53	-1.35	1.3
PCA $_{(3 \times 2)}^{Y_1 Y_2}$	-15.55	-15.56	-14.09	-6.18	-5.69	6.9
PCA $_{(2 \times 3)}^{Y_1 Y_2}$	-16.14	-16.29	-14.10	-3.38	-2.74	7.4
ICA $_{(3 \times 2)}^{Y_1 Y_2}$	-15.78	-15.58	-14.10	-3.39	-2.74	11.2
ICA $_{(2 \times 3)}^{Y_1 Y_2}$	-15.35	-15.41	-14.10	-6.18	-5.70	11.1
CQMoM $_{(3 \times 2)}^{X_2 X_1}$	-16.01	-15.73	-15.76	-5.36	-4.75	2.8
CQMoM $_{(3 \times 2)}^{X_1 X_2}$	-16.01	-15.73	-15.76	-5.36	-4.75	2.9
DPCM $_{(3 \times 3)}$	-15.31	-2.11	-1.76	-1.53	-1.35	1.8
PCA $_{(3 \times 3)}$	-15.68	-15.80	-14.10	-13.84	-13.66	7.9
ICA $_{(3 \times 3)}$	-15.22	-15.25	-14.10	-13.84	-13.67	12.1
TPM $_{(3 \times 3)}^{X_2 X_1}$	-16.01	-16.05	-16.01	-12.66	-12.16	8.3
TPM $_{(3 \times 3)}^{X_1 X_2}$	-16.01	-16.05	-16.01	-12.66	-12.16	8.3
CQMoM $_{(3 \times 3)}^{X_2 X_1}$	Unrealizable conditional moment set					-
CQMoM $_{(3 \times 3)}^{X_1 X_2}$	Unrealizable conditional moment set					-



**Fig. 1.** Probability density functions: (a) rotated Gaussian distribution (Eq. (20)), (b) Gumbel distribution (Eq. (21)), (c) multimodal distribution with platykurtic and mesokurtic modes (Eq. (22)), (d) bimodal distribution with leptokurtic and mesokurtic modes (Eq. (23)), (e) multimodal Gaussian distribution (Eq. (24)) and (f) bimodal distribution with separated platykurtic and mesokurtic modes (Eq. (25)).

by the PCA and ICA are almost equal. The results for the TPM with both variables permutations were quite similar to those obtained using the PCA or ICA.

The CQMoM presented problems with nonrealizable conditional moment sets for the cases with equal number of quadrature points for both variables. For the cases with  $N_1 \neq N_2$ , the CQMoM results were similar to those obtained by the PCA or ICA. Both variables permutations gave similar results in CQMoM.

Regarding the CPU times, the computational effort of the considered methods increased in the following order: CQMoM, PCA, TPM and ICA. This behavior was also observed for all distributions analyzed in this work. The higher CPU time of the ICA compared

with the PCA may be attributed to the necessity of additional pre-processing steps and the time to perform the CuBICA algorithm.

#### 6.1.2. Gumbel distribution

The distribution function shown in Fig. 1(b) is known in the literature as the Gumbel distribution and it is given by Eq. (21) in Table 1. Differently from the Gaussian distribution, the high order statistics are important to its characterization. The resulting cumulative errors for each one of the methods are shown in Table 3. As can be seen from this table, the DCPM gives good results for this case, demonstrating that the original coordinate frame does not have to be transformed. The PCA results are not good for the

**Table 3**

Cumulative relative errors  $\log(\epsilon_0)$  in the recovered moments of the Gumbel distribution (Eq. (21)) up to a given order, O, using PCA, ICA, TPM and CQMoM.

Method	$\log \epsilon_1$	$\log \epsilon_2$	$\log \epsilon_3$	$\log \epsilon_4$	$\log \epsilon_5$	CPU (μs)
DPCM <sub>(2×2)</sub>	-15.44	-14.70	-14.57	-2.64	-2.01	0.8
PCA <sub>(2×2)</sub>	-∞	-16.33	-2.37	-1.85	-1.50	2.3
ICA <sub>(2×2)</sub>	-15.63	-15.64	-14.71	-2.64	-2.01	11.1
CQMoM <sub>(2×2)</sub> <sup>x<sub>2</sub> x<sub>1</sub></sup>	-16.04	-16.10	-14.78	-2.64	-2.01	1.9
CQMoM <sub>(2×2)</sub> <sup>x<sub>1</sub> x<sub>2</sub></sup>	-16.04	-16.10	-14.78	-2.64	-2.01	1.7
DPCM <sub>(3×2)</sub> <sup>x<sub>1</sub>,x<sub>2</sub></sup>	-15.55	-14.66	-14.51	-2.79	-2.16	1.3
DPCM <sub>(2×3)</sub> <sup>x<sub>1</sub>,x<sub>2</sub></sup>	-15.54	-14.67	-14.52	-2.79	-2.16	1.2
PCA <sub>(3×2)</sub> <sup>y<sub>1</sub>,y<sub>2</sub></sup>	-15.95	-15.92	-2.37	-1.86	-1.53	6.8
PCA <sub>(2×3)</sub> <sup>y<sub>1</sub>,y<sub>2</sub></sup>	-15.19	-15.20	-2.37	-1.87	-1.52	7.1
ICA <sub>(3×2)</sub> <sup>y<sub>1</sub>,y<sub>2</sub></sup>	-15.89	-15.65	-14.70	-2.79	-2.16	11.1
ICA <sub>(2×3)</sub> <sup>y<sub>1</sub>,y<sub>2</sub></sup>	-15.80	-15.76	-14.70	-2.79	-2.16	11.4
CQMoM <sub>(3×2)</sub> <sup>x<sub>2</sub> x<sub>1</sub></sup>	-15.42	-15.49	-15.50	-2.79	-2.16	2.9
CQMoM <sub>(3×2)</sub> <sup>x<sub>1</sub> x<sub>2</sub></sup>	-15.42	-15.49	-15.50	-2.79	-2.16	2.7
DPCM <sub>(3×3)</sub>	-14.99	-14.60	-14.46	-14.45	-14.43	1.6
PCA <sub>(3×3)</sub>	-15.19	-15.18	-2.37	-1.88	-1.55	7.4
ICA <sub>(3×3)</sub>	-16.04	-16.10	-14.69	-14.10	-13.89	12.1
TPM <sub>(3×3)</sub> <sup>x<sub>2</sub> x<sub>1</sub></sup>	-15.89	-15.66	-15.63	-14.79	-13.86	7.9
TPM <sub>(3×3)</sub> <sup>x<sub>1</sub> x<sub>2</sub></sup>	-15.89	-15.66	-15.63	-14.79	-13.86	7.9
CQ MoM <sub>(3×3)</sub> <sup>x<sub>2</sub> x<sub>1</sub></sup>	-15.69	-15.38	-15.37	-14.90	-14.03	3.9
CQMoM <sub>(3×3)</sub> <sup>x<sub>1</sub> x<sub>2</sub></sup>	-15.69	-15.38	-15.37	-14.90	-14.03	3.8

moments higher than 2nd order, which can be explained by the fact that the principal coordinate frame found by PCA is  $45^\circ$  rotated in relation to the original coordinate frame. Although in this new rotated frame the variables are uncorrelated, there are variable dependence effects in the high order moments, explaining the bad results obtained by the PCA.

The ICA did not rotate the original coordinate frame, achieving 3rd and 5th-order accuracy for the  $2 \times 2$  and  $3 \times 3$ -point quadratures, respectively. In other words, these quadrature accurately reconstructed all moments that were used in their determination. The accuracy of the  $2 \times 3$  and  $3 \times 2$ -point quadratures were not improved over that of the  $2 \times 2$ -point quadrature. This example shows that the ICA can, in some cases, give better results than the PCA. Also, it shows that when there is no need for transforming the coordinate frame, the ICA does a much better job than the PCA.

The results for the CQMoM are very similar to those for the ICA for all quadratures, with the quadrature points obtained by both methods being almost identical. The results for the TPM are also similar to those achieved using the ICA.

### 6.1.3. Three-modal distribution

The three-modal distribution given by Eq. (22) in Table 1 is shown in Fig. 1(c) and presents platykurtic and mesokurtic modes. The resulting cumulative errors are presented in Table 4. For this distribution the DCPM only reproduces accurately the zero and first order moments. The PCA, ICA and CQMoM gave  $2 \times 2$ -point quadratures that reproduce well all the moments up to second order. An increase in the number of quadrature points did not improve the results obtained by the ICA and PCA. For this case, the directions found by the ICA are nonorthogonal showing its additional flexibility and explaining why its results are somewhat better than those of PCA.

On the other hand, the  $3 \times 3$ -point quadrature obtained by the TPM achieved 3rd-order accuracy, being better than those obtained with PCA and ICA. In this case, the solution of the linear system of equations to determine the quadrature weights increased the quadrature order.

The CQMoM results are even better, achieving 3rd-order accuracy even for the 6-point quadratures. The  $3 \times 3$ -point quadrature obtained by the CQMoM shows only a modest improvement in the reproduction of the higher order moments. This example shows

**Table 4**

Cumulative relative errors  $\log(\epsilon_0)$  in the recovered moments of the three-modal distribution (Eq. (22)) up to a given order, O, using PCA, ICA, TPM and CQMoM.

Method	$\log \epsilon_1$	$\log \epsilon_2$	$\log \epsilon_3$	$\log \epsilon_4$	$\log \epsilon_5$	CPU (μs)
DPCM <sub>(2×2)</sub>	-15.38	-2.18	-1.76	-1.47	-1.26	1.0
PCA <sub>(2×2)</sub>	-15.56	-15.55	-2.37	-1.85	-1.52	2.4
ICA <sub>(2×2)</sub>	-15.33	-15.35	-2.77	-2.35	-2.04	11.3
CQMoM <sub>(2×2)</sub> <sup>x<sub>2</sub> x<sub>1</sub></sup>	-15.60	-15.63	-2.56	-2.09	-1.74	1.9
CQMoM <sub>(2×2)</sub> <sup>x<sub>1</sub> x<sub>2</sub></sup>	-16.01	-15.60	-2.72	-2.25	-1.88	1.9
DPCM <sub>(3×2)</sub> <sup>x<sub>1</sub>,x<sub>2</sub></sup>	-15.15	-2.18	-1.76	-1.47	-1.27	1.3
DPCM <sub>(2×3)</sub> <sup>x<sub>1</sub>,x<sub>2</sub></sup>	-16.07	-2.18	-1.76	-1.47	-1.27	1.3
PCA <sub>(3×2)</sub> <sup>y<sub>1</sub>,y<sub>2</sub></sup>	-∞	-16.31	-2.37	-1.85	-1.53	6.9
PCA <sub>(2×3)</sub> <sup>y<sub>1</sub>,y<sub>2</sub></sup>	-16.19	-16.34	-2.37	-1.87	-1.56	6.9
ICA <sub>(3×2)</sub> <sup>y<sub>1</sub>,y<sub>2</sub></sup>	-15.74	-15.71	-2.77	-2.34	-2.01	11.5
ICA <sub>(2×3)</sub> <sup>y<sub>1</sub>,y<sub>2</sub></sup>	-15.27	-15.26	-2.77	-2.30	-1.99	11.5
CQMoM <sub>(3×2)</sub> <sup>x<sub>2</sub> x<sub>1</sub></sup>	-15.51	-15.58	-15.61	-3.10	-2.55	2.8
CQMoM <sub>(3×2)</sub> <sup>x<sub>1</sub> x<sub>2</sub></sup>	-15.56	-15.53	-15.55	-2.77	-2.21	2.7
DPCM <sub>(3×3)</sub>	-15.38	-2.18	-1.76	-1.48	-1.27	1.6
PCA <sub>(3×3)</sub>	-15.67	-15.71	-2.37	-1.88	-1.57	7.2
ICA <sub>(3×3)</sub>	-15.54	-15.51	-2.77	-2.29	-1.97	12.2
TPM <sub>(3×3)</sub> <sup>x<sub>2</sub> x<sub>1</sub></sup>	-15.97	-15.82	-15.68	-3.66	-3.14	8.0
TPM <sub>(3×3)</sub> <sup>x<sub>1</sub> x<sub>2</sub></sup>	-15.97	-15.98	-15.90	-3.70	-3.13	7.9
CQMoM <sub>(3×3)</sub> <sup>x<sub>2</sub> x<sub>1</sub></sup>	-15.17	-15.28	-15.37	-3.77	-3.30	3.8
CQMoM <sub>(3×3)</sub> <sup>x<sub>1</sub> x<sub>2</sub></sup>	-15.84	-15.95	-15.79	-3.78	-3.25	4.1

that CQMoM obtains good results for the cases where there is no problem of unrealizable conditional moment sets.

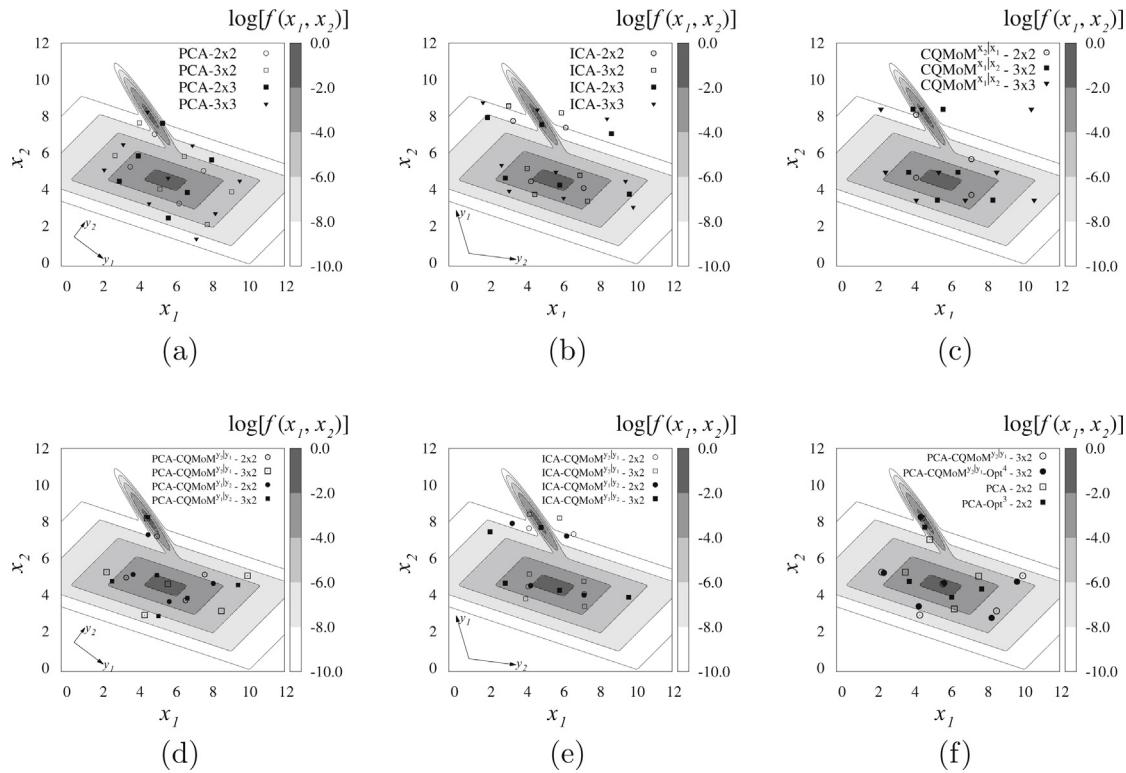
### 6.1.4. Bimodal distribution – Gaussian/Laplace modes

This distribution function have two modes corresponding to mesokurtic and leptokurtic ones as given in Eq. (23) in Table 1 and shown in Fig. 1(d). Table 5 presents the cumulative moment errors. It is clear that the DCPM generates poor results for this case. The PCA and ICA produced  $2 \times 2$ -point quadratures that are 2nd-order accurate. Nevertheless, the directions found by the ICA, Fig. 2(b), are quite different from those found by the PCA, Fig. 2(a). Besides, it can also be seen that the directions found by the ICA are nonorthogonal. The TPM  $3 \times 3$ -point quadrature could not be obtained as the linear system solution gave negative weights for both variable permutations.

**Table 5**

Cumulative relative errors  $\log(\epsilon_0)$  in the recovered moments of the bimodal Gaussian/Laplace distribution (Eq. (23)) up to a given order, O, using PCA, ICA, TPM and CQMoM.

Method	$\log \epsilon_1$	$\log \epsilon_2$	$\log \epsilon_3$	$\log \epsilon_4$	$\log \epsilon_5$	CPU (μs)
DPCM <sub>(2×2)</sub>	-16.02	-1.91	-1.53	-1.26	-1.05	0.9
PCA <sub>(2×2)</sub>	-15.72	-15.70	-2.13	-1.64	-1.33	2.3
ICA <sub>(2×2)</sub>	-15.91	-15.75	-2.63	-2.10	-1.70	11.2
CQMoM <sub>(2×2)</sub> <sup>x<sub>2</sub> x<sub>1</sub></sup>	-15.54	-15.59	-3.26	-2.37	-1.86	1.7
CQMoM <sub>(2×2)</sub> <sup>x<sub>1</sub> x<sub>2</sub></sup>						-
DPCM <sub>(3×2)</sub> <sup>x<sub>1</sub>,x<sub>2</sub></sup>	-15.48	-1.91	-1.53	-1.26	-1.04	1.3
DPCM <sub>(2×3)</sub> <sup>x<sub>1</sub>,x<sub>2</sub></sup>	-15.85	-1.91	-1.53	-1.26	-1.04	1.2
PCA <sub>(3×2)</sub> <sup>y<sub>1</sub>,y<sub>2</sub></sup>	-15.80	-15.86	-2.13	-1.66	-1.36	6.9
PCA <sub>(2×3)</sub> <sup>y<sub>1</sub>,y<sub>2</sub></sup>	-15.58	-15.62	-2.13	-1.65	-1.34	6.9
ICA <sub>(3×2)</sub> <sup>y<sub>1</sub>,y<sub>2</sub></sup>	-15.98	-15.85	-2.63	-2.10	-1.71	11.5
ICA <sub>(2×3)</sub> <sup>y<sub>1</sub>,y<sub>2</sub></sup>	-15.32	-15.21	-2.63	-2.29	-1.99	11.5
CQMoM <sub>(3×2)</sub> <sup>x<sub>2</sub> x<sub>1</sub></sup>						-
CQMoM <sub>(3×2)</sub> <sup>x<sub>1</sub> x<sub>2</sub></sup>	-15.72	-15.81	-15.55	-2.58	-2.01	2.7
DPCM <sub>(3×3)</sub>	-15.42	-1.91	-1.53	-1.26	-1.04	1.6
PCA <sub>(3×3)</sub>	-15.51	-15.46	-2.13	-1.66	-1.36	7.3
ICA <sub>(3×3)</sub>	-15.48	-15.49	-2.63	-2.28	-2.04	12.1
TPM <sub>(3×3)</sub> <sup>x<sub>2</sub> x<sub>1</sub></sup>						-
TPM <sub>(3×3)</sub> <sup>x<sub>1</sub> x<sub>2</sub></sup>						-
CQMoM <sub>(3×3)</sub> <sup>x<sub>2</sub> x<sub>1</sub></sup>						-
CQMoM <sub>(3×3)</sub> <sup>x<sub>1</sub> x<sub>2</sub></sup>	-16.02	-15.86	-15.77	-4.02	-3.43	4.0



**Fig. 2.** Quadrature abscissas for the bimodal Gaussian/Laplace distribution (Eq. 23) using (a) PCA, (b) ICA, (c) CQMoM, (d) PCA-CQMoM, (e) ICA-CQMoM methods, and (f) PCA-Optimization for the moments up to 3rd order and 4 quadrature points and PCA-CQMoM-Optimization for the moments up to 4th order and 6 quadrature points.

The CQMoM presented unrealizable conditional moment sets for some variable permutations. However, when this did not happen, the 6 and 9-point quadratures achieved 3rd-order accuracy. The quadrature points obtained by CQMoM are shown in Fig. 2(c).

#### 6.1.5. Multimodal Gaussian distribution

Eq. (24) in Table 1 gives the multimodal Gaussian distribution shown in Fig. 1(e). The cumulative moment errors are shown in Table 6. The DCPM led to quadratures that are only first-order accurate and the PCA and ICA generate quadratures that are only 2nd-order accurate, but the ICA  $\epsilon_3$  values are about an order of magnitude smaller than the corresponding PCA values. The effect of increasing the number of quadrature points is negligible for the PCA and ICA. On the other hand, the TPM obtained a 3rd-order accurate quadrature for both variable permutations.

The CQMoM produced 4 and 6-point quadratures that are 2nd and 3rd-order accurate for this distribution. However, it was not possible to obtain a 3 × 3-point quadrature with the CQMoM due to the occurrence of unrealizable conditional moment sets. Fig. 3(a), (b) and (c) shows the abscissas obtained using the PCA, ICA and CQMoM, respectively. It can be seen that the ICA found nonorthogonal independent directions.

#### 6.1.6. Bimodal distribution with separated modes

Fig. 1(f) shows the bimodal distribution with totally separated platykurtic and mesokurtic modes that is given by Eq. (25) in Table 1. The cumulative errors are presented in Table 7. Again, the DCPM led to quadratures that are only first-order accurate and the PCA and ICA give quadratures that are only 2nd-order accurate. Fig. 4(a) and (b) shows the values of the abscissas using PCA and ICA, respectively. It can be seen that most of the new quadrature points in the 6 and 9-quadrature rules are located in regions where the distribution has very low values, which might explain

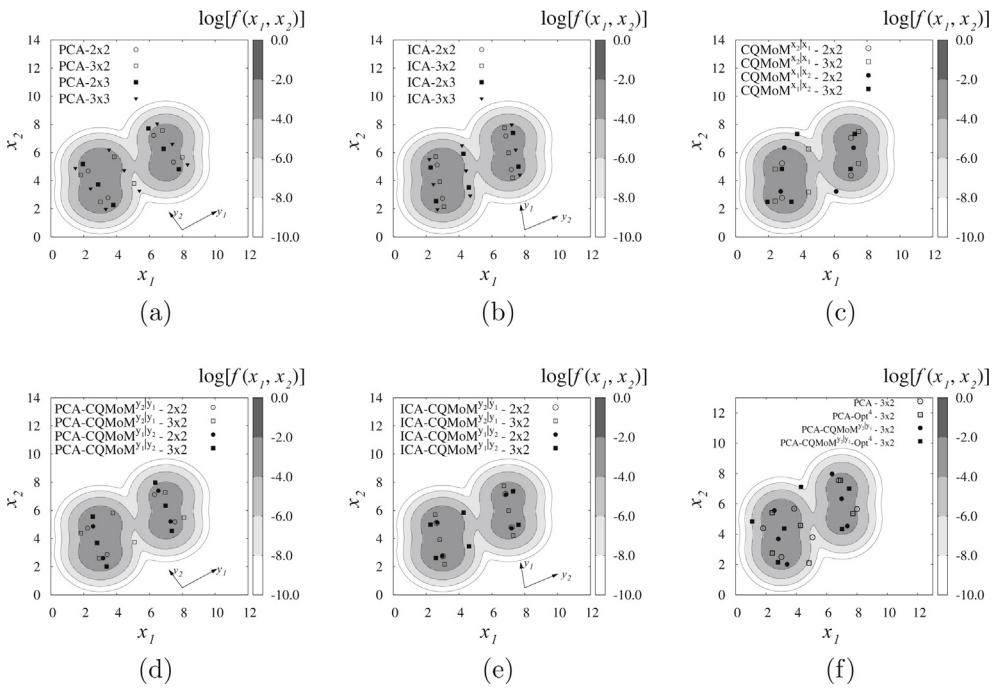
this behavior. The TPM could not obtain a 3 × 3-point quadrature because negative weights were obtained in the system solution.

The CQMoM could not obtain any quadrature rule for this distribution due to the occurrence of unrealizable conditional moment sets for both variable permutations. This is an example that shows that CQMoM can fail completely.

**Table 6**

Cumulative relative errors  $\log(\epsilon_O)$  in the recovered moments of the multi-modal Gaussian distribution (Eq. (24)) up to a given order, O, using PCA, ICA, TPM and CQMoM.

Method	$\log \epsilon_1$	$\log \epsilon_2$	$\log \epsilon_3$	$\log \epsilon_4$	$\log \epsilon_5$	CPU ( $\mu\text{s}$ )
DPCM $_{(2\times 2)}$	$-\infty$	-1.48	-1.16	-0.96	-0.82	0.9
PCA $_{(2\times 2)}$	-15.96	-16.00	-2.63	-2.19	-1.86	2.4
ICA $_{(2\times 2)}$	-15.69	-15.59	-3.78	-2.55	-2.03	11.2
CQMoM $^{X_2}_{(2\times 2)}$	-15.76	-15.65	-3.13	-2.45	-1.99	1.8
CQMoM $^{X_1}_{(2\times 2)}$	-15.79	-15.72	-2.72	-2.13	-1.72	1.8
DPCM $^{X_1}_{(3\times 2)}$	-15.54	-1.48	-1.16	-0.96	-0.83	1.2
DPCM $^{X_2}_{(3\times 2)}$	-15.25	-1.48	-1.16	-0.96	-0.83	1.2
PCA $^{Y_1}_{(3\times 2)}$	-15.38	-15.38	-2.63	-2.31	-2.15	7.0
PCA $^{Y_2}_{(2\times 3)}$	-15.96	-16.00	-2.63	-2.18	-1.86	7.0
ICA $^{Y_1}_{(3\times 2)}$	-15.42	-15.35	-3.78	-2.57	-2.05	11.2
ICA $^{Y_2}_{(2\times 3)}$	-15.24	-15.29	-3.78	-3.07	-2.62	11.3
CQMoM $^{X_2}_{(3\times 2)}$	-15.57	-15.57	-15.54	-3.06	-2.56	2.6
CQMoM $^{X_1}_{(3\times 2)}$	-15.28	-15.41	-15.48	-2.79	-2.28	2.6
DPCM $_{(3\times 3)}$	-15.16	-1.48	-1.16	-0.96	-0.83	1.5
PCA $_{(3\times 3)}$	-15.32	-15.37	-2.63	-2.30	-2.13	7.3
ICA $_{(3\times 3)}$	-15.28	-15.22	-3.78	-3.53	-3.39	12.3
TPM $^{X_2}_{(3\times 3)}$	-15.77	-15.79	-15.71	-2.63	-2.06	8.0
TPM $^{X_1}_{(3\times 3)}$	-15.93	-15.78	-15.73	-3.10	-2.56	7.9
CQMoM $^{X_2}_{(3\times 3)}$				Unrealizable conditional moment set		-
CQMoM $^{X_1}_{(3\times 3)}$				Unrealizable conditional moment set		-



**Fig. 3.** Quadrature abscissas for the multimodal Gaussian distribution (Eq. 24) using (a) PCA, (b) ICA, (c) CQMoM, (d) PCA-CQMoM, (e) ICA-CQMoM and (f) PCA-Optimization and PCA-CQMoM-Optimization for the moments up to 4th order and 6 quadrature points.

## 6.2. Results for PCA-CQMoM and ICA-CQMoM

The results in the previous section show that CQMoM and TPM are superior to those obtained by PCA and ICA if the former methods are able to obtain the multivariate quadrature.

Since both PCA and ICA generate transformed variables that are more independent, they can be used in combination with the CQMoM, which works perfectly if the distribution variables are actually independent, to increase its robustness. These combined methods are called PCA-CQMoM and ICA-CQMoM. First, the PCA or ICA transformation are obtained and used to transform the moments to the PCA or ICA coordinate frame. Then, the

CQMoM uses these moments to get the quadrature points, which are transformed back to the original coordinate frame using the same procedure that was described previously for the PCA and ICA methods. The aim of this analysis is to verify if the CQMoM problems of unrealizable conditional moment set could be solved or mitigated.

Since all moments of a given order are necessary to transform any moment of this order, only the 4 and 6-point quadratures can be generated by the combined methods using the moments up to fifth order. Therefore, in order to evaluate the combined methods, only these quadrature rules were determined for the last three distributions given in Table 1.

### 6.2.1. Bimodal distribution – Gaussian/Laplace modes

Table 8 presents the cumulative errors in the moments corresponding to the distribution illustrated in Fig. 1(d). As it can be seen, the transformation of the moments to the PCA or ICA coordinate frame solved the problem of unrealizable conditional moment set found previously by the CQMoM on the original coordinate frame (Section 6.1.4 and Table 5). Similarly to those cases where the CQMoM was successfully applied, the  $2 \times 2$  and  $3 \times 2$ -point quadratures reconstructed accurately all moments up to second and third

**Table 7**

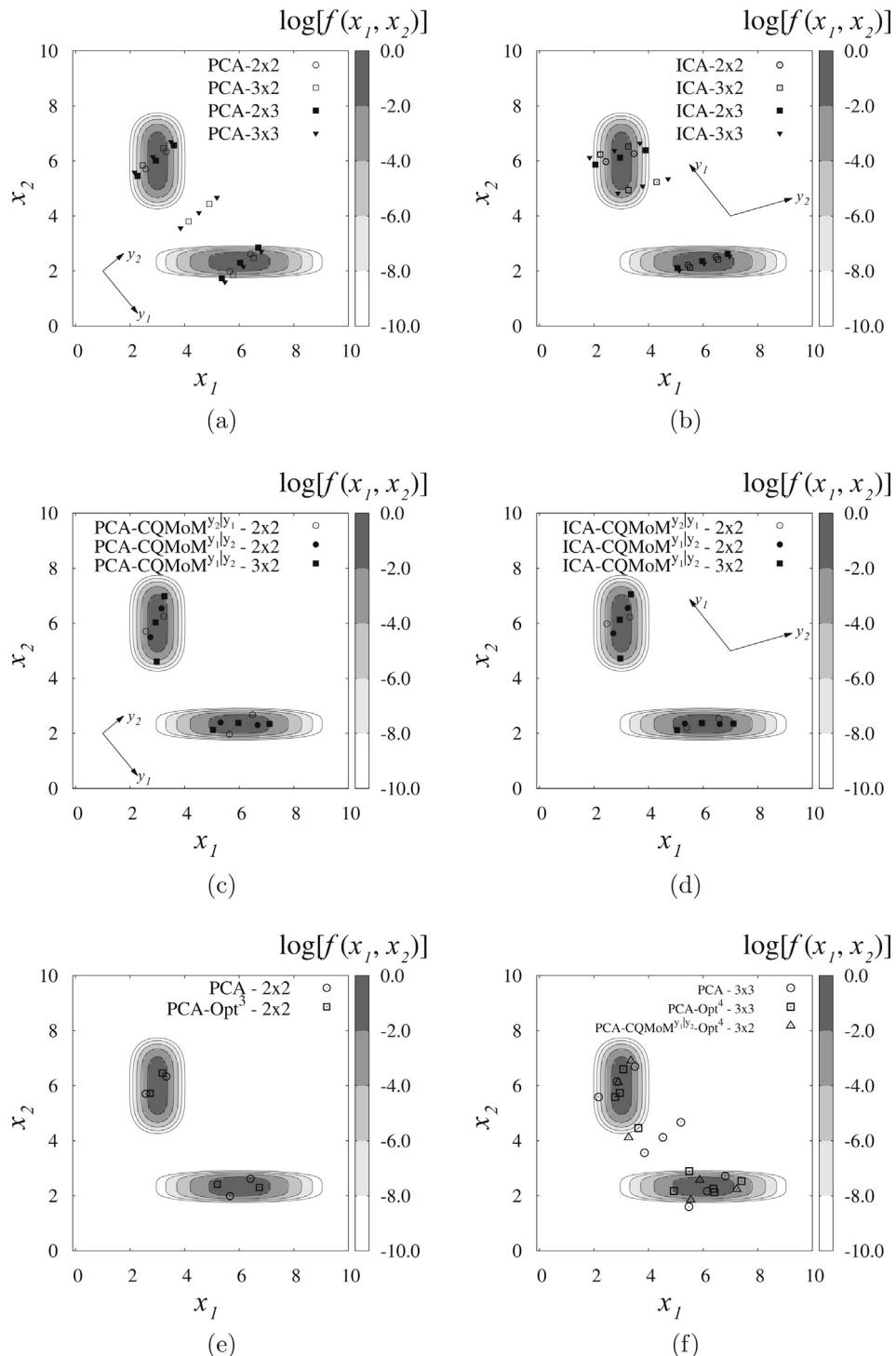
Cumulative relative errors  $\log(\epsilon_0)$  in the recovered moments of the bimodal distribution with separated modes (Eq. (25)) up to a given order, O, using PCA, ICA, TPM and CQMoM.

Method	$\log \epsilon_1$	$\log \epsilon_2$	$\log \epsilon_3$	$\log \epsilon_4$	$\log \epsilon_5$	CPU ( $\mu\text{s}$ )
DPCM $_{(2 \times 2)}$	-15.91	-1.15	-0.79	-0.52	-0.30	0.9
PCA $_{(2 \times 2)}$	-15.75	-15.90	-2.40	-1.99	-1.72	2.4
ICA $_{(2 \times 2)}$	-15.77	-15.81	-2.33	-1.97	-1.73	11.1
CQMoM $_{(2 \times 2)}^{x_2 x_1}$	unrealizable conditional moment set					-
CQMoM $_{(2 \times 2)}^{x_1 x_2}$	unrealizable conditional moment set					-
DPCM $_{(3 \times 2)}^{x_1,x_2}$	-15.44	-1.15	-0.79	-0.52	-0.30	1.2
DPCM $_{(2 \times 3)}^{x_1,x_2}$	-15.35	-1.15	-0.79	-0.52	-0.30	1.1
PCA $_{(2 \times 2)}^{y_1,y_2}$	-15.15	-15.16	-2.40	-1.99	-1.73	6.9
PCA $_{(2 \times 3)}^{y_1,y_2}$	-15.27	-15.24	-2.40	-1.99	-1.72	6.9
ICA $_{(3 \times 2)}^{y_1,y_2}$	-15.75	-15.83	-2.33	-1.97	-1.73	11.3
ICA $_{(2 \times 3)}^{y_1,y_2}$	$-\infty$	-16.16	-2.33	-1.97	-1.73	11.4
CQMoM $_{(3 \times 2)}^{x_2 x_1}$	Unrealizable conditional moment set					-
CQMoM $_{(3 \times 2)}^{x_1 x_2}$	Unrealizable conditional moment set					-
DPCM $_{(3 \times 3)}$	-15.07	-1.15	-0.79	-0.52	-0.30	1.5
PCA $_{(3 \times 3)}$	-14.95	-14.96	-2.40	-1.99	-1.73	7.3
ICA $_{(3 \times 3)}$	-15.54	-15.52	-2.33	-1.97	-1.73	12.3
TPM $_{(3 \times 3)}^{x_2 x_1}$	Negative weights					-
TPM $_{(3 \times 3)}^{x_1 x_2}$	Negative weights					-
CQMoM $_{(3 \times 3)}^{x_2 x_1}$	Unrealizable conditional moment set					-
CQMoM $_{(3 \times 3)}^{x_1 x_2}$	Unrealizable conditional moment set					-

**Table 8**

Cumulative relative errors  $\log(\epsilon_0)$  in the recovered moments of the Gaussian/Laplace distribution (Eq. (23)) up to a given order, O, using PCA-CQMoM and ICA-CQMoM.

Method	$\log \epsilon_1$	$\log \epsilon_2$	$\log \epsilon_3$	$\log \epsilon_4$	$\log \epsilon_5$	CPU ( $\mu\text{s}$ )
PCA-CQMoM $_{(2 \times 2)}^{y_2 y_1}$	-15.80	-15.81	-2.27	-1.75	-1.41	3.9
PCA-CQMoM $_{(2 \times 2)}^{y_1 y_2}$	-15.42	-15.36	-2.40	-1.88	-1.56	3.9
ICA-CQMoM $_{(2 \times 2)}^{y_2 y_1}$	-16.02	-15.85	-3.49	-2.39	-1.85	8.4
ICA-CQMoM $_{(2 \times 2)}^{y_1 y_2}$	-15.81	-15.79	-2.67	-2.11	-1.70	8.3
PCA-CQMoM $_{(3 \times 2)}^{y_2 y_1}$	-15.98	-16.03	-15.49	-3.29	-2.81	8.1
PCA-CQMoM $_{(3 \times 2)}^{y_1 y_2}$	-15.28	-15.29	-15.31	-2.84	-2.40	7.9
ICA-CQMoM $_{(3 \times 2)}^{y_2 y_1}$	-15.42	-15.43	-15.42	-2.44	-1.89	14.1
ICA-CQMoM $_{(3 \times 2)}^{y_1 y_2}$	-15.36	-15.40	-15.41	-2.94	-2.32	14.2



**Fig. 4.** Quadrature abscissas for the bimodal distribution with separated modes (Eq. 25) using (a) PCA, (b) ICA, (c) PCA-CQMoM, (d) ICA-CQMoM, (e) PCA-Optimization for the moments up to 3rd order and 4 quadrature points and (f) PCA-CQMoM-Optimization for the moments up to 4th order and 6 quadrature points and PCA-Optimization for the moments up to 4th order and 9 quadrature points.

order, respectively. Therefore, the robustness of the CQMoM was improved without loss of accuracy. Moreover, the results obtained by the PCA-CQMoM or ICA-CQMoM for the  $2 \times 2$ -point quadrature are very close to those obtained by the PCA and ICA alone.

Fig. 2(d) and (e) illustrates the values of the abscissas obtained using PCA-CQMoM and ICA-CQMoM, respectively. It is clear that, unlike the pure PCA and ICA results, the abscissas are not placed in parallel to the PCA or ICA coordinate frame.

### 6.2.2. Multi-modal Gaussian distribution

This distribution is given by Eq. (24) and represented in Fig. 1(e). As it was shown in Section 6.1.5, the CQMoM had no problems with non realizable moment set in calculating the  $2 \times 2$  and  $3 \times 2$ -point quadrature rules. This function is used here to show that the combined PCA-CQMoM and ICA-CQMoM give similar results to those of the CQMoM when there is no problem of non realizable moment set. The cumulative moment errors are presented in Table 9. The

**Table 9**

Cumulative relative errors  $\log(\epsilon_0)$  in the recovered moments of the multi-modal Gaussian distribution (Eq. (24)) up to a given order, O, using PCA-CQMoM and ICA-CQMoM.

Method	$\log \epsilon_1$	$\log \epsilon_2$	$\log \epsilon_3$	$\log \epsilon_4$	$\log \epsilon_5$	CPU (μs)
PCA-CQMoM $^{y_2 y_1}_{(2\times 2)}$	−∞	−16.04	−2.63	−2.19	−1.87	4.0
PCA-CQMoM $^{y_1 y_2}_{(2\times 2)}$	−15.90	−15.89	−3.61	−2.54	−2.02	4.0
ICA-CQMoM $^{y_2 y_1}_{(2\times 2)}$	−15.79	−15.83	−5.65	−2.57	−2.04	8.4
ICA-CQMoM $^{y_1 y_2}_{(2\times 2)}$	−15.69	−15.52	−3.79	−2.54	−2.02	8.4
PCA-CQMoM $^{y_2 y_1}_{(3\times 2)}$	−15.69	−15.67	−15.59	−2.74	−2.22	7.8
PCA-CQMoM $^{y_1 y_2}_{(3\times 2)}$	−15.26	−15.27	−15.27	−2.57	−2.03	7.9
ICA-CQMoM $^{y_2 y_1}_{(3\times 2)}$	−15.90	−15.90	−15.66	−2.58	−2.05	14.2
ICA-CQMoM $^{y_1 y_2}_{(3\times 2)}$	−15.38	−15.33	−15.32	−3.11	−2.60	14.2

**Table 10**

Cumulative relative errors  $\log(\epsilon_0)$  in the recovered moments of the bimodal distribution with separated modes (Eq. (25)) up to a given order, O, using PCA-CQMoM and ICA-CQMoM.

Method	$\log \epsilon_1$	$\log \epsilon_2$	$\log \epsilon_3$	$\log \epsilon_4$	$\log \epsilon_5$	CPU (μs)
PCA-CQMoM $^{y_2 y_1}_{(2\times 2)}$	−15.68	−15.63	−2.41	−1.99	−1.73	4.0
PCA-CQMoM $^{y_1 y_2}_{(2\times 2)}$	−15.91	−15.89	−3.15	−2.65	−2.33	4.0
ICA-CQMoM $^{y_2 y_1}_{(2\times 2)}$	−15.86	−15.89	−2.41	−2.06	−1.82	8.3
ICA-CQMoM $^{y_1 y_2}_{(2\times 2)}$	−16.19	−16.08	−2.96	−2.49	−2.19	8.2
PCA-CQMoM $^{y_2 y_1}_{(3\times 2)}$	Unrealizable conditional moment set					−
PCA-CQMoM $^{y_1 y_2}_{(3\times 2)}$	−15.41	−15.42	−15.35	−3.18	−2.86	7.9
ICA-CQMoM $^{y_2 y_1}_{(3\times 2)}$	Unrealizable conditional moment set					−
ICA-CQMoM $^{y_1 y_2}_{(3\times 2)}$	−15.58	−15.65	−15.69	−3.19	−2.85	14.2

abscissas obtained by the PCA-CQMoM and ICA-CQMoM are shown in Fig. 3(d) and (e), respectively. Comparing the results shown in Tables 6 and 9, it is clear that the quadrature rules obtained by the CQMoM and by any of the combined methods have similar accuracy.

### 6.2.3. Bimodal distribution with separated modes

Table 10 presents the cumulative moment errors corresponding to the distribution given by Eq. (25) and illustrated in Fig. 1(f). Using the original coordinate frame the CQMoM was not able to determine any quadrature rule, as pointed out in Section 6.1.6. Using the PCA or ICA transformed moments, the CQMoM could calculate the  $2 \times 2$ -point quadrature rules, which have 2nd-order accuracy. The  $3 \times 2$ -point quadrature rule could be successfully calculated only for one of the variable permutations for both the PCA-CQMoM and ICA-CQMoM, being 3rd-order accurate. This confirms that these combined methods are more robust than the CQMoM.

The abscissas obtained by the PCA-CQMoM and ICA-CQMoM are shown in Fig. 4(c) and (d), respectively, where it is clear that, differently of those results shown in Fig. 4(a) and (b), all abscissas are on regions where the distribution values are far from zero. The abscissas obtained by both combined methods are very similar for this case.

**Table 11**

Cumulative relative errors  $\log(\epsilon_0)$  in the recovered moments of the Gaussian/Laplace distribution (Eq. (23)) up to a given order, O, using optimization for the moments up to 3rd and 4th orders.

Method	$\log \epsilon_1$	$\log \epsilon_2$	$\log \epsilon_3$	$\log \epsilon_4$	$\log \epsilon_5$	CPU (h)
PCA-Opt $^3_{(2\times 2)}$	−14.49	−14.52	−14.44	−2.44	−1.92	0.004
PCA-Opt $^3_{(3\times 2)}$	−15.56	−15.50	−15.60	−2.54	−2.01	0.21
PCA-Opt $^3_{(3\times 3)}$	−6.53	−6.67	−6.78	−3.33	−2.61	2.47 <sup>a</sup>
PCA-Opt $^4_{(2\times 2)}$	−8.69	−3.15	−3.09	−3.18	−2.40	0.13
PCA-Opt $^4_{(3\times 2)}$	−9.24	−4.11	−4.22	−4.31	−3.37	0.64
PCA-Opt $^4_{(3\times 3)}$	−6.52	−4.71	−4.82	−4.91	−3.20	2.85 <sup>a</sup>
PCA-CQMoM $^{y_2 y_1}_{(3\times 2)}$ -Opt $^4_{(3\times 2)}$	−14.75	−14.78	−14.69	−14.69	−3.34	0.15

<sup>a</sup> Reached the maximum number of function evaluations.

### 6.3. Solution refinement using optimization

The use of local optimization generated poor results for the quadrature rule determination for all cases analyzed in this work. This and the large dependency on the initial guess indicate the existence of several local optimal solutions. It was verified that even for simple distributions, like unimodal exponential functions, it was not possible to solve the nonlinear system of equations given by Eq. (3) by optimization using Eq. (4). Accurate moment inversion was only achieved for an overdetermined problem and using global optimization. The optimization behavior depends on the generated random population and it was noticed that a good initial guess facilitates the algorithm convergence. Therefore, the following hybrid methods were investigated: (i) the quadrature rule provided by the PCA is used as the initial guess to optimize moments up to 3rd order, called PCA-Opt<sup>3</sup> and (ii) the quadrature rule provided by the PCA or PCA-CQMoM was used as the initial guess to optimize moments up to 4th order, called the PCA-Opt<sup>4</sup> and PCA-CQMoM-Opt<sup>4</sup>, respectively. The same functions used in Section 6.2 were used to test these methods.

#### 6.3.1. Bimodal distribution – Gaussian/Laplace modes

The cumulative moment errors are shown in Table 11. Comparing the results of Tables 5, 8 and 11 it can be seen that the order of the quadrature accuracy could be improved for some cases. The 4-point quadrature rule obtained by PCA-Opt<sup>3</sup> is 3rd-order accurate and the 6-point quadrature rule obtained by PCA-CQMoM-Opt<sup>4</sup> is 4th-order accurate. Fig. 2(f) shows that only a small change of the abscissa values was necessary to improve the quadratures provided by the PCA and PCA-CQMoM.

However, the PCA-Opt<sup>3</sup> results for the  $2 \times 3$ -point quadrature is almost identical to that of the 4-point quadrature obtained by PCA-Opt<sup>3</sup>. The results obtained by PCA-Opt<sup>3</sup> for the  $3 \times 3$ -point quadrature shown poor accuracy and presented non negligible errors for the lower order moments. The inclusion of the 4th order moments in the PCA-Opt<sup>4</sup> generated quadrature rules of low accuracy, which reconstruct the lower order moments (zeroth, first and second orders) with considerable errors. The  $3 \times 3$  quadrature rule demanded high CPU time and reached the maximum number of evaluations used in the optimization algorithm.

#### 6.3.2. Multi-modal Gaussian distribution

Table 12 shows the cumulative moment errors for the quadrature rules obtained for this distribution. The PCA-Opt<sup>3</sup> gave again a 4-point quadrature that is 3rd order accurate. Good results were obtained using both the PCA-Opt<sup>4</sup> and PCA-CQMoM-Opt<sup>4</sup> for their 6-point quadrature rules that are basically 4th order accurate. The quadrature points assignment for these cases is shown in Fig. 3(f).

#### 6.3.3. Bimodal distribution with separated modes

Table 13 presents the cumulative moment errors for this case. Similarly to the results obtained in the previous two sections, the

**Table 12**

Cumulative relative errors  $\log(\epsilon_0)$  in the recovered moments of the multi-modal Gaussian distribution (Eq. (24)) up to a given order, O, using optimization for the moments up to 3rd and 4th orders.

Method	$\log \epsilon_1$	$\log \epsilon_2$	$\log \epsilon_3$	$\log \epsilon_4$	$\log \epsilon_5$	CPU (h)
PCA-Opt <sup>3</sup> <sub>(2×2)</sub>	-14.74	-14.87	-14.74	-2.54	-1.99	0.0039
PCA-Opt <sup>3</sup> <sub>(3×2)</sub>	-14.77	-9.77	-9.83	-2.87	-2.34	1.58
PCA-Opt <sup>3</sup> <sub>(3×3)</sub>	-14.19	-6.32	-6.25	-2.86	-2.34	2.44 <sup>a</sup>
PCA-Opt <sup>4</sup> <sub>(2×2)</sub>	-13.29	-3.01	-3.10	-3.12	-2.49	0.014
PCA-Opt <sup>4</sup> <sub>(3×2)</sub>	-12.83	-12.98	-11.61	-11.34	-3.41	0.2
PCA-Opt <sup>4</sup> <sub>(3×3)</sub>	-15.18	-7.69	-7.58	-7.46	-3.77	2.84 <sup>a</sup>
PCA-CQMoM <sup>y1,y2</sup> -Opt <sup>4</sup> <sub>(3×2)</sub>	-12.75	-10.14	-10.02	-10.11	-3.07	0.48

<sup>a</sup> Reached the maximum number of function evaluations.

**Table 13**

Cumulative relative errors  $\log(\epsilon_0)$  in the recovered moments of the bimodal distribution with separated modes (Eq. (25)) up to a given order, O, using optimization for the moments up to 3rd and 4th orders.

Method	$\log \epsilon_1$	$\log \epsilon_2$	$\log \epsilon_3$	$\log \epsilon_4$	$\log \epsilon_5$	CPU (h)
PCA-Opt <sup>3</sup> <sub>(2×2)</sub>	-15.68	-15.04	-15.07	-3.10	-2.66	0.0042
PCA-Opt <sup>3</sup> <sub>(3×2)</sub>	-9.21	-9.36	-9.47	-3.35	-3.04	1.55
PCA-Opt <sup>3</sup> <sub>(3×3)</sub>	-7.72	-6.16	-6.27	-3.09	-2.76	2.43 <sup>a</sup>
PCA-Opt <sup>4</sup> <sub>(2×2)</sub>	-4.13	-4.28	-3.27	-3.35	-2.83	0.11
PCA-Opt <sup>4</sup> <sub>(3×2)</sub>	-7.69	-6.23	-6.34	-6.42	-3.70	2.20 <sup>a</sup>
PCA-Opt <sup>4</sup> <sub>(3×3)</sub>	-12.51	-12.29	-12.12	-12.21	-3.64	2.60
PCA-CQMoM <sup>y1,y2</sup> -Opt <sup>4</sup> <sub>(3×2)</sub>	-10.81	-10.41	-10.43	-10.51	-3.57	1.64

<sup>a</sup> Reached the maximum number of function evaluations.

PCA-Opt<sup>3</sup> led to a 4-point quadrature that is 3rd order accurate with small changes of the abscissa values, which are shown in Fig. 4(e).

Moreover, the PCA-CQMoM-Opt<sup>4</sup> generated a 6-point quadrature rule that is basically 4th-order accurate. The corresponding abscissas are shown in Fig. 4(f). Note that all the optimized abscissas are in regions where the distribution is far from zero.

For this distribution, the PCA-Opt<sup>4</sup> obtained a 4th-order accurate 9-point quadrature, which is more accurate than those obtained by the PCA-CQMoM<sup>3</sup><sub>(3×3)</sub> and PCA-CQMoM<sup>4</sup><sub>(3×3)</sub>.

## 7. Conclusions

Several methods were compared for multivariate moment inversion. The ICA was introduced and some combined methods were proposed to improve robustness and accuracy.

It was shown that the DPCM is not suitable when there are any kind of variable dependence on the original coordinate frame. This method is only first order accurate. The PCA and ICA were robust, obtaining the quadratures for all test distributions, but they are generally second order accurate. However, the ICA has shown to be somewhat more accurate than the PCA. The TPM 9-point quadrature showed to be third-order accurate but the method suffers from lack of robustness. The CQMoM gave 6 and 9-point quadratures that were also third-order accurate when the method did not fail. The combined methods PCA-CQMoM and ICA-CQMoM inherited the CQMoM accuracy but are more robust, mitigating this CQMoM deficiency.

Global optimization was effective in improving the accuracy of the two-dimensional quadratures obtained by PCA or PCA-CQMoM. For the former, 4-point, 3rd order accurate quadratures could be consistently obtained. For the latter, 6-point, 4th order accurate quadratures were obtained. The computation cost of global optimization is sufficiently large to preclude its use but in direct moment methods.

As only the linear ICA was analyzed in this work, its non-linear versions should be investigated in the future.

## Acknowledgements

Paulo L.C. Lage acknowledges the financial support from CNPq, grants nos. 302963/2011-1 and 476268/2009-5 and from FAPERJ, grant no. E-26/111.361/2010.

## Appendix A. Details of the DCPM

Let  $\{x_{j_i}, \omega_{j_i}\}$ ,  $j_i = 1, \dots, N_i$  be the unidimensional quadrature obtained from the first  $2N_i$  pure  $x_i$  moments,  $\{\mu_0, \dots, k_i, \dots, 0\}$ ,  $k_i = 0, \dots, 2N_i - 1$ , which is accurate to the  $2N_i - 1$  order. Then, considering a normalized distribution, the multidimensional quadrature has  $N = \prod_{i=1}^h N_i$  points that are obtained by:

$$\begin{aligned} \mathbf{x}_j &= \mathbf{x}_{j_1, j_2, \dots, j_h} = [x_{1, j_1}, x_{2, j_2}, \dots, x_{h, j_h}] \quad \text{and} \\ \omega_j &= \omega_{j_1, j_2, \dots, j_h} = \prod_{i=1}^h \omega_{j_i}, \quad j_i = 1, \dots, N_i, \quad i = 1, \dots, h \end{aligned} \quad (\text{A.1})$$

where  $j$  is given by any reordering of the  $j_i$  indexes as, for instance,  $j = j_1 + \sum_{i=2}^h (j_i - 1) \prod_{k=1}^{i-1} N_k$ ,  $h > 1$ .

## Appendix B. Details of the PCA

Without any loss of generality, the PCA is commonly applied using normalized distributions ( $\mu_0 = 1$ ) and centered moments which are defined as:

$$\tilde{\mu}_{k_1, k_2, \dots, k_h} = \tilde{\mu}_{\mathbf{k}} = \int \int \cdots \int \left[ \prod_{i=1}^h (x_i - \mu_{\mathbf{e}_i})^{k_i} \right] f(\mathbf{x}) d\mathbf{x} \quad (\text{B.1})$$

where  $\tilde{\mu}$  is the central moment and  $\mathbf{e}_i$  is the  $h$ -dimensional unit vector on the  $i$  direction. The covariance matrix,  $\Sigma$ , can be written in terms of the centered moments as:

$$\Sigma = \begin{bmatrix} \tilde{\mu}_{2,0,\dots,0} & \tilde{\mu}_{1,1,\dots,0} & \dots & \tilde{\mu}_{1,0,\dots,1} \\ \tilde{\mu}_{1,1,\dots,0} & \tilde{\mu}_{0,2,\dots,0} & \dots & \tilde{\mu}_{0,1,\dots,1} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mu}_{1,0,\dots,1} & \tilde{\mu}_{0,1,\dots,1} & \dots & \tilde{\mu}_{0,0,\dots,2} \end{bmatrix}_{h \times h} \quad (\text{B.2})$$

The key step of the PCA method is to solve an eigenvalue problem where the following decomposition is applied on the covariance matrix:

$$\Sigma = \mathbf{H} \mathbf{D} \mathbf{H}^T \quad (\text{B.3})$$

where  $\mathbf{D}$  is a diagonal matrix whose elements are the non-negative eigenvalues of  $\Sigma$  ordered according to their decreasing values,  $\lambda_1 \geq \lambda_2 \dots \geq \lambda_h$ , and the columns of the matrix  $\mathbf{H}$  correspond to the eigenvectors following the same ordering. The matrix  $\mathbf{H}$  is also known as the Hotelling transformation matrix. This procedure is straightforward if the singular value decomposition (SVD) is used (Chambers, Hand, & Härdle, 1995, 2009).

The centered moments of order  $s$ , where  $s = \sum_{i=1}^h k_i$ , can be used to define a symmetric tensor,  $\mathbf{T}$ , with rank  $s$  on the space  $\mathbf{R}^h$  (McCullagh, 1987; Yoon & McGraw, 2004a) that stores all the statistics of a given order  $s$ .

The tensor built using the centered second order moments correspond to the covariance matrix and it can be written as  $T_{ij} = \Sigma_{ij}$ ,  $i, j = 1, \dots, h$ , denoting a symmetric matrix with dimension  $h \times h$ . Generalizing this concept, the value of an element of the statistical tensor having order  $s$ ,  $T_{i_1, i_2, \dots, i_s}$ , in a problem in the space  $\mathbf{R}^h$ , where  $i_1, i_2, \dots, i_s \in \{1, 2, \dots, h\}$ , is given by  $\tilde{\mu}_{k_1, k_2, \dots, k_s}$ , where  $k_j$  is the number of occurrences of the integer  $j$  on the set  $\{i_1, i_2, \dots, i_s\}$ .

The Hotelling matrix  $\mathbf{H}$  can be used to transform the multivariate moment tensors to the principal coordinate frame,  $\mathbf{y}$ . For example, the forth order tensor is transformed to the principal coordinate frame by:

$$T'_{mnp} = \sum_{i,j,k,l=1}^h H_{im} H_{jn} H_{ko} H_{lp} T_{ijkl} \quad (\text{B.4})$$

where  $\mathbf{T}'$  is the statistical tensor in the new coordinate frame. As it can be seen, in order to transform a moment of a given order  $s$  it is necessary to use all the moments of this same order.

Once the multivariate moments are known in the principal coordinates, the DCPM is applied using just the pure moments in the principal directions. The abscissas of the obtained multidimensional quadrature  $\mathbf{y}_j$  are then transformed back to the original coordinate frame,  $\mathbf{x}_j$ , using Eq. (B.5), which is called the method of back projection (Yoon & McGraw, 2004a):

$$\mathbf{x} = \mathbf{H}\mathbf{y} + \boldsymbol{\mu} \quad (\text{B.5})$$

where  $\boldsymbol{\mu} = [\mu_{\mathbf{e}_i}]$  is the mean vector.

### Appendix C. Details of CQMoM

Consider the bivariate distribution given by Eq. (8) and the conditional moments defined by Eq. (9). The CQMoM uses a moment set  $\mu_{kl} = \langle x_1^k x_2^l \rangle$  to obtain a bivariate quadrature that can be represented by the following discretized form for Eq. (8):

$$f(x_1)f(x_2|x_1) = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \omega_{ij} \delta(x_1 - x_{1,i}) \delta(x_2 - x_{2,j}) \quad (\text{C.1})$$

The calculation of the quadrature points can be summarized as follows:

- 1 Using the pure moments  $\mu_{k0} = \langle x_1^k \rangle$ ,  $k = 0, \dots, 2N_1 - 1$ , a univariate moment inversion method is applied to obtain the quadrature rule in the  $x_1$  variable, which is represented by its weights and abscissas  $\{\omega_i, x_{1,i}\}$ ,  $i = 1, \dots, N_1$ .
- 2 For each  $l = 0, \dots, 2N_2 - 1$ , this quadrature is applied to the definition of  $\mu_{kl}$  using Eq. (8) to generate the linear system

$$\mu_{kl} = \langle x_1^k x_2^l \rangle = \sum_{i=1}^{N_1} \omega_i x_{1,i}^k \langle x_2^l | x_{1,i} \rangle, \quad k = 0, \dots, N_1 - 1$$

which can be solved for  $\langle x_2^l | x_{1,i} \rangle$ ,  $i = 1, \dots, N_1$ . It should be noted that for  $l = 0$ ,  $\mu_{k0} = \langle x_2^0 | x_{1,i} \rangle = 1$ ,  $\forall i$ .

- 3 For each  $i = 1, \dots, N_1$ , the conditional moments  $\langle x_2^l | x_{1,i} \rangle$ ,  $l = 0, \dots, 2N_2 - 1$ , are inverted by a univariate moment inversion method to obtain the quadrature represented by  $\{\omega_{ij}, x_{2,j}\}$ ,  $j = 1, \dots, N_2$ .

Therefore, for the assignment of  $N = N_1 N_2$  quadrature points in the two-dimensional domain, the following moment sets are used:

$$\begin{aligned} \{\mu_{k,0}\}, \quad k = 0, \dots, 2N_1 - 1, & \quad \text{and} \\ \{\mu_{k,l}\}, \quad k = 0, \dots, N_1 - 1, l = 1, \dots, 2N_2 - 1 & \end{aligned} \quad (\text{C.2})$$

Thus, a total number of  $N_1(2N_2 + 1)$  moments is used to calculate the two-dimensional quadrature. This is smaller than the expected degrees of freedom,  $3N_1 N_2$ , because all the two-dimensional abscissas derived from the conditional moments have the same  $x_{1,i}$  coordinate, which corresponds to  $(N_2 - 1)N_1$  restrictions.

Differently from PCA, Eq. (C.2) shows that the CQMoM does not use a symmetric set of moments. For instance, for assignment of a 3-point quadrature in each direction, the PCA needs all moments up to fifth order, which means that 21 mixed moments are necessary for a bivariate case. The CQMoM also needs 21 moments, but the moment set used is different, including all the moments up to third order, four moments of fourth order, four moments of fifth order, two moments of sixth order and one moment of seventh order.

### Appendix D. Fundamentals of the ICA

#### D.1. Cumulants

Cumulants were first introduced on astronomy area by the mathematician Thorvald N. Thiele (Lauritzen, 2002) and they are an alternative to represent the statistical information of a distribution function (Amblard & Brossier, 1995; Blaschke & Wiskott, 2004; Cardoso, 1998; McCullagh, 1987; Wyss, 1980).

For a univariate normalized distribution function, the cumulants,  $\kappa_k$ , are directly related to the centered moments  $\tilde{\mu}_k$ , being possible to obtain the cumulants using the moments of the same or lower order and vice-versa. The first, second and third order cumulants are equal to the corresponding order central moments. The cumulants with order  $k \geq 4$  are related to the central moments,  $\tilde{\mu}_i$ ,  $i \leq k$ , by:

$$\kappa_k = \tilde{\mu}_k - \sum_{i=2}^{k-2} \binom{k-1}{i-1} \kappa_i \tilde{\mu}_{k-i}, \quad k = 4, \dots \quad (\text{D.1})$$

The definition of cumulants can be generalized to multivariate distribution functions and, like the moments, it can be represented by symmetric tensors, being the cumulant tensor  $\mathbf{K}$  related to the

moment tensor  $\mathbf{T}$ . For example, the second, third and fourth order cumulants can be written as:

$$K_{ij} = T_{ij} = \Sigma_{ij} \quad (\text{D.2})$$

$$K_{ijk} = T_{ijk} \quad (\text{D.3})$$

$$K_{ijkl} = T_{ijkl} - T_{ij}T_{kl} - T_{ik}T_{jl} - T_{il}T_{jk} \quad (\text{D.4})$$

where  $i, j, k, l = 1, \dots, h$  and  $\Sigma_{ij}$  are the components of the covariance matrix.

## D.2. Differential entropy and negentropy

The Shannon entropy or differential entropy of a probability density function is defined as (Haykin, 1994):

$$\Xi(f_{\mathbf{x}}) = -E\{\log[f(\mathbf{x})]\} = -\int f(\mathbf{x}) \log[f(\mathbf{x})] d\mathbf{x} \quad (\text{D.5})$$

A fundamental result from the information theory is that, among all zero mean distributions with equal covariance matrix, the corresponding Gaussian distribution has the largest entropy:

$$\Xi(g_{\mathbf{x}}) \geq \Xi(f_{\mathbf{x}}) \quad (\text{D.6})$$

where  $g(\mathbf{x})$  is a Gaussian distribution which has the same covariance matrix  $\Sigma$  of  $f(\mathbf{x})$ , whose differential entropy is given by:

$$\Xi(\phi_{\mathbf{x}}) = \frac{1}{2} \{h[1 + \ln(2\pi)] + \ln[\det \Sigma]\} \quad (\text{D.7})$$

Therefore, a measure of nongaussianity, known as negentropy, can be defined as:

$$J(f_{\mathbf{x}}) = \Xi(g_{\mathbf{x}}) - \Xi(f_{\mathbf{x}}) \quad (\text{D.8})$$

By definition,  $J(f_{\mathbf{x}}) \geq 0$ , being zero only for Gaussian distributions (Comon, 1994; Hyvärinen & Oja, 2000). Another interesting result for the differential entropy is its behavior for a linear variable transformation:

$$\Xi(f_{A\mathbf{x}+c}) = \Xi(f_{\mathbf{x}}) + \ln|\det A| \quad (\text{D.9})$$

which leads to

$$J(f_{A\mathbf{x}+c}) = J(f_{\mathbf{x}}) \quad (\text{D.10})$$

## D.3. Mutual information and variable independence

A direct measure of the independence of the variables  $x_i \in \mathbf{x}$  whose distribution is  $f(\mathbf{x})$  is given by the mutual information between  $f(\mathbf{x})$  and the product of the marginal distributions  $f_i(x_i)$ :

$$I(f_{\mathbf{x}}) = \int_{\mathbf{x}} f(\mathbf{x}) \ln \left( \frac{f(\mathbf{x})}{\prod_{i=1}^h f_i(x_i)} \right) d\mathbf{x} = \sum_{i=1}^h \Xi(f_i) - \Xi(f_{\mathbf{x}}) \quad (\text{D.11})$$

It can be shown that  $I(f_{\mathbf{x}}) \geq 0$ , being equal to zero only when the variables are independent (Haykin, 1994). Using Eqs. (D.8) and (D.7), it can be shown that

$$I(f_{\mathbf{x}}) = J(f_{\mathbf{x}}) - \sum_{i=1}^h J(f_{x_i}) + \frac{1}{2} \ln \left[ \frac{\prod_{i=1}^h \Sigma_{ii}}{\det \Sigma} \right] \quad (\text{D.12})$$

## D.4. Whitening and mutual information

Before applying an ICA algorithm, some preprocessing steps are necessary. The first step is known as centering and consists of transforming the distribution to have zero-mean as described in B (Eq. (B.1)). The second step is to whiten (or sphere) the variables, which consists of using a linear transformation to generate new variables with a unitary covariance matrix, i.e., the variables are uncorrelated and have unit variances. Therefore, the whitening process can be achieved using a decomposition of the covariance matrix similar to that used in the PCA for the covariance matrix diagonalization:

$$\mathbf{W} = \mathbf{H}^T \mathbf{D}^{-1/2} \mathbf{H} \quad (\text{D.13})$$

where  $\mathbf{W}$  is the whitening matrix that is used to transform both the variables and the moment or cumulant tensors to the whitened coordinate frame. For example, the forth order cumulant tensor is transformed by:

$$K_{b_{mnop}} = \sum_{i,j,k,l=1}^h W_{mi} W_{nj} W_{ok} W_{pl} K_{\tilde{x}_{ijkl}} \quad (\text{D.14})$$

and the white variables,  $\mathbf{b}$  are given by:

$$\mathbf{b} = \mathbf{W}\mathbf{x} = \mathbf{W}(\mathbf{x} - \boldsymbol{\mu}) = \mathbf{W}\mathbf{x} + \mathbf{c} \quad (\text{D.15})$$

Accordingly with Eq. (D.9), the differential entropy of the white distribution is:

$$\Xi(f_{\mathbf{b}}) = \Xi(g_{\mathbf{b}}) + \ln |\det \mathbf{D}^{-1/2}| \quad (\text{D.16})$$

because  $|\det \mathbf{H}| = 1$ . Since  $g(\mathbf{x})$  and  $f(\mathbf{x})$  have the same covariance matrix, the negentropy is unaltered in the whitening process:

$$J(f_{\mathbf{b}}) = \Xi(g_{\mathbf{b}}) - \Xi(f_{\mathbf{b}}) = \Xi(g_{\mathbf{x}}) - \Xi(f_{\mathbf{x}}) = J(f_{\mathbf{x}}) \quad (\text{D.17})$$

Besides, the negentropy is also invariant if a orthogonal transformation,  $\mathbf{Q}$  ( $|\det \mathbf{Q}| = 1$ ), is applied to the white variables:

$$J(f_{\mathbf{z}}) = J(f_{\mathbf{b}}) = J(f_{\mathbf{x}}), \quad \mathbf{z} = \mathbf{Q}\mathbf{b} \quad (\text{D.18})$$

Therefore, from Eq. (D.12) it can be shown that:

$$I(f_{\mathbf{z}}) = J(f_{\mathbf{z}}) - \sum_{i=1}^h J(f_i(z_i)) = J(f_{\mathbf{x}}) - \sum_{i=1}^h J(f_i(z_i)) \quad (\text{D.19})$$

## References

- Almeida, L. (2000). Linear and nonlinear ICA based on mutual information. In *The IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium (AS-SPCC)* (pp. 117–122).
- Amblard, P.-O., & Brossier, J.-M. (1995). Adaptive estimation of the fourth-order cumulant of a white stochastic process. *Signal Processing*, 42, 37–43.
- Attarakih, M. M., Bart, H.-J., & Faqir, N. M. (2006). Numerical solution of the bivariate population balance equation for the interacting hydrodynamics and mass transfer in liquid–liquid extraction columns. *Chemical Engineering Science*, 61, 113–123.
- Blaschke, T., & Wiskott, L. (2004). CuBICA: Independent component analysis by simultaneous third- and fourth-order cumulant diagonalization. *IEEE Transactions on Signal Processing*, 52, 1250–1256.
- Buffo, A., Vanni, M., & Marchisio, D. (2012). Multidimensional population balance model for the simulation of turbulent gas–liquid systems in stirred tank reactors. *Chemical Engineering Science*, 70, 31–44.
- Cardoso, J. F. (1998). Blind signal separation: Statistical principles. *Proceedings of the IEEE*, 86, 2009–2025.
- Cardoso, J. F. (1999). High-order contrasts for independent component analysis. *Neural Computation*, 11, 157–192.
- Cardoso, J.-F. (2011). JADE in C. URL <http://perso.telecom-paristech.fr/cardoso/guidesepso.html>
- Cardoso, J.-F., & Souloumiac, A. (1993). Blind beamforming for non gaussian signals. *IEE Proceedings*, 140, 362–370.
- Chambers, J., Hand, D., & Härdle, W. (1995). *SVD and signal processing III: Algorithms, architectures, and applications*. New York: Elsevier Science.
- Chambers, J., Hand, D., & Härdle, W. (2009). *Computational statistics. Statistics and computing*. New York: Springer Dordrecht Heidelberg.

- Cheng, J. C., & Fox, R. O. (2010). Kinetic modeling of nanoprecipitation using CFD coupled with a population balance. *Industrial & Engineering Chemistry Research*, *49*, 10651–10662.
- Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, *36*, 287–314.
- Favero, J. L., & Lage, P. L. C. (2012). The dual-quadrature method of generalized moments using automatic integration packages. *Computers & Chemical Engineering*, *38*, 1–10.
- Fox, R. O. (2006). Bivariate direct quadrature method of moments for coagulation and sintering of particle populations. *Journal of Aerosol Science*, *37*, 1562–1580.
- Fox, R. O. (2008). A quadrature-based third-order moment method for dilute gas-particle flows. *Journal of Computational Physics*, *227*, 6313–6350.
- Fox, R. O. (2009a)]. Higher-order quadrature-based moment methods for kinetic equations. *Journal of Computational Physics*, *228*, 7771–7791.
- Fox, R. O. (2009b)]. Optimal moment sets for multivariate direct quadrature method of moments. *Industrial & Engineering Chemistry Research*, *48*, 9686–9696.
- Friedman, J., & Tukey, J. (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*, *23*, 881–890.
- Gautschi, W. (1994). Algorithm 726: ORTHPOL – A package of routines for generating orthogonal polynomials and gauss-type quadrature rules. *ACM Transactions on Mathematical Software*, *20*.
- Gautschi, W. (2004). *Orthogonal polynomials: Computation and approximation. Numerical Mathematics and Scientific Computation*. Oxford, Great Britain: Oxford University Press.
- Gordon, R. (1968). Error bounds in equilibrium statistical mechanics. *Journal of Mathematical Physics*, *9*, 655–663.
- Haykin, S. (1994). *Neural networks. A comprehensive foundation* (2nd ed.). New York: Macmillan College Publishing.
- Hill, P. J., & Ng, K. M. (1996). New discretization procedure for the agglomeration equation. *AIChE Journal*, *42*, 727–741.
- Hulbert, H. M., & Akiyama, T. (1969). Liouville equations for agglomeration and dispersion processes. *Industrial & Engineering Chemistry Fundamentals*, *8*, 319–324.
- Hulbert, H. M., & Katz, S. (1964). Some problems in particle technology: A statistical mechanical formulation. *Chemical Engineering Science*, *19*, 555–574.
- Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent Component Analysis*. John Wiley & Sons.
- Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, *13*, 411–430.
- Hyvärinen, A., & Pajunen, P. (1999). Nonlinear independent component analysis: Existence and uniqueness results. *Neural Networks*, *12*, 429–439.
- Iriarri, R. (2008). Fast Monte Carlo methodology for multivariate particulate systems-I: Point ensemble Monte Carlo. *Chemical Engineering Science*, *63*, 95–110.
- John, V., & Thein, F. (2012). On the efficiency and robustness of the core routine of the quadrature method of moments (QMOM). *Chemical Engineering Science*, *75*, 327–333.
- Johnson, S. G. (2012). The NLopt nonlinear-optimization package. URL <http://ab-initio.mit.edu/nlopt>
- Jones, M. C., & Sibson, R. (1987). What is projection pursuit? *Journal of the Royal Statistical Society. Series A (General)*, *150*, 1–37.
- Jutten, C., & Karhunen, J. (2004). Advances in blind source separation (BSS) and independent component analysis (ICA) for nonlinear mixtures. *International Journal of Neural Systems*, *14*, 267–292.
- Kaelo, P., & Ali, M. M. (2006). Some variants of the controlled random search algorithm for global optimization. *Journal Optimization Theory & Applications*, *130*, 253–264.
- Krallis, A., Meimargolou, D., & Kiparissides, C. (2008). Dynamic prediction of the bivariate molecular weight-copolymer composition distribution using sectional-grid and stochastic numerical methods. *Chemical Engineering Science*, *63*, 4342–4360.
- Kumar, S., & Ramkrishna, D. (1996). On the solution of population balance equations by discretization-I. A fixed pivot technique. *Chemical Engineering Science*, *51*, 1311–1332.
- Lage, P. L. C. (2011). On the representation of QMOM as a weighted-residual method – The dual-quadrature method of generalized moments. *Computers & Chemical Engineering*, *35*, 2186–2203.
- Lauritzen, S. (2002). *Thiele, pioneer in statistics*. Oxford, New York: Oxford University Press.
- Lemanowicz, M., & Gierczycki, A. T. (2010). Assessment of QMOM and DQMOM calculation efficiency. *Chemical and Process Engineering*, *31*, 589–602.
- Liang, Y. C., Lee, H. P., Lim, S. P., Lin, W. Z., Lee, K. H., & Wu, C. G. (2002). Proper orthogonal decomposition and its applications – Part I: Theory. *Journal of Sound and Vibration*, *252*, 527–544.
- Lister, J. D., Smit, D. J., & Hounslow, M. J. (1995). Adjustable discretized population balance for growth and aggregation. *AIChE Journal*, *41*, 591–603.
- Marchisio, D. L. (2009). On the use of bi-variate population balance equations for modelling barium titanate nanoparticle precipitation. *Chemical Engineering Science*, *64*, 697–708.
- Marchisio, D. L., & Fox, R. O. (2005). Solution of population balance equations using the direct quadrature method of moments. *Journal of Aerosol Science*, *36*, 43–73.
- Maplesoft Inc. (2008). Maple v.12. URL <http://www.maplesoft.com/>
- McCullagh, P. (1987). *Tensor methods in statistics*. London: Chapman and Hall.
- McGraw, R. (1997). Description of the aerosol dynamics by the quadrature method of moments. *Aerosol Science and Technology*, *27*, 255–265.
- Meimargolou, D., & Kiparissides, C. (2007). Monte carlo simulation for the solution of the bi-variate dynamic population balance equation in batch particulate systems. *Chemical Engineering Science*, *62*, 5295–5299.
- Miller, R. E. (1999). *Optimization: Foundations and applications*. John Wiley & Sons, Inc.
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Journal of Aerosol Science*, *2*, 559–572.
- Pierre, D. (1987). *Optimization theory with applications*. Dover Books on Mathematics Series. Dover Publications.
- Pinto, J. C., & Lage, P. L. C. (2001). *Métodos Numéricos em Problemas de Engenharia Química*. Brasil, RJ: E-papers.
- Powell, M. J. D. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives*.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C: The art of scientific computing* (2nd ed.). New York, NY, USA: Cambridge University Press.
- Ramkrishna, D. (1971). Solution of population balance equations. *Chemical Engineering Science*, *26*, 1134–1136.
- Ramkrishna, D. (2000). *Population balances – Theory and applications to particulate systems in engineering*. San Diego: Academic Press.
- Schölkopf, B., Smola, A., & Müller, K.-R. (1996). *Nonlinear component analysis as a kernel eigenvalue problem*.
- Shannon, C. (1964). *The mathematical theory of communication*. Urbana: University of Illinois Press.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423.
- Silva, L. F. L. R., & Lage, P. L. C. (2011). Development and implementation of a polydisperse multiphase flow model in OpenFOAM. *Computers & Chemical Engineering*, *35*, 2653–2666.
- Silva, L. F. L. R., Rodrigues, R. C., Mitre, J. F., & Lage, P. L. C. (2010). Comparison of the accuracy and performance of quadrature-based methods for population balance problems with simultaneous breakage and aggregation. *Computers & Chemical Engineering*, *34*, 286–297.
- Subramanianand, G., & Ramkrishna, D. (1971). On the solution of statistical models of cell populations. *Mathematical Biosciences*, *10*, 1–23.
- Weise, T. (2009). Global Optimization Algorithms – Theory and Application. URL <http://www.it-weise.de/projects/book.pdf> (published on-line).
- Wright, D. L., McGraw, R., & Rosner, D. E. (2001). Bivariate extension of the quadrature method of moments for modeling simultaneous coagulation and sintering of particle populations. *Journal of Colloid and Interface Science*, *236*, 242–251.
- Wyss, W. (1980). Cumulants and moments. *Reports on Mathematical Physics*, *18*, 79–86.
- Yoon, C., & McGraw, R. (2004a)]. Representation of generally mixed multivariate aerosols by the quadrature method of moments: I. Statistical foundation. *Journal of Aerosol Science*, *35*, 561–576.
- Yoon, C., & McGraw, R. (2004b)]. Representation of generally mixed multivariate aerosols by the quadrature method of moments: II. Aerosol dynamics. *Journal of Aerosol Science*, *35*, 577–598.
- Yuan, C., & Fox, R. O. (2011). Conditional quadrature method of moments for kinetic equations. *Journal of Computational Physics*, *230*, 8216–8246.
- Zucca, A., Marchisio, D. L., Vanni, M., & Barresi, A. A. (2007). Validation of bivariate DQMOM for nanoparticle processes simulation. *AIChE Journal*, *53*, 918–931.