Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 226 (2007) 621–644
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

An improved algorithm for the multidimensional moment-constrained maximum entropy problem

Rafail V. Abramov *

Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, 851 S. Morgan Street, Chicago, IL 60607, USA

Received 20 February 2007; received in revised form 19 April 2007; accepted 24 April 2007
Available online 22 May 2007

Abstract

The maximum entropy principle is a versatile tool for evaluating smooth approximations of probability density functions with the least bias beyond specified constraints. In the recent paper we introduced new computational framework for the moment-constrained maximum entropy problem in a multidimensional domain, and developed a simple numerical algorithm capable of computing maximum entropy problem in a two-dimensional domain with moment constraints of order up to 4. Here we design an improved numerical algorithm for computing the maximum entropy problem in a two- and higher-dimensional domain with higher order moment constraints. The algorithm features multidimensional orthogonal polynomial basis in the dual space of Lagrange multipliers to achieve numerical stability and rapid convergence of Newton iterations. The new algorithm is found to be capable of solving the maximum entropy problem in the two-dimensional domain with moment constraints of order up to 8, in the three-dimensional domain with moment constraints of order up to 6, and in the four-dimensional domain with moment constraints of order up to 4, corresponding to the total number of moment constraints of 44, 83 and 69, respectively. The two- and higher-dimensional maximum entropy test problems in the current work are based upon long-term statistics of numerical simulation of the real-world geophysical model for wind stress driven oceanic currents such as the Gulf Stream and the Kuroshio.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Maximum entropy problem; Multidimensional maximum entropy problem; Numerical methods

1. Introduction

The moment-constrained maximum entropy problem yields an estimate of a probability density with highest uncertainty among all densities satisfying supplied moment constraints. Generally, other methods to obtain smooth approximations are available, such as kernel density estimates for statistical ensembles [24,28,29]. Here, however, we assume instead that the moment data are available as, for example, in a variety of settings in solid state physics [4,15,16,31], econometrics [23,32], and geophysical applications such as weather and

* Tel.: +1 312 413 7945; fax: +1 312 996 1491.
E-mail address: abramov@math.uic.edu

0021-9991/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2007.04.026

climate prediction [2,3,14,17,19,25,27]. In addition, maximum entropy problem provides rigorous upper bound on Shannon entropy under specified constraints, which sometimes is a highly desirable feature not, for example, readily available for kernel density estimates. The approximation itself is obtained by maximizing Shannon entropy under the constraints established by measured moments [22] within the mathematical framework known as the moment problem. Generally one distinguishes between three types of moment problems:

- Hausdorff moment problem: The target probability density is supported on the closed interval $[a,b]$, often remapped into $[0,1]$ for simplicity.
- Stieltjes moment problem: The target probability density is supported on the half-open interval $[0,+\infty)$.
- Hamburger moment problem: The target probability density is supported on the real line.

For details, see, for example [8,30] and references therein. Here we restrict ourselves to the setting of the Hamburger moment problem, although generalization to the Hausdorff and Stieltjes moment problems is straightforward. A standard formalism [33] transforms the constrained maximum entropy problem into the concave minimization problem of the Lagrangian function. Conventional methods to resolve unconstrained optimization problems are very well studied, and include the gradient descent method, the Newton algorithm [7], and the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm [5].

The new maximum entropy algorithm is intended to be used mainly in geophysical applications, such as medium range weather forecasting and climate prediction. The idea of applying the rigorous framework of information theory for ensemble forecasting was proposed in [27], utilizing Shannon entropy as a measure of the lack of predictive information in a forecast. It later evolved into a concept of measuring the lack of information in an average long-term statistical state (climate) relative to a forecast via relative entropy [17,25]. The latter was formalized in [19] by developing a hierarchy of rigorous lower bounds for relative entropy. Successful development of a simple maximum entropy algorithm in [1], capable of carrying out computations on the two-dimensional domain with moments of order up to four, turned the theoretical framework in [19] into a practical predictability tool in [2] via applying relative entropy for information measurements in higher moments of forecast ensembles (non-Gaussianity) and in cross-moments of two-dimensional probability densities, represented by mutual information. This framework was successfully used in [3] for a triangular T21 spherical truncation of barotropic flow with realistic Earth-like topography. The same relative entropy framework was employed in [14] for rigorous estimates of the credibility of small-sampled statistical ensembles. Key results of the aforementioned work on predictability of forecast ensembles through information theory are also summarized in Chapter 1 of [18].

As follows from the context of applications in weather and climate prediction [2,3], the capability of computing the maximum entropy problem in a two- and higher-dimensional domain is highly desirable. On the other hand, computationally costly geophysical simulations with realistic models often do not yield enough statistical information to reliably compute higher order moments for an operational forecast. With that in mind, here we develop an improved maximum entropy algorithm, capable of computing the problem in a multidimensional domain with moment constraints of low and moderate order, which is a vast improvement over the simple two-dimensional four-moment algorithm, developed previously in [1]. We have to mention, however, that, although conditions for determinacy of a moment sequence in a multidimensional setting were established (see, for example [9] and references therein), there is no way to tell *a priori*, unlike in the one-dimensional Hamburger moment problem with Hankel conditions [8], whether or not the variety of probability measures corresponding to a given set of moment constraints is nonempty and includes its upper entropy bound. Yet, in numerous geophysical applications with no observable probability density state (when, for instance, the state is represented by a forecast ensemble), it is crucial to have a smooth representation of the density of a state with maximum uncertainty under given constraints, which suggests the maximum entropy problem as a natural approach for such an approximation.

The manuscript is organized as follows. In Section 2.1 an optimization problem for the Shannon entropy is formulated for a set of moment constraints in a domain of arbitrary dimension. In Section 2.2 we introduce a suitable orthogonal polynomial basis which helps to avoid numerical instabilities in the Newton iterations for higher-order constraints due to different sensitivity of the Lagrangian function to changes in low- and high-order Lagrange multipliers. Section 2.3 details the modified Gram–Schmidt algorithm, capable of maintaining

orthogonality of the polynomial basis subject to changing iterates of the target probability density function (PDF). Section 2.4 outlines the step-by-step schematics of the new algorithm for easy practical implementation. Section 3.1 displays the results of systematic numerical testing of the new algorithm for the most basic one-dimensional 4-moment maximum entropy test problem with a broad range of moment constraints and different resolutions of the Gauss–Hermite quadrature, used to calculate entries of the gradient and Hessian matrix of the Lagrangian function. In Section 3.2 we show how the algorithm approximates two test PDFs with qualitatively different shape features in the one-dimensional setting with successive order of the finite moment sequence of power up to 2, 4, 8 and 16. Section 3.3 contains the results of tests for the two-, three- and four-dimensional setting with moments of order up to 8, 6 and 4, respectively, where the test maximum entropy problems are based on the long-term statistics of the 1.5-layer quasigeostrophic double gyre model, which describes the motion of large scale oceanic currents under wind stress, such as the Gulf Stream and the Kuroshio. Section 4 summarizes the results of this work.

2. The algorithm

In this section we formulate the multidimensional maximum entropy problem and detail various technical aspects of novel numerical algorithm capable of computing the maximum entropy problem in the two- and higher-dimensional domain with moment constraints of low and moderate order.

2.1. The multidimensional maximum entropy problem

We here start with formulation of the moment-constrained maximum entropy principle in a multidimensional domain. Here we follow notations introduced earlier in [1], where for $\vec{x} \in \mathbb{R}^N$, with N being the dimension of the domain, an arbitrary moment of $\vec{x}$ (the product of arbitrary powers of components of $\vec{x}$) is concisely written as

$$\vec{x}^{\vec{i}} = \prod_{k=1}^{N} x_k^{i_k}, \quad \vec{i} \in \mathbb{I}^N \tag{1}$$

such that the moment order $|\vec{i}|$ is the total power of all vector components, i.e.

$$|\vec{i}| = \sum_{k=1}^{N} i_k. \tag{2}$$

Using the above notation, for a probability density ρ we write the set of moment constraints of the total power up to M as

$$\mu_0 = \int_{\mathbb{R}^N} \rho(\vec{x})\,\mathrm{d}\vec{x} = 1, \tag{3a}$$

$$\mu_{\vec{i}} = \bar{x}_i = \int_{\mathbb{R}^N} \vec{x}^{\vec{i}}\rho(\vec{x})\,\mathrm{d}\vec{x}, \quad |\vec{i}| = 1, \tag{3b}$$

$$\mu_{\vec{i}} = \int_{\mathbb{R}^N} (\vec{x} - \bar{\vec{x}})^{\vec{i}}\rho(\vec{x})\,\mathrm{d}\vec{x}, \quad |\vec{i}| = 2 \ldots M, \tag{3c}$$

where $\bar{x}_i$ denotes the ith component of the mean state vector $\bar{\vec{x}}$ for ρ, while moments of second and higher orders are centered at the mean state. According to the maximum entropy principle, we are looking for the probability density ρ^* which maximizes the Shannon entropy

$$S(\rho) = -\int_{\mathbb{R}^N} \rho(\vec{x}) \ln \rho(\vec{x})\,\mathrm{d}\vec{x} \tag{4}$$

(i.e. such that $S(\rho^*) = \max_\rho S(\rho)$) over all probability densities ρ which satisfy the moment constraints in (3). In the case of (3) with $N = 1$, the problem of finding a probability density function supported on the real line and satisfying the constraints is called the Hamburger moment problem [8,30], where the maximum entropy solution apparently belongs to the set of Hamburger solutions. The existence of solutions to the Hamburger

moment problem is well studied, and depends on positive-definiteness of the corresponding set of Hankel determinants [8].

It is illustrated in [1], that the formulated above constrained maximum entropy problem is further reduced to the following unconstrained optimization problem of minimizing the Lagrangian function

$$\mathscr{L}(\vec{\lambda}) = \ln \int_{\mathbb{R}^N} \exp\left(\sum_{|\vec{i}|=1}^{M} \lambda_{\vec{i}} \vec{x}^{\vec{i}} \right) \mathrm{d}\vec{x} - \sum_{|\vec{i}|=2}^{M} \lambda_{\vec{i}} \mu_{\vec{i}} \tag{5}$$

over the set of the Lagrange multipliers $\lambda_{\vec{i}}$, with corresponding optimal probability density

$$\rho(\vec{x}) = \frac{\exp\left(\sum_{|\vec{i}|=1}^{M} \lambda_{\vec{i}} (\vec{x} - \bar{\vec{x}})^{\vec{i}} \right)}{\int_{\mathbb{R}^N} \exp\left(\sum_{|\vec{i}|=1}^{M} \lambda_{\vec{i}} (\vec{x} - \bar{\vec{x}})^{\vec{i}} \right) \mathrm{d}\vec{x}}. \tag{6}$$

Due to the normalization requirement of the optimal probability density ρ in (6), the order of constraints M is restricted to even values, but is otherwise arbitrary. It is shown in [1] that the problem of unconstrained minimization of (5) over the set of Lagrange multipliers $\lambda_{\vec{i}}$ is concave, and thus if its minimum exists, then it is unique.

2.2. Improved polynomial basis for optimization

Although the minimization problem in (5) is concave, the straightforward optimization of (5) via an iterative technique like the Newton method [7], or the BFGS algorithm [5], directly over its Lagrange multipliers encounters numerical difficulty due to different sensitivity of the value of the Lagrangian function $\mathscr{L}(\vec{\lambda})$ to changes in different Lagrange multipliers $\lambda_{\vec{i}}$. Indeed, it is easy to see that the value of the Lagrangian function is not likely to respond as much to a change in a first-level Lagrange multiplier $\lambda_{\vec{i}}$ with $|i| = 1$, as it is likely to respond to a change in, say, a fifth-level Lagrange multiplier with $|i| = 5$, due to the fact that the exponential function in the integrand of the phase-space integral in (5) is more sensitive to changes in higher powers of $\vec{x}$. As a result, optimizing over the Lagrange multipliers directly often fails when the maximum moment order M exceeds the value of 6 or even 4.

Thus, in order to successfully iterate the optimization problem in (5), one has to replace the set of monomials $\vec{x}^{\vec{i}}$ with a different basis, in which the Lagrangian function in (5) has roughly same sensitivity to changes in any of the coordinates of such a basis. A simple solution is to replace basis monomials $\vec{x}^{\vec{i}}$ with a set of Mth order polynomials $p_k(\vec{x})$,

$$\begin{aligned} &\{\vec{x}^{\vec{i}}, \lambda_{\vec{i}}\}, \quad \vec{i} \in \mathbb{I}^N, \ 0 \leqslant |i| \leqslant M \ \rightarrow \ \{p_k(\vec{x}), \gamma_k\}, \quad 1 \leqslant k \leqslant K, \\ &K = \frac{(M+N)!}{M!N!} - 1, \end{aligned} \tag{7}$$

where γ_k are the Lagrange multipliers of the new basis. Since each basis polynomial $p_k(\vec{x})$ is of Mth order, the Lagrangian function should have comparable sensitivity to changes in different γ_k. Note that the constant (zero-power) term of each p_k in (7) is zero, since no constant terms enter the argument of the exponential function in (5). The Lagrangian function (5) is written in the new polynomial coordinates as

$$\mathscr{L}(\vec{\lambda}) = \ln \int_{\mathbb{R}^N} \exp\left(\sum_{k=1}^{K} \gamma_k p_k(\vec{x}) \right) \mathrm{d}\vec{x} - \sum_{k=1}^{K} \gamma_k p_k(\vec{\mu}), \tag{8}$$

where $p_k(\vec{\mu})$ above in (8) denotes the polynomial $p_k(\vec{x})$ where all the powers $\vec{x}^{\vec{i}}$ are replaced with corresponding constraints $\mu_{\vec{i}}$ from (3), except for first-order constraints, which are set to zero to comply with the sum in (5) starting at $|\vec{i}| = 2$. The corresponding optimal probability density function is now written as

$$\rho(\vec{x}) = \frac{\exp\left(\sum_{k=1}^{K} \gamma_k p_k(\vec{x} - \bar{\vec{x}}) \right)}{\int_{\mathbb{R}^N} \exp\left(\sum_{k=1}^{K} \gamma_k p_k(\vec{x} - \bar{\vec{x}}) \right) \mathrm{d}\vec{x}}. \tag{9}$$

For the one-dimensional setup (i.e. when $\vec{x}$ is a scalar), it is common to use the shifted Chebyshev polynomials [4] or the Lagrange interpolation polynomials with suitably spaced roots [31], so that comparable sensitivity of the Lagrangian function to changes in new coordinates is ensured. Although it might be possible to generalize the Chebyshev polynomials or the Lagrange interpolants to the multidimensional setup, here we instead use an adaptive system of K general orthogonal polynomials, tailored for the optimization problem.

In order to determine key properties of a suitable polynomial system, we write the formula of the Newton iterations for (8):

$$\vec{\gamma}^{n+1} = \vec{\gamma}^{n} - H^{-1}\nabla\mathscr{L}|_{\vec{\gamma}^{n}}, \tag{10}$$

where $\vec{\gamma}^{n}$ denotes the vector of Lagrange multipliers in the new basis at nth Newton iteration. Here the entries of the gradient $\nabla\mathscr{L}$ and Hessian H are given by

$$(\nabla\mathscr{L})_k = \frac{\partial\mathscr{L}}{\partial\gamma_k} = \frac{Q[p_k]}{Q[1]} - p_k(\vec{\mu}), \tag{11a}$$

$$(H)_{kl} = \frac{\partial^2\mathscr{L}}{\partial\gamma_k\partial\gamma_l} = \frac{Q[p_k p_l]}{Q[1]} - \frac{Q[p_k]Q[p_l]}{(Q[1])^2}, \tag{11b}$$

where the building block $Q[g]$ for an arbitrary function $g(\vec{x})$ is computed as

$$Q[g] = \int_{\mathbb{R}^N} g(\vec{x}) \exp\left(\sum_{k=1}^{K} \gamma_k p_k\right) \mathrm{d}\vec{x}. \tag{12}$$

The integral in (12) is computed through the Gauss–Hermite quadrature, as suggested in [1]. It is easy to see that the optimization problem in new coordinates remains concave: one writes the Hessian in (11b) in the form

$$H = \frac{1}{Q[1]} Q\left[\left(p_k - \frac{Q[p_k]}{Q[1]}\right) \otimes \left(p_l - \frac{Q[p_l]}{Q[1]}\right)\right],$$

and thus the inner product $\vec{v}^T H\vec{v}$ for an arbitrary vector $\vec{v}$ is

$$\vec{v}^T H\vec{v} = \frac{1}{Q[1]} Q\left[\left(\sum_k v_k \left(p_k - \frac{Q[p_k]}{Q[1]}\right)\right)^2\right] \geqslant 0.$$

In order to improve numerical stability of the Newton iterations in (10), we require the following orthogonality condition for the polynomials p_k at each step of the Newton iterations:

$$Q[p_k p_l] = Q[1]\delta_{pl}, \tag{13}$$

where δ_{pl} is the usual Kronecker delta-symbol. The orthogonality requirement in (13) also provides roughly same sensitivity of the Lagrangian function in (8) to changes in different γ_k due to comparable curvatures of the second-order surface approximation to $\mathscr{L}$ in all directions at the current iteration point, and turns the formula in (11b) for the Hessian matrix into

$$(H)_{kl} = \frac{\partial^2\mathscr{L}}{\partial\gamma_k\partial\gamma_l} = \delta_{kl} - \frac{Q[p_k]Q[p_l]}{Q[1]^2}. \tag{14}$$

With (14), the inverse Hessian matrix H^{-1} is computed explicitly via the Sherman–Morrison formula

$$(H^{-1})_{kl} = \delta_{kl} + \frac{Q[p_k]Q[p_l]}{(Q[1])^2 - \sum_{m=1}^{K}(Q[p_m])^2} \tag{15}$$

such that the next iterate in (10) is readily available. For details on the Sherman–Morrison formula see, for example, [13].

2.3. Orthogonalization of the polynomial basis

In the previous section we have formulated the orthogonality condition in (13) for the basis polynomials in (7) in order to improve numerical stability of the Newton iterations in (10) for the minimization of the Lagrangian function in (8). However, the quadratures Q in (12) are weighted by the probability density function of the form (9) which changes between Newton iterations, and thus the orthogonal relation in (13) computed at nth iteration is not necessarily preserved at $(n+1)$th iteration. Therefore, a Gram–Schmidt [26] type of polynomial reorthogonalization of p_k is required at each step of Newton iteration. Poor numerical stability of the classical Gram–Schmidt reorthogonalization is widely known, and common numerically stable tools to reorthogonalize a set of Euclidean vectors usually involve the Householder reflections or Givens rotations [13]. However, it is not clear whether an analog of Householder or Givens decomposition exists for polynomials of an arbitrary dimension and order, and a suitable stabilized modification of the Gram–Schmidt algorithm will be used instead.

Recent works [6,10–12] demonstrate that the modified Gram–Schmidt algorithm yields a good precision for vectors which are not "too ill-conditioned". In particular, it is demonstrated in [11] that the modified Gram–Schmidt algorithm with reorthogonalization yields errors which are small multiples of the machine round-off error. Here, since the reorthogonalization of p_k has to be carried out at each Newton iteration to preserve orthogonality of polynomials, it is reasonable to assume that the metric defined by (12) does not change between iterations as much so that an orthogonal system of polynomials at the current step of the Newton iterations becomes too ill-conditioned at the next step.

According to the classical Gram–Schmidt method, given K arbitrary linearly independent polynomials a_k, their orthogonal in the sense of (13) counterparts p_k are computed as

$$p_k = (Q[1])^{1/2} \frac{a_k - \sum_{l=1}^{k-1} \frac{Q[a_k p_l]}{Q[1]} p_l}{\left(Q \left[a_k - \sum_{l=1}^{k-1} \frac{Q[a_k p_l]}{Q[1]} p_l \right]^2 \right)^{1/2}}, \quad 1 \leqslant k \leqslant K. \tag{16}$$

As mentioned before, the classical Gram–Schmidt method is numerically unstable, and here we use a suitably modified version of the Gram–Schmidt method with reorthogonalization from [12], tailored for a polynomial basis. It is not convenient to illustrate the modified Gram–Schmidt method as a formula in (16) due to recursive nature of its computational implementation, and instead we give a step-by-step program of the algorithm as in [12] in a fashion which resembles some modern computer languages:

Modified Gram–Schmidt algorithm with reorthogonalization

```
for k = 1,...,K {
    for l = 1,2 {
      a_k^l = a_k^{l-1}
      for m = 1,...,k-1
        a_k^l = a_k^l - Q[a_k^l p_m] p_m
    }
    p_k = (Q[(p_m)^2])^{-1/2} a_k
}
for k = 1,...,K
    p_k = (Q[1])^{1/2} p_k
```

Although the classical Gram–Schmidt algorithm in (16) is twice as fast as the modified Gram–Schmidt algorithm with reorthogonalization described above, the latter has much better numerical stability, which makes it more suitable for our purposes.

2.4. Schematic outline of the algorithm

This section illustrates a schematic step-by-step outline of the algorithm, suitable as a set of general guidelines for its practical numerical implementation. The preconditioning of input constraints, choice of starting

point for Newton iterations and numerical algorithm for computing Gauss–Hermite quadratures in (12) are the same as in [1].

(1) Precondition input constraints by setting zero mean and identity covariance as in [1].
(2) Generate a set of K random linearly independent polynomials p_k of Mth order with all their zero-power terms set to zero.
(3) Choose the set of Lagrange multipliers γ_k corresponding to the Gaussian distribution with zero mean and identity covariance for the starting point ρ_0 of Newton iterations as in [1] in the new polynomial basis p_k.
(4) Reorthogonalize the set of polynomials p_k according to (13) with respect to the current iterate of ρ, using the Gauss–Hermite quadratures to compute integrals Q in (12) as in [1], and recompute the set of Lagrange multipliers γ_k with respect to the reorthogonalized basis.
(5) Compute the gradient and inverse Hessian of the new iterate of the Lagrangian function $\mathscr{L}$ via the Sherman–Morrison formula in (15).
(6) Perform one step of the modified (guarded) Newton iteration according to (10) with inexact line search as in [1], using the new gradient and inverse Hessian matrix.
(7) If the minimum of the Lagrangian function is not reached, return to step 4; otherwise, recompute the optimal γ_k into the set of standard Lagrange multipliers $\lambda_{\vec{i}}$ for the monomial basis $\vec{x}^{\vec{i}}$, to match the format of the input constraints.

The algorithm is implemented for an arbitrary phase space dimension N and an arbitrary order of input constraints M, using the object-oriented style of the C++ programming language. However, practical limitations on N and M are imposed by computational speed of machine-specific hardware.

3. Numerical experiments

Here we show the results of the systematic testing of the new algorithm in several practical situations. In particular, first we present an extensive study of convergence properties of the algorithm in a simplest 4-moment one-dimensional set-up for different resolutions of the Gauss–Hermite quadrature, followed by a number of higher-order multidimensional tests, based on a real-world geophysical model for wind stress driven oceanic currents, to demonstrate practical computational ability of the new algorithm.

3.1. Basic test for maximum entropy: the one-dimensional 4-moment problem

The 4-moment, one-dimensional maximum entropy problem is a simplest set-up which requires numerical simulation to compute the optimal set of Lagrange multipliers; an optimal solution to the 2-moment problem is automatically the Gaussian distribution with prescribed mean state and variance. Hence, the 4-moment problem is chosen as a basic maximum entropy test due to its computational simplicity and the fact that for a fixed mean state and variance it has only two degrees of freedom, the third and fourth moments, also called skewness and flatness, respectively:

$$\begin{aligned} \text{Skewness} &= \mu_3 = \int_{\mathbb{R}} (x-\bar{x})^3 \rho(x)\,\mathrm{d}x, \\ \text{Flatness} &= \mu_4 = \int_{\mathbb{R}} (x-\bar{x})^4 \rho(x)\,\mathrm{d}x. \end{aligned} \tag{17}$$

Without loss of generality we set the mean state, $\bar{x}$, to zero, and variance, μ_2, to 1, because it is demonstrated in [1] that the multidimensional maximum entropy problem of an arbitrarily high order can be explicitly rescaled to an equivalent problem with zero mean state and identity covariance matrix. The fact that we have only two variable parameters, skewness and flatness, allows us to study basic convergence properties of the algorithm for a broad range of these parameters.

3.1.1. Area of convergence for the 4-moment one-dimensional problem

It has been shown in Section 2.2 that the Lagrangian minimization problem in polynomial coordinates (8) is concave, which means that it has at most one optimal solution. However, an optimal solution does not necessarily exist for a set of arbitrarily specified moment constraints. The existence of solutions to the M-moment (for even M) Hamburger moment problem is specified by a set of Hankel conditions

$$\Delta_k = \det\begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_{k/2} \\ \mu_1 & \mu_2 & \cdots & \mu_{k/2+1} \\ \vdots & \vdots & & \vdots \\ \mu_{k/2} & \mu_{k/2+1} & \cdots & \mu_k \end{pmatrix} > 0, \quad k = 0, 2, 4, \ldots, M. \tag{18}$$

For the 4-moment Hamburger problem with zero mean state and unit variance Δ_0 and Δ_2 are automatically positive, and thus the existence condition is given by a single Hankel determinant

$$\Delta_4 = \det\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & \mu_3 \\ 1 & \mu_3 & \mu_4 \end{pmatrix} > 0, \tag{19}$$

which imposes the following relation between the skewness μ_3 and flatness μ_4:

$$\text{4-moment maximum entropy solution exists if } \mu_4 > \mu_3^2 + 1. \tag{20}$$

Therefore, here we study convergence of the new algorithm over the two-dimensional area inside the parabolic test curve defined by

$$\mu_4 = \mu_3^2 + 1. \tag{21}$$

In Figs. 1 and 2, we show the results of the 4-moment one-dimensional convergence test for the new maximum entropy algorithm on the $\mu_3 \times \mu_4$ domain with fixed $\mu_1 = 0$ and $\mu_2 = 1$. The test is performed for the Gauss–Hermite quadratures of (12) with four different resolutions: 20, 30, 40 and 50 discretization points. All convergence tests are performed within the area of existing solutions given by (20), and designated by an appropriate curve in Figs. 1 and 2. One can observe that the algorithm fails to converge at a few points in close proximity to the convergence area boundary limited by (21), and the number of failures is between 4.7% and 5.9% of the total number of simulations (systematically decreasing as resolution of the Gauss–Hermite quadrature increases). Overall, the results show a reasonably good practical range of convergence of the new algorithm, with failures only at "pathological" points near the boundary (21) defined by the existence condition in (20). The Gauss–Hermite quadrature with 50 discretization points in each dimension is used for further numerical experiments in the current work.

3.1.2. Integrability of optimal 4-moment PDFs

An optimal one-dimensional 4-moment probability density function (PDF) with zero mean state is defined by a set of four Lagrange multipliers $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ as

$$\begin{aligned} &\rho_4(x) = \exp(\lambda_0 + \lambda_1 x + \lambda_2 x^2 + \lambda_3 x^3 + \lambda_4 x^4) \\ &\text{with} \\ &\lambda_0 = -\ln \int_{\mathbb{R}} \exp\left(\lambda_1 x + \lambda_2 x^2 + \lambda_3 x^3 + \lambda_4 x^4\right) \mathrm{d}x. \end{aligned} \tag{22}$$

It is clear that for $\lambda_4 > 0$ the probability density $\rho_4(x)$, given by (22), is not integrable on $\mathbb{R}$ due to its exponential growth as $x \to +\infty$ or $x \to -\infty$. On the other hand, the Gauss–Hermite quadratures in (12) are computed over a finite set of discretization points (roots of an appropriate Gauss–Hermite polynomial) for which $\rho_4(x)$ even with $\lambda_4 > 0$ does not achieve infinite values. Therefore, it is quite possible for the new algorithm to converge at an "optimal" set of Lagrange multipliers with $\lambda_4 > 0$ which indeed maximizes entropy within the context of finite Gauss–Hermite quadratures in (12), but is not valid for the 4-moment Hamburger problem on $\mathbb{R}$. Thus, we need to find the subset of the set of existing solutions given by (20) which corresponds to optimal $\rho_4(x)$ with $\lambda_4 < 0$, and check whether the new algorithm provides valid solutions within this domain.

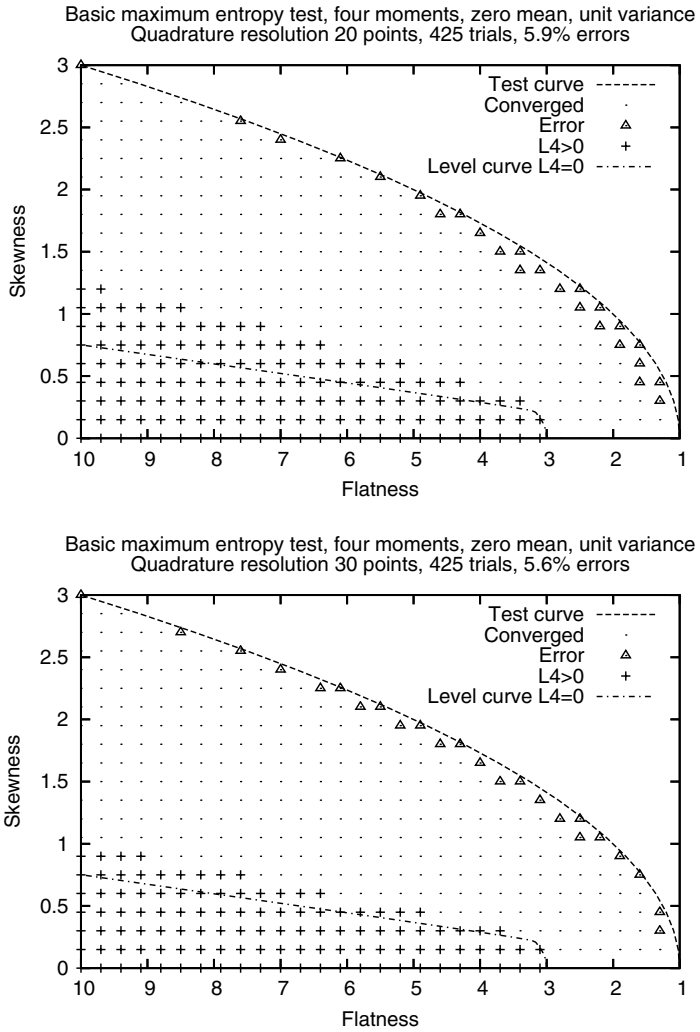


Fig. 1. The domain of convergence of the maximum entropy algorithm for the 4-moment one-dimensional maximum-entropy problem with zero mean and unit variance. The amplitudes of skewness and flatness are limited at 3 and 10, respectively. The resolution of Gauss–Hermite quadratures is 20 and 30 points. The dashed test curve designates the area of existence for optimal solutions. Non-normalizable on $\mathbb{R}$ optimal solutions ($\lambda_4 > 0$) are denoted as "+". Level curve $\lambda_4 = 0$ is drawn in a dot-dashed style.

In order to distinguish between the sets of optimal solutions $\rho_4(x)$ with positive and negative λ_4 in the $\mu_3 \times \mu_4$ plane (with μ_1 and μ_2 fixed at 0 and 1, respectively), one has to determine a level curve in this domain which corresponds to $\lambda_4 = 0$ and separates the regions of positive and negative λ_4 from each other. Taking into account the form of the optimal probability density in (6), or, equivalently, in (22), the following relation holds between the moment μ_i and the Lagrange multiplier λ_j:

$$\frac{\partial \mu_i}{\partial \lambda_j} = \mu_{i+j}, \tag{23}$$

and, therefore, for the 4-moment one-dimensional problem one can write

$$\begin{pmatrix} \mathrm{d}\mu_1 \\ \mathrm{d}\mu_2 \\ \mathrm{d}\mu_3 \\ \mathrm{d}\mu_4 \end{pmatrix} = \begin{pmatrix} \mu_2 & \mu_3 & \mu_4 & \mu_5 \\ \mu_3 & \mu_4 & \mu_5 & \mu_6 \\ \mu_4 & \mu_5 & \mu_6 & \mu_7 \\ \mu_5 & \mu_6 & \mu_7 & \mu_8 \end{pmatrix} \begin{pmatrix} \mathrm{d}\lambda_1 \\ \mathrm{d}\lambda_2 \\ \mathrm{d}\lambda_3 \\ \mathrm{d}\lambda_4 \end{pmatrix}. \tag{24}$$

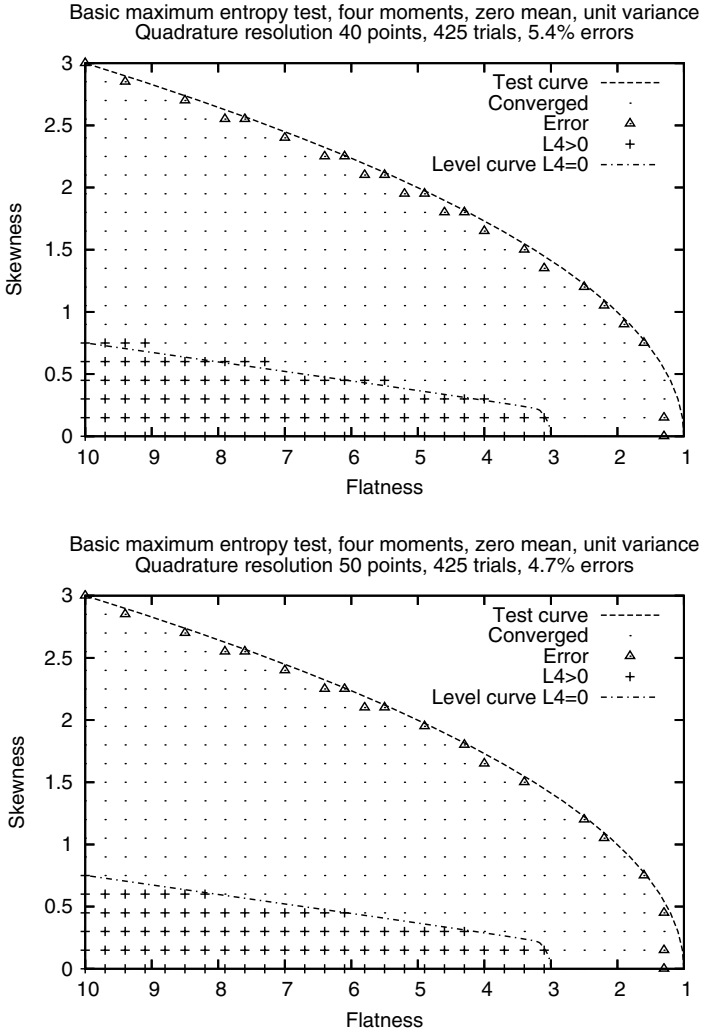


Fig. 2. The domain of convergence of the maximum entropy algorithm for the 4-moment one-dimensional maximum-entropy problem with zero mean and unit variance. The amplitudes of skewness and flatness are limited at 3 and 10, respectively. The resolution of Gauss–Hermite quadratures is 40 and 50 points. The dashed test curve designates the area of existence for optimal solutions. Non-normalizable on $\mathbb{R}$ optimal solutions ($\lambda_4 > 0$) are denoted as "+". Level curve $\lambda_4 = 0$ is drawn in a dot-dashed style.

By setting $d\lambda_4 = 0$ (since we are looking for its level curve) and $d\mu_1 = 0$, $d\mu_2 = 0$ (since μ_1 and μ_2 are fixed at 0 and 1, respectively), from (24) we obtain

$$\begin{aligned} &\mu_2 d\lambda_1 + \mu_3 d\lambda_2 + \mu_4 d\lambda_3 = 0, \\ &\mu_3 d\lambda_1 + \mu_4 d\lambda_2 + \mu_5 d\lambda_3 = 0, \end{aligned} \tag{25}$$

which yields the following set of ordinary differential equations

$$\begin{aligned} \frac{d\lambda_2}{d\lambda_1} &= \frac{\mu_3\mu_4 - \mu_2\mu_5}{\mu_3\mu_5 - \mu_4^2}, \\ \frac{d\lambda_3}{d\lambda_1} &= \frac{\mu_2\mu_4 - \mu_3^2}{\mu_3\mu_5 - \mu_4^2}, \end{aligned} \tag{26}$$

where λ_1 plays the role of the independent variable. At this point, taking into account (24) and (26), one writes the equation for the level curve of λ_4 in the $\mu_3 \times \mu_4$ domain as

$$\frac{d\mu_4}{d\mu_3} = \frac{\mu_3\mu_5^2 - \mu_4^2\mu_5 + \mu_3\mu_4\mu_6 - \mu_2\mu_5\mu_6 + \mu_2\mu_4\mu_7 - \mu_3^2\mu_7}{2\mu_3\mu_4\mu_5 - \mu_4^3 - \mu_2\mu_5^2 + \mu_2\mu_4\mu_6 - \mu_3^2\mu_6}. \tag{27}$$

In practice, of course, we do not solve (27) directly: instead, the system in (26) is numerically solved with the initial condition $\lambda_1 = \lambda_3 = \lambda_4 = 0$, $\lambda_2 = -0.5$, which corresponds to the Gaussian distribution with zero mean state and unit variance. Then, the solution of (26) is mapped onto the $\mu_3 \times \mu_4$ domain via (17).

In Figs. 1 and 2 we show the solution curve of (26) which corresponds to the fixed fourth Lagrange multiplier $\lambda_4 = 0$, and also record the points were the new algorithm has converged, but the obtained optimal 4-moment probability density is not integrable on $\mathbb{R}$ due to positive λ_4. As one can observe, for relatively low resolution 20 and 30 point Gauss–Hermite quadratures there are a few optimal points outside the domain bounded by (26) which yield positive λ_4, however, as resolution increases to 40 and 50 points, the number of registered outcomes with positive λ_4 decreases rapidly to zero.

3.1.3. Errors in PDFs for different Gauss–Hermite quadrature resolutions

Even if the new algorithm converges for a given set of constraints (i.e. the stopping criterion for the Newton iterations is satisfied within the machine-precision tolerance), the computed optimal Lagrange multipliers are not necessarily precise due to errors in finite quadrature. While it is true that for a Gaussian probability density the Gauss–Hermite quadrature yields exact values of moments of up to the order of the quadrature itself, the quadrature errors must develop when the optimal probability density departs from Gaussian. Thus, the optimal Lagrange multipliers might not be computed precisely, and, therefore, the shape of thus computed optimal probability density function should differ from the "true" optimal PDF. One, of course, cannot compute the optimal PDFs precisely due to finite resolution of the Gauss–Hermite quadrature. However, we solve the maximum entropy problem for identical sets of constraints with successively increasing quadrature resolutions, and observe how the shape of the optimal PDF changes with better quadrature precision. Here we pick four sets of constraints for the test problem with zero mean state $\mu_1 = 0$ and unit variance $\mu_2 = 1$:

(A) Skewness $\mu_3 = 0.15$ and flatness $\mu_4 = 1.4$ – this set corresponds to a slightly skewed and strongly bimodal sub-Gaussian distribution with two relatively high and narrow peaks;
(B) Skewness $\mu_3 = 0.6$ and flatness $\mu_4 = 2$ – this is a skewed and bimodal distribution, although not as bimodal as a strongly bimodal example above;
(C) Skewness $\mu_3 = 1.4$ and flatness $\mu_4 = 5$ – this is a strongly skewed but unimodal distribution with a long and weighty "tail" on the right-hand side of the origin;
(D) Skewness $\mu_3 = 0$ and flatness $\mu_4 = 3$ – this is a Gaussian distribution, for which the Gauss–Hermite quadrature happens to be exact. It is shown as a control test to demonstrate the machine precision of the Gauss–Hermite quadrature for the Gaussian distribution regardless of order.

We show the results of the Gauss–Hermite quadrature precision test in Figs. 3 and 4, and also summarize the L_2-errors for computed PDFs in Table 1. The L_2-errors between two PDFs ρ_1 and ρ_2 are computed as follows:

$$\|\rho_2 - \rho_1\|_2 = \sqrt{\int_{\mathbb{R}} (\rho_2 - \rho_1)^2 \, dx}. \tag{28}$$

One can observe in Figs. 3 and 4 that, from visual perspective, the largest errors due to different Gauss–Hermite quadrature resolutions develop for the two bimodal examples (A) and (B), defined above. Apparently, this happens due to the fact that high and narrow peaks of the displayed PDFs are hard to capture with a low-resolution quadrature, because the discretization points of the quadrature might not fall directly onto a peak. Thus, an increase in Gauss–Hermite quadrature resolution leads to a substantial improvement in the shape of PDFs. On the contrary, there are no visible discrepancies between different Gauss–Hermite quadrature resolutions for the test case (C), which refers to a strongly skewed but unimodal distribution. Apparently, even low-resolution quadratures are able to capture the shape of the distribution in (C) quite precisely due to its smoothness and despite the fact that it is strongly non-Gaussian. For the test case (D), which is a Gaussian

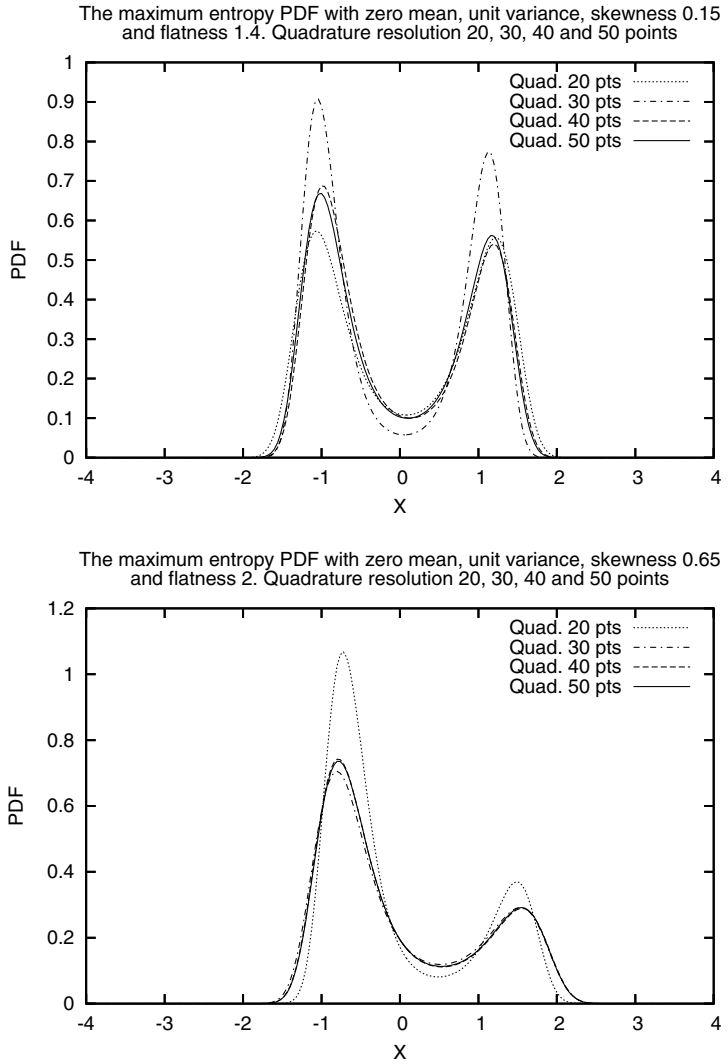


Fig. 3. The 4-moment one-dimensional optimal PDFs for values of skewness and flatness set to 0.15 and 1.4, and 0.65 and 2, respectively, computed at different quadrature resolutions.

distribution, no visible differences between PDFs are observed, as expected. The summary in Table 1 essentially confirms the results in Figs. 3 and 4: namely, the largest L_2-errors for computed PDFs can be observed for strongly bimodal distributions, while a strongly skewed but unimodal distribution can be computed relatively precisely even by low-resolution quadratures. The errors for a Gaussian distribution are on the level of machine-precision error, as expected.

3.2. Higher-moment one-dimensional problems

Naturally, moments of high order are not the best set of constraints to detail an approximation of a probability density function via maximum entropy principle, and there are better ways to increase resolution of probability density approximation in the one-dimensional setting like recursion coefficients (see, for example [15,16,31] and references therein). Although the new algorithm is designed to operate mainly in the generalized setting of two- and higher-dimensional phase space with low to moderate order of constraints, in this section we intend to demonstrate that the algorithm is numerically robust in the one-dimensional setting with relatively high order of moment constraints.

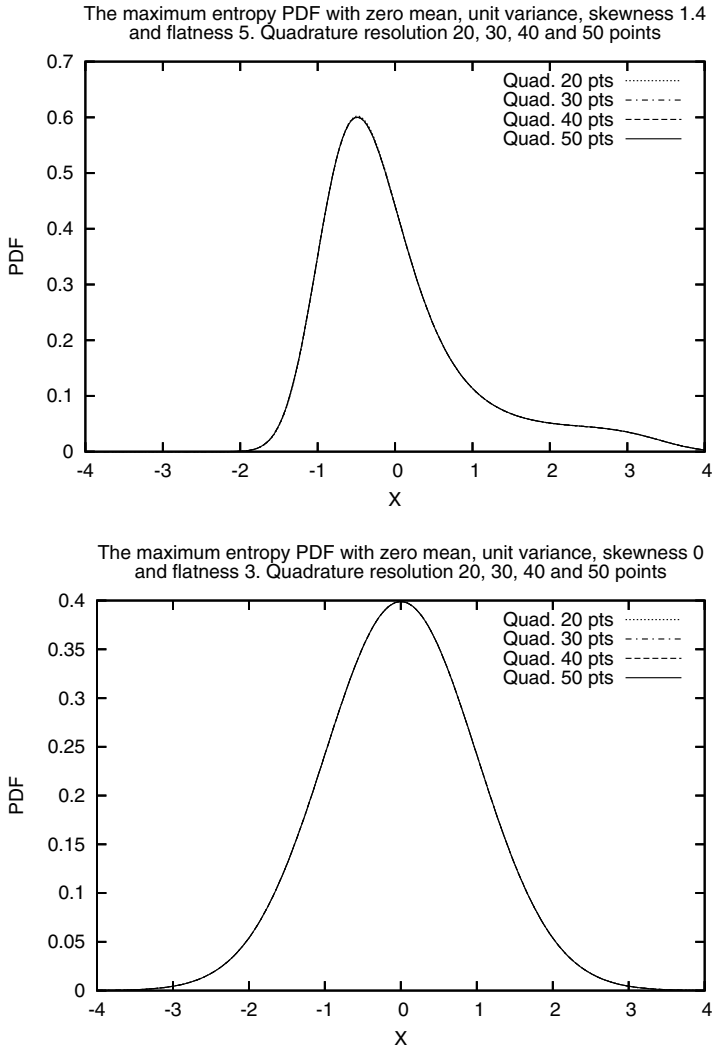
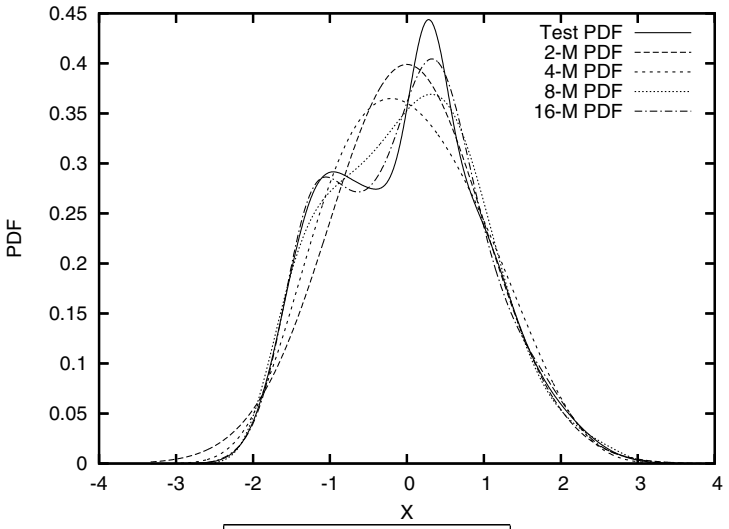


Fig. 4. The 4-moment one-dimensional optimal PDFs for values of skewness and flatness set to 1.4 and 5, and 0 and 3, respectively, computed at different quadrature resolutions.

Table 1
The L_2-errors between 4-moment one-dimensional optimal PDFs computed at different quadrature resolution, for different values of skewness and flatness

Quad. res., pts.	20 vs 50	30 vs 50	40 vs 50
Skew = 0.15, flat = 1.4	0.09709	0.1922	0.0452
Skew = 0.65, flat = 2	0.2246	0.03372	0.7318×10^{-2}
Skew = 0, flat = 3	0.9105×10^{-15}	1.459×10^{-16}	0.69×10^{-15}
Skew = 1.4, flat = 5	2.804×10^{-3}	3.113×10^{-4}	0.541×10^{-4}

For the testing of the new algorithm we pick two probability density functions with moderately complex shapes. The first test probability density, displayed in Fig. 5, is a superposition of six Gaussian curves with various positions, widths and heights, and is a largely unimodal and positively skewed with a small bump on the left-hand slope. The second test probability density, shown in Fig. 6 is qualitatively different: it is a



L_2-error table

μ_{max}	error
2	0.09731
4	0.0917
8	0.05484
16	3.053×10^{-2}

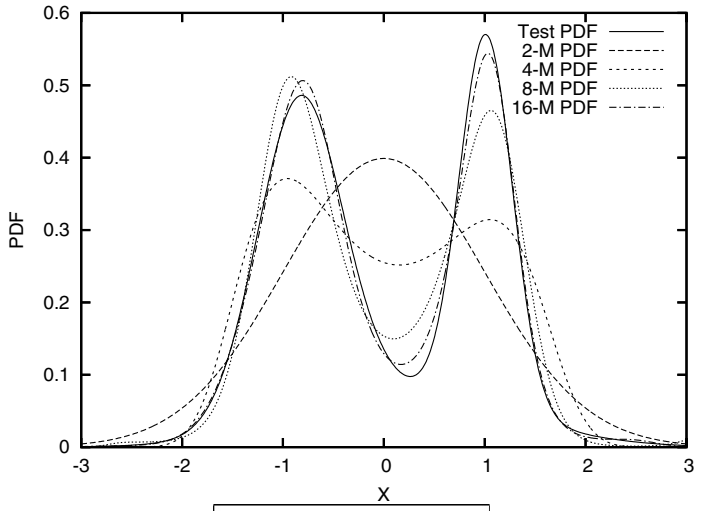
Fig. 5. The first test PDF and its 2-, 4-, 8- and 16-moment approximations, with the table of errors.

superposition of only two Gaussian curves, however those curves are sufficiently distant from each other to give a distinctively bimodal shape to the second test probability density. We then construct the 2-, 4-, 8- and 16-moment maximum entropy approximations to each of these probability density functions and measure the L_2-errors between the test probability density functions and their approximations in the same manner as in (28).

The maximum entropy approximations and their L_2-errors for the first test PDF are shown in Fig. 5. One can see that the increase in the order of moment constraints visually improves the similarity between the test PDF and the corresponding approximation. This trend is also confirmed in the table of L_2-errors in the same figure. For the second test PDF in Fig. 6 the situation is similar, namely, an improvement in maximum entropy approximation is observed for increased order of moment constraints both in visual shape and L_2-error. However, one can see that the improvement from eighth order to 16th order (which means having twice as many constraints) yields merely a moderate improvement in the approximation despite supplying a lot more information in the latter case and vastly increasing numerical complexity thereof. This observation is an indirect confirmation of the fact that moment constraints naturally do not supply detailed information about finer details of the shape of a PDF.

3.3. Two- and higher-dimensional problems

In this section we demonstrate the ability of the algorithm to compute optimal maximum entropy probability density functions for two- and higher-dimensional problems. Here, rather than generating synthetic tests PDFs like those in the previous section for the one-dimensional problems, we show a few illustrative examples from a real-world geophysical model. The model we consider simulates large scale oceanic currents which are



L_2-error table	
μ_{max}	error
2	0.3418
4	0.2133
8	0.09573
16	$3.147 \text{x} 10^{-2}$

Fig. 6. The second test PDF and its 2-, 4-, 8- and 16-moment approximations, with the table of errors.

formed by the boundary layer separation at eastern boundaries of continents in the Northern hemisphere due to wind stress-driven ocean circulation, namely the Gulf Stream, which separates from the eastern boundary of Northern America, and the Kuroshio, originating at the eastern boundary of the Asian continent. The model consists of the 1.5-layer quasigeostrophic double gyre equation, suggested by McCalpin and Haidvogel [20,21] for studies of the wind-driven ocean circulation. The simulation is performed for the reference set of physical parameters set forth by McCalpin and Haidvogel [20,21] and corresponds to the wind-driven flow in a 3600×2800-km rectangular basin at 30° Northern latitude, which is, of course, an idealized setting for complex flows such as the Gulf Stream, which are also governed by large scale features of the nearby coastline. The statistical quantities observed are so-called empirical orthogonal functions (EOFs) and their principal components (PCs), whose physical meaning is explained below.

Operational high-resolution models, designed to simulate real-world geophysical phenomena, may have tens and hundreds of thousands spatial discretization points in order to capture essential features of the phenomenon which is being simulated. In particular, in the 1.5-layer quasigeostrophic double gyre model used in the current paper as an example, the numerical solution is computed at 25,200 discretization points for each time step. As a result, for such a sophisticated model it is highly desirable to have a suitable coordinate system which captures essential features of large-scale behavior along just a few carefully chosen directions. In the atmosphere/ocean science community, these directions are called the empirical orthogonal functions (EOFs) and are computed as appropriately normalized eigenvectors of the long-term covariance matrix of the model. EOFs are sorted in descending order of magnitude of their corresponding eigenvalues, such that the first EOF is aligned with the direction of largest variability in the system, the second EOF is aligned with the direction of next largest variability which is orthogonal to the first, and so forth. As a result, EOFs are highly efficient in capturing large scale features of the model, and usually first five or six EOFs explain more the 99% of variability in the model. Note that EOFs are useful mainly from diagnostic standpoint; truncating the model itself

to the first few EOFs to simplify computations normally does not yield consistent behavior except, perhaps, for very few special cases. Coordinates of a solution in the EOF basis are called principal components (PCs). Conventionally, EOFs are normalized in such a way that the long-term covariance matrix of their principal components is the identity matrix.

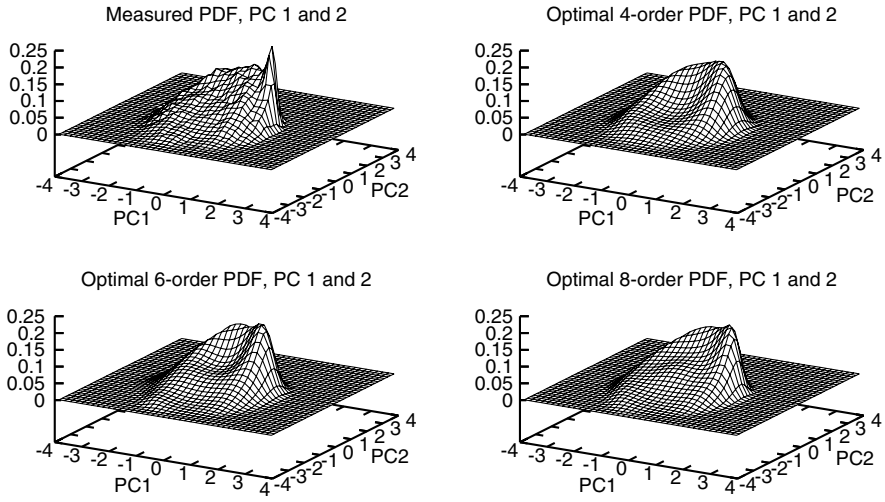


Fig. 7. The two-dimensional probability density functions of principal components 1 and 2 for the 1.5-layer quasigeostrophic double gyre model. Upper-left – actual PDF recorded from model simulation by bin-counting, upper-right – 4-moment optimal PDF (14 constraints), lower-left – 6-moment optimal PDF (27 constraints), lower-right – 8-moment optimal PDF (44 constraints).

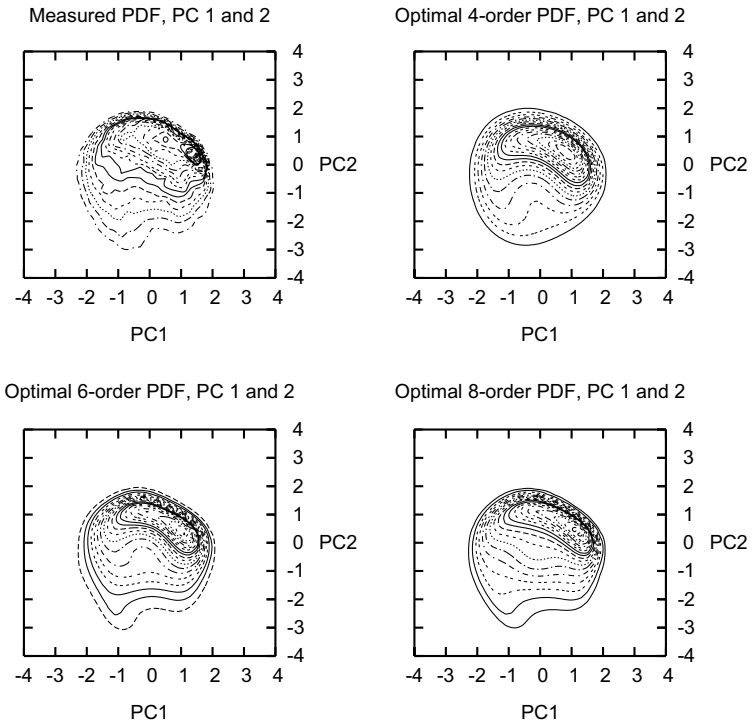


Fig. 8. The two-dimensional probability density functions of principal components 1 and 2 for the 1.5-layer quasigeostrophic double gyre model. Upper-left – actual PDF recorded from model simulation by bin-counting, upper-right – 4-moment optimal PDF (14 constraints), lower-left – 6-moment optimal PDF (27 constraints), lower-right – 8-moment optimal PDF (44 constraints). Height difference between successive contours is 0.01.

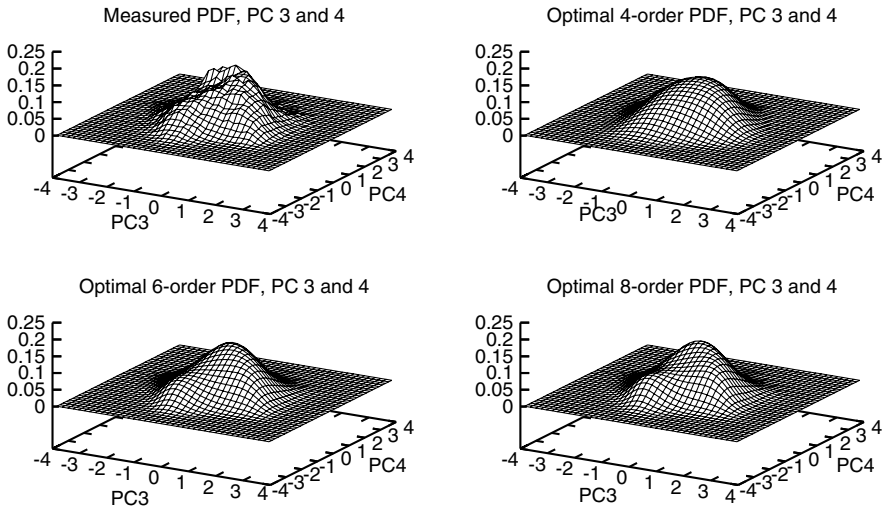


Fig. 9. The two-dimensional probability density functions of principal components 3 and 4 for the 1.5-layer quasigeostrophic double gyre model. Upper-left – actual PDF recorded from model simulation by bin-counting, upper-right – 4-moment optimal PDF (14 constraints), lower-left – 6-moment optimal PDF (27 constraints), lower-right – 8-moment optimal PDF (44 constraints).

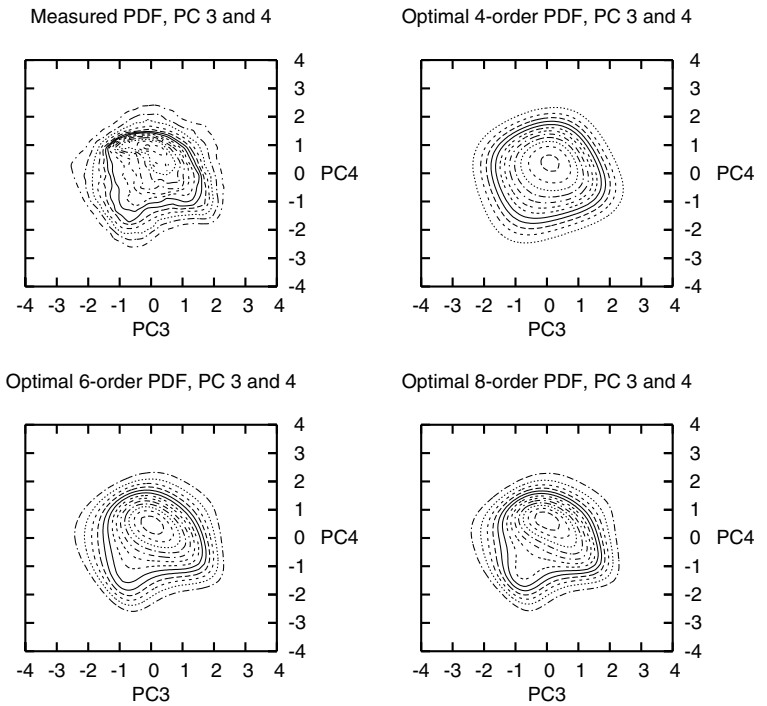


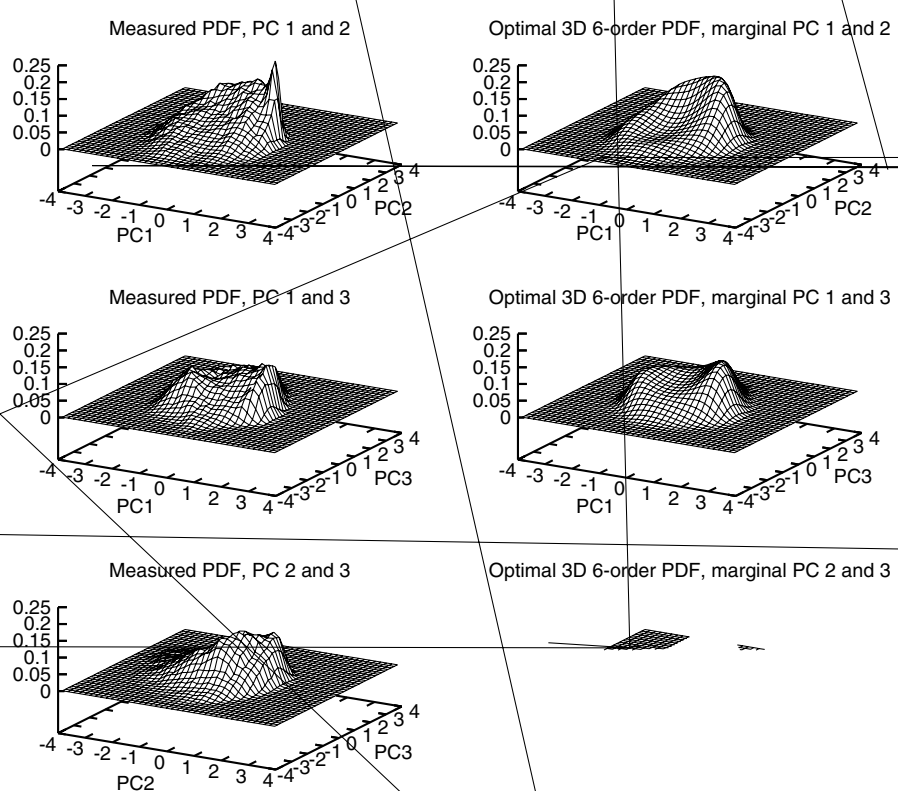
Fig. 10. The two-dimensional probability density functions of principal components 3 and 4 for the 1.5-layer quasigeostrophic double gyre model. Upper-left – actual PDF recorded from model simulation by bin-counting, upper-right – 4-moment optimal PDF (14 constraints), lower-left – 6-moment optimal PDF (27 constraints), lower-right – 8-moment optimal PDF (44 constraints). Height difference between successive contours is 0.01.

3.3.1. Two-dimensional problems

In Figs. 7 and 8 we show the recorded, via a standard bin-counting algorithm, long-term two-dimensional joint probability density functions for principal components PC1 and PC2 of the EOF basis for the 1.5-layer quasigeostrophic double gyre model, together with optimal maximum entropy PDFs obtained for the sets of moments of order up to 4, 6 and 8. Due to the fact that the maximum entropy problems here are two-dimensional, the corresponding number of constraints for these problems, according to the formula in (7), is 14, 27 and 44, respectively. Observe that while neither of the optimal PDFs approximates small scale spikes of the

Table 2
L_2-errors between optimal and actual PDFs, and values of entropy of optimal PDFs for different sets of moments

Order	Error	Entropy
	PDF errors and entropy, PC1 and 2	
4	0.0544	2.646
6	0.04639	2.629
8	0.03449	2.62
	PDF errors and entropy, PC3 and 4	
4	0.03445	2.801
6	2.858×10^{-2}	2.785
8	2.45×10^{-2}	2.783



actual recorded histogram, the amount and quality of reproduced large and moderate scale features systematically increases with increasing order of constraints. Same types of PDFs, but for different joint distribution of principal components PC3 and PC4, are shown in Figs. 9 and 10, where one can observe the same trends in approximation and detailing of optimal maximum entropy PDFs as in Figs. 7 and 8. For systematic comparison of optimal PDFs of different order, in Table 2 we display L_2-errors between optimal maximum entropy PDFs and the actual recorded PDFs, as well as values of the Shannon entropy for optimal PDFs. Observe that while errors between actual and optimal PDFs systematically decrease as more constraints are used in corresponding maximum entropy problems, the values of entropy also decrease with increasing order of constraints. The latter occurs due to the fact that introducing more constraints can only decrease entropy by capturing more information about the same PDF [2,3,19]. Also note that the errors do not decrease significantly with increasing order of constraints. Doubling the order (4–8), which results in tripling the number

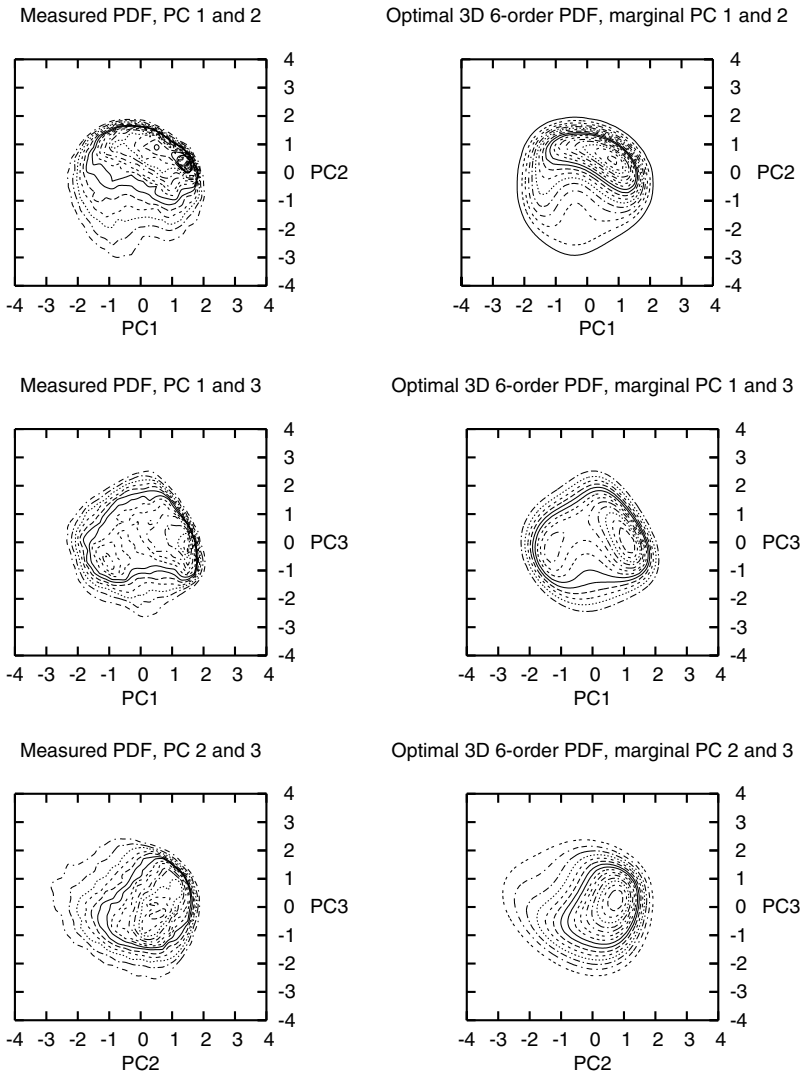
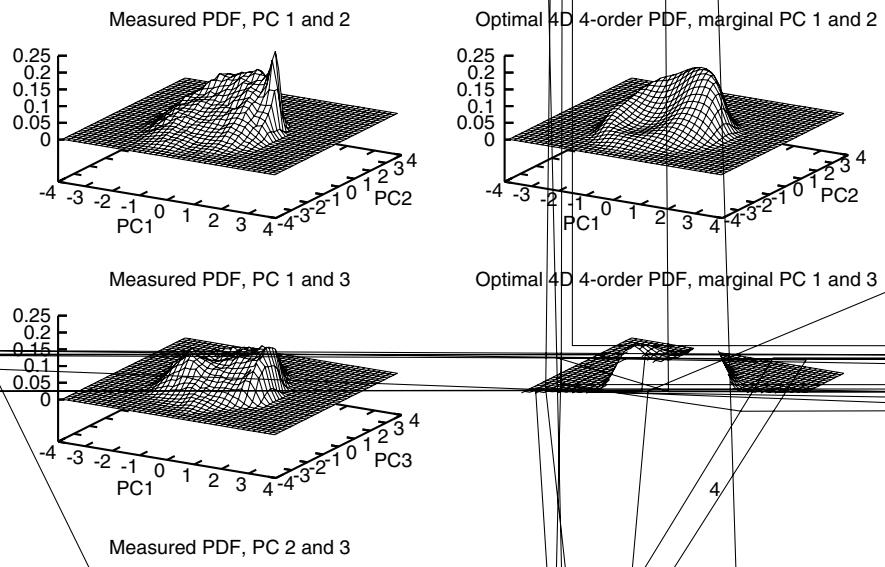


Fig. 12. Comparison of three-dimensional 6-order (83 constraints) optimal PDF with actual histograms recorded from the model via bin-counting. Left column – the recorded joint distributions of PC1 and PC2, PC1 and PC3, PC2 and PC3. Right column – the corresponding two-dimensional marginal distributions of the single three-dimensional optimal PDF, spanning PC1, PC2 and PC3. Height difference between successive contours is 0.01.



of constraints (14–44), typically yields error decrease by only 30–40%. This is an evidence of slow convergence of the moment sequence in the multidimensional Hamburger moment problem. For more details, see [9].

3.3.2. Three- and four-dimensional problems

In addition to the tests of the new algorithm in the two-dimensional domain, we compute maximum entropy problem for the same set-up in three- and four-dimensional domains. For the three-dimensional moment problem we choose the moment constraints of the principal components PC1, PC2 and PC3. The optimal three-dimensional probability density function is computed for the moment constraints of up to order 6, totaling 83 constraints. Since it is hard to visualize a three-dimensional probability density function, instead we show its two-dimensional marginal distributions (obtained via averaging over the third coordinate) for each pair of the three principal components, PC1–PC2, PC1–PC3 and PC2–PC3. These three marginal distributions are displayed in Figs. 11 and 12, together with the joint probability density functions recorded directly from the model simulation via bin-counting. Observe that the marginals of the optimal three-dimensional maximum entropy PDF reproduce large and moderate scale features of the actual recorded PDFs remarkably well, which can be seen on both mesh and contour plots, although fail to notice small scale spikes (see, for example, marginal PDF of principal components PC1–PC2). For the four-dimensional moment problem we choose the moment constraints of the principal components PC1, PC2, PC3 and PC4. The optimal four-dimensional probability density function is computed for the moment constraints of up to order 4, totaling 69 constraints. Like before, we show the two-dimensional marginal distributions (obtained via averaging over the third and fourth coordinates) for each pair of the four principal components, PC1–PC2, PC1–PC3, PC2–PC3, PC1–PC4, PC2–PC4 and PC3–PC4. These six marginal distributions are displayed in Figs. 13–16, together with the joint probability density functions recorded directly from the model simulation via

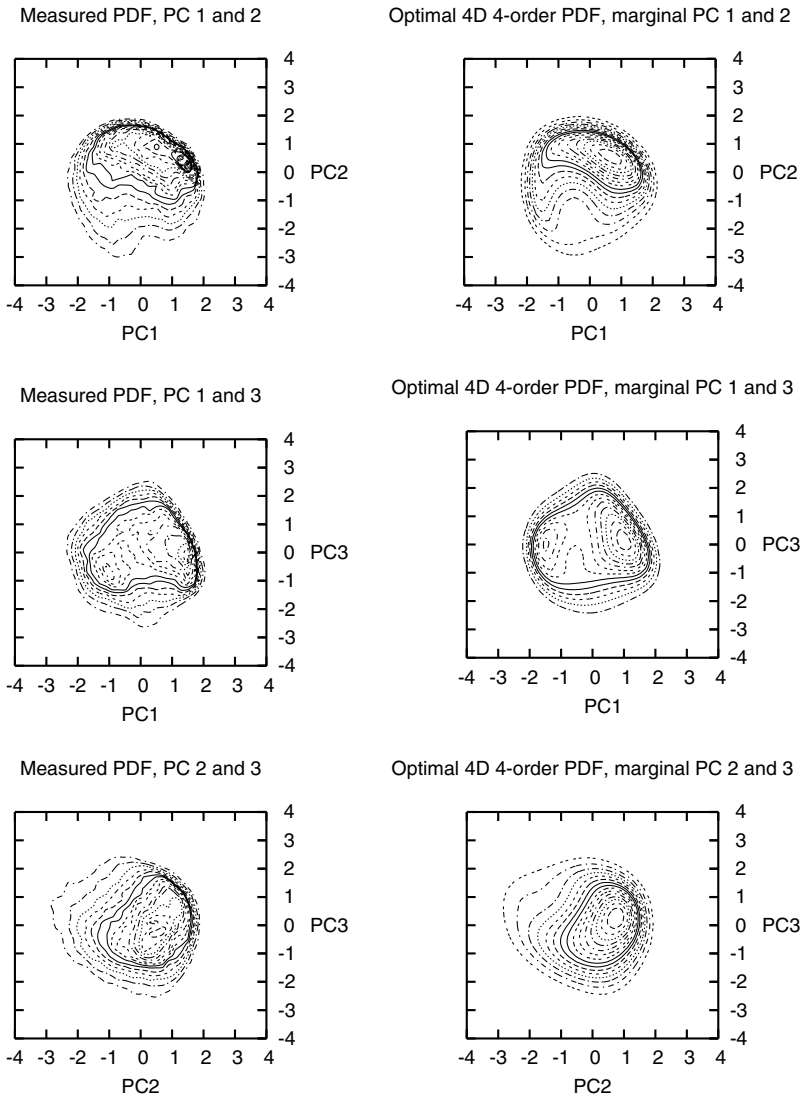
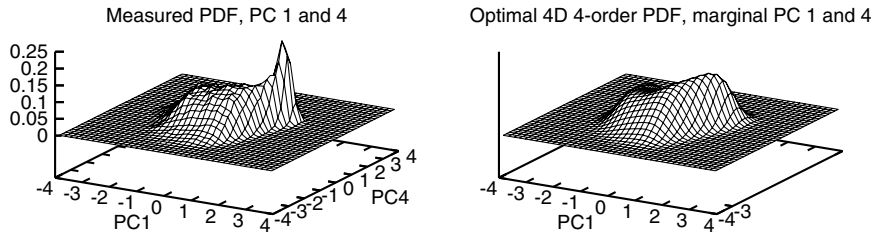


Fig. 14. Comparison of four-dimensional 4-order (69 constraints) optimal PDF with actual histograms recorded from the model via bin-counting. Left column – the recorded joint distributions of PC1 and PC2, PC1 and PC3, PC2 and PC3. Right column – the corresponding two-dimensional marginal distributions of the single four-dimensional optimal PDF, spanning PC1, PC2, PC3 and PC4. Height difference between successive contours is 0.01.

bin-counting. Again, observe that the marginals of the optimal three-dimensional maximum entropy PDF reproduce large and moderate scale features of the actual recorded PDFs rather well, although fail to notice small scale spikes (see, for example, marginal PDFs of principal components PC1–PC2 or PC1–PC4).

4. Summary

The current work further develops the practical computational approach for the moment-constrained maximum entropy principle in the multidimensional domain, set forth in [1]. Here we suggest the maximum entropy algorithm which is designed to operate in a domain of arbitrary dimension, limited mainly by computational speed of modern computer hardware. In order to avoid numerical instabilities in the Newton



iterations due to different sensitivity of the Lagrangian function to changes in different Lagrange multipliers, we introduce a suitable orthogonal polynomial basis, which also facilitates computation of the Hessian matrix for the Newton iterations via the Sherman–Morrison formula. The modified stable Gram–Schmidt algorithm maintains orthogonality of the polynomial basis with respect to the inner product generated by varying iterates of the target probability density function. The domain integrals are computed via the high-order Gauss–Hermite quadratures. We systematically test the new algorithm for the most basic one-dimensional 4-moment maximum entropy test problem with a broad range of moment constraints and different Gauss–Hermite quadrature resolutions. It is observed that the impact of Gauss–Hermite quadrature resolution is more profound for PDFs with tall and narrow spikes, which is expected because spikes are harder to capture with low-order resolution. It is also demonstrated that the new algorithm fails to converge only for a few test points near the boundary of the domain of convergence for the 4-order Hamburger moment problem, while converging everywhere else within the domain. For the one-dimensional maximum entropy problem of higher order we demonstrate how the algorithm approximates two test PDFs with qualitatively different shape features with successive order of finite moment sequence up to 2, 4, 8 and 16. It is found that the errors between the test PDF and maximum entropy PDFs systematically decrease when the order of moment constraints is increased. For the maximum entropy tests in two- and higher-dimensional domain, we use the long-term statistics of principal components of first four empirical orthogonal functions of the 1.5-layer quasigeostrophic double gyre model, which is used to describe oceanic currents driven by the wind stress with boundary layer separation, such as the Gulf Stream and the Kuroshio currents. Remarkably, the new algorithm is found to be capable of computing the maximum entropy problem in a two-dimensional domain with moment constraints of order up to 8 (44 moment constraints in total), in a three-dimensional domain with moment constraints of order up to 6 (83 moment constraints in total), and in a four-dimensional domain with moment constraints

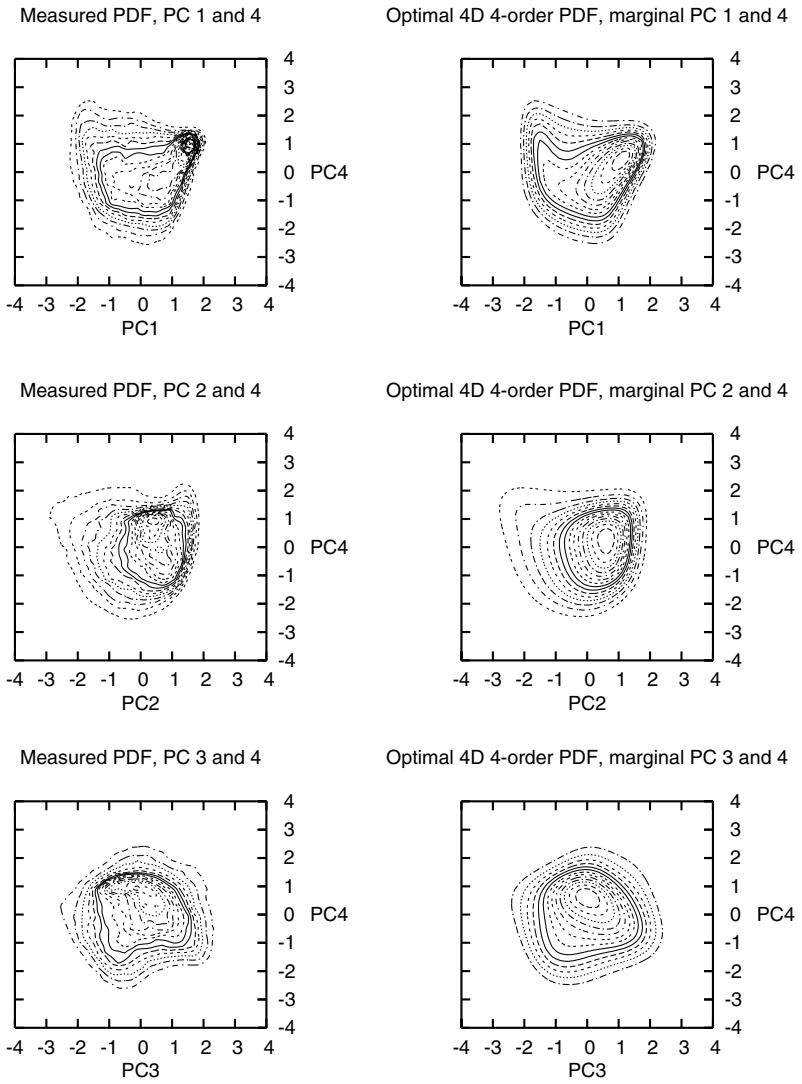


Fig. 16. Comparison of four-dimensional 4-order (69 constraints) optimal PDF with actual histograms recorded from the model via bin-counting. Left column – the recorded joint distributions of PC1 and PC4, PC2 and PC4, PC3 and PC4. Right column – the corresponding two-dimensional marginal distributions of the single four-dimensional optimal PDF, spanning PC1, PC2, PC3 and PC4. Height difference between successive contours is 0.01.

of order up to 4 (69 moment constraints in total). It is observed that computed maximum entropy approximations systematically approach the actual test PDF as the order of moment constraints is increased.

Acknowledgments

The author thanks Prof. Andrew Majda for pointing the author towards the 1.5-layer quasigeostrophic double gyre model, currently used in their joint project on predictability of large scale oceanic currents, which was also the main motivation for developing the new maximum entropy algorithm. Prof. John McCalpin and Prof. Dale Haidvogel are thanked for providing the Fortran code of the 1.5-layer quasigeostrophic double gyre model to the author, along with the detailed description of the model. The author also thanks Prof. Alexander Furman and Dr. Anton Leykin for encouraging conversations, and Prof. Bent Fuglede for the preprint of his early work on the multidimensional moment problem. The author is supported by the National Science Foundation grant DMS-0608984 and the Office of Naval Research grant N00014-06-1-0286.

References

[1] R. Abramov, A practical computational framework for the multidimensional moment-constrained maximum entropy principle, J. Comp. Phys. 211 (1) (2006) 198–209.
[2] R. Abramov, A. Majda, Quantifying uncertainty for non-Gaussian ensembles in complex systems, SIAM J. Sci. Comp. 26 (2) (2003) 411–447.
[3] R. Abramov, A. Majda, R. Kleeman, Information theory and predictability for low frequency variability, J. Atmos. Sci. 62 (1) (2005) 65–87.
[4] K. Bandyopadhyay, A. Bhattacharya, P. Biswas, D. Drabold, Maximum entropy and the problem of moments: a stable algorithm, Phys. Rev. E 71 (5) (2005).
[5] R. Byrd, P. Lu, J. Nocedal, A limited memory algorithm for bound-constrained optimization, SIAM J. Sci. Comp. 16 (6) (1995) 1190–1208.
[6] A. Dax, A modified Gram–Schmidt algorithm with iterative orthogonalization and column pivoting, Linear Alg. Appl. 310 (2000) 25–42.
[7] R. Fletcher, Practical Methods of Optimization, Wiley, New York, 1987.
[8] M. Frontini, A. Tagliani, Maximum entropy in the finite Stieltjes and Hamburger moment problem, J. Math. Phys. 35 (12) (1994) 6748–6756.
[9] B. Fuglede, The multidimensional moment problem, Expo. Math. 1 (1983) 47–65.
[10] L. Giraud, J. Langou, When modified Gram–Schmidt generates a well-conditioned set of vectors, IMA J. Numer. Anal. 22 (2002) 521–528.
[11] L. Giraud, J. Langou, M. Rozložník, On the round-off error analysis of the Gram–Schmidt algorithm with reorthogonalization, Technical Report TR/PA/02/33, CERFACS, 2002.
[12] L. Giraud, J. Langou, M. Rozložník, On the loss of orthogonality in the Gram–Schmidt orthogonalization process, Technical Report TR/PA/03/25, CERFACS, 2003.
[13] G. Golub, C. Van Loan, Matrix Computations, third ed., Johns Hopkins University Press, Baltimore, MD, 1996.
[14] K. Haven, A. Majda, R. Abramov, Quantifying predictability through information theory: small sample estimation in a non-Gaussian framework, J. Comp. Phys. 206 (2005) 334–362.
[15] R. Haydock, C. Nex, Comparison of quadrature and termination for estimating the density of states within the recursion method, J. Phys. C: Solid State Phys. 17 (1984) 4783–4789.
[16] R. Haydock, C. Nex, A general terminator for the recursion method, J. Phys. C: Solid State Phys. 18 (1985) 2235–2248.
[17] R. Kleeman, Measuring dynamical prediction utility using relative entropy, J. Atmos. Sci. 59 (2002) 2057–2072.
[18] A. Majda, R. Abramov, M. Grote, Information Theory and Stochastics for Multiscale Nonlinear Systems, CRM Monograph Series of Centre de Recherches Mathématiques, vol. 25, Université de Montréal, American Mathematical Society, 2005, ISBN 0-8218-3843-1.
[19] A. Majda, R. Kleeman, D. Cai, A framework for predictability through relative entropy, Meth. Appl. Anal. 9 (2002) 425–444.
[20] J. McCalpin, The statistics and sensitivity of a double-gyre model: the reduced gravity, quasigeostrophic case, J. Phys. Oceanogr. 25 (1995) 806–824.
[21] J. McCalpin, D. Haidvogel, Phenomenology of the low-frequency variability in a reduced-gravity, quasigeostrophic double-gyre model, J. Phys. Oceanogr. 26 (1996) 739–752.
[22] L. Mead, N. Papanicolaou, Maximum entropy in the problem of moments, J. Math. Phys. 25 (1984) 2404–2417.
[23] D. Ormoneit, H. White, An efficient algorithm to compute maximum entropy densities, Econ. Rev. 18 (2) (1999) 127–140.
[24] P. Rao, Nonparametric Functional Estimation, Academic Press, New York, 1983.
[25] M. Roulston, L. Smith, Evaluating probabilistic forecasts using information theory, Mon. Wea. Rev. 130 (2002) 1653–1660.
[26] E. Schmidt, Über die Auflösung linearer Gleichungen mit unendlich vielen Unbekannten, Rend. Circ. Mat. Palermo, Ser. I 25 (1908) 53–77.
[27] T. Schneider, S. Griffies, A conceptual framework for predictability studies, J. Clim. 12 (1999) 3133–3155.
[28] D. Scott, Multivariate Density Estimation: Theory, Practice and Visualization, Wiley, New York, 1992.
[29] B. Silverman, Density Estimation for Statistics and Data Analysis, Chapman and Hall, London, 1986.
[30] A. Tagliani, Hausdorff moment problem and maximum entropy: a unified approach, Appl. Math. Comp. 105 (1999) 291–305.
[31] I. Turek, A maximum-entropy approach to the density of states with the recursion method, J. Phys. C 21 (1988) 3251–3260.
[32] X. Wu, Calculation of maximum entropy densities with application to income distribution, J. Econ. 115 (2003) 347–354.
[33] Z. Wu, G. Phillips, R. Tapia, Y. Zhang, A fast Newton algorithm for entropy maximization in phase determination, SIAM Rev. 43 (4) (2001) 623–642.