

The Multivariate Watson Distribution: Maximum-Likelihood Estimation and other Aspects

Suvrit Sra*

Max Planck Institute for Intelligent Systems
Tübingen, Germany
suvrit@tuebingen.mpg.de

Dmitrii Karp

Institute of Applied Mathematics
Vladivostok, Russian Federation
dimkrp@gmail.com

Abstract

This paper studies fundamental aspects of modelling data using multivariate Watson distributions. Although these distributions are natural for modelling axially symmetric data (i.e., unit vectors where $\pm \mathbf{x}$ are equivalent), for high-dimensions using them can be difficult—largely because for Watson distributions even basic tasks such as maximum-likelihood are numerically challenging. To tackle the numerical difficulties some approximations have been derived. But these are either grossly inaccurate in high-dimensions (*Directional Statistics*, Mardia & Jupp, 2000) or when reasonably accurate (*J. Machine Learning Research, W.& C.P., v2*, Bijral *et al.*, 2007, pp. 35–42), they lack theoretical justification. We derive asymptotically precise two-sided bounds for the maximum-likelihood estimates which lead to new approximations. Our approximations are theoretically well-defined, numerically accurate, and easy to compute. We build on our parameter estimation and discuss mixture-modelling with Watson distributions; here we uncover a hitherto unknown connection to the “diametrical clustering” algorithm of Dhillon *et al.* (*Bioinformatics*, 19(13), 2003, pp. 1612–1619).

Keywords: Watson distribution, Kummer function, Confluent hypergeometric function, Directional statistics, Diametrical clustering, Special function, Hypergeometric identity

1 Introduction

Life on the surface of the unit hypersphere is more twisted than you might imagine: designing elegant probabilistic models is easy but using them is often not. This difficulty usually stems from the complicated normalising constants associated with directional distributions. Nevertheless, owing to their powerful modelling capabilities, distributions on hyperspheres continue finding numerous applications—see e.g., the excellent book *Directional Statistics* (Mardia and Jupp, 2000).

A fundamental directional distribution is the von Mises-Fisher (vMF) distribution, which models data concentrated around a mean-direction. But for data that have additional structure, vMF can be inappropriate: in particular, for axially symmetric data it is more natural to prefer the (Dimroth-Scheidegger)-Watson distribution (Mardia and Jupp, 2000; Watson, 1965). And this distribution is the focus of our paper.

Three main reasons motivate our study of the multivariate Watson (mW) distribution, namely: (i) it is fundamental to directional statistics; (ii) it has not received much attention in modern data-analysis setups involving high-dimensional data; and (iii) it provides a theoretical basis to “diametrical clustering”, a procedure developed for gene-expression analysis (Dhillon *et al.*, 2003).

Somewhat surprisingly, for high-dimensional settings, the mW distribution seems to be fairly understudied. One reason might be that the traditional domains of directional statistics are low-dimensional, e.g., circles or spheres. Moreover, in low-dimensions numerical difficulties that are rife in high-dimensions

*This work was largely done when the first author was affiliated with the Max Planck Institute for Biological Cybernetics

are not so pronounced. This paper contributes theoretically and numerically to the study of the mW distribution. We hope that these contributions and the connections we make to established applications help promote wider use of the mW distribution.

1.1 Related Work

Beyond their use in typical applications of directional statistics (Mardia and Jupp, 2000), directional distributions gained renewed attention in data-mining, where the vMF distribution was first used by (Banerjee et al., 2003, 2005), who also derived some *ad-hoc* parameter estimates; Non *ad-hoc* parameter estimates for the vMF case were obtained by Tanabe et al. (2007).

More recently, the Watson distribution was considered in (Bijral et al., 2007) and also in (Sra, 2007). Bijral et al. (2007) used an approach similar to that of (Banerjee et al., 2003) to obtain an *ad-hoc* approximation to the maximum-likelihood estimates. We eliminate the *ad-hoc* approach and formally derive tight, two-sided bounds which lead to parameter approximations that are accurate and efficiently computed.

Our derivations are based on carefully exploiting properties (several *new* ones are derived in this paper) of the confluent hypergeometric function, which arises as a part of the normalisation constant. Consequently, a large body of classical work on special functions is related to our paper. But to avoid detracting from the main message limitations, we relegate highly technical details to the appendix.

Another line of related work is based on mixture-modelling with directional distributions, especially for high-dimensional datasets. In (Banerjee et al., 2005), mixture-modelling using the Expectation Maximisation (EM) algorithm for mixtures of vMFs was related to cosine-similarity based K-means clustering. Specifically, Banerjee et al. (2005) showed how the cosine based K-means algorithm may be viewed as a limiting case of the EM algorithm for mixtures of vMFs. Similarly, we investigate mixture-modelling using Watson distributions, and connect a limiting case of the corresponding EM procedure to a clustering algorithm called “diametrical clustering” (Dhillon et al., 2003). Our viewpoint provides a new interpretation of the (discriminative) diametrical clustering algorithm and also lends generative semantics to it. Consequently, using a mixture of Watson distributions we also obtain a clustering procedure that can provide better clustering results than plain diametrical clustering alone.

2 Background

Let $\mathbb{S}^{p-1} = \{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^p, \|\mathbf{x}\|_2 = 1\}$ be the $(p-1)$ -dimensional unit hypersphere centred at the origin. We focus on axially symmetric vectors, i.e., $\pm\mathbf{x} \in \mathbb{S}^{p-1}$ are equivalent; this is also denoted by $\mathbf{x} \in \mathbb{P}^{p-1}$, where \mathbb{P}^{p-1} is the projective hyperplane of dimension $p-1$. A natural choice for modelling such data is the multivariate Watson distribution (Mardia and Jupp, 2000). This distribution is parametrised by a *mean-direction* $\boldsymbol{\mu} \in \mathbb{P}^{p-1}$, and a *concentration* parameter $\kappa \in \mathbb{R}$; its probability density function is

$$W_p(\mathbf{x}; \boldsymbol{\mu}, \kappa) = c_p(\kappa) e^{\kappa(\boldsymbol{\mu}^T \mathbf{x})^2}, \quad \mathbf{x} \in \mathbb{P}^{p-1}. \quad (2.1)$$

The normalisation constant $c_p(\kappa)$ in (2.1) is given by

$$c_p(\kappa) = \frac{\Gamma(p/2)}{2\pi^{p/2} M(\frac{1}{2}, \frac{p}{2}, \kappa)}, \quad (2.2)$$

where M is Kummer’s confluent hypergeometric function defined as ((Erdélyi et al., 1953, formula 6.1(1)) or (Andrews et al., 1999, formula (2.1.2))

$$M(a, c, \kappa) = \sum_{j=0}^{\infty} \frac{a^{\overline{j}} \kappa^j}{c^{\overline{j}} j!}, \quad a, c, \kappa \in \mathbb{R}, \quad (2.3)$$

and $a^{\overline{0}} = 1$, $a^{\overline{j}} = a(a+1) \cdots (a+j-1)$, $j \geq 1$, denotes the *rising-factorial*.

Observe that for $\kappa > 0$, the density concentrates around $\boldsymbol{\mu}$ as κ increases, whereas for $\kappa < 0$, it concentrates around the great circle orthogonal to $\boldsymbol{\mu}$. Observe that $(\mathbf{Q}\boldsymbol{\mu})^T \mathbf{Q}\mathbf{x} = \boldsymbol{\mu}^T \mathbf{x}$ for any orthogonal matrix \mathbf{Q} . In particular for $\mathbf{Q}\boldsymbol{\mu} = \boldsymbol{\mu}$, $\boldsymbol{\mu}^T(\mathbf{Q}\mathbf{x}) = \boldsymbol{\mu}^T \mathbf{x}$; thus, the Watson density is rotationally symmetric about $\boldsymbol{\mu}$.

2.1 Maximum Likelihood Estimation

We now consider the basic and apparently simple task of maximum-likelihood parameter estimation for mW distributions: this task turns out to be surprisingly difficult.

Let $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{P}^{p-1}$ be i.i.d. points drawn from $W_p(\mathbf{x}; \boldsymbol{\mu}, \kappa)$, the Watson density with mean $\boldsymbol{\mu}$ and concentration κ . The corresponding log-likelihood is

$$\ell(\boldsymbol{\mu}, \kappa; \mathbf{x}_1, \dots, \mathbf{x}_n) = n(\kappa \boldsymbol{\mu}^T \mathbf{S} \boldsymbol{\mu} - \log M(1/2, p/2, \kappa) + \gamma), \quad (2.4)$$

where $\mathbf{S} = n^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T$ is the sample *scatter matrix*, and γ is a constant term that we can ignore. Maximising (2.4) leads to the following parameter estimates (Mardia and Jupp, 2000, Sec. 10.3.2) for the mean vector

$$\hat{\boldsymbol{\mu}} = \mathbf{s}_1 \quad \text{if } \hat{\kappa} > 0, \quad \hat{\boldsymbol{\mu}} = \mathbf{s}_p \quad \text{if } \hat{\kappa} < 0, \quad (2.5)$$

where $\mathbf{s}_1, \dots, \mathbf{s}_p$ are normalised eigenvectors ($\in \mathbb{P}^{p-1}$) of the scatter matrix \mathbf{S} corresponding to the eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. The concentration estimate $\hat{\kappa}$ is obtained by solving¹

$$g(\tfrac{1}{2}, \tfrac{p}{2}; \hat{\kappa}) := \frac{M'(\tfrac{1}{2}, \tfrac{p}{2}, \hat{\kappa})}{M(\tfrac{1}{2}, \tfrac{p}{2}, \hat{\kappa})} = \hat{\boldsymbol{\mu}}^T \mathbf{S} \hat{\boldsymbol{\mu}} := r \quad (0 \leq r \leq 1), \quad (2.6)$$

where M' denotes the derivative with respect to $\hat{\kappa}$. Notice that (2.5) and (2.6) are coupled—so we need some way to decide whether to solve $g(1/2, p/2; \hat{\kappa}) = \lambda_1$ or to solve $g(1/2, p/2; \hat{\kappa}) = \lambda_p$ instead. An easy choice is to solve both equations, and select the solution that yields a higher log-likelihood. Solving these equations is much harder.

One could solve (2.6) using a root-finding method (e.g. Newton-Raphson). But, the situation is not that simple. For reasons that will soon become clear, an out-of-the-box root-finding approach can be unduly slow or even fraught with numerical peril, effects that become more pronounced with increasing data dimensionality. Let us, therefore, consider a slightly more general equation (we also drop the accent on κ):

Solve for κ

$$g(a, c; \kappa) := \frac{M'(a, c; \kappa)}{M(a, c; \kappa)} = r \quad (2.7)$$

$$c > a > 0, \quad 0 \leq r \leq 1.$$

3 Solving for κ

In this section we present two different solutions to (2.7). The first is the “obvious” method based on a Newton-Raphson root-finder. The second method is the key numerical contribution of this paper: a method that computes a closed-form approximate solution to (2.7), thereby requiring merely a few floating-point operations!

3.1 Newton-Raphson

Although we establish this fact not until Section 3.2, suppose for the moment that (2.7) *does* have a solution. Further, assume that by bisection or otherwise, we have bracketed the root κ to be within an interval and are thus ready to invoke the Newton-Raphson method.

Starting at κ_0 , Newton-Raphson solves the equation $g(a, c; \kappa) - r = 0$ by iterating

$$\kappa_{n+1} = \kappa_n - \frac{g(a, c; \kappa_n) - r}{g'(a, c; \kappa_n)}, \quad n = 0, 1, \dots \quad (3.1)$$

¹More precisely, we need $\lambda_1 > \lambda_2$ to ensure a unique m.l.e. for positive κ , and $\lambda_{p-1} > \lambda_p$, for negative κ .

This iteration may be simplified by rewriting $g'(a, c; \kappa)$. First note that

$$g'(a, c; \kappa) = \frac{M''(a, c; \kappa)}{M(a, c; \kappa)} - \left(\frac{M'(a, c; \kappa)}{M(a, c; \kappa)} \right)^2, \quad (3.2)$$

then, recall the following two identities

$$M''(a, c; \kappa) = \frac{a(a+1)}{c(c+1)} M(a+2, c+2; \kappa); \quad (3.3)$$

$$M(a+2, c+2; \kappa) = \frac{(c+1)(-c+\kappa)}{(a+1)\kappa} M(a+1, c+1; \kappa) + \frac{(c+1)c}{(a+1)\kappa} M(a, c; \kappa). \quad (3.4)$$

Now, use both (3.3) and (3.4) to rewrite the derivative (3.2) as

$$g'(a, c; \kappa) = (1 - c/\kappa)g(a, c; \kappa) + (a/\kappa) - (g(a, c; \kappa))^2. \quad (3.5)$$

The main consequence of these simplifications is that iteration (3.1) can be implemented with only *one* evaluation of the ratio $g(a, c; \kappa_n) = M'(a, c; \kappa_n)/M(a, c; \kappa_n)$. Efficiently computing this ratio is a *non-trivial* task in itself; an insight into this difficulty is offered by observations in (Gautschi, 1977; Gil et al., 2007). In the worst case, one may have to compute the numerator and denominator separately (using multi-precision floating point arithmetic), and then divide. Doing so can require several million extended precision floating point operations, which is very undesirable.

3.2 Closed-form Approximation for (2.7)

We now derive two-sided bounds which will lead to a closed-form approximation to the solution of (2.7). This approximation, while marginally less accurate than the one via Newton-Raphson, should suffice for most uses. Moreover, it is incomparably faster to compute as it is in closed-form.

Before proceeding to the details, let us look at a little history. For 2–3 dimensional data, or under very restrictive assumptions on κ or r , some approximations had been previously obtained (Mardia and Jupp, 2000). Due to their restrictive assumptions, these approximations have limited applicability, especially for high-dimensional data, where these assumptions are often violated (Banerjee et al., 2005). Recently Bijral et al. (2007) followed the technique of Banerjee et al. (2005) to essentially obtain the *ad-hoc* approximation (actually particularly for the case $a = 1/2$)

$$BBG(r) := \frac{cr - a}{r(1 - r)} + \frac{r}{2c(1 - r)}, \quad (3.6)$$

which they observed to be quite accurate. However, (3.6) lacks theoretical justification; other approximations were presented in (Sra, 2007), though again only *ad-hoc*.

Below we present new approximations for κ that are theoretically well-motivated and also numerically more accurate. Key to obtaining these approximations are a set of bounds localizing κ , and we present these in a series of theorems below. The proofs are given in the appendix.

3.2.1 Existence and Uniqueness

The following theorem shows that the function $g(a, c; \kappa)$ is strictly increasing.

Theorem 3.1. *Let $c > a > 0$, and $\kappa \in \mathbb{R}$. Then the function $\kappa \rightarrow g(a, c; \kappa)$ is monotone increasing from $g(a, c; -\infty) = 0$ to $g(a, c; \infty) = 1$.*

Proof. Since $g(a, c; \kappa) = (a/c)f_1(\kappa)$, where f_μ is defined in (A.11), this theorem is a direct consequence of Theorem A.4. \square

Hence the equation $g(a, c; \kappa) = r$ has a unique solution for each $0 < r < 1$. This solution is negative if $0 < r < a/c$ and positive if $a/c < r < 1$. Let us now localize this solution to a narrow interval by deriving tight bounds on it.

3.2.2 Bounds on the solution κ

Deriving tight bounds for κ is key to obtaining our new theoretically well-defined numerical approximations; moreover, these approximations are easy to compute because the bounds are given in closed form.

Theorem 3.2. *Let the solution to $g(a, c; \kappa) = r$ be denoted by $\kappa(r)$. Consider the following three bounds:*

$$(lower\ bound) \quad L(r) = \frac{rc - a}{r(1 - r)} \left(1 + \frac{1 - r}{c - a} \right), \quad (3.7)$$

$$(bound) \quad B(r) = \frac{rc - a}{2r(1 - r)} \left(1 + \sqrt{1 + \frac{4(c + 1)r(1 - r)}{a(c - a)}} \right), \quad (3.8)$$

$$(upper\ bound) \quad U(r) = \frac{rc - a}{r(1 - r)} \left(1 + \frac{r}{a} \right). \quad (3.9)$$

Let $c > a > 0$, and $\kappa(r)$ be the solution (2.7). Then, we have

1. for $a/c < r < 1$,

$$L(r) < \kappa(r) < B(r) < U(r), \quad (3.10)$$

2. for $0 < r < a/c$,

$$L(r) < B(r) < \kappa(r) < U(r). \quad (3.11)$$

3. and if $r = a/c$, then $\kappa(r) = L(a/c) = B(a/c) = U(a/c) = 0$.

All three bounds (L , B , and U) are also asymptotically precise at $r = 0$ and $r = 1$.

Proof. The proofs of parts 1 and 2 are given in Theorems A.5 and A.6 (see Appendix), respectively. Part 3 is trivial. It is easy to see that $\lim_{r \rightarrow 0,1} U(r)/L(r) = 1$, so from inequalities (3.10) and (3.11), it follows that

$$\lim_{r \rightarrow 0,1} \frac{L(r)}{\kappa(r)} = \lim_{r \rightarrow 0,1} \frac{B(r)}{\kappa(r)} = \lim_{r \rightarrow 0,1} \frac{U(r)}{\kappa(r)} = 1. \quad \square$$

More precise asymptotic characterizations of the approximations L , B and U are given in section 3.2.4.

3.2.3 BBG approximation

Our bounds above also provide some insight into the previous heuristically motivated κ -approximation $BBG(r)$ of Bijral et al. (2007) given by (3.6). Specifically, we check whether $BBG(r)$ satisfies the lower and upper bounds from Theorem 3.2.

To see when $BBG(r)$ violates the lower bound, solve $L(r) > BBG(r)$ for r to obtain

$$\frac{2c^2 + a - \sqrt{(2c^2 - a)(2c^2 - a - 8ac)}}{2(2c^2 - a + c)} < r < \frac{2c^2 + a + \sqrt{(2c^2 - a)(2c^2 - a - 8ac)}}{2(2c^2 - a + c)}.$$

For the Watson case $a = 1/2$; this means that $BBG(r)$ violates the lower bound and *underestimates* the solution for $r \in (0.11, 0.81)$ if $c = 5$; for $r \in (0.0528, 0.904)$ if $c = 10$; for $r \in (0.00503, 0.99)$ if $c = 100$; for $r \in (0.00050025, 0.999)$ if $c = 1000$. This fact is also reflected in Figure 2.

To see when $BBG(r)$ violates the upper bound, solve $BBG(r) > U(r)$ for r to obtain

$$r < \frac{2ac}{2c^2 - a}.$$

For the Watson case $a = 1/2$; this means that $BBG(r)$ violates the upper bound and *overestimates* the solution for $r \in (0, 0.1)$ if $c = 5$; for $r \in (0, 0.05)$ if $c = 10$; for $r \in (0, 0.005)$ if $c = 100$; for $r \in (0, 0.0005)$ if $c = 1000$.

What do these violations imply? They show that a combination of $L(r)$ and $U(r)$ is guaranteed to give a better approximation than $BBG(r)$ for nearly all $r \in (0, 1)$ except for a very small neighbourhood of the point where $BBG(r)$ intersects $\kappa(r)$.

3.2.4 Asymptotic precision of the approximations

Let us now look more precisely at how the various approximations behave at limiting values of r . There are three points where we can compute asymptotics: $r = 0$, $r = a/c$, and $r = 1$. First, we assess how $\kappa(r)$ itself behaves.

Theorem 3.3. *Let $c > a > 0$, $r \in (0, 1)$; let $\kappa(r)$ be the solution to $g(a, c; \kappa) = r$. Then,*

$$\kappa(r) = -\frac{a}{r} + (c - a - 1) + \frac{(c - a - 1)(1 + a)}{a}r + O(r^2), \quad r \rightarrow 0, \quad (3.12)$$

$$\kappa(r) = \left(r - \frac{a}{c}\right) \left\{ \frac{c^2(1 + c)}{a(c - a)} + \frac{c^3(1 + c)^2(2a - c)}{a^2(c - a)^2(c + 2)} \left(r - \frac{a}{c}\right) + O\left(\left(r - \frac{a}{c}\right)^2\right) \right\}, \quad r \rightarrow \frac{a}{c} \quad (3.13)$$

$$\kappa(r) = \frac{c - a}{1 - r} + 1 - a + \frac{(a - 1)(a - c - 1)}{c - a}(1 - r) + O((1 - r)^2), \quad r \rightarrow 1. \quad (3.14)$$

This theorem is given in the appendix with detailed proof as Theorem A.7.

We can compute asymptotic expansions for the various approximations by standard Laurent expansion. For $L(r)$ we obtain:

$$L(r) = -\frac{a(c - a + 1)}{(c - a)r} + (c/(c - a) + c - a) + O(r), \quad r \rightarrow 0,$$

$$L(r) = \frac{c^2(c + 1)}{a(c - a)}(r - a/c) + \frac{c^3(ac - (c + 1)(c - a))}{a^2(c - a)^2}(r - a/c)^2 + O((r - a/c)^3), \quad r \rightarrow a/c,$$

$$L(r) = \frac{c - a}{1 - r} + (1 - a) + \frac{a(a - c - 1)}{c - a}(1 - r) + O((1 - r)^2), \quad r \rightarrow 1.$$

For $U(r)$ we get:

$$U(r) = -\frac{a}{r} + (c - a - 1) + \frac{(c - a)(a + 1)}{a}r + O(r^2), \quad r \rightarrow 0,$$

$$U(r) = \frac{c^2(c + 1)}{a(c - a)}(r - a/c) + \frac{c^3(2ac + a - c^2)}{a^2(c - a)^2}(r - a/c)^2 + O((r - a/c)^3), \quad r \rightarrow a/c,$$

$$U(r) = \frac{((c/a) - a + c - 1)}{1 - r} - ((c/a) + a) + O(1 - r), \quad r \rightarrow 1.$$

For $B(r)$ the expansions are:

$$B(r) = -\frac{a}{r} + \frac{a^2 - 2ac + c^2 - c - 1}{c - a} + O(r), \quad r \rightarrow 0,$$

$$B(r) = \frac{c^2(1 + c)}{a(c - a)}(r - a/c) + \frac{c^3(1 + c)^2(2a - c)}{a^2(c - a)^2(c + 2)}(r - a/c)^2 + O((r - a/c)^3), \quad r \rightarrow a/c,$$

$$B(r) = \frac{c - a}{1 - r} + \frac{c + 1 - a^2}{a} + O(1 - r), \quad r \rightarrow 1.$$

Finally, for the approximation (3.6) we have the following expansions:

$$BBG(r) = -\frac{a}{r} + (c - a) + O(r), \quad r \rightarrow 0,$$

$$BBG(r) = \frac{a}{2c(c - a)} + O(r - a/c), \quad r \rightarrow a/c,$$

$$BBG(r) = \frac{2ac - 2c^2 - 1}{2c(1 - r)} - \frac{2ac + 1}{2c} + O(1 - r), \quad r \rightarrow 1.$$

We summarize the results in Table 1 below.

Table 1 uses the following terminology: (i) we call an approximation $f(r)$ to be *incorrect* around $r = \alpha$, if $f(r)/\kappa(r) \rightarrow 0, \infty$ as $r \rightarrow \alpha$; (ii) we say $f(r)$ is *correct of order 1* around $r = \alpha$, if $f(r)/\kappa(r) \rightarrow C$ such that $C \neq 0, \infty$ as $r \rightarrow \alpha$; (iii) we say $f(r)$ is *correct of order 2* around $r = \alpha$ if $f(r)/\kappa(r) = 1 + O(r - \alpha)$ as $r \rightarrow \alpha$; and (iv) $f(r)$ is *correct of order 3* around $r = \alpha$ if $f(r)/\kappa(r) = 1 + O((r - \alpha)^2)$ as $r \rightarrow \alpha$.

No matter how we count the total "order of correctness" it is clear from Table 1 that our approximations are superior to that of (Bijral et al., 2007).

The table shows that actually $L(r)$ and $U(r)$ can be viewed as three-point $[2/2]$ Padé approximations to $\kappa(r)$ at $r = 0$ and $r = a/c$ and $r = 1$ with different orders at different points, while $B(r)$ is a special non-rational three point approximation with even higher total order of contact.

Moreover, since we not only give the order of correctness but also prove the inequalities, we always know exactly which approximation underestimates $\kappa(r)$ and which overestimates $\kappa(r)$. Such information might be important to some applications. The approximation of (Bijral et al., 2007) is clearly less precise and does not satisfy such inequalities. Also, note that all the above facts are equally true in the Watson case $a = 1/2$.

Approx. Point	$L(r)$	$B(r)$	$U(r)$	$BBG(r)$
$r = 0$	Correct of order 1	Correct of order 2	Correct of order 3	Correct of order 2
$r = a/c$	Correct of order 2	Correct of order 3	Correct of order 2	Incorrect
$r = 1$	Correct of order 3	Correct of order 2	Correct of order 1	Correct of order 1

Table 1: Summary of various approximations

4 Application to Mixture Modelling and Clustering

Now that we have shown how to compute maximum-likelihood parameter estimates, we proceed onto *mixture-modelling* for mW distributions.

Suppose we observe the set $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{P}^{p-1}\}$ of i.i.d. samples. We wish to model this set using a mixture of K mW distributions. Let $W_p(\mathbf{x}|\boldsymbol{\mu}_j, \kappa_j)$ be the density of the j -th mixture component, and π_j its prior ($1 \leq j \leq K$) – then, for observation \mathbf{x}_i we have the density

$$f(\mathbf{x}_i|\boldsymbol{\mu}_1, \kappa_1, \dots, \boldsymbol{\mu}_K, \kappa_K) = \sum_{j=1}^K \pi_j W_p(\mathbf{x}_i|\boldsymbol{\mu}_j, \kappa_j).$$

The corresponding log-likelihood for the entire dataset \mathcal{X} is given by

$$\mathcal{L}(\mathcal{X}; \boldsymbol{\mu}_1, \kappa_1, \dots, \boldsymbol{\mu}_K, \kappa_K) = \sum_{i=1}^n \log \left(\sum_{j=1}^K \pi_j W_p(\mathbf{x}_i|\boldsymbol{\mu}_j, \kappa_j) \right). \quad (4.1)$$

To maximise the log-likelihood, we follow a standard Expectation Maximisation (EM) procedure (Dempster et al., 1977). To that end, first bound \mathcal{L} from below as

$$\mathcal{L}(\mathcal{X}; \boldsymbol{\mu}_1, \kappa_1, \dots, \boldsymbol{\mu}_K, \kappa_K) \geq \sum_{ij} \beta_{ij} \log \frac{\pi_j W_p(\mathbf{x}_i|\boldsymbol{\mu}_j, \kappa_j)}{\beta_{ij}}, \quad (4.2)$$

where β_{ij} is the *posterior* probability (for \mathbf{x}_i , given component j), and it is defined by the *E-Step*:

$$\beta_{ij} = \frac{\pi_j W_p(\mathbf{x}_i|\boldsymbol{\mu}_j, \kappa_j)}{\sum_l \pi_l W_p(\mathbf{x}_i|\boldsymbol{\mu}_l, \kappa_l)}. \quad (4.3)$$

```

Input:  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n : \text{where each } \mathbf{x}_i \in \mathbb{P}^{p-1}\}$ ,  $K$ : number of components
Output: Parameter estimates  $\pi_j$ ,  $\boldsymbol{\mu}_j$ , and  $\kappa_j$ , for  $1 \leq j \leq K$ 
Initialise  $\pi_j, \boldsymbol{\mu}_j, \kappa_j$  for  $1 \leq j \leq K$ 
while not converged do
    {Perform the E-step of EM}
    foreach  $i$  and  $j$  do
        | Compute  $\beta_{ij}$  using (4.3) (or via (4.6) if using hard-assignments)
    end
    {Perform the M-step of EM}
    for  $j = 1$  to  $K$  do
        |  $\pi_j \leftarrow \frac{1}{n} \sum_{i=1}^n \beta_{ij}$ 
        | Compute  $\boldsymbol{\mu}_j$  using (4.4)
        | Compute  $\kappa_j$  using (4.5)
    end
end

```

Algorithm 1: EM Algorithm for mixture of Watson (moW)

Maximising the lower-bound (4.2) subject to $\boldsymbol{\mu}_j^T \boldsymbol{\mu}_j = 1$, yields the *M-Step*:

$$\boldsymbol{\mu}_j = \mathbf{s}_1^j \quad \text{if } \kappa_j > 0, \quad \boldsymbol{\mu}_j = \mathbf{s}_p^j \quad \text{if } \kappa_j < 0, \quad (4.4)$$

$$\kappa_j = g^{-1}(1/2, p/2, r_j), \quad \text{where } r_j = \boldsymbol{\mu}_j^T \mathbf{S}^j \boldsymbol{\mu}_j, \quad (4.5)$$

$$\pi_j = \frac{1}{n} \sum_i \beta_{ij},$$

where \mathbf{s}_i^j denotes the eigenvector corresponding to eigenvalue λ_i (where $\lambda_1 \geq \dots \geq \lambda_p$) of the *weighted-scatter matrix*:

$$\mathbf{S}^j = \frac{1}{\sum_i \beta_{ij}} \sum_i \beta_{ij} \mathbf{x}_i \mathbf{x}_i^T.$$

Now we can iterate between (4.3), (4.4) and (4.5) to obtain an EM algorithm. Pseudo-code for such a procedure is shown below as Algorithm 1.

Note: Hard Assignments. We note that as usual, to reduce the computational burden, we can replace can E-step (4.3) by the standard *hard-assignment* heuristic:

$$\beta_{ij} = \begin{cases} 1, & \text{if } j = \operatorname{argmax}_{j'} \log \pi_{j'} + \log W_p(\mathbf{x}_i | \boldsymbol{\mu}_{j'}, \kappa_{j'}), \\ 0, & \text{otherwise.} \end{cases} \quad (4.6)$$

The corresponding *M-Step* also simplifies considerably. Such hard-assignments maximize a lower-bound on the incomplete log-likelihood, and yield *partitional-clustering* algorithms (in fact, we show experimental results in Section 5.2 where we cluster data using a partitional-clustering algorithm based on this hard-assignment heuristic).

4.1 Diametrical Clustering

We now turn to the diametrical clustering algorithm of Dhillon et al. (2003), and show that it is merely a special case of the mixture-model described above. Diametrical clustering is motivated by the need to group together correlated and anti-correlated data points (see Figure 1 for an illustration). For data normalised to have unit euclidean norm, such clustering treats *diametrically* opposite points equivalently. In other words, \mathbf{x} lies on the projective plane. Therefore, a natural question is whether diametrical clustering is related to Watson distributions, and if so, how?

The answer to this question will become apparent once we recall the diametrical clustering algorithm (shown as Algorithm 2) of (Dhillon et al., 2003). In Algorithm 2 we have labelled the “E-Step” and the “M-Step”. These two steps are simplified instances of the E-step (4.3) (alternatively 4.6) and

Input: $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n : \mathbf{x}_i \in \mathbb{P}^{p-1}\}$, K : number of clusters
Output: A partition $\{\mathcal{X}_j : 1 \leq j \leq K\}$ of \mathcal{X} , and centroids $\boldsymbol{\mu}_j$
 Initialise $\boldsymbol{\mu}_j$ for $1 \leq j \leq K$
while not converged **do**
 E-step:
 Set $\mathcal{X}_j \leftarrow \emptyset$ for $1 \leq j \leq K$
 for $i = 1$ to n **do**
 $\mathcal{X}_j \leftarrow \mathcal{X}_j \cup \{\mathbf{x}_i\}$ where $j = \operatorname{argmax}_{1 \leq h \leq K} (\mathbf{x}_i^T \boldsymbol{\mu}_h)^2$
 end
 M-step:
 for $j = 1$ to K **do**
 $\mathbf{A}_j = \sum_{\mathbf{x}_i \in \mathcal{X}_j} \mathbf{x}_i \mathbf{x}_i^T$
 $\boldsymbol{\mu}_j \leftarrow \mathbf{A}_j \boldsymbol{\mu}_j / \|\mathbf{A}_j \boldsymbol{\mu}_j\|$
 end
end

Algorithm 2: Diametrical Clustering

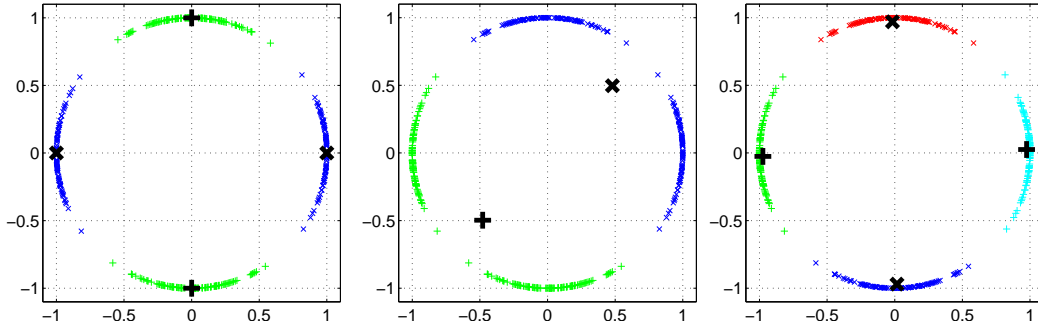


Figure 1: The left panel shows axially symmetric data that has two clusters (centroids are indicated by '+' and 'x'). The middle and right panel shows clustering yielded by (Euclidean) K-means (note that the centroids fail to lie on the circle in this case) with $K = 2$ and $K = 4$, respectively. Diametrical clustering recovers the true clusters in the left panel.

M-step (4.4). To see why, consider the E-step (4.3). If $\kappa_j \rightarrow \infty$, then for each i , the corresponding posterior probabilities $\beta_{ij} \rightarrow \{0, 1\}$; the particular β_{ij} that tends to 1 is the one for which $(\boldsymbol{\mu}_j^T \mathbf{x}_i)^2$ is maximised – this is precisely the choice used in the E-step of Algorithm 2. With binary values for β_{ij} , the M-Step (4.4) also reduces to the version followed by Algorithm 2.

An alternative, perhaps better view is obtained by regarding diametrical clustering as a special case of mixture-modelling where a hard-assignment rule is used. Now, if all mixture components have the *same, positive* concentration parameter κ , then while computing β_{ij} via (4.6) we may ignore κ altogether, which reduces Algorithm 1 to Algorithm 2.

Given this interpretation of diametrical clustering, it is natural to expect that the additional modelling power offered by mixtures of Watson distributions might lead to better clustering. This is indeed the case, as indicated by some of our experiments in Section 5.2 below, where we show that merely including the concentration parameter κ can lead to improved clustering accuracies, or to clusters with higher quality (in a sense that will be made more precise below).

5 Experiments

We now come to numerical results to assess the methods presented. We divide our experiments into two groups. The first group comprises numerical results that illustrate accuracy of our approximation to κ . The second group supports our claim that the extra modelling power offered by moWs also translates

into better clustering results.

5.1 Estimating κ

We show two representative experiments to illustrate the accuracy of our approximations. The first set (§5.1.1) compares our approximation with that of Bijral et al. (2007), as given by (3.6). This set considers the Watson case, namely $a = 1/2$ and varying dimensionality $c = p/2$. The second set (§5.1.2) of experiments shows a sampling of results for a few values of c and κ as the parameter a is varied. This set illustrates how well our approximations behave for the general nonlinear equation (2.7).

5.1.1 Comparison with the BBG approximation for the Watson case

Here we fix $a = 1/2$, and vary c on an exponentially spaced grid ranging from $c = 10$ to $c = 10^4$. For each value of c , we generate geometrically spaced values of the “true” κ_* in the range $[-200c, 200c]$. For each choice of κ_* picked within this range, we compute the ratio $r = g(1/2, c, \kappa_*)$ (using MATHEMATICA for high precision). Then, given $a = 1/2$, c , and r , we estimate κ_* by solving $\kappa \approx g^{-1}(1/2, c, r)$ using $BBG(r)$, $L(r)$, $B(r)$, and $U(r)$, given by (3.6), (3.7), (3.8), and (3.9), respectively.

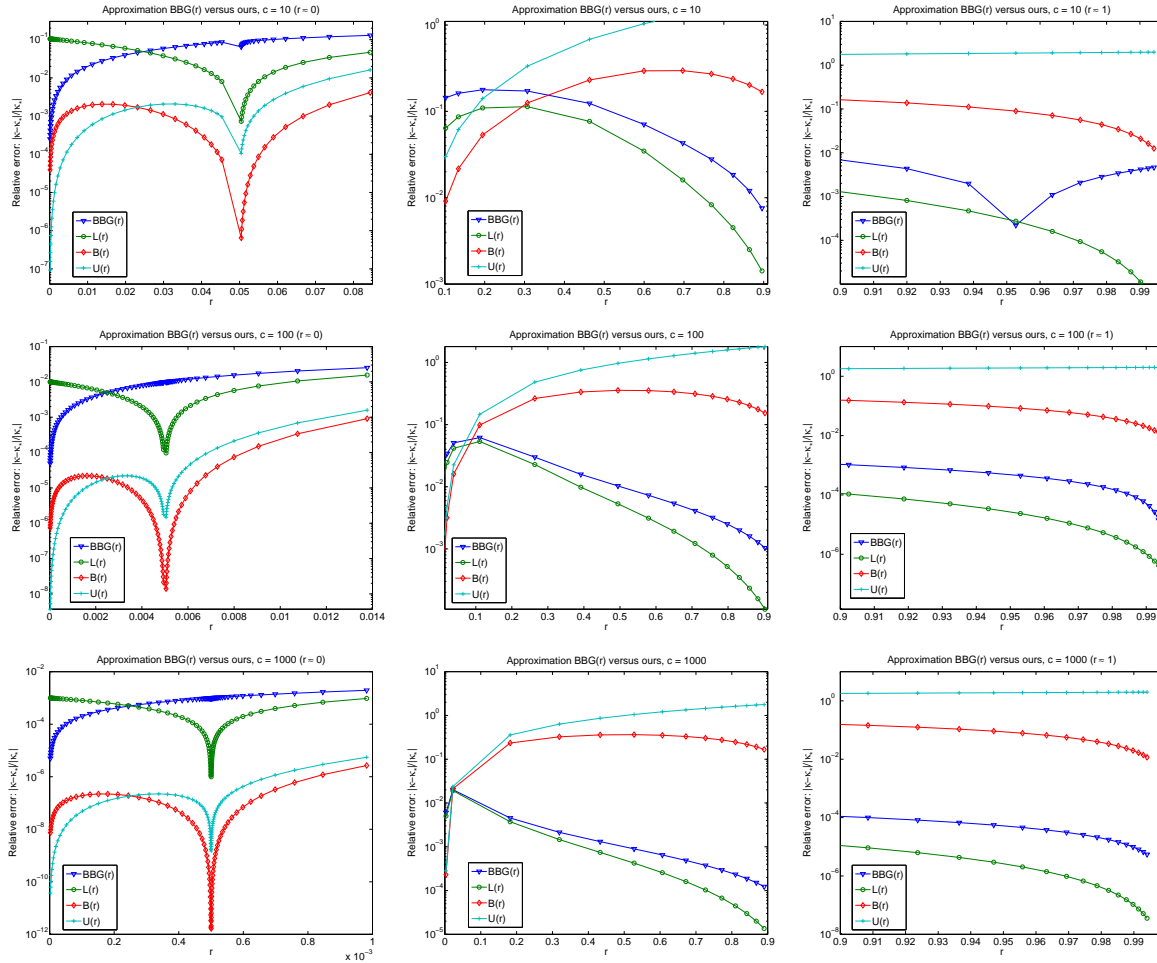


Figure 2: Relative errors $|\hat{\kappa} - \kappa_*|/|\kappa_*|$ of $BBG(r)$, $L(r)$, $B(r)$, and $U(r)$ for $c \in \{10, 100, 1000, 10000\}$ as r varies between $(0, 1)$. The left column shows errors for “small” r (i.e., r close to 0), the middle column shows errors for “mid-range” r , and the last column shows errors for the “high” range ($r \approx 1$).

Figure 2 shows the results of computing these approximations. Two points are immediate from

the plots: (i) approximation $L(r)$ is more accurate than $BBG(r)$ across almost the whole range of dimensions and r values; and (ii) for small r , $BBG(r)$ can be more accurate than $L(r)$, but in this case both $U(r)$ and $B(r)$ are much more accurate.

5.1.2 Comparisons of the approximation for fixed c and varying a

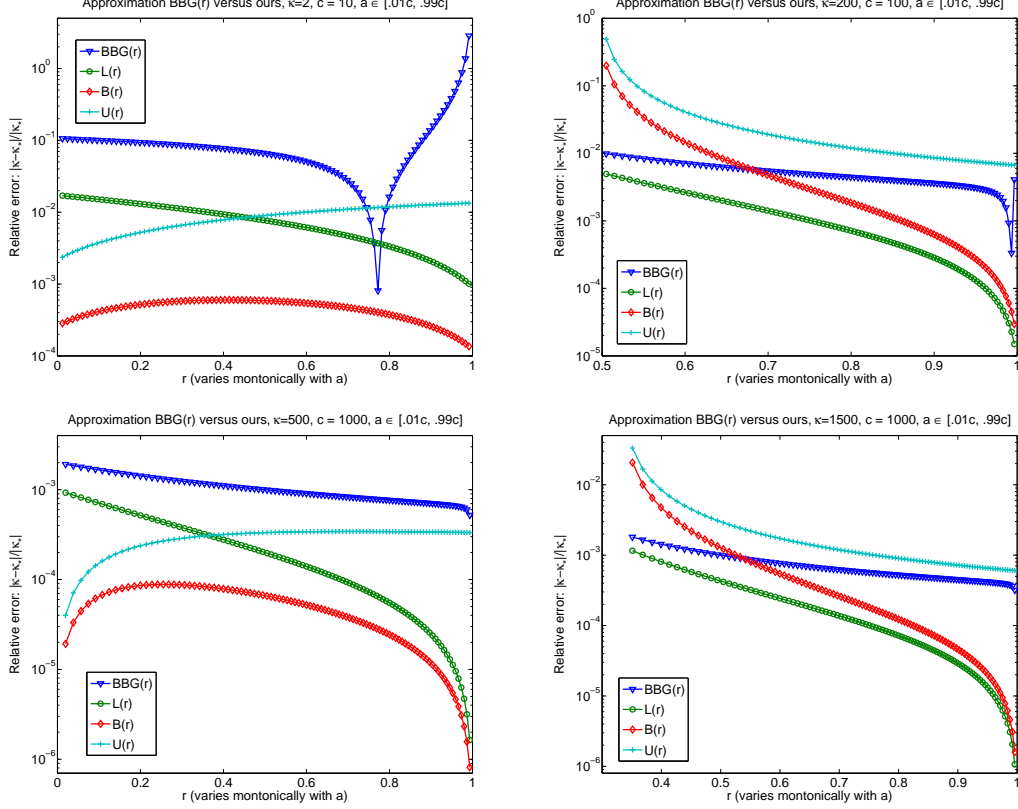


Figure 3: Relative errors of $BBG(r)$, $L(r)$, $B(r)$, and $U(r)$ for different sets of c and κ values, as a is varied from $0.01c$ to $0.99c$.

In our next set of experiments, we chose a few values of c and κ (see Figure 3), and varied a linearly to lie in the range $[0.01c, 0.99c]$. Figure 3 reports the relative errors of approximation incurred by the various approximations.

From the plots it is clear that one of $L(r)$, $B(r)$, or $U(r)$ always yields results more accurate than $BBG(r)$. The various results suggest the following rough rule-of-thumb: prefer $U(r)$ for $0 < r < a/(2c)$, prefer $B(r)$ for $a/(2c) \leq r < 2a/\sqrt{c}$ and prefer $L(r)$ for $2a/\sqrt{c} \leq r < 1$.

5.2 Clustering using mW distributions

Now we turn to our second set of experiments. Below we show results of two experiments: (i) with synthetic data, where a desired “true-clustering” is known; and (ii) with gene expression data for which previously axially symmetric clusters have been considered.

For both our experiments, we compare moW (Algorithm 1 with (4.6) for the E-step) against the diametrical clustering procedure of Dhillon et al. (2003). The key aim of the experiments is to show that the extra modelling power offered by a mixture of mW distributions can provide clustering results better than plain diametrical clustering.

Table 2: Percentages of accurately clustered points for diametrical clustering vs. moW (over 10 runs). Since this is simulated data, we knew the cluster labels. The accuracy is then computed by matching the predicted labels with the known ones. In line with the theory, with increasing concentration the modelling power offered by moW shows a clear advantage over ordinary diametrical clustering.

κ_2	Diametrical (avg/best/worst)-%	moW (avg/best/worst)-%
3	52.65 / 56.50 / 51.50	51.65 / 53.50 / 50.50
10	52.75 / 56.00 / 50.50	54.10 / 57.00 / 50.00
20	57.60 / 64.00 / 51.50	74.45 / 87.00 / 63.50
50	66.00 / 78.50 / 50.00	99.50 / 99.50 / 99.50
100	71.20 / 81.00 / 55.00	100.00 / 100.00 / 100.00

5.2.1 Synthetic data

We generated data that merely exhibit axial symmetry and have varying degrees of concentration around given mean directions. Since both the diametrical method as well as moW model axial symmetry they can be fairly compared on this data. The distinction comes, however, where moW further models concentration (via κ), and in case the generated data is sufficiently concentrated, this modelling translates into empirically superior performance. Naturally, to avoid unfairly skewing results in favour of moW, we do not compare it against diametrical clustering on synthetic data sampled from a mixture of mW distributions as moW explicitly optimises such a model.

For our data generation we need to sample points from $W_p(\kappa, \boldsymbol{\mu})$, for which we invoke a simplified version of the powerful Gibbs sampler of (Hoff, 2009) that can simulate Bingham-von Mises-Fisher distributions. We note here that Bingham distribution is parametrised by a matrix \mathbf{A} , and to use it for sampling Watson distributions, we merely need to realise that $\mathbf{A} = \kappa \boldsymbol{\mu} \boldsymbol{\mu}^T$.

With the sampling code in hand, we generate synthetic datasets with varying concentration as follows. First, two random unit vectors $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{P}^{29}$ are selected. Then, we fix $\kappa_1 = 3$ and sample 200 points from $W_3(\kappa_1, \boldsymbol{\mu}_1)$. Next, we vary κ_2 in the set $\{3, 10, 20, 50, 100\}$, and generate 200 points for each value of κ_2 by sampling from $W_p(\kappa_2, \boldsymbol{\mu}_2)$. Finally, by mixing the κ_1 component with each of the five κ_2 components we obtain five datasets \mathcal{X}_t ($1 \leq t \leq 5$).

Each of these five datasets is then clustered into two clusters, using moW and diametrical clustering. Both algorithms are run ten times each to smooth out the effect of random initializations. Table 2 shows the results of clustering by displaying the accuracy which measures the percentage of data points that were assigned to the “true” clusters (i.e., the true components in the mixture). The accuracies strongly indicate that explicit modelling of concentration leads to better clustering as κ_2 increases. In other words, larger κ_2 makes points from the second cluster more concentrated around $\pm \boldsymbol{\mu}_2$, thereby allowing easier separation between the clusters.

5.2.2 Real Data

We now compare clustering results of moW with those of diametrical clustering on three gene microarray datasets that were also used in the original diametrical clustering paper (Dhillon et al., 2003). These datasets are: (i) Human Fibroblasts (Iyer et al., 1999); (ii) Yest Cell Cycle (Spellman et al., 1998); and (iii) Rosetta yeast (Hughes et al., 2000). The respective matrix sizes that we used were: (i) 517×12 ; (ii) 696×82 ; and (iii) 900×300 (these 900 genes were randomly selected from the original 5245).

Since we do not have ground-truth clusterings for these datasets, we validate our results using internal measures. Specifically, we compute two scores: *homogeneity* and *separation*, which are defined below by H_{avg} and S_{avg} , respectively. Let $\mathcal{X}_j \subset \mathcal{X}$ denote cluster j ; then we define

$$H_{\text{avg}} = \frac{1}{n} \sum_{j=1}^K \sum_{\mathbf{x}_i \in \mathcal{X}_j} (\mathbf{x}_i^T \boldsymbol{\mu}_j)^2, \quad (5.1)$$

$$S_{\text{avg}} = \frac{1}{\sum_{j \neq l} |\mathcal{X}_j| |\mathcal{X}_l|} \sum_{j \neq l} |\mathcal{X}_j| |\mathcal{X}_l| \min(\boldsymbol{\mu}_j^T \boldsymbol{\mu}_l, -\boldsymbol{\mu}_j^T \boldsymbol{\mu}_l). \quad (5.2)$$

Table 3: Clustering accuracy on gene-expression datasets (over 10 runs). Noticeable differences (i.e., > 0.02) between the algorithms are highlighted in bold.

Method	Diametrical (avg/best/worst)	moW (avg/best/worst)
Yeast-4		
Homogeneity	0.38 / 0.38 / 0.38	0.37 / 0.37 / 0.37
Separation	-0.00 / -0.23 / 0.24	-0.04 / -0.23 / 0.20
Yeast-6		
Homogeneity	0.41 / 0.41 / 0.40	0.41 / 0.41 / 0.40
Separation	-0.06 / -0.15 / 0.14	-0.07 / -0.20 / 0.13
Rosetta-2		
Homogeneity	0.16 / 0.17 / 0.16	0.16 / 0.17 / 0.16
Separation	0.24 / 0.08 / 0.28	-0.20 / -0.28 / 0.09
Rosetta-4		
Homogeneity	0.23 / 0.23 / 0.23	0.23 / 0.23 / 0.23
Separation	-0.01 / -0.08 / 0.16	-0.03 / -0.09 / 0.12
Fibroblast-2		
Homogeneity	0.70 / 0.70 / 0.70	0.70 / 0.70 / 0.70
Separation	0.26 / -0.65 / 0.65	-0.01 / -0.65 / 0.65
Fibroblast-5		
Homogeneity	0.78 / 0.78 / 0.78	0.76 / 0.76 / 0.75
Separation	-0.05 / -0.28 / 0.40	-0.12 / -0.30 / 0.35

We note a slight departure from the standard in our definitions above. In (5.1), instead of summing over $\mathbf{x}_i^T \boldsymbol{\mu}_j$, we sum over their squares, while in (5.2), instead of $\boldsymbol{\mu}_j^T \boldsymbol{\mu}_l$, we use $\min(\boldsymbol{\mu}_j^T \boldsymbol{\mu}_l, -\boldsymbol{\mu}_j^T \boldsymbol{\mu}_l)$ because for us $+\boldsymbol{\mu}_j$ and $-\boldsymbol{\mu}_j$ represent the same cluster.

We note that diametrical clustering optimises precisely the criterion (5.1), and is thus favoured by our criterion. Higher values of H_{avg} mean that the clusters have higher intra-cluster cohesiveness, and thus are “better” clusters. In contrast, lower values of S_{avg} mean that the inter-cluster dissimilarity is high, i.e., better separated clusters.

Table 3 shows results yielded by diametrical clustering and moW on the three different gene datasets. For each dataset, we show results for two values of K . The H_{avg} values indicate that moW yields clusters having approximately the same intra-cluster cohesiveness as diametrical. However, moW attains better inter-cluster separation as it more frequently leads to lower S_{avg} values.

6 Conclusions

We studied the multivariate Watson distribution, a fundamental tool for modelling axially symmetric data. We solved the difficult nonlinear equations that arise in maximum-likelihood parameter estimation. In high-dimensions these equations pose severe numerical challenges. We derived tight two-sided bounds that led to approximate solutions to these equations; we also showed our solutions to be accurate. We applied our results to mixture-modelling with Watson distributions and consequently uncovered a connection to the diametrical clustering algorithm of (Dhillon et al., 2003). Our experiments showed that for clustering axially symmetric data, the additional modelling power offered by mixtures of Watson distributions can lead to better clustering. Further refinements to the clustering procedure, as well as other applications of Watson mixtures in high-dimensional settings is left as a task for the future.

Acknowledgements

The first author thanks Prateek Jain for initial discussions related to Watson distributions. The second author acknowledges support of the Russian Basic Research Fund (grant 11-01-00038-a).

References

References

- Andrews, G. E., Askey, R., Roy, R., 1999. Special functions. Cambridge University Press.
- Banerjee, A., Dhillon, I. S., Ghosh, J., Sra, S., 2003. Generative model-based clustering of directional data. In: Proceedings of The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD-2003). pp. 19–28.
- Banerjee, A., Dhillon, I. S., Ghosh, J., Sra, S., Sep 2005. Clustering on the Unit Hypersphere using von Mises-Fisher Distributions. *J. Machine Learning Research* 6, 1345–1382.
- Bijral, A., Breitenbach, M., Grudic, G. Z., 2007. Mixture of Watson Distributions: A Generative Model for Hyperspherical Embeddings. In: Artificial Intelligence and Statistics (AISTATS 2007). pp. 35–42.
- Cuyt, A., Petersen, V. B., Verdonk, B., Waadeland, H., Jones, W. B., 2008. Handbook of Continued Fractions for Special Functions. Springer.
- Dempster, A., Laird, N., Rubin, D., 1977. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society* 39.
- Dhillon, I. S., Marcotte, E. M., Roshan, U., 2003. Diametrical clustering for identifying anti-correlated gene clusters. *Bioinformatics* 19 (13), 1612–1619.
- Erdélyi, A., Magnus, W., Oberhettinger, F., Tricomi, F. G., 1953. Higher transcendental functions. Vol. 1. McGraw Hill.
- Gautschi, W., 1977. Anomalous Convergence of a Continued Fraction for Ratios of Kummer Functions. *Mathematics of Computation* 31 (140), 994–999.
- Gil, A., Segura, J., Temme, N. M., 2007. Numerical Methods for Special Functions. Cambridge University Press.
- M.O. Gonzalez, Classical Complex Analysis (Pure and Applied Mathematics), CRC Press, 1991.
- A. Hurwitz, R. Courant, Vorlesungen über allgemeine Funktionentheorie und Elliptische Funktionen, Second Edition, Verlag von Julius Springer, Berlin, 1925.
- R. L. Graham, D. E. Knuth, and O. Patashnik, 1998. Concrete Mathematics. Addison Wesley.
- Hoff, P. D., 2009. Simulation of the Matrix Bingham–von Mises–Fisher Distribution, With Applications to Multivariate and Relational Data. *Journal of Computational and Graphical Statistics* 18 (2), 438–456.
- Hughes, T. R., Marton, M. J., Jones, A. R., Roberts, C. J., Stoughton, R., Armour, C. D., Bennett, H. A., Coffey, E., Dai, H., Shoemaker, D. D., Gachotte, D., Chakraburttty, K., Simon, J., Bard, M., Friend, S. H., 2000. Functional discovery via a compendium of expression profiles. *Cell* 102, 109–126.
- Iyer, V. R., Eisen, M. B., Ross, D. T., Schuler, G., Moore, T., Lee, J. C. F., Trent, J. M., Staudt, L. M., Hudson, J., Boguski, M. S., Lashkari, D., Shalon, D., Botstein, D., Brown, P. O., 1999. The Transcriptional Program in the Response of Human Fibroblasts to Serum. *Science* 283 (5398), 83–87.
- Karp, D., 2011. Turán’s inequality for the Kummer function of the phase shift of two parameters. *Journal of Mathematical Sciences* 178 (2), 178–186.
- Karp, D., Sitnik, S. M., 2010. Log-convexity and log-concavity of hypergeometric-like functions. *Journal of Mathematical Analysis and Applications* 364, 384–394.
- Mardia, K. V., Jupp, P., 2000. Directional Statistics, 2nd Edition. John Wiley & Sons.
- Spellman, P. T., Sherlock, G., Zhang, M., Iyer, V. R., Anders, K., Eisen, M., Brown, P. O., Botstein, D., Futcher, B., 1998. Comprehensive identification of cell cycle regulated gene of the yeast *Saccharomyces Cerevisia* by microarray hybridization. *Mol. Bio. Cell* 9, 3273–3297.
- Sra, S., 2007. Matrix Nearness Problems in Data Mining. Ph.D. thesis, Univ. of Texas at Austin.
- Tanabe, A., Fukumizu, K., Oba, S., Takenouchi, T., Ishii, S., 2007. Parameter estimation for von Mises-Fisher distributions. *Computational Statistics* 22 (1), 145–157.
- Watson, G. S., 1965. Equatorial distributions on a sphere. *Biometrika* 52 (1-2), 193–201.

A Mathematical Details

This appendix includes mathematical details supporting the technical material of the main text. While many of the facts are classic knowledge, some might be found only in specialised literature. Thus, we have erred on the side of including too much rather than too little.

A.1 Hypergeometric functions

Hypergeometric functions provide one of the richest classes of functions in analysis. Indeed, any series with ratio of neighbouring terms equal to a rational function of the summation index is a constant multiple of the *generalised hypergeometric function* ${}_pF_q$ defined by the power-series

$${}_pF_q(a_1, \dots, a_p; c_1, \dots, c_q; z) = \sum_{k \geq 0} \frac{a_1^{\bar{k}} \dots a_p^{\bar{k}}}{c_1^{\bar{k}} \dots c_q^{\bar{k}}} \frac{z^k}{k!}, \quad (\text{A.1})$$

where $a^{\bar{k}} = a(a+1) \dots (a+k-1)$ is the *rising factorial* (often also denoted by the Pochhammer symbol $(a)_k$). Hypergeometric functions arise naturally as solutions to certain differential equations; for a gentle introduction to hypergeometric functions we refer the reader to (Graham et al., 1998), while for a more advanced treatment the reader may find (Andrews et al., 1999) valuable.

In this paper, we restrict our attention to Kummer's confluent hypergeometric function: ${}_1F_1$, which is also denoted as M . Moreover, we limit our attention to the case of real valued arguments.

A.1.1 Some useful identities for $M(a, c, x)$

We list below some identities for M that we will need for our analysis. To ease the notational burden, we also use the shorthand $M_i \equiv M(a+i, c+i, x)$; e.g., $M_0 \equiv M(a, c, x)$.

$$\frac{d^n}{dx^n} M_0 = \frac{a^{\bar{n}}}{c^{\bar{n}}} M_n. \quad (\text{A.2})$$

$$M_1 = \frac{c(1-c+x)}{ax} M_0 + \frac{c(c-1)}{ax} M_{-1}. \quad (\text{A.3})$$

$$(c-a)M(a+1, c+2, x) = (c+1)M_1 - (a+1)M_2 \quad (\text{A.4})$$

$$(c-a)xM(a+2, c+3, x) = (c+1)(c+2)[M_2 - M_1]; \quad (\text{A.5})$$

$$(a+1)xM_2 = (c+1)(x-c)M_1 + c(c+1)M_0 \quad (\text{A.6})$$

$$xM(a+2, c+3, x) = (c+2)[M_2 - M(a+1, c+2, x)], \quad (\text{A.7})$$

Identity (A.2) follows inductively; (A.3) from (Cuyt et al., 2008, 16.1.9c); (A.4) from (Erdélyi et al., 1953, formula 6.4(4)); (A.5) on combining (Erdélyi et al., 1953, formula 6.4(5)) with (Erdélyi et al., 1953, formula 6.4(4)); (A.6) from (A.3) by replacing $c \rightarrow c+1$, $a \rightarrow a+1$; and (A.7) from (Erdélyi et al., 1953, formula 6.4(5)).

Now we build on the above identities to introduce a technical but crucial lemma.

Lemma A.1. *The following identity holds for the Kummer function:*

$$M_1^2 - M_2 M_0 = \frac{(c-a)x}{c+1} \left[\frac{1}{c+1} M(a+1, c+2, x)^2 - \frac{1}{c+2} M(a+2, c+3, x) M(a, c+1, x) \right. \\ \left. + \frac{1}{c(c+1)} M(a+1, c+2, x) M(a+2, c+2, x) \right]. \quad (\text{A.8})$$

Proof. Application of (A.4) and (A.5) yields after collecting terms

$$\begin{aligned} & \frac{1}{c+1}M(a+1, c+2, x)^2 - \frac{1}{c+2}M(a+2, c+3, x)M(a, c+1, x) + \frac{1}{c(c+1)}M(a+1, c+2, x)M_2 \\ &= \frac{a(a+1)}{c(c-a)^2}M_2^2 - \frac{c(c+1)}{(c-a)^2x}M_2M_0 - \frac{(c+1)(a-x)}{(c-a)^2x}M_1^2 \\ & \quad + \frac{c(c+1)}{(c-a)^2x}M_1M_0 + \frac{ac(c+1) - x(a+c+2ac)}{(c-a)^2cx}M_2M_1. \quad (\text{A.9}) \end{aligned}$$

Application of this formula allows us to write the difference between the left-hand and right-hand sides of (A.8) as

$$\text{lhs} - \text{rhs} = \frac{cM_1 - aM_2}{c(c+1)(c-a)}(x(a+1)M_2 - (c+1)(x-c)M_1 - c(c+1)M_0) = 0,$$

where the equality to 0 is due to (A.6). \square

A.2 The Kummer ratio

The central object of study in this paper is the *Kummer-ratio*:

$$g(x) = g(a, c; x) := \frac{M'(a, c, x)}{M(a, c, x)} = \frac{a}{c} \frac{M(a+1, c+1, x)}{M(a, c, x)}. \quad (\text{A.10})$$

This ratio satisfies many fascinating properties; but of necessity, we must content ourselves with only the essential properties. In particular, our analysis focuses on the following: (i) proving that g is monotonic, and thereby invertible; and (ii) obtaining bounds on the root of $g(x) - r = 0$. In the sequel, it will be useful to use the slightly more general function

$$f_\mu(x) := \frac{M(a+\mu, c+\mu, x)}{M(a, c, x)}, \quad \mu > 0, \quad (\text{A.11})$$

so that $g(x) = (a/c)f_1(x)$. Before proving monotonicity of g , we derive two useful lemmas.

Lemma A.2 (Log-convexity). *Let $c > a > 0$ and $x \geq 0$. Then the function*

$$\mu \mapsto \frac{\Gamma(a+\mu)}{\Gamma(c+\mu)}M(a+\mu, c+\mu, x) = \sum_{k=0}^{\infty} \frac{\Gamma(a+\mu+k)}{\Gamma(c+\mu+k)} \frac{x^k}{k!} =: h_{a,c}(\mu; x)$$

is strictly log-convex on $[0, \infty)$ (note that h is a function of μ).

Proof. Write the power-series expansion in x for $h_{a,c}(\mu; x)$ as

$$h_{a,c}(\mu; x) = \sum_{k=0}^{\infty} h_k(a, c, \mu) \frac{x^k}{k!}, \quad h_k(a, c, \mu) = \frac{\Gamma(a+\mu+k)}{\Gamma(c+\mu+k)}.$$

Since log-convexity is additive it is sufficient to prove that $\mu \mapsto h_k(a, c, \mu)$ is log-convex. For this we compute the second-derivative

$$\frac{\partial^2}{\partial \mu^2} \log h_k(a, c, \mu) = \psi'(a+\mu+k) - \psi'(c+\mu+k),$$

where ψ is the logarithmic derivative of the gamma function. We need to show that this expression is positive when $c > a > 0$, $k \geq 0$ and $\mu \geq 0$. According to the Gauss formula (Andrews et al., 1999, Theorem 1.6.1)

$$\psi(x) = \int_0^{\infty} \left(\frac{e^{-t}}{t} - \frac{e^{-tx}}{1-e^{-t}} \right) dt,$$

so that

$$\psi''(x) = - \int_0^\infty \frac{t^2 e^{-tx}}{1 - e^{-t}} dt < 0.$$

Hence the function $\psi'(x)$ is decreasing and our claim follows. \square \square

Lemma A.3. *Let $c > a > 0$, and $x \geq 0$. Then the function*

$$\mu \mapsto \frac{\Gamma(a + \mu)}{\Gamma(c + \mu)} M(c - a, c + \mu, x) =: \hat{h}_{a,c}(\mu; x)$$

is strictly log-convex on $[0, \infty)$.

Proof. Using precisely the same argument as in the proof of Lemma A.2 we see that $\mu \mapsto \Gamma(a + \mu)/\Gamma(c + \mu)$ is log-convex. Next, the log-convexity of $\mu \mapsto M(c - a; c + \mu; x)$ has been proved by many authors (see, for instance, Karp and Sitnik (2010) and references therein). Thus multiplicativity of log-convexity completes the proof. \square \square

With these two lemmas in hand we are ready to prove the first main theorem.

Theorem A.4 (Monotonicity). *Let $c > a > 0$. The function $x \mapsto f_\mu(x)$ is monotone increasing on $(-\infty, \infty)$, with $f_\mu(-\infty) = 0$ and $f_\mu(\infty) = \Gamma(c + \mu)\Gamma(a)/(\Gamma(c)\Gamma(a + \mu))$.*

Proof. We divide the proof into two cases: (i) $x \geq 0$, and (ii) $x < 0$.

Case i: Let $x \geq 0$. It follows from (A.2) that

$$\frac{d}{dx} M(a, c, x) = \frac{a}{c} M(a + 1, c + 1, x),$$

whereby (using our compact notation) we have

$$M_0^2 f'_\mu(x) = \frac{a + \mu}{c + \mu} M_{\mu+1} M_0 - \frac{a}{c} M_\mu M_1.$$

We need to show that the above expression is positive, which amounts to showing

$$\frac{a + \mu}{c + \mu} \frac{M_{\mu+1}}{M_\mu} > \frac{a}{c} \frac{M_1}{M_0}, \tag{A.12}$$

or equivalently

$$\frac{[\Gamma(a + \mu + 1)/\Gamma(c + \mu + 1)] M_{\mu+1}}{[\Gamma(a + \mu)/\Gamma(c + \mu)] M_\mu} > \frac{[\Gamma(a + 1)/\Gamma(c + 1)] M_1}{[\Gamma(a)/\Gamma(c)] M_0}.$$

The last inequality follows from Lemma A.2. To see how, recall that if $\mu \mapsto h(\mu)$ is log-convex, then the function $\mu \mapsto h(\mu + \delta)/h(\mu)$ is increasing² for each fixed $\delta > 0$. Thus, in particular applying this property to $h_{a,c}(\mu; x)$ with $\delta = 1$ we have

$$\frac{h_{a,c}(\mu + 1; x)}{h_{a,c}(\mu; x)} > \frac{h_{a,c}(1; x)}{h_{a,c}(0; x)},$$

which is precisely the required inequality. This establishes the monotonicity. The value of $f_\mu(\infty)$ follows from the asymptotic formula (Andrews et al., 1999, Corollary 4.2.3):

$$M(a, c, x) \sim \frac{\Gamma(c)}{\Gamma(a)} \frac{e^x}{x^{c-a}} {}_2F_0(c - a, 1 - a; -; 1/x), \quad x \rightarrow \infty. \tag{A.13}$$

²Easily verified by noting that when h is log-convex, its logarithmic derivative $h'(\mu)/h(\mu)$ is increasing, which immediately implies that the derivative of $h(\mu + \delta)/h(\mu)$ is positive.

Case ii: Let $x < 0$. Like in Case (i) we need to show that

$$\frac{[\Gamma(a + \mu + 1)/\Gamma(c + \mu + 1)]M_{\mu+1}}{[\Gamma(a + \mu)/\Gamma(c + \mu)]M_\mu} > \frac{[\Gamma(a + 1)/\Gamma(c + 1)]M_1}{[\Gamma(a)/\Gamma(c)]M_0}. \quad (\text{A.14})$$

but this time for $x < 0$. Apply the Kummer transformation $M(a; c; x) = e^x M(c - a; c; -x)$ and write $y = -x > 0$ to get

$$\frac{[\Gamma(a + \mu + 1)/\Gamma(c + \mu + 1)]M(c - a; c + \mu + 1; y)}{[\Gamma(a + \mu)/\Gamma(c + \mu)]M(c - a; c + \mu; y)} > \frac{[\Gamma(a + 1)/\Gamma(c + 1)]M(c - a; c + 1; y)}{[\Gamma(a)/\Gamma(c)]M(c - a; c; y)}.$$

Using the notation introduced in Lemma A.3 the last inequality becomes

$$\frac{\hat{h}_{a,c}(\mu + 1; x)}{\hat{h}_{a,c}(\mu; x)} > \frac{\hat{h}_{a,c}(1; x)}{\hat{h}_{a,c}(0; x)},$$

which holds as a consequence of the log-convexity of $\mu \mapsto \hat{h}_{a,c}(\mu; x)$. Finally from the Kummer transformation and formula (A.13) we have

$$f_\mu(x) = \frac{M(c - a; c + \mu; -x)}{M(c - a; c; -x)} \sim \frac{\Gamma(c + \mu)}{\Gamma(c)} \frac{1}{(-x)^\mu} \frac{{}_2F_0(a + \mu, 1 + a - c; -; -1/x)}{{}_2F_0(a, 1 + a - c; -; -1/x)} \rightarrow 0 \text{ as } x \rightarrow -\infty. \quad \square$$

(A.15)

A.3 Bounds on κ

We now derive bounds on $\kappa(r)$, the solution to the equation $g(a, c; \kappa) = r$. For ease of exposition we divide the bounds into two cases: (i) positive $\kappa(r)$ and (ii) negative $\kappa(r)$.

Theorem A.5 (Positive κ). *Let $\kappa(r)$ be the solution to (2.7); $c > a > 0$, and $r \in (a/c, 1)$. Then, we have the bounds*

$$\frac{rc - a}{r(1 - r)} \left(1 + \frac{1 - r}{c - a} \right) < \kappa(r) < \frac{rc - a}{2r(1 - r)} \left(1 + \sqrt{1 + \frac{4(c + 1)r(1 - r)}{a(c - a)}} \right) < \frac{rc - a}{r(1 - r)} \left(1 + \frac{r}{a} \right). \quad (\text{A.16})$$

Proof. Lower-bound. To simplify notation we use $x = \kappa(r)$ below. Denote $r_1 = g(a + 1, c + 1; x)$. Now, replace $a \leftarrow a + 1$, $c \leftarrow c + 1$ and divide by M_1 in identity (A.3) to obtain

$$x = \frac{cr - a}{r(1 - r_1)}, \quad (\text{A.17})$$

where as before $r = g(a, c, x)$. To prove the lower bound in (A.16) we need to show that

$$\frac{cr - a}{r(1 - r_1)} > \frac{cr - a}{r(1 - r)} + \frac{cr - a}{r(c - a)}.$$

Elementary calculation reveals that this inequality is equivalent to

$$\frac{(c - a - 1)r + 1}{c - a + 1 - r} < r_1,$$

once we account for $cr - a > 0$ by our hypothesis. Plugging in the definitions of r and r_1 we get:

$$\frac{(c - a - 1)aM_1 + cM_0}{(c - a + 1)cM_0 - aM_1} < \frac{(a + 1)M_2}{(c + 1)M_1},$$

where as before we use $M_i = M(a + i, c + i, x)$. Cross-multiplying, we obtain

$$h(x) := c(c - a + 1)(a + 1)M_2M_0 - (c + 1)(c - a - 1)aM_1^2 - c(c + 1)M_1M_0 - a(a + 1)M_2M_1 > 0.$$

Now on noticing that $c(c-a+1)(a+1) = ac(c-a) + c(c+1)$ and $(c-a-1)(c+1)a = ac(c-a) - a(a+1)$, we can regroup $h(x)$ to get

$$h(x) = ac(c-a)[M_2M_0 - M_1^2] + (M_2 - M_1)[c(c+1)M_0 - a(a+1)M_1].$$

Now identity (A.5) yields

$$M_2 - M_1 = \frac{x(c-a)}{(c+1)(c+2)}M(a+2, c+3, x),$$

and a simple calculation shows that

$$c(c+1)M_0 - a(a+1)M_1 = \frac{ax(c-a)}{c+1}M(a+1, c+2, x) + (c-a)(c+a+1)M(a, c+1, x).$$

Substituting these formulae into $h(x)$ and using identity (A.8) we get a rather complicated term

$$\begin{aligned} h(x) = & -\frac{ac(c-a)^2x}{c+1} \left[\frac{1}{c+1}M(a+1, c+2, x)^2 - \frac{1}{c+2}M(a+2, c+3, x)M(a, c+1, x) \right. \\ & \left. + \frac{1}{c(c+1)}M(a+1, c+2, x)M_2 \right] + \frac{a(c-a)^2x^2}{(c+1)^2(c+2)}M(a+2, c+3, x)M(a+1, c+2, x) \\ & + \frac{(c-a)^2(c+a+1)x}{(c+1)(c+2)}M(a+2, c+3, x)M(a, c+1, x). \end{aligned} \quad (\text{A.18})$$

However, on invoking the contiguous relation (A.7), $h(x)$ can be simplified considerably to yield

$$\frac{(c+1)h(x)}{(c-a)^2x} = \frac{(a+1)(c+1)}{c+2}M(a, c+1, x)M(a+2, c+3, x) - aM(a+1, c+2, x)^2. \quad (\text{A.19})$$

Therefore, the condition $h(x) > 0$ is equivalent to (after introducing the notation $c' = c+1$)

$$\frac{c'}{c'+1}M(a, c', x)M(a+2, c'+2, x) - \frac{a}{a+1}M(a+1, c'+1, x)^2 > 0, \quad (\text{A.20})$$

or, in other words

$$\frac{(a+1)}{(c'+1)} \frac{M(a+2, c'+2, x)}{M(a+1, c'+1, x)} > \frac{a}{c'} \frac{M(a+1, c'+1, x)}{M(a, c', x)}.$$

But this final inequality follows from Theorem A.4 by using $\mu = 1$ in (A.12).

Upper-bound. To prove the upper bound, first for brevity introduce the notation $b = c-a$, and $q = 1-r$. The lower-bound in (A.16) can be then rewritten as ($b-cq = cr-a > 0$)

$$\frac{b-cq}{q(1-q)} \left(1 + \frac{q}{b} \right) < x,$$

which in turn can be rearranged to

$$q^2(x - c/b) - q(x + c - 1) + b < 0.$$

Note first that the equation $q^2(x - c/b) - q(x + c - 1) + b = 0$ has two distinct real roots since the discriminant (upon using $b = c-a$)

$$D = (x + c - 1)^2 - 4(c-a)(x - c/(c-a)) = (1 - x - c)^2 + 4ax + 4c(1 - x) = (1 - x + c)^2 + 4ax > 0.$$

We need to consider three cases:

(1) $x - c/b > 0$ implies q lies between the roots

$$\frac{b(x + c - 1)}{2(bx - c)} - \frac{b\sqrt{D}}{2(bx - c)} < q < \frac{b(x + c - 1)}{2(bx - c)} + \frac{b\sqrt{D}}{2(bx - c)}.$$

(2) $x - c/b < 0$ implies q is smaller than the smaller root or bigger than the bigger root, i.e.,

$$q < \frac{b(x+c-1)}{2(bx-c)} + \frac{b\sqrt{D}}{2(bx-c)}, \quad \text{or} \quad \frac{b(x+c-1)}{2(bx-c)} - \frac{b\sqrt{D}}{2(bx-c)} < q.$$

(3) $x - c/b = 0$ implies $b/(x+c-1) < q$.

Since

$$\lim_{x \rightarrow c/b} \left(\frac{b(x+c-1)}{2(bx-c)} - \frac{b\sqrt{D}}{2(bx-c)} \right) = \frac{b}{x+c-1},$$

in all three situations we have

$$\frac{b(x+c-1) - b\sqrt{D}}{2(bx-c)} < q.$$

Changing $a \rightarrow a+1$ and $c \rightarrow c+1$ here (recall that $b = c - a$) we get

$$0 < \frac{b(x+c) - b\sqrt{(x+c)^2 - 4(bx-c-1)}}{2(bx-c-1)} < 1 - r_1, \quad (\text{A.21})$$

where as before $r_1 = g(a+1, c+1, x)$. The positivity is clear on inspection for both $bx - c - 1 > 0$ and $bx - c - 1 < 0$. Next, after suitably rewriting (A.17), we have

$$x = \frac{b - cq}{(1-q)(1-r_1)}.$$

Applying inequality (A.21) here, we obtain

$$x < \frac{2(b-cq)(bx-c-1)}{(1-q)b(x+c-\sqrt{(x+c)^2-4(bx-c-1)})}.$$

Squaring and simplifying we get the inequality

$$(bx-c-1)(x^2q(1-q)b(c-b) - xb(b-cq)(c-b) - (c+1)(b-cq)^2) < 0$$

for $bx - c - 1 > 0$ and the inequality

$$(bx-c-1)(x^2q(1-q)b(c-b) - xb(b-cq)(c-b) - (c+1)(b-cq)^2) > 0$$

for $bx - c - 1 < 0$. Hence both situations reduce to the single inequality

$$x^2q(1-q)b(c-b) - xb(b-cq)(c-b) - (c+1)(b-cq)^2 < 0,$$

which on plugging $q = 1 - r$ and $b = c - a$ becomes

$$x^2r(1-r)a(c-a) - xa(c-a)(cr-a) - (c+1)(cr-a)^2 < 0.$$

Since the coefficient at x^2 is clearly positive x must lie between the roots, in particular it should be smaller than the bigger root, which is the upper bound in (A.16).

The rightmost bound. Verifying the rightmost inequality is an exercise in high-school algebra. \square

Theorem A.6 (Negative κ). *Let $\kappa(r)$ be the solution to (2.7), $c > a > 0$, and r lie in $(0, a/c)$. Then, we have the following bounds:*

$$\frac{rc-a}{r(1-r)} \left(1 + \frac{1-r}{c-a} \right) < \frac{rc-a}{2r(1-r)} \left(1 + \sqrt{1 + \frac{4(c+1)r(1-r)}{a(c-a)}} \right) < \kappa(r) < \frac{rc-a}{r(1-r)} \left(1 + \frac{r}{a} \right) \quad (\text{A.22})$$

Proof. Upper bound: To simplify notation we write as before $x = \kappa(r)$. Recall that $r_1 = g(a+1, c+1; x)$; according to (A.14), the value $r_1 > r$; so in view of $cr - a < 0$ and (A.17), we have the inequality

$$x < \frac{cr - a}{r(1 - r)},$$

which is equivalent to (noting that $x < 0$)

$$r^2 + \left(\frac{c}{x} - 1\right)r - \frac{a}{x} < 0.$$

Thus, r must lie between the roots of the quadratic

$$r^2 + \left(\frac{c}{x} - 1\right)r - \frac{a}{x} = 0.$$

Straightforward analysis shows that the discriminant is positive for all real x if $c > a > 0$, and we have two distinct real roots so that

$$\frac{1}{2} - \frac{c}{2x} - \frac{1}{2}\sqrt{1 + \frac{2(2a - c)}{x} + \frac{c^2}{x^2}} < r < \frac{1}{2} - \frac{c}{2x} + \frac{1}{2}\sqrt{1 + \frac{2(2a - c)}{x} + \frac{c^2}{x^2}}. \quad (\text{A.23})$$

We will use the lower bound above written here for $r_1 = g(a+1, c+1; x)$

$$0 < \frac{1}{2} - \frac{c+1}{2x} - \frac{1}{2}\sqrt{1 + \frac{2(2a - c + 1)}{x} + \frac{(c+1)^2}{x^2}} < r_1.$$

Applying this lower bound for r_1 to (A.17) and dividing by $2x < 0$, we obtain the bound

$$x < \frac{2(cr - a)}{r(1 + (c+1)/x + \sqrt{1 + 2(2a - c + 1)/x + (c+1)^2/x^2}}.$$

By high school algebra we immediately conclude that the denominator is positive, so that

$$(-x)\sqrt{1 + 2(2a - c + 1)/x + (c+1)^2/x^2} < (x + 1 - c) + 2a/r$$

Squaring and rearranging we obtain the desired upper bound in (A.22).

Lower bounds: First we prove the leftmost bound in (A.22) (i.e., we show that it less than x). Since $r \in (0, a/c)$ we have $cr - a < 0$; so, following the line of proof of Theorem A.5 we get

$$\frac{(c - a - 1)r + 1}{c - a + 1 - r} > r_1$$

which leads to $h(x) < 0$ where h is as defined in the course of proof of Theorem A.5. However, since $x < 0$ for $r \in (0, a/c)$ we must again show that (where $c' = c + 1$)

$$\frac{(a+1)M(a+2; c'+2; x)}{(c'+1)M(a+1; c'+1; x)} > \frac{aM(a+1; c'+1; x)}{c'M(a; c'; x)},$$

but this time for $x < 0$. Applying the Kummer transformation to the inequality above leads to

$$\frac{[\Gamma(a + \mu + 1)/\Gamma(c + \mu + 1)]M(c - a; c + \mu + 1; y)}{[\Gamma(a + \mu)/\Gamma(c + \mu)]M(c - a; c + \mu; y)} > \frac{[\Gamma(a + 1)/\Gamma(c' + 1)]M(c' - a; c + 1; y)}{[\Gamma(a)/\Gamma(c')]M(c' - a; c'; y)},$$

with $\mu = 1$ and $y = -x > 0$. This inequality follows immediately from Theorem A.3 as we have demonstrated in the proof of the previous theorem.

Proving the second (and tighter) lower bound in (A.22) requires more work. Introduce thus the notation $b = c - a$ and $q = 1 - r$. The leftmost bound in (A.22) can be rewritten as

$$\frac{b - cq}{q(1 - q)} \left(1 + \frac{q}{b}\right) < x,$$

which can be rearranged to

$$q^2(x - c/b) - q(x + c - 1) + b < 0.$$

The discriminant of the corresponding quadratic equation equals

$$D = (x + c - 1)^2 - 4(c - a)(x - c/(c - a)) = (x + c - 1)^2 - 4(c - a)x + 4c > 0,$$

since $c - a > 0$ and $x < 0$. For $x - c/b < 0$ this implies that q must be smaller than the smaller root or bigger than the bigger root of $q^2(x - c/b) - q(x + c - 1) + b = 0$. That is,

$$q < \frac{b(x + c - 1)}{2(bx - c)} + \frac{b\sqrt{(x + c - 1)^2 - 4(bx - c)}}{2(bx - c)},$$

or

$$\frac{b(x + c - 1)}{2(bx - c)} - \frac{b\sqrt{(x + c - 1)^2 - 4(bx - c)}}{2(bx - c)} < q.$$

Changing $a \rightarrow a + 1$ and $c \rightarrow c + 1$ here (recall that $b = c - a$), from the second inequality we obtain

$$0 < \frac{b[(x + c) - \sqrt{(x + c)^2 - 4(bx - c - 1)}]}{2(bx - c - 1)} < 1 - r_1,$$

where as before $r_1 = g(a + 1, c + 1, x)$. Positivity follows by separately considering $x + c \geq 0$ and $x + c < 0$. Next, a simple manipulation of (A.17) shows that

$$x = \frac{b - cq}{(1 - q)(1 - r_1)}.$$

Applying the above inequality concerning r_1 here, we obtain the bound (since $b - cq < 0$)

$$x > \frac{2(b - cq)(bx - c - 1)}{b(1 - q)[(x + c) - \sqrt{(x + c)^2 - 4(bx - c - 1)}]}.$$

Here $b - cq < 0$ (since $r < a/c$), $bx - c - 1 < 0$ (since $x < 0$, $b > 0$) and $(x + c) - \sqrt{(x + c)^2 - 4(bx - c - 1)} < 0$ (since the second term is bigger than the first). So the above bound on x is negative as expected. By simple algebra and in view of the signs we have

$$x + c - \frac{2(b - cq)(bx - c - 1)}{xb(1 - q)} > \sqrt{(x + c)^2 - 4(bx - c - 1)} > 0.$$

Squaring yields

$$\begin{aligned} & x^2 b^2 (1 - q)^2 (x + c)^2 - 4(b - cq)(bx - c - 1)xb(1 - q)(x + c) + 4(b - cq)^2 (bx - c - 1)^2 \\ & > ((x + c)^2 - 4(bx - c - 1))x^2 b^2 (1 - q)^2, \end{aligned}$$

whereby reducing similar terms and on dividing by $-4(bx - c - 1) > 0$ we obtain

$$(b - cq)xb(1 - q)(x + c) - (b - cq)^2 (bx - c - 1) - x^2 b^2 (1 - q)^2 > 0.$$

Simplifying we get

$$x^2 q(1 - q)b(c - b) - xb(b - cq)(c - b) - (c + 1)(b - cq)^2 < 0,$$

so that after plugging in $q = 1 - r$ and $b = c - a$ we have the inequality

$$x^2 r(1 - r)a(c - a) - xa(c - a)(cr - a) - (c + 1)(cr - a)^2 < 0.$$

Thus, x must be between the roots of this quadratic. The fact that it is bigger than the smallest root is precisely the inequality between x and the middle term in (A.22).

Finally, comparing the two lower bounds in (A.22) leads to the inequality

$$\frac{1}{2} + \frac{1 - r}{c - a} > \frac{1}{2} \sqrt{1 + \frac{4(1 - r)r(c + 1)}{(c - a)a}},$$

which upon squaring and simplifying reduces to $r < a/c$, the hypothesis of the theorem. \square

\square

In the next theorem we find the asymptotic expansions of the solution to $g(a, c; x) = r$ around $r = 0$, $r = a/c$ and $r = 1$.

Theorem A.7. *Let $c > a > 0$, $r \in (0, 1)$; let $x(r)$ be the solution to $g(a, c; x) = r$. Then,*

$$x(r) = -\frac{a}{r} + (c - a - 1) + \frac{(c - a - 1)(1 + a)}{a}r + O(r^2), \quad r \rightarrow 0, \quad (\text{A.24})$$

$$x(r) = \left(r - \frac{a}{c}\right) \left\{ \frac{c^2(1+c)}{a(c-a)} + \frac{c^3(1+c)^2(2a-c)}{a^2(c-a)^2(c+2)} \left(r - \frac{a}{c}\right) + O\left(\left(r - \frac{a}{c}\right)^2\right) \right\}, \quad r \rightarrow \frac{a}{c} \quad (\text{A.25})$$

$$x(r) = \frac{c-a}{1-r} + 1 - a + \frac{(a-1)(a-c-1)}{c-a}(1-r) + O((1-r)^2), \quad r \rightarrow 1. \quad (\text{A.26})$$

Proof. The first step is the standard division of power series (see, for instance, (Gonzalez , 1991, Theorem 8.8) or (Hurwitz, Courant , 1925, Chapter 2.3)) which yields in a neighbourhood of $x = 0$:

$$g(x) = \frac{aM(a+1, c+1; x)}{cM(a, c; x)} = \frac{a}{c} + \frac{a(c-a)}{c^2(1+c)}x + \frac{a(a-c)(2a-c)}{c^3(1+c)(2+c)}x^2 + O(x^3). \quad (\text{A.27})$$

Next, the division of the asymptotic formulas (A.13) used with appropriate parameters gives in the neighbourhood of $x = \infty$:

$$g(x) = \frac{aM(a+1, c+1; x)}{cM(a, c; x)} = 1 + (a-c)\frac{1}{x} + (a-c)(1-a)\frac{1}{x^2} + (1-a)(c-a)(2a-c-2)\frac{1}{x^3} + O(x^{-4}). \quad (\text{A.28})$$

The inversion of a power series in a neighbourhood of a finite point x_0 is achieved via the following formula of Lagrange (Hurwitz, Courant , 1925, Chapter 7, Theorem 1*):

$$x(r) = x_0 + \sum_{n=1}^{\infty} \lim_{x \rightarrow x_0} \left[\frac{d^{n-1}}{dx^{n-1}} \left(\frac{x - x_0}{g(x) - r_0} \right)^n \right] \frac{(r - r_0)^n}{n!},$$

where $r_0 = g(x_0)$ and $g'(x_0) \neq 0$. Introducing the notation

$$g(x) - r_0 = g_1(x - x_0) + g_2(x - x_0)^2 + g_3(x - x_0)^3 + \dots$$

for the Taylor coefficients of g , we obtain (see (Gonzalez , 1991, (8.14.11))):

$$x(r) - x_0 = \frac{1}{g_1}(r - r_0) - \frac{g_2}{g_1^3}(r - r_0)^2 + \frac{2g_2^2 - g_1g_3}{g_1^5}(r - r_0)^3 + O((r - r_0)^4). \quad (\text{A.29})$$

In our case we have expansions in the neighbourhoods of $x_0 = 0$ and $x_0 = \infty$. For the case $x_0 = 0$ formula (A.29) immediately reveals:

$$x(r) = (r - a/c) \left\{ \frac{c^2(1+c)}{a(c-a)} + \frac{c^3(1+c)^2(2a-c)}{a^2(c-a)^2(c+2)}(r - a/c) + O((r - a/c)^2) \right\} \quad (\text{A.30})$$

which is precisely formula (A.25). For the point at infinity the Lagrange formula does not have the form given above. To compute the correct expression introduce the new variable $y = 1/x$ and rewrite the expansion (A.28) in the form:

$$r = \tilde{f}(y) = f(1/y) = q_0 + q_1y + q_2y^2 + \dots$$

Then according to (A.29):

$$y = \frac{1}{q_1}(r - q_0) - \frac{q_2}{q_1^3}(r - q_0)^2 + \frac{2q_2^2 - q_1q_3}{q_1^5}(r - q_0)^3 + O((r - q_0)^4).$$

Again applying standard division of power series (see, for instance (Gonzalez , 1991, Theorem 8.8) or (Hurwitz, Courant , 1925, Chapter 2.3)) we get:

$$x(r) = \frac{1}{y(r)} = \frac{q_1}{r - q_0} + \frac{q_2}{q_1} + \frac{1}{q_1} \left(\frac{q_3}{q_1} - \frac{q_2^2}{q_1^2} \right) (r - q_0) + O((r - q_0)^2)$$

Substituting here the values of q_i from (A.28), we obtain the expansion:

$$x(r) = \frac{1}{1-r} \left\{ (c-a) + (1-a)(1-r) + \frac{(a-1)(a-c-1)}{c-a} (1-r)^2 + O((1-r)^3) \right\}. \quad (\text{A.31})$$

This proves formula (A.26).

From formula (A.15) we immediately derive

$$g(a; c; x) = \frac{a}{-x} + \frac{a(1+a-c)}{(-x)^2} + \frac{a(1+a-c)(2+2a-c)}{(-x)^3} + O(1/(-x)^4)$$

Introducing $y = -1/x$ we rewrite this as

$$r = ay + a(1+a-c)y^2 + a(1+a-c)(2+2a-c)y^3 + O(y^4) = q_0 + q_1y + q_2y^2 + q_3y^3 + O(y^4)$$

Then according to (A.29):

$$y = \frac{1}{q_1}(r - q_0) - \frac{q_2}{q_1^3}(r - q_0)^2 + \frac{2q_2^2 - q_1q_3}{q_1^5}(r - q_0)^3 + O((r - q_0)^4).$$

Again applying standard division of power series (see, for instance (Gonzalez , 1991, Theorem 8.8) or (Hurwitz, Courant , 1925, Chapter 2.3)) we get:

$$x(r) = \frac{-1}{y(r)} = -\frac{q_1}{r - q_0} - \frac{q_2}{q_1} - \frac{1}{q_1} \left(\frac{q_3}{q_1} - \frac{q_2^2}{q_1^2} \right) (r - q_0) + O((r - q_0)^2)$$

or

$$\begin{aligned} x(r) &= \frac{-1}{y(r)} = -\frac{a}{r} - \frac{a(1+a-c)}{a} - \frac{1}{a} \left(\frac{a(1+a-c)(2+2a-c)}{a} - \frac{a^2(1+a-c)^2}{a^2} \right) r + O(r^2) \\ &= -\frac{a}{r} + (c-a-1) + \frac{(c-a-1)(1+a)}{a} r + O(r^2), \end{aligned}$$

which completes the demonstration of (A.24). □