Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Asymptotic analysis and approximations of parameterized univariate Gaussian splitting

Athena Xiourouppa

athena.xiourouppa@adelaide.edu.au

Adelaide University

John Maclean

Adelaide University

Dmitry Mikhin

Acacia

Research Article

Keywords: Finite mixture model, Gaussian approximation, Asymptotic analysis, Optimisation, Euclidean norm

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11461511/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Asymptotic analysis and approximations of parameterized univariate Gaussian splitting

Athena Xiourouppa[1,2*], John Maclean[2] and Dmitry Mikhin[1]

[1]Research and Development, Acacia, 5 Butler Drive, Hendon, 5014, SA, Australia.

[2*]School of Mathematical Sciences, Adelaide University, North Terrace, Adelaide, 5005, SA, Australia.

*Corresponding author(s). E-mail(s): athena.xiourouppa@adelaide.edu.au;
Contributing authors: john.maclean@adelaide.edu.au; dmitry.mikhin@acaciasystems.com.au;

Abstract

Splitting a Gaussian into a Gaussian mixture is a common tactic for managing the impact of a non-linear transformation on the Gaussian. To obtain a mixture model that approximates the Gaussian accurately we minimise the squared Euclidean norm between the two. For homoscedastic mixtures with fixed spacing between mixands, we obtain asymptotic approximations for this statistical distance under various limits of the mixture parameters. Based on these expressions, we derive an approximation for the mixand spacing that yields the minimal statistical distance. Extensive numerical investigations show our asymptotic results have good predictive power on the behaviour of the splitting algorithm. We further show the asymptotic results can be used to design and parameterize a splitting algorithm with close to optimal accuracy.

Keywords: Finite mixture model, Gaussian approximation, Asymptotic analysis, Optimisation, Euclidean norm

Acknowledgements. The authors are deeply grateful to our colleagues Sanjeev Arulampalam, Kyle Talbot, Pei Leong, Melissa Humphries, and Elena Kupriyanova for their valuable feedback on the earlier versions of this manuscript. We also appreciate the support from Acacia who are sponsoring the post-graduate studies of Athena and quietly shouldered this major distraction by Dmitry.

1 Introduction

Estimating the location and uncertainty of some object of interest by a Gaussian is a ubiquitous strategy in the fields of signal processing, uncertainty quantification, and machine learning, and used in applications as diverse as orbit determination, object tracking, numerical weather forecasting and epidemiology [1–10]. In these application domains, the estimate is sequentially and repeatedly transformed, either under the action of a forecast model, or as new data become available. As the transform is often non-linear, standard methods in this area transform the Gaussian when required via linearisation, the unscented transform (UT) [1], or ensemble methods [4].

A common issue, and the main inspiration for the methods in this paper, is that the Gaussian approximation may become insufficient to describe the quantity of interest under a sufficiently nonlinear transformation. Our focus is on a set of methods, complementary to the standard methods above, that 'split' the Gaussian into a sum of weighted Gaussians with smaller variances (or covariance matrices), such that linearisation or the UT perform better when applied to each narrow mixand. In non-linear estimation contexts, splitting is integral to Gaussian Sum Filters (GSF) [11–14]. We now summarise the decisions involved in implementing a splitting algorithm, as well as what is known or recommended in the literature.

Selecting the number of components in a mixture model appropriately is critical. GSFs often evaluate the identifiability of components to either further split mixands with large uncertainty or merge those with similar features to reduce redundancy [14–16]. In [17], it was found that fitting models with too many mixands deteriorates convergence rates. The impact of over-fitted mixtures is also seen in [18, 19], but on the posterior distribution's asymptotic behaviour, finding that surplus components attain asymptotically shrinking weights, negating their impact.

Another important parameter is the spacing between mixands. We generally require mixands to be sufficiently spaced to cover the probability regions of the data or distribution being approximated. Conditions on mixand separation for the convergence of the Expectation-Maximisation (EM) algorithm [20], often used to fit mixture models to data are proven in [21–23]. Alternatives to the EM algorithm are discussed in [24, 25]. Overlapping mixands, a result of very small mixand spacing, is problematic due to a loss of identifiability. In a GSF, one may quantify overlap to determine when merging components is appropriate [5, 26].

Finally, we arrive at the specific problem addressed in this paper: how to find a good splitting. To develop a splitting algorithm that will yield a mixture with optimal parameterization, one usually considers some distance between it and the original Gaussian. Distances include the squared Euclidean, or L^2, norm [27–30] or Kullback-Leibler (KL) divergence [31, 32]. The L^2 distance is usually preferred as it offers symmetry and closed-form analytical solutions for Gaussian mixture models [30, 33, 34].

Proceeding along the paper, Section 2 constructs a Gaussian mixture with the placement of M mixands at fixed spacing h (as done in [5, 35, 36]) and with identical variances σ^2 (as done in [3, 36–38]), and solves for the minimal L^2 distance to a reference Gaussian. We allow σ^2 to be set by the user, unlike the fixed values considered in [3, 5, 35]. Then in Section 3 we obtain asymptotic expressions for very high number

of mixands M, small and large mixand spacing h, and a heuristic for large M at moderate Mh. Numerics in Section 4 confirm that the asymptotic results are useful predictions of the optimal behaviour of this splitting algorithm. In particular, we show that one can construct an accurate splitting using only our asymptotic results; we compare with the algorithm of [30] carefully and show that our approximation achieves close to theoretically optimal accuracy across a broad range of parameters. We discuss implications and extensions in Section 5, including the generalisation of our results beyond a reference Gaussian measure.

2 The archetypal splitting problem

Consider a variable x, which is described by a standard Gaussian distribution $p(x) \sim \widetilde{\mathcal{N}}(0,1)$. Other univariate cases are covered by a linear change in coordinate, *i.e.*, offset and scaling. While practical scenarios generally involve multivariate systems, many multivariate splitting algorithms take parameters from univariate splitting libraries as input, [3, 39–41].

We wish to *split* the Gaussian into a best approximating sum of M Gaussians with means μ_m, variances σ_m^2, and weights w_m,

$$\widetilde{\mathcal{N}}(x;0,1) \approx \widetilde{\mathcal{Q}}(x) = \sum_{m=0}^{M-1} w_m \mathcal{N}(x;\mu_m,\sigma_m^2). \tag{1}$$

Here $\mathcal{N}(x;\mu,\sigma^2)$ is the probability density function (PDF) of a univariate normal distribution with the mean μ and variance σ^2.

The weights must be normalised as

$$\sum_{m=0}^{M-1} w_m = 1, \tag{2}$$

because the right-hand side of Eq. (1) must be a valid PDF. We require that all weights are positive, however, as long as the sum of terms in Eq. (1) is non-negative, negative weights may be permitted.

To obtain the best approximation in Eq. (1), we minimise the squared L^2 norm of the mismatch,

$$L^2 = \int_{-\infty}^{+\infty} \left[\widetilde{\mathcal{N}}(x;0,1) - \widetilde{\mathcal{Q}}(x)\right]^2 dx. \tag{3}$$

We simplify the problem by assuming homoscedasticity, such that $\sigma_m^2 = \sigma^2$, determined by the user, and setting equidistant means $\mu_m = mh + h_M$, where h is the step between the means, and h_M is the initial offset. From the desired symmetry of the approximation we get

$$h_M = \begin{cases} 0 & M \text{ is odd}, \\ h/2 & M \text{ is even}. \end{cases} \tag{4}$$

The conditions for the L^2 minimum through the partial derivatives are

$$\frac{\partial L^2}{\partial w_m} = 0, \text{ for } m = 0, \ldots, M-1, \text{ and } \frac{\partial L^2}{\partial h} = 0. \tag{5}$$

It follows from Eq. (3) that L^2 is quadratic in w_m, and therefore, the first M of equations (5) form a linear system for the weights w_m. Solving the system allows expressing the weights as functions of h. Then L^2 becomes a univariate function of h, and the last of Eq. (5) becomes a univariate non-linear equation for h [30].

For M total components, we define

$$\widetilde{M} = \begin{cases} (M-1)/2, & M \text{ is odd}, \\ M/2, & M \text{ is even}, \end{cases} \tag{6}$$

and renumber the weight indices so that they satisfy $m \in [-\widetilde{M}, \widetilde{M}]$ if M is odd, or $m \in [-\widetilde{M}, \widetilde{M}-1]$ if M is even. The symmetric mixands have the same weights, and one more weight is eliminated using normalisation via Eq. (2), leaving only the 'independent' weights with $m \in [1, \widetilde{M}]$ if M is odd and $m \in [1, \widetilde{M}-1]$ if M is even. Then differentiation over w_m produces a linear system of equations for the weights [30]

$$\boldsymbol{A}\boldsymbol{w} = \boldsymbol{b}, \quad \text{where } \boldsymbol{b} = [b_1, \ldots, b_{\widetilde{M}}]^{\mathsf{T}}, \ \mathbf{A} = \|a_{m,k}\|, \tag{7}$$

while Eq. (3) simplifies to

$$L^2 = \|F\|_2 - 2\boldsymbol{w}^{\mathsf{T}}\boldsymbol{b} + \boldsymbol{w}^{\mathsf{T}}\mathbf{A}\boldsymbol{w} = \|F\|_2 - \boldsymbol{w}^{\mathsf{T}}\boldsymbol{b}. \tag{8}$$

From [30], for odd M,

$$\begin{aligned} \|F\|_2 &= (2\sqrt{\pi})^{-1} - 2d_0 + c_{0,0}, \\ b_m &= 2(c_{0,0} - c_{0,m} - d_0 + d_m), \\ a_{m,k} &= 2(c_{k,m} + c_{-k,m} - 2c_{0,m} - 2c_{0,k} + 2c_{0,0}). \end{aligned} \tag{9}$$

For even M,

$$\begin{aligned} \|F\|_2 &= (2\sqrt{\pi})^{-1} - 2d_{1/2} + \big(c_{1/2,-1/2} + c_{1/2,1/2}\big)/2, \\ b_m &= 2\big(d_{m+1/2} - d_{1/2}\big) - c_{-1/2,m+1/2} - c_{1/2,m+1/2} + c_{-1/2,1/2} + c_{1/2,1/2}, \\ a_{m,k} &= 2\big(c_{m+1/2,k+1/2} + c_{m+1/2,-k-1/2}\big) - 2\big(c_{-1/2,k+1/2} + c_{-1/2,m+1/2}\big) \\ &\quad - 2\big(c_{1/2,k+1/2} + c_{1/2,m+1/2}\big) + 2\big(c_{-1/2,1/2} + c_{1/2,1/2}\big). \end{aligned} \tag{10}$$

In either case,

$$d_\alpha := \langle \widetilde{\mathcal{N}}, \mathcal{N}_\alpha \rangle = \frac{1}{\sqrt{2\pi(1+\sigma^2)}} \exp\left(-\frac{\alpha^2 h^2}{2(1+\sigma^2)}\right),$$
$$c_{\alpha,\beta} := \langle \mathcal{N}_\alpha, \mathcal{N}_\beta \rangle = \frac{1}{2\sigma\sqrt{\pi}} \exp\left(-\frac{(\alpha-\beta)^2 h^2}{4\sigma^2}\right), \tag{11}$$

valid for $\alpha, \beta \in \mathbb{R}$.

This solution does not cover the special case of $M = 2$. However, the weights for this corner case are immediately available from the symmetry of approximation: $w_{-1} = w_0 = 1/2$. Then the mismatch norm equals $\|F\|_2$ from Eq. (10).

Remark 1 (Generalisation of minimisation problem) For the problem of minimising the L^2 distance between some distribution $\Phi(x)$ and a mixture model $\tilde{\Phi}(x)$ with M weighted mixands and constant periodicity h,

- if $\Phi(x)$ is Gaussian and the family of $\tilde{\Phi}(x)$ is Gaussian, we obtain the well-formed theoretical framework from [30], restated above in Eqs. (7) to (11),
- if $\Phi(x)$ is *not* Gaussian and the family of $\tilde{\Phi}(x)$ is Gaussian, $\boldsymbol{A}$ is still the same as above, but $\boldsymbol{b}$ changes; it may be available in closed-form depending on the inner products $\langle \Phi, \mathcal{N}_k \rangle$, $k \in [0, M]$, though quadratures are available for integration if this is not the case, and,
- regardless of whether $\Phi(x)$ is Gaussian, if the family of $\tilde{\Phi}(x)$ is *not* Gaussian, both $\boldsymbol{A}$ and $\boldsymbol{b}$ are affected, but we can still obtain a solution by appropriately replacing the base norm $\|F\|_2$ and the inner products in Eq. (11).

Having established the problem of interest, we move to consider the behaviour of L^2 and $\boldsymbol{w}$ under special circumstances of M and h, and how that impacts our ability to solve for the optimal mixture.

3 Asymptotic solutions and analysis

We now derive asymptotic, closed-form results. Throughout this section, we make the practical assumption that the user-chosen σ^2 is not close to 1. In Section 3.1, we investigate the case of $M \to \infty$, the most notable as it well-approximates the optimal L^2. We then examine the case when M is large, but balanced by small h (Section 3.2); the solution for such intermediate Mh values allows us to approximate the h that yields the optimal L^2. Then, in Section 3.3, we examine the case of $h \to 0$ and in Section 3.4, of $h \to \infty$. These cases are less relevant to finding the optimal L^2, but still yield useful approximations. Finally, in Section 3.5, we summarise the interactions of these solutions and their potential.

3.1 Large M

The limit of $M \to \infty$ corresponds to a case where excessive mixands entirely cover the space of the original Gaussian. We find the following results for the weights and L^2.

Theorem 1 *As $M \to \infty$, the asymptotic weights take the Gaussian form:*

$$w_m \approx \begin{cases} \mathcal{N}_m(0, (1-\sigma^2)/h^2) & M \text{ is odd,} \\ \mathcal{N}_m(-1/2, (1-\sigma^2)/h^2) & M \text{ is even.} \end{cases} \tag{12}$$

The asymptotic L^2 solution is:

$$\lim_{M\to\infty} L^2 \approx \frac{1}{\sqrt{\pi}} \exp\left(-4\pi^2 \frac{\sigma^2(1-\sigma^2)}{h^2}\right)\left[1 + \exp\left(-4\pi^2 \frac{(1-\sigma^2)^2}{h^2}\right)\right]. \tag{13}$$

Proof First,[1] we observe that

$$\mathcal{N}_x(mh, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(m-x/h)^2}{2(\sigma/h)^2}\right] = \frac{\mathcal{N}_m(x/h, \sigma^2/h^2)}{h}. \tag{14}$$

We seek the asymptotic expression for the weights in the form

$$w_m \approx C\mathcal{N}_m(\gamma, \tau^2) \tag{15}$$

for some C, γ, and τ to be determined.

The product of the two Gaussians, now functions of m, is

$$w_m = \begin{cases} C\mathcal{N}_x\Big(\gamma h, \sigma^2 + \tau^2 h^2\Big)\mathcal{N}_m(m_c, \sigma_c^2) & M \text{ is odd,} \\ C\mathcal{N}_x\Big((\gamma + 1/2)h, \sigma^2 + \tau^2 h^2\Big)\mathcal{N}_m(m_c, \sigma_c^2) & M \text{ is even,} \end{cases} \tag{16}$$

where the mean and variance in m are

$$m_c = \sigma_c^2\left(\frac{\gamma}{\tau^2} + \frac{xh}{\sigma^2}\right) \quad \text{and} \quad \sigma_c^2 = \left(\frac{1}{\tau^2} + \frac{h^2}{\sigma^2}\right)^{-1}. \tag{17}$$

We substitute Eq. (16) into Eq. (1) and apply the Euler-Maclaurin formula [42, p. 16] (without the derivative terms) to replace the sum with an integral over the same variable m. Being a PDF, $\mathcal{N}_m(m_c, \sigma_c^2)$ integrates to one, such that we obtain the parameters

$$C = 1, \quad \tau = \sqrt{1-\sigma^2}/h, \quad \gamma = \begin{cases} 0 & M \text{ is odd,} \\ -1/2 & M \text{ is even,} \end{cases}$$

and also validate our choice of the Gaussian expression for weights in Eq. (15). The expressions (17) for the mean and variance simplify to

$$m_c = \frac{x(1-\sigma^2)}{h} + \gamma \quad \text{and} \quad \sigma_c^2 = \frac{\sigma^2(1-\sigma^2)}{h^2}. \tag{18}$$

Note that σ_c^2 is a constant, while m_c is a linear function of x.

To compute the asymptotic value of L^2 as $M \to \infty$ for odd M, we substitute Eq. (16) into Eq. (3) and take the limit:

$$\lim_{M\to\infty} L^2 = \int_{-\infty}^{+\infty} \left[\mathcal{N}_x(0,1) - C\mathcal{N}_x(0,1) \lim_{M\to\infty} \sum_{m=-\widetilde{M}}^{\widetilde{M}} \mathcal{N}_m(m_c, \sigma_c^2)\right]^2 dx$$

[1] Note: for this section, we use the shorter subscript form to describe Gaussian distributions, *i.e.*, $\mathcal{N}_x(\mu, \sigma^2) \equiv \mathcal{N}(x; \mu, \sigma^2)$.

$$= \int_{-\infty}^{+\infty} \mathcal{N}_x(0,1)^2 \left[1 - \sum_{m=-\infty}^{+\infty} \mathcal{N}_m(m_c, \sigma_c^2)\right]^2 dx. \tag{19}$$

For even M, the process is the same, however, the upper limit on the sum becomes $\widetilde{M} - 1$, which has no impact since $\widetilde{M} - 1 \to \infty$ as $M \to \infty$. The other difference is the value of m_c by Eq. (18).

To find the infinite sum, consider the following result from [43, p. 481]:

$$\sum_{k=-\infty}^{+\infty} \exp\left(-\frac{(k+t)^2}{n}\right) = \sqrt{\pi n}\left[1 + 2\sum_{m=1}^{+\infty} \exp\left(-\pi^2 n m^2\right)\cos\left(2\pi m t\right)\right]. \tag{20}$$

The left-hand side of Eq. (20) matches $\mathcal{N}_m(m_c, \sigma_c^2)$ for $k = m$, $t = -m_c$ and $n = 2\sigma_c^2$. The exponential term in each element decays rapidly as $m \to \infty$, while the cosine term is bounded on $[-1, 1]$. Thus, we approximate the series by its first element as:

$$\sum_{m=-\infty}^{+\infty} \mathcal{N}_m(m_c, \sigma_c^2) \approx 1 + 2\exp\left(-2\pi^2\sigma_c^2\right)\cos\left(2\pi m_c\right) + \mathcal{O}\left(\exp\left(-8\pi^2\sigma_c^2\right)\right).$$

Substituting this expression into Eq. (19) and omitting the Big-$\mathcal{O}$ terms yields:

$$\lim_{M\to\infty} L^2 = 4\exp\left(-4\pi^2\sigma_c^2\right)\int_{-\infty}^{+\infty} \mathcal{N}_x(0,1)^2\cos^2(2\pi m_c)\, dx. \tag{21}$$

Using the cosine double angle identity and applying [42, eq. 7.4.6], we take this integral and obtain the asymptotic value of L^2 by Eq. (13).

We conclude that up to the next-order terms, the asymptotic L^2 errors for $M \to \infty$ are equal for odd and even M, which is intuitively expected. □

For large, but finite M, the approximation obtained in Theorem 1 is not applicable for small h. Looking at the derivation details, M has to be large, but the product Mh shall not be too small, otherwise, the mixands would fail to cover the original distribution. This more restrictive assumption fails if $h \to 0$. Let us now consider this case.

3.2 Large M, small h heuristic

To address the case of intermediate Mh, consider the following heuristic approach. For odd M, the highest-order mixand has mean $\widetilde{M}h$. Having many mixands from $-\widetilde{M}h$ to $+\widetilde{M}h$, we can expect good approximation in that interval, so assume that its contribution to L^2 is zero. On the other hand, for $|x| > \widetilde{M}h$ we only have the tails of the mixands that, for $\sigma < 1$, decay faster than the original Gaussian.

One option is to assume that these tails provide no approximation at all, and the value of the Gaussian sum is zero. Under such assumption, we can express the overall L^2 mismatch through erf, the Gauss error function [42, Sect. 7.1]. Possibly more accurately, we can assume that the Gaussian sum beyond $\widetilde{M}h$ is represented by the single highest-order component with the weight $w_{\widetilde{M}}$. This way, for odd M, we obtain a new minimisation target of a single variable $w' = w_{\widetilde{M}}$:

$$L^2_{\text{heur}} = 2\int_{\widetilde{M}h}^{+\infty} \left[\widetilde{\mathcal{N}} - \frac{w'}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\widetilde{M}h)^2}{2\sigma^2}\right)\right]^2 dx\,. \tag{22}$$

For even M, the highest-order mixand has mean $(\widetilde{M}-1/2)h$. For large Mh, the factor of $-1/2$ is negligible. There are many mixands from zero to $(\widetilde{M}-1/2)h$, so assume that the contribution of this interval to L^2 is zero, while for $|x|>(\widetilde{M}-1/2)h$ assume that the Gaussian sum is represented by the single highest-order component with the weight $w'=w_{\widetilde{M}-1}$. This yields

$$L^2_{\text{heur}} = 2\int_{(\widetilde{M}-1/2)h}^{+\infty}\left[\widetilde{\mathcal{N}}-\frac{w'}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-(\widetilde{M}-1/2)h)^2}{2\sigma^2}\right)\right]^2 dx. \tag{23}$$

Theorem 2 *The solution to Eqs. (22) and (23) is*

$$w' = \frac{\sigma\sqrt{2}}{\sqrt{1+\sigma^2}}\exp\left(-\eta^2\right)\left[1-\operatorname{erf}(\sigma\eta)\right], \tag{24}$$

where $\eta=(\widetilde{M}+\gamma)h/\sqrt{2(1+\sigma^2)}$ *and*

$$\gamma=\begin{cases}0 & M \text{ is odd},\\ -1/2 & M \text{ is even},\end{cases}$$

which matches the γ *from the proof of Theorem 1.*

The L^2 *solution is*

$$L^2_{heur} = \frac{\left[1-\operatorname{erf}\left((\widetilde{M}+\gamma)h\right)\right]}{2\sqrt{\pi}} + \frac{w'^2}{2\sigma\sqrt{\pi}} - \frac{w'\sqrt{2}}{\sqrt{\pi(1+\sigma^2)}}\exp\left(-\eta^2\right)\left[1-\operatorname{erf}(\sigma\eta)\right], \tag{25}$$

Proof Expand Eq. (22), assuming odd M:

$$L^2_{\text{heur}} = \frac{1}{2\sqrt{\pi}}\left[1-\operatorname{erf}\left(\widetilde{M}h\right)\right] + \frac{w'^2}{2\sigma\sqrt{\pi}} - \frac{w'\sqrt{2}}{\sqrt{\pi(1+\sigma^2)}}\exp\left(-\eta^2\right)\left[1-\operatorname{erf}(\sigma\eta)\right], \tag{26}$$

where $\eta=\widetilde{M}h/\sqrt{2(1+\sigma^2)}$. Eq. (26) is a second-order polynomial in w', and solving for its minimum, we obtain Eq. (24).

For even M, we observe that the only difference between Eq. (22) and Eq. (23) is the substitution of $\widetilde{M}$ with $\widetilde{M}-1/2$, that is, the value of γ. Therefore, we can reuse the proof by applying this replacement in the final Eqs. (24) and (25). □

We may further simplify the error function via the approximation

$$1-\operatorname{erf}(x)=\operatorname{erfc}(x)\approx\frac{\exp\left(-x^2\right)}{x\sqrt{\pi}}, \tag{27}$$

Corollary 3 *For moderate* Mh*, the simplified heuristic solution to the weights is*

$$w'\approx\frac{\sqrt{2}\exp\left[-\eta^2(1+\sigma^2)\right]}{\eta\sqrt{\pi(1+\sigma^2)}}=\frac{2}{(\widetilde{M}+\gamma)h\sqrt{\pi}}\exp\left(-(\widetilde{M}+\gamma)^2h^2/2\right). \tag{28}$$

The simplified heuristic solution for L^2 *is*

$$L^2_{heur}\approx\frac{\exp\left(-(\widetilde{M}+\gamma)^2h^2\right)}{2(\widetilde{M}+\gamma)h\pi}\left(1-\frac{4}{\sigma(\widetilde{M}+\gamma)h\sqrt{\pi}}\right). \tag{29}$$

We see that as the product Mh increases, the leading term decreases rapidly due to the exponent.

The preceding two results, concerning large M and a heuristic at large M with moderate Mh, are the most practically relevant for the design of splitting methods. The following two results, at small and large h respectively, allow us to complete the asymptotic picture for the analysis of our algorithm.

3.3 Small h

The limit $h \to 0$ represents high overlap of mixands, such that the central density area is covered in excess, while outer tails are neglected. As the derivations for this scenario are elaborate, we only present results for odd M, and refer to [44] for complete details.

Theorem 4 *Let M be odd. The first two terms of the power series of the L^2 distance over $h_* = h/\sigma$*

$$L^2 = \sum_{j=0}^{\infty} L_{2j}^2 h_*^{2j} \,, \tag{30}$$

are

$$L_0^2 = \|F\|_2 - \sum_{n=1}^{\widetilde{M}} \beta_{2n} C_{2n}^{(-2n)}, \tag{31}$$

$$L_2^2 = -\sum_{n=1}^{\widetilde{M}} \beta_{2n} C_{2n}^{(-2(n-1))} - \beta_{2(\widetilde{M}+1)} C_{2\widetilde{M}}^{(-2\widetilde{M})} \langle \boldsymbol{p}_{2(\widetilde{M}+1)}, \tilde{\boldsymbol{p}}_{2\widetilde{M}} \rangle. \tag{32}$$

The coefficients $C_{2n}^{(-2(n-m))}$ for $n = 1, \ldots, \widetilde{M}$ and $m = 0, 1$ are obtained by solving

$$\boldsymbol{G}\boldsymbol{s}_{2m} = \boldsymbol{r}_{2m}, \tag{33}$$

where

$$\begin{aligned}
\boldsymbol{G} &= \left\| \alpha_{2i,2j}^{(2(i+j))} \right\| \quad \text{for } i,j = 1, \ldots, \widetilde{M}\,, \\
\boldsymbol{s}_{2m} &= \left[C_2^{(-2(1-m))}, C_4^{(-2(2-m))}, \ldots, C_{2\widetilde{M}}^{(-2(\widetilde{M}-m))} \right], \\
\boldsymbol{r}_0 &= \left[\beta_2, \beta_4, \ldots, \beta_{2\widetilde{M}} \right], \\
\boldsymbol{r}_2 &= \left(-C_{2\widetilde{M}}^{(-2\widetilde{M})} \left[\alpha_{2,2(\widetilde{M}+1)}^{(2(\widetilde{M}+2))}, \alpha_{4,2(\widetilde{M}+1)}^{(2(\widetilde{M}+3))}, \ldots, \alpha_{2\widetilde{M},2(\widetilde{M}+1)}^{(2(2\widetilde{M}+1))} \right] + \right. \\
&\quad \left. \left[0, \ldots, 0, \beta_{2(\widetilde{M}+1)} - \textstyle\sum_{k=1}^{\widetilde{M}} \alpha_{2(\widetilde{M}+1),2k}^{2(k+\widetilde{M}+1)} C_{2k}^{(-2k)} \right] \right) \langle \boldsymbol{p}_{2(\widetilde{M}+1)}, \tilde{\boldsymbol{p}}_{2\widetilde{M}} \rangle.
\end{aligned} \tag{34}$$

The coefficients $C_{2k}^{(-2k)}$ in $\boldsymbol{r}_2$ are already known from $\boldsymbol{s}_0$, the solution for $m = 0$. In Eq. (34)

$$\beta_{2k} = \frac{(-1)^k}{k!\, 2^{2k} \sigma \sqrt{\pi}} \left(\frac{2^{k+1/2} \sigma^{2k+1}}{(1+\sigma^2)^{k+1/2}} - 1 \right), \tag{35}$$

$$\boldsymbol{p}_{2k} = \left[1, 2^{2k}, \ldots, \widetilde{M}^{2k} \right], \forall k \geq 1, \tag{36}$$

and

$$\alpha_{2j,2(n-j)}^{(2n)} = \frac{(-1)^n}{n!\, 2^{2n-1} \sigma \sqrt{\pi}} \binom{2n}{2j}. \tag{37}$$

Proof We seek the even terms L_{2j}^2 in the power series of L^2 as the odd terms are zero by symmetry. To do so, we need the corresponding power series for $\boldsymbol{w}$ and $\boldsymbol{b}$, and to obtain the former, we need the power series of $\boldsymbol{A}$. Differentiating the elements b_m from Eq. (9) with respect to h_* and evaluating at $h_* = 0$, we obtain

$$\boldsymbol{b} = \sum_{k=1}^{\infty} \beta_{2k}^{(2k)} \boldsymbol{p}_{2k} h_*^{2k}, \tag{38}$$

where

$$\beta_{2k} = \frac{(-1)^k}{k!\, 2^{2k} \sigma \sqrt{\pi}} \left(\frac{2^{k+1/2} \sigma^{2k+1}}{(1+\sigma^2)^{k+1/2}} - 1 \right),$$

and the vectors are given by

$$\boldsymbol{p}_{2k} = \left[1, 2^{2k}, \ldots, \widetilde{M}^{2k}\right], \forall k \geq 1.$$

Similarly, differentiation of $a_{m,k}$ from Eq. (9) with respect to h_* at $h_* = 0$ yields the power series for the matrix $\boldsymbol{A}$:

$$\boldsymbol{A} = \sum_{n=2}^{\infty} \boldsymbol{A}_{2n} h_*^{2n}, \tag{39}$$

where the matrices $\boldsymbol{A}_{2n}$ are expressed as

$$\boldsymbol{A}_{2n} = \sum_{j=1}^{n-1} \alpha_{2j,2(n-j)}^{(2n)} \left(\boldsymbol{p}_{2j} \otimes \boldsymbol{p}_{2(n-j)} \right), \tag{40}$$

the $\otimes$ symbol denotes the outer product of vectors, and

$$\alpha_{2j,2(n-j)}^{(2n)} = \frac{(-1)^n}{n!\, 2^{2n-1} \sigma \sqrt{\pi}} \binom{2n}{2j}.$$

The second of the lower indices in $\alpha_{2j,2(n-j)}^{(2n)}$ is not strictly necessary, as it is defined by the first lower and the upper indices. However, it makes for convenient notation relating the indices of α with the indices of $\boldsymbol{p}$ in the outer product.

When $h \to 0$, by Eqs. (38) and (39) the matrix $\boldsymbol{A}$ decreases faster than the right-hand side $\boldsymbol{b}$, and therefore, the solution $\boldsymbol{w}$ must grow on the order of at least h^{-2}. The sum of these increasing weights equals one, hence, some of them must be negative and no all-positive solution may exist for very small h.

The weights $\boldsymbol{w}$ may grow faster than h_*^{-2} if they are orthogonal to $\boldsymbol{b}$ to make L^2 finite. In [44], we demonstrate that the lowest power of h_* is $-2\widetilde{M}$. The power series of the weights is:

$$\boldsymbol{w} = \sum_{m=-\widetilde{M}}^{\infty} \boldsymbol{w}_{2m} h_*^{2m} = \sum_{m=-\widetilde{M}}^{\infty} h_*^{2m} \sum_{j=1}^{\widetilde{M}} C_{2j}^{(2m)} \tilde{\boldsymbol{p}}_{2j}. \tag{41}$$

The complementary vectors $\tilde{\boldsymbol{p}}_{2m}$, $m = 1, \ldots, \widetilde{M}$ are designed to be orthogonal to all $\boldsymbol{p}_{2j}$ for $j < \widetilde{M}$ and $j \neq m$, and normalised such that $\langle \tilde{\boldsymbol{p}}_{2m}, \boldsymbol{p}_{2m} \rangle = 1$, where angular brackets denote the inner product. The $\tilde{\boldsymbol{p}}_{2m}$ vectors, like their $\boldsymbol{p}_{2m}$ counterparts, are independent, but not mutually orthogonal; they are no longer orthogonal to $\boldsymbol{p}_{2j}$ when $j > \widetilde{M}$.

Substituting the series (38), (39) and (41) into the system $\boldsymbol{A}\boldsymbol{w} = \boldsymbol{b}$ and equating the terms with the same powers of h_*, we obtain a linear system for $C_{\ldots}^{(\ldots)}$. Its solution is intricate because the lower-order matrices $\boldsymbol{A}_{2n}$ are not full-rank: the rank is defined by the order n and dimensionality $\widetilde{M}$. For $\boldsymbol{A}_4$, all rows are multiples of the first, and therefore, $\boldsymbol{A}_4$ has rank one. Similarly, $\boldsymbol{A}_6$ has rank two, $\boldsymbol{A}_8$ has rank three (assuming $\widetilde{M} \geq 3$), etc. For arbitrary n, the rank of $\boldsymbol{A}_{2n}$ is $\min(n-1, \widetilde{M})$.

To obtain all coefficients in L_{2t}^2, $t \geq 0$, we decompose Eq. (8) into a power series and equate the h_*^{2q} terms of $\boldsymbol{Aw} = \boldsymbol{b}$ for $q \geq -\widetilde{M} + 2$; the right-hand side is zero for $q < 1$. The equation of order h_*^{2q} contains the coefficients $C_{2l}^{(2(q-j))}$ for $j \in [2, q + M]$, $l \in [1, \widetilde{M}]$. To solve for the unknown $C_{\cdots}^{(\cdots)}$, we multiply the h_*^{2q} equation by some $\tilde{p}_{2n}$, where $n \in [1, \widetilde{M}]$. We consider the groups of these $2q/2n$ equations with $q = -\widetilde{M} + s + n$ for $s \geq 1$ and $n \in [1, \min(s, \widetilde{M})]$. Progressively taking such groups for $s = 1, 2, \ldots, \widetilde{M} - 1$, we find each time that the right-hand side is zero, and therefore, the coefficients $C_{2m}^{(2(-M+s-m))}$, $m \in [1, s]$ are zeros as well. Then for $s = \widetilde{M}$ and $s = \widetilde{M} + 1$ we obtain Eq. (34) for $m = 0, 1$, respectively, that define all the coefficients used in Eq. (31) for L_0^2 and Eq. (32) for L_2^2. □

The L_0^2 is positive, as this is the limit of L^2, a positive quantity, for $h \to 0$. It is natural to expect that L_2^2 is negative, as having some spread of mixands must provide a better approximation than having no spread at all. However, we currently do not have a formal proof of this conjecture.

Remark 2 For even M, the explicit form of the power series of L^2 is more unwieldy than in Eq. (34). The left-hand side is the same. However, in the right-hand side, we now have

$$\boldsymbol{b} = \sum_{k=1}^{+\infty} h_*^{2k} \sum_{j=1}^{k} \beta_{2k}^{(2j)} \boldsymbol{p}_{2j}, \tag{42}$$

where

$$\beta_{2k}^{(2j)} = \begin{cases} \frac{(-1)^{k+1}}{k!\, 2^{4k-2j}\sigma\sqrt{\pi}} \binom{2k}{2j} & j < k, \\ \beta_{2k} & j = k, \end{cases}$$

and

$$\boldsymbol{p}_{2k} = \left[(1 + 1/2)^{2k} - (1/2)^{2k}, \ldots, (\widetilde{M} + 1/2)^{2k} - (1/2)^{2k}\right], \forall k \geq 1. \tag{43}$$

The full set of vectors $\boldsymbol{p}_{2j}$, $j = 1, \ldots, k$ in Eq. (42) as compared to Eq. (38) results in additional nested sums in $\boldsymbol{r}_0$, $\boldsymbol{r}_2$ of the even-case equivalent of Eq. (34). Thus, we omit these equations here and refer to [44] for the details, as mentioned.

Remark 3 Eq. (33) holds for higher-order terms in the power series of L^2, where the left-hand side $\boldsymbol{G}\boldsymbol{s}_{2m}$ remains the same and only the right-hand side $\boldsymbol{r}_{2m}$ differs. Proof is given in [44].

Motivated by Remark 3, the matrix $\boldsymbol{G}$ is of interest as it remains constant with respect to the index of the power series, and therefore, we can invert it once and re-use the result to find higher-order terms. We will demonstrate invertibility of $\boldsymbol{G}$ to verify that we can find all elements of the power series L_{2j}^2.

3.3.1 Analysis of the matrix $\boldsymbol{G}$

Lemma 5 *The matrix $\boldsymbol{G}$ from Eq. (34) is symmetric positive definite (SPD).*

Proof Consider the elements of $\boldsymbol{G}$ from Eq. (37):

$$\alpha_{2i,2j}^{(2(i+j))} = \frac{(-1)^{i+j}}{(i+j)!2^{2i+2j-1}\sigma\sqrt{\pi}} \quad \frac{(2(i+j))!}{(2i)!(2j)!}$$

$$= \frac{2}{\sigma\sqrt{\pi}} \; \frac{(-1)^i}{2^{2i}(2i)!} \; \frac{(-1)^j}{2^{2j}(2j)!} \; \frac{(2(i+j))!}{(i+j)!}. \tag{44}$$

The second and third terms of Eq. (44) depend either on the row index i or the column index j, but not both. Therefore, the matrix $\boldsymbol{G}$ can be factorised as

$$\boldsymbol{G} = \frac{2}{\sigma\sqrt{\pi}} \, \boldsymbol{S}\boldsymbol{H}\boldsymbol{S}^{\mathsf{T}}. \tag{45}$$

Here $\boldsymbol{S}$ is a diagonal matrix with non-zero elements

$$s_i = \frac{(-1)^i}{2^{2i}(2i)!}, \tag{46}$$

while $\boldsymbol{H}$ is a Hankel matrix with constant skew-diagonals made of quadruple factorial numbers [45]

$$h_{i,j} = \frac{(2(i+j))!}{(i+j)!} = \frac{(2n)!}{n!}, \quad \text{for } n = 2, \ldots, 2\widetilde{M}. \tag{47}$$

By its symmetry and Corollary 7, $\boldsymbol{H}$ is SPD. Therefore, $\boldsymbol{G}$ is SPD. □

Proposition 6 *The matrix*

$$\boldsymbol{H}_{full} = \begin{bmatrix} 1 & \boldsymbol{v} \\ \boldsymbol{v}^{\mathsf{T}} & \boldsymbol{H} \end{bmatrix}, \tag{48}$$

where $\boldsymbol{H}$ is from Eq. (47) and $\boldsymbol{v}_j = (2j)!/j!$ for $j \in [1, \widetilde{M}]$, is symmetric positive definite.

Proof The matrix $\boldsymbol{H}_{\text{full}}$ has the same structure as $\boldsymbol{H}$, just starting the indices i, j at zero, and the quadruple factorial numbers at $n = 0$. For this sequence of numbers there is a closed-form expression [46] of its Hankel transform [47]:

$$T\{h_n\} = 8^{\frac{(n+1)n}{2}} \prod_{k=1}^{n} k! \prod_{k=1}^{n} (2k-1)!! \tag{49}$$

By definition of the Hankel transform, Eq. (49) gives the determinant of $\boldsymbol{H}_{\text{full}}$ for $n = 2\widetilde{M}$.

The Sylvester's criterion says that a matrix is positive definite if and only if the determinants of all leading principal minors of a matrix are positive [48, Theorem 7.2.5]. We can obtain the determinant of any sub-matrix of $\boldsymbol{H}_{\text{full}}$ by the transformation in Eq. (49). For the n^{th} sub-matrix $\boldsymbol{H}_{\text{full}n}$, $T\{h_{n-1}\}$ will give its determinant, for $n \in [1, \widetilde{M}+1]$. Clearly, for these values of n, $T\{h_{n-1}\} > 0$, such that $\boldsymbol{H}_{\text{full}}$ must be positive definite. □

Corollary 7 *From Proposition 6, since $\boldsymbol{H}$ is a sub-matrix of $\boldsymbol{H}_{full}$, specifically the $\widetilde{M}^{th}$ reverse leading principal minor, $\boldsymbol{H}$ is also positive definite.*

Corollary 8 *From Lemma 5, the mixand variance σ appears only as a constant multiplier in the first term of Eq. (44). Therefore, we can take it out and move to the right-hand side of Eq. (33). This way we would perform the matrix inversion once and apply the result for arbitrary σ.*

By Lemma 5, Eq. (33) and higher-order elements of the series Eq. (30) are guaranteed to be solvable since $\boldsymbol{G}$ is constant with respect to the element of the power series.

We now consider the final case, $h \to \infty$.

3.4 Large h

The limit $h \to \infty$ corresponds to highly dispersed mixands, which results in coverage 'gaps' for approximating the Gaussian.

Theorem 9 *As $h \to \infty$, the weights have the asymptotic solution:*

$$w_m = \begin{cases} \frac{1}{1+2\widetilde{M}}\left(1 + \frac{4\widetilde{M}\sigma}{\sqrt{2(1+\sigma^2)}}\right) & M \text{ is odd}, \quad m = 0, \\ \frac{1}{1+2\widetilde{M}}\left(1 - \frac{2\sigma}{\sqrt{2(1+\sigma^2)}}\right), & M \text{ is odd}, \quad m \neq 0, \\ \frac{1}{2\widetilde{M}} & M \text{ is even}, \quad \forall m. \end{cases} \tag{50}$$

The asymptotic L^2 solution is

$$L^2_\infty = \begin{cases} \frac{1}{2\sqrt{\pi}} - \frac{2}{\sqrt{2\pi(1+\sigma^2)}} + \frac{1}{2\sigma\sqrt{\pi}} - \frac{\widetilde{M}\sigma\sqrt{\pi}}{1+2\widetilde{M}} b_\infty^2 & M \text{ is odd}, \\ \frac{1}{2\sqrt{\pi}} + \frac{1}{4\widetilde{M}\sigma\sqrt{\pi}} & M \text{ is even}, \end{cases} \tag{51}$$

where

$$b_\infty = \frac{1}{\sigma\sqrt{\pi}}\left(1 - \frac{2\sigma}{\sqrt{2(1+\sigma^2)}}\right).$$

Proof For odd M, the $c_{\alpha,\beta}$ factors tend to zero unless $\alpha = \beta$, and the d_α factors tend to zero unless $\alpha = 0$. Therefore, Eqs. (9) simplify to:

$$\begin{aligned} b_m &\to 2(c_{0,0} - d_0) && = \frac{1}{\sigma\sqrt{\pi}}\left(1 - \frac{2\sigma}{\sqrt{2(1+\sigma^2)}}\right) =: b_\infty, \\ a_{m,k} &\to 2(c_{k,m}\delta_{k,m} + 2c_{0,0}) && = \frac{1}{\sigma\sqrt{\pi}}(2 + \delta_{k,m}), \end{aligned}$$

for all m and k; here $\delta_{k,m}$ is the Kronecker delta symbol. Thus, we have

$$\mathbf{A} \to \frac{1}{\sigma\sqrt{\pi}}(2\mathbf{J} + \mathbf{I}) \quad \text{and} \quad \boldsymbol{b} \to b_\infty \boldsymbol{j},$$

where $\mathbf{J}$ is a matrix of all ones, $\mathbf{I}$ is the identity matrix, and $\boldsymbol{j}$ is a vector of all ones, all with length $\widetilde{M}$. From the normalisation of the weights, we have $\mathbf{J}\boldsymbol{w} = (1 - w_0)\boldsymbol{j}/2$, and therefore,

$$\mathbf{A}\boldsymbol{w} \approx \frac{1}{\sigma\sqrt{\pi}}(2\mathbf{J} + \mathbf{I})\boldsymbol{w} = \frac{1}{\sigma\sqrt{\pi}}\left((1 - w_0)\boldsymbol{j} + \boldsymbol{w}\right) \approx b_\infty \boldsymbol{j}.$$

Solving this equation, the large-step approximation for the weights is

$$\boldsymbol{w} \approx \left(w_0 - \frac{2\sigma}{\sqrt{2(1+\sigma^2)}}\right)\boldsymbol{j}.$$

Substituting it back into the normalisation condition and solving for w_0, we find w_0 and w_m satisfy the first two cases of Eq. (50).

All asymptotic weights are non-negative, as $\sigma \leq 1$. Then the L^2 norm is found from Eq. (7), yielding the odd case of Eq. (51).

For even M, recall that the lower subscripts α, β in d_α, $c_{\alpha,\beta}$ are not just indices but *multipliers* of h^2 per Eq. (11). The multipliers were integers in the odd case, but not here.

Thus, in the limit of $h \to \infty$, we have $c_{\alpha,\beta} \to 0$ unless $\alpha = \beta$, $d_\alpha \to 0$ unless $\alpha = 0$, and the terms in Eq. (10) simplify to

$$\begin{aligned} b_m &\to c_{1/2,1/2} &&= \frac{1}{2\sigma\sqrt{\pi}}, \\ a_{m,k} &\to 2c_{m+1/2,k+1/2}\delta_{m,k} + 2c_{1/2,1/2} &&= \frac{1+\delta_{m,k}}{\sigma\sqrt{\pi}}, \end{aligned}$$

or

$$\mathbf{A} = \frac{1}{\sigma\sqrt{\pi}}(\mathbf{J}+\mathbf{I}) \quad \text{and} \quad \boldsymbol{b} = \frac{1}{2\sigma\sqrt{\pi}}\boldsymbol{j}.$$

Here, $\mathbf{J}$, $\mathbf{I}$, and $\boldsymbol{j}$ have length $\widetilde{M} - 1$. From the normalisation condition, we have $\mathbf{J}\boldsymbol{w} = (1/2 - w_0)\boldsymbol{j}$, and therefore, the equation for $\boldsymbol{w}$ reduces to $\mathbf{I}\boldsymbol{w} = w_0\boldsymbol{j}$. Thus, all weights are the same, as stated in the even case of Eq. (50). The L^2 norm is then the even case of Eq. (51). □

Intuitively, in the limit of $h \to \infty$, all the even-case mixands "move" to infinity and do not overlap with the original distribution. The squared L^2 norm of the mismatch is then the squared norm of the original distribution itself plus the squared norm of a mixand multiplied by its weight (squared) and taken $2\widetilde{M}$ times, yielding Eq. (51).

Corollary 10 *From Theorem 9, as $h \to \infty$, all weights are non-negative, since $\sigma \leq 1$. However, from Theorem 4, the weights are negative as $h \to 0$. Therefore, for each M there must be some minimal 'cutoff' value h_{cut}, such that a non-negative solution to the weights is possible only for $h \geq h_{cut}$.*

We will briefly investigate the cutoff values numerically in Section 4.

We now analyse the significance and interactions of the asymptotic solutions obtained.

3.5 Summary of asymptotic solutions

We considered the four special cases of Sections 3.1 to 3.4 in isolation under different assumptions. We do not know if their domains of applicability overlap. If so, we can yield approximations not only for L^2, but also for the h_{opt} that yields ${L^2}_{\text{opt}}$.

For example, Eq. (13) from Theorem 1 suggests that the mismatch L^2 tends to zero for $h \to 0$. However, in Eq. (31) we found the actual limit is finite. We also know that Eq. (13) is not applicable for very small h. Still, if we assume that the areas of validity for the small-h and large-M approximations overlap, we can combine these results to find an approximation for h_{opt}. At small h we represent L^2 as the second-order power series $L_0^2 + L_2^2 h^2$, which is expected to decrease with h. On the other hand, the L^2 approximation by Eq. (13) tends to zero when $h \to 0$, and then grows with h. Therefore, as h increases the two functions must intersect, and their intersection would yield an approximation to the exact optimal solution h_{opt}.

The solution is not possible in closed form, but can be approximated. Given that the large-M prediction of L^2 is exponentially small, we can replace it with zero. Then the hypothetical point of optimum becomes

$$h^\dagger_{\text{opt}} = -(L_0^2/L_2^2)^{1/2}. \tag{52}$$

Having $h^\dagger_{\text{opt}}$, we then can obtain the corresponding weights and L^2 either by substituting $h^\dagger_{\text{opt}}$ into the exact form of $L^2(h)$ or using the large-M asymptotic expressions.

Similarly, the heuristic approximation of L^2 by Eq. (25) from Theorem 2 starts at a finite value for $h = 0$ and then decreases. Again, we can solve for the intersection of this solution with the large-M expression (13) from Theorem 1, to find another approximation h^*_{opt} to h_{opt}, although there is no closed-form solution or even approximation. Then we find the corresponding weights and L^2 from either the original $L^2(h)$ equation or its large-M form.

Regarding $h \to \infty$, the large-h expression for L^2 in Eq. (51) from Theorem 9 has no dependence on h, so it is less suitable for estimating h_{opt} and ${L^2}_{\text{opt}}$.

We now examine these asymptotic results and approximations numerically to further analyse the interactions discussed.

4 Numerical analysis

We start by examining the behaviour of the L^2 mismatch on h and M in Figs. 1 to 3. Throughout this section, the mixand standard deviation is $\sigma = 0.2$ unless specified otherwise.

Fig. 1 reproduces [30, fig. 1] and demonstrates that for all considered values of M, the functions $L^2(h)$ are unimodal with unique minima h_{opt} (dots). On top of that figure, we superimpose the asymptotic result for large M from Eq. (13) (Theorem 1). As we can see, this solution provides a very accurate approximation for the exact L^2 for $h \gtrsim h_{\text{opt}}$. Compared to the large-h approximation, which only provides the limit, not the dependency on h, the large-M curve is likely to be more suitable and useful for approximating ${L^2}_{\text{opt}}$ for moderately large h. At the same time, the large-M approximation is "one-sided": it predicts ever decreasing values of L^2 as h gets smaller, while the exact solution has a minimum at the optimal h_{opt} and then increases again.

Fig. 2 also reproduces [30, fig. 1], now superimposing the two-term Maclaurin series for small h, shown as thin, coloured dashed lines and a crosshair for L^2_0 at $h = 0$. The L^2 approximation by this series loses accuracy and drops to zero well to the left from h_{opt}, indicating that the small-h and large-M approximations cannot be combined to estimate h_{opt}. Instead, such an estimate is apparently achieved using the heuristic solutions $L^2_{\text{heur}}(h)$ (dashed lines in Fig. 3) that intersect the large-M curve very close to the locations of the true minima. Despite the empirical nature of this approximation, it is evidently quite accurate for moderate Mh. The L^2 values at the intersection points are usually within one decimal order from the exact solutions. We solve numerically for the intersections of the large-M and heuristic Mh curves, since we cannot obtain them in closed form, as discussed previously, and plot them in Fig. 3. The intersection points of the curves h^*_{opt} become closer to h_{opt} for increasing M. We present two approximations to the optimal L^2: ${L^2}^*_{\text{opt}} = L^2(\boldsymbol{w}(h^*_{\text{opt}}))$ (circular outlines in Fig. 3) computed from the exact function $L^2(h)$ and weights defined in [30], implemented here without high-precision computing, and ${L^2}^{**}_{\text{opt}} = L^2(\boldsymbol{w}^*(h^*_{\text{opt}}))$ (diamonds in Fig. 3), which computes L^2 using the large-M approximation for the weights from Theorem 1. Both approximations are quite accurate for the presented range of M values. Solving

for the weights requires an extra computational effort, but yields significantly superior accuracy, with all calculations being done in standard precision.

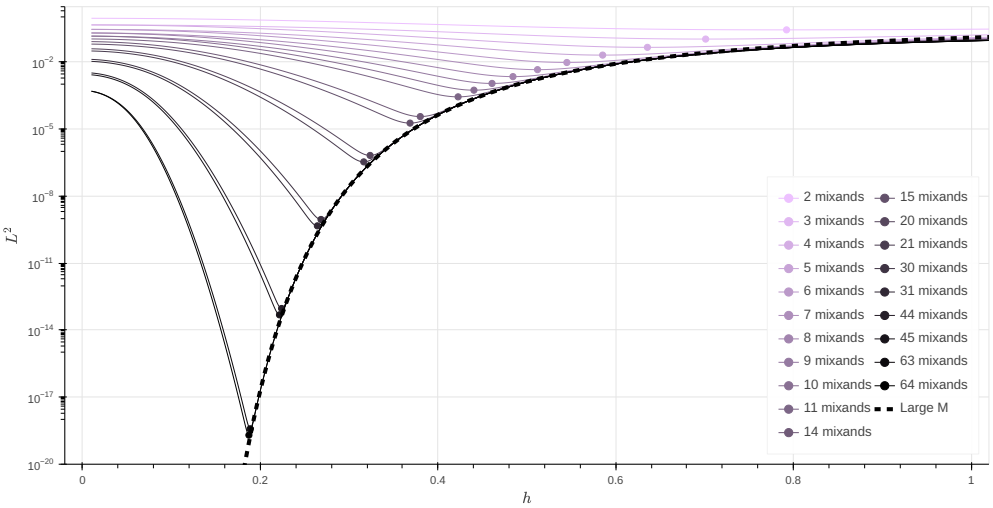


Fig. 1: The approximation mismatch L^2 as a function of the mean step h, part 1. Solid lines: exact solutions for various M. Dots: points of optima. Thick dashed line: large-M asymptotic solution from Theorem 1, which passes very close to the exact solutions to the right of h_{opt} starting with $M \geq 8$.

The overall layout of the mixand weights for various M is presented in Fig. 4a. The values on the x axis are the corresponding positions of the mixand means. As demonstrated in Section 3.1, for large M, the weights' dependence on the off-centre mixand index (and hence, the mean) approaches a Gaussian curve, which is confirmed by the presented plot. Fig. 4b displays the absolute difference between the precise weights and their large-M asymptotic approximations. The asymptotic solution is most accurate when close to the centre of the original distribution, and, as expected, becomes more precise as M increases.

In [30], we considered the error in normalised moment approximation at orders 2, 4, 6, 8, and 10, and KL divergence between the mixture and original Gaussian for various M to assess the splitting at h_{opt}. Replications of these figures with h^*_{opt} were remarkably similar, so we refer the reader to [30, figs. 5-7]. Instead, in Fig. 5, we compute the moment errors and KL divergence, as well as L^2 (seen in Fig. 3), at h_{opt} and h^*_{opt}, and plot the differences. In Fig. 5a, we solve for the weights exactly, using Eq. (7) in standard precision, and in Fig. 5b, we use the large-M approximation for the weights. Understandably, the errors are larger at h^*_{opt} than at h_{opt}, especially for higher-order moments, and generally decrease with M. Furthermore, Fig. 5b exhibits larger differences than Fig. 5a, although as M increases, this disparity becomes less noticeable.

Fig. 6 examines the dependence of the various special values of h on M. The optimal mean step h_{opt} (solid line) decreases with M, as expected. However, it is

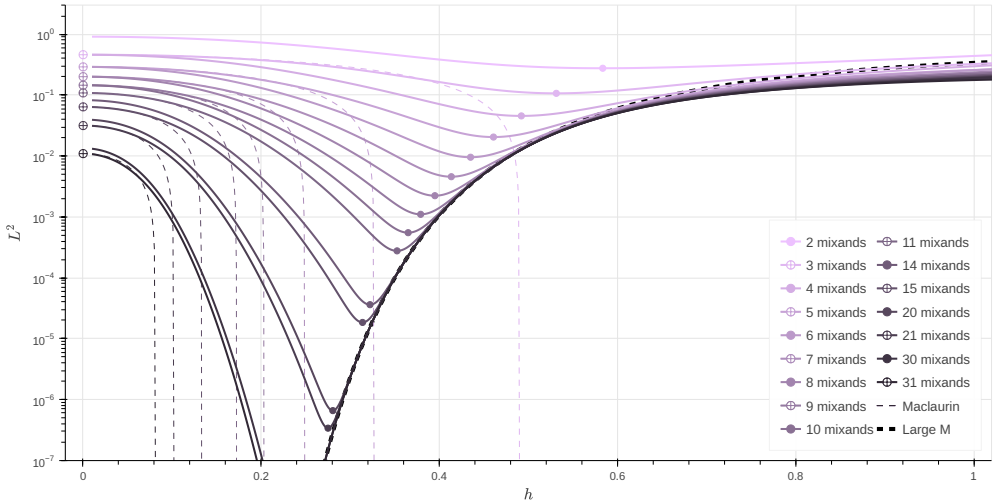


Fig. 2: The approximation mismatch L^2 as a function of the mean step h, part 2. Solid lines, dots, and thick dashed line: as in Fig. 1. Thin dashed lines: small-h solutions from Theorem 4, odd case only. Crosshairs: first element of the L^2 power series L_0^2. The small-h solutions maintain a greater range of accuracy across h for smaller M, though all eventually drop to zero well before reaching h_{opt}.

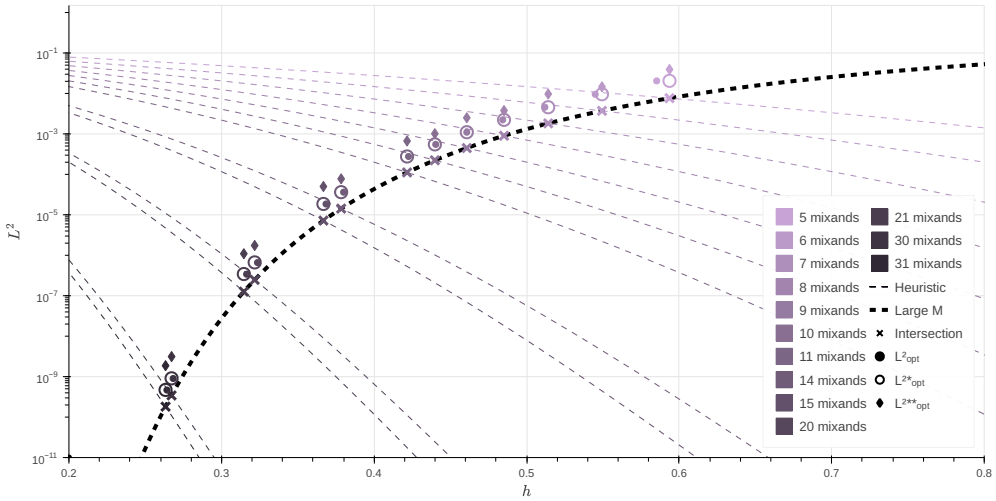
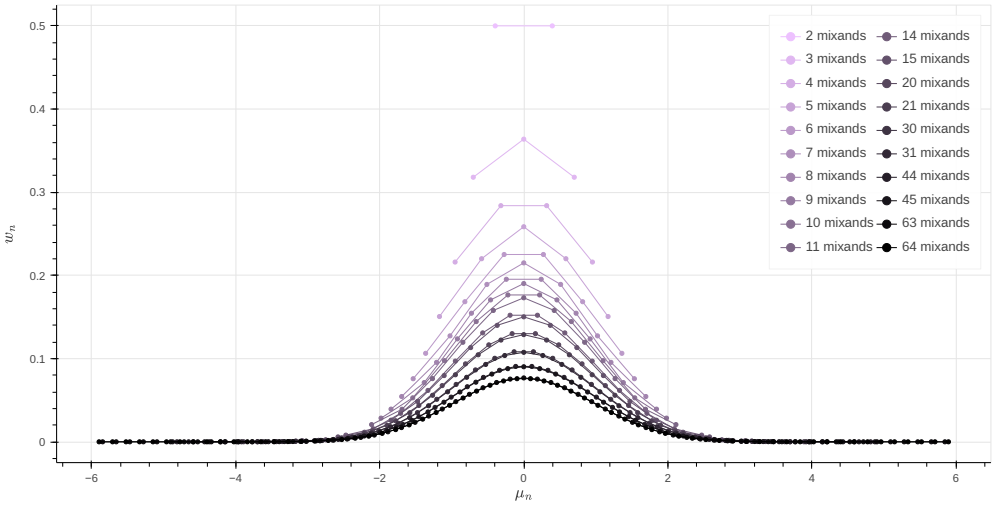
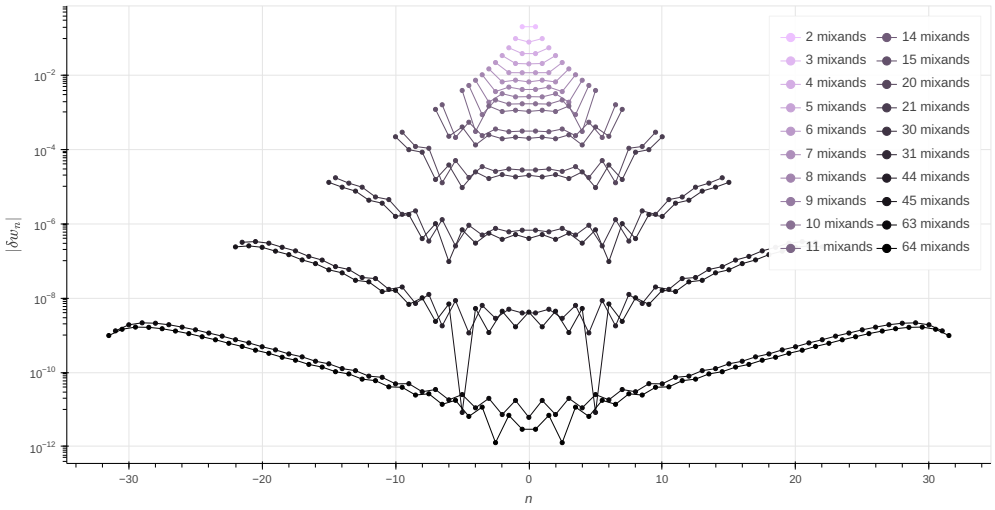


Fig. 3: The approximation mismatch L^2 as a function of the mean step h, part 3. Solid dots and thick dashed line: as in Fig. 1. Thin dashed lines: heuristic solutions L^2_{heur} from Theorem 2. Crosses: intersections of the curves at h^*_{opt}. Circular outlines: L^{2*}_{opt}. Diamonds: L^{2**}_{opt}. Both results are close to the true L^2_{opt} (dots), although the former is understandably better, as it makes fewer approximations.

unknown whether it converges to zero or to some finite limit, which would be a function of the mixand standard deviation σ. This exact solution is closely followed by the



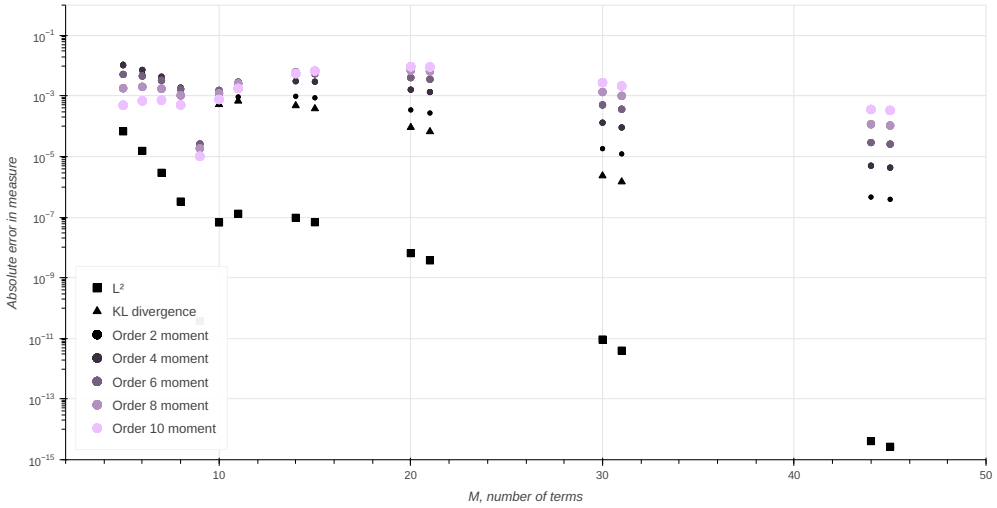
(a) Dependence of the mixand weights on the position of their means.



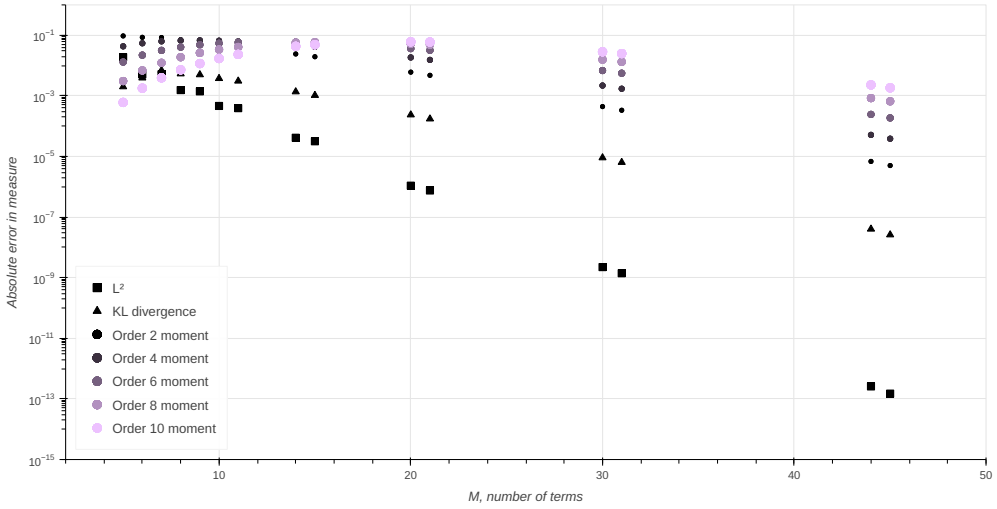
(b) The absolute error in the approximate weights from Eq. (15) as compared to the exact solutions.

Fig. 4: The mixand weights (top) and the errors of the large-M weight approximation (bottom). The weights are progressively more accurate with increasing M, supporting the conclusions from Figs. 1 and 3.

heuristic approximation h^*_{opt} (dashed line); however, being just a heuristic, it does not approach the true solution asymptotically. In fact, the difference between the two exhibits a small increase closer to the right side of the plot. Still, the agreement is good for the presented medium to large values of M. The small-h approximation $h^\dagger_{\text{opt}}$, in comparison, is nowhere near the exact solution (dash-dots). The dotted lines with



(a) The differences computed with weights $\boldsymbol{w}$ found from Eq. (7); $M = 9$ achieves an incidentally good error, likely an artefact related to the choice of σ.



(b) The differences computed with large-M weights $\boldsymbol{w}^*$.

Fig. 5: The absolute difference in L^2, KL divergence, and normalised moment errors when evaluated at h_{opt} versus h^*_{opt}, with weights $\boldsymbol{w}$ computed found from Eq. (7) (top) and weights $\boldsymbol{w}^*$ computed from the large-M approximation (bottom). The size and colour of the dots correlates with the order of the moments. Both approximations show unstable behaviour for $M \leq 10$, but then decrease consistently as M grows. The differences are smaller when the exact weights are used (top plot), which is expected.

triangle markers show the cutoff limits h_{cut}, defined in Corollary 10. Within the plot

limits, they are strictly below the optimal steps. However, it is not clear if these lines would intersect for very large M.

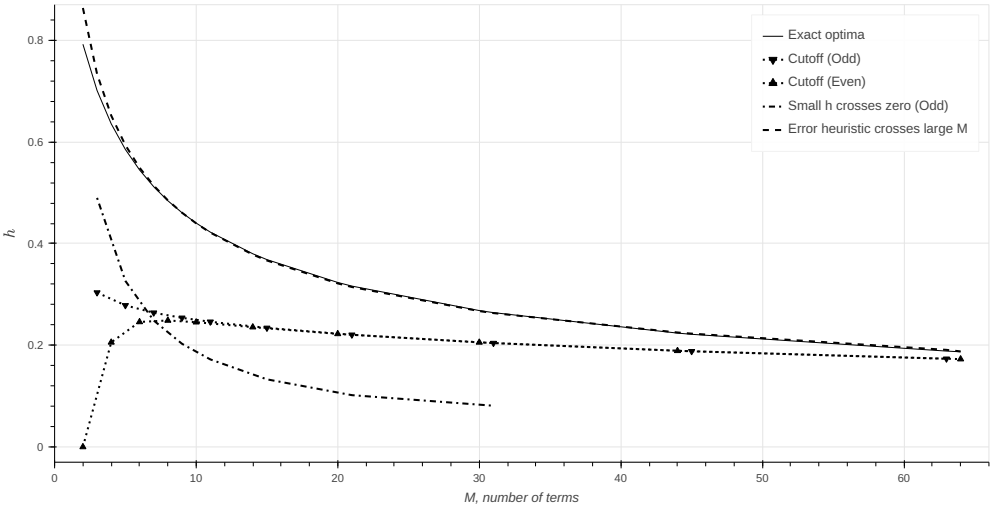


Fig. 6: Relationship between h and M for the exact optima h_{opt} (solid line), cutoff limits h_{cut} (dotted lines with different triangle symbols for the odd and even cases), the intersection $h^{\dagger}_{\text{opt}}$ between the small-h approximation and zero (dash-dotted line), and the intersection h^{*}_{opt} between the heuristic solution and the large-M asymptotic curve (dashed line). We only plot small-h up to $M = 31$ as the elements of the Hankel matrix in Eq. (47) grow as a factorial, necessitating the use of higher numerical precision for large M; we skip it given the poor performance of this method. h^{*}_{opt} approaches h_{opt} very closely, even at $M < 10$, though it subtly changes from under- to over-estimating the true optimum around $M = 40$. The cutoff and optima curves also become closer together for large M, indicating that for sufficiently large M, we may start to see optima with negative weights.

4.1 Summary of numerical analysis

The presented results demonstrate that an accurate splitting algorithm can be constructed with remarkably little effort. At any moderately large user-specified number of mixands M,

- set the mixand spacing $h = h^{*}_{\text{opt}}$, as described in Section 3.5, and
- use this h to find the weights by solving Eq. (7).

For step values near the optima, sufficiently accurate solutions can be found using regular double-precision floating-point for up to about $M = 50$; for more terms, it is necessary to use an extended-precision implementation. Alternatively, it is possible to use directly the large-M asymptotic weights from Theorem 1, essentially reducing our algorithm to the method of [49]. However, the accuracy of the L^2 approximation using such weights is at least an order of magnitude worse, as seen in Fig. 3.

Overall, this simplified algorithm avoids the need for high-precision arithmetic (except for very high M) and optimisers that were employed in the exact solution [30], while incurring only small additional errors, as examined in Fig. 5. The differences in L^2, KL divergence, and normalised moment errors are particularly small for $M > 10$, validating the quality of the splitting at the approximate optimal h with either method for computing weights. Regarding the user-specified mixand variance σ^2, we generally expect these mismatches to decrease with σ (see [30, fig. 4]). However, as mentioned in the preface of Section 3, σ shall not be close to 1, otherwise our asymptotic approximations would become less reliable and a separate analysis would be required.

5 Discussion

We derived asymptotic solutions and approximations to the optimal representation of a Gaussian by a Gaussian mixture model, particularly motivated by the case of applying a non-linear transformation to the Gaussian. Our analysis assumed symmetry, homoscedasticity, and fixed mixand spacing, as employed in [3, 5, 29, 30, 36, 37]. Our results are based on the framework from [30], which minimises the L^2 distance between the distributions to achieve an optimal mixture. The solution in [30], however, is purely numerical. This paper presents an in-depth analytic exploration of the problem, and comparison to the original solution.

In particular, for the cases of very large number of mixands M, very small or large mixand spacing h, and the moderately large combination Mh, we derived closed-form expressions for the weights and L^2. None of these expressions, in isolation, approximate the optimal L^2 or its corresponding h_{opt}. However, by solving for the intersection of the large-M solution from Theorem 1 and the heuristic method of Theorem 2, we obtained an accurate approximation to the mixand weights (Fig. 4), the exact minimum of L^2 (Fig. 3), as well as the moments and the KL divergence (Fig. 5). We established accuracy numerically for the broad range of M, but it cannot be guaranteed for an arbitrary number of terms. In particular, this approach cannot be used to examine the asymptotic behaviour of h_{opt} for very large M, which remains unknown. However, for moderate-to-large M, which are arguably most practically important, we introduced and analysed a reliable splitting algorithm.

Let us pause to discuss some explicit connections to the literature. Previously, interpolation with Gaussian kernels was discussed in [49], who yielded a quasi-interpolation formula to a prescribed order of error. The large-M asymptotic L^2 curve of Theorem 1 reproduces their quasi-interpolation formula with striking similarity. Our solution is not valid for small h, and their quasi-interpolation does not converge for small h. Convergence is demonstrated for some schemes [50] with variable σ decreasing with h, which for our case may indicate that very large values of M require correspondingly smaller σ for the solution to be valid.

The parameterization and minimisation of the L^2 between distributions is not new: what is new is the generality of the problem framing here and in the companion paper [30]. For instance, L^2 minimisation was also done in [29], up to $M = 39$ mixands, albeit only for three specific choices of mixand variance σ^2. The method of [40] uses the

beta divergence to measure the distance between distributions, but with specific selections of parameters used in [40] that measure reduces to the L^2 distance. Meanwhile, methods such as [12, 35, 36] do not necessarily minimise the L^2 but have compatible fixed mixand spacing and homoscedastic variance, although they set mixand variance based on the optimised weights and spacing as opposed to allowing the user to select it, as done here. Limitations arise when methods are restricted to small numbers of mixands, such as [35], which only uses $M = 2$ or $M = 3$.

Consider now some avenues for future work. The asymptotic solution of the weights and L^2 for small h in Theorem 4 was yielded as a power series decomposition. We provided the first two elements, and through Remark 3 and Corollary 8, we can solve for all higher-order elements, independent of σ. Furthermore, by Remark 1, only the right-hand side is impacted by a change in the kernel of the original distribution, meaning one could reuse the results for approximating *any* distribution by a Gaussian mixture with the parameterization of [30] or others similar. While we did not confirm convergence of the power series beyond second order, techniques such as Padé approximants [51] can increase the radius of convergence significantly. These techniques would also bolster the moderately limited use of the large-h asymptotic solutions in Theorem 9, which, unlike the small-h solutions, are independent of h. For instance, one could compute the small-h series to a higher order, take an $[L \,/\, L]$ Padé form, and fix the expression at both extremities: at zero (using small-h) and at infinity (large-h).

There are relationships between the L^2 and distances such as Hellinger and KL divergence, and their bounds, discussed for instance in [52, Sect. 2.4]. For algorithms that minimise one of these other distances but maintain the same parameterization, such as [7, 53], it may be possible to convert the expressions or at the very least define lower and upper bounds [54, Chapter 11].

As discussed in Remark 1, the framework established here is generalisable to minimising L^2 between non-Gaussian, well-behaved distributions and Gaussian mixtures since only the right-hand side of many equations changes and we can keep the results of the left-hand side. This was clear in Remarks 2 and 3 for small h, where we can obtain power series for other special functions that describe the optimal fit to that distribution. Based on the results of [49], which uses Gaussians to approximate arbitrary functions, the large-M solution can also plausibly generalise to non-Gaussians. Then, the heuristic approach of Theorem 2 can at least be replicated, although we believe some of the existing expressions may be directly usable.

The importance of other distributions notwithstanding, this paper is focused on Gaussian distributions. We examined in depth the earlier proposed technique [30] and found an approximate solution to the minimal L^2, which was demonstrated to preserve most of the favourable properties of the original exact solution. Combined with the mentioned extensions, we have a versatile framework for deriving asymptotic and approximate solutions to the Gaussian splitting problem.

Statements and Declarations

Funding. This work has been partially funded by Acacia and the Defence Trailblazer Project.

Competing interests. The authors have no competing interests to declare that are relevant to the content of this article.

Ethics approval and consent to participate. Not applicable.

Consent for publication. Not applicable.

Data availability. No data was created or analyzed in the research conducted for this article.

Materials availability. Not applicable.

Code availability. Code is not currently available due to tight coupling with proprietary code base; we're working to isolate and separate the code, to be able to open-source it.

Author contribution. D.M. conceptualized the research, performed initial formal investigation, and supervized the project. D.M. and A.X. both performed analysis, validation, and visualization. D.M., A.X., and J.M. wrote and reviewed the manuscript.

References

[1] Julier, S., Uhlmann, J., Durrant-Whyte, H.F.: A new method for the nonlinear transformation of means and covariances in filters and estimators. IEEE Transactions on Automatic Control **45**(3), 477–482 (2000) https://doi.org/10.1109/9.847726

[2] Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press, Cambridge (2002). https://doi.org/10.1017/cbo9780511802270

[3] Hanebeck, U.D., Briechle, K., Rauh, A.: Progressive Bayes: a new framework for nonlinear state estimation. In: Dasarathy, B.V. (ed.) Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2003, pp. 256–267. SPIE, Orlando, FL (2003). International Society for Optics and Photonics

[4] Evensen, G.: Data Assimilation: The Ensemble Kalman Filter. Springer, Berlin (2009). Includes bibliographical references and index.

[5] Leong, P.H., Arulampalam, S., Lamahewa, T.A., Abhayapala, T.D.: A Gaussian-sum based Cubature Kalman filter for bearings-only tracking. IEEE Transactions on Aerospace and Electronic Systems **49**(2), 1161–1176 (2013)

[6] Shorten, D.P., Humphries, M., Maclean, J., Yang, Y., Roughan, M.: Optimal proposal particle filters for detecting anomalies and manoeuvres from two line element data. Acta Astronautica **228**, 709–723 (2025) https://doi.org/10.1016/j.actaastro.2024.11.033

[7] Zanetti, R., DeMars, K.J., Tuggle, D., Michaelson, K., Gupta, M.: Uncertainty quantification using directional splitting and Gaussian Mixture models with applications to orbital dynamics. In: Proceedings of the AAS/AIAA Astrodynamics Specialist Conference, AAS 25, Boston, MA, pp. 10–14 (2025)

[8] Zhang, B., Shin, Y.C.: An adaptive Gaussian mixture method for nonlinear uncertainty propagation in neural networks. Neurocomputing **458**, 170–183 (2021)

[9] Xu, T., Zhang, Z., Han, H.: Adaptive Gaussian Mixture Model for Uncertainty Propagation Using Virtual Sample Generation. Applied Sciences **13**(5), 3069 (2023)

[10] Straka, O., Hanebeck, U.D.: Efficient Gaussian Mixture Filters based on Transition Density Approximation. arXiv (2025). https://doi.org/10.48550/arXiv.2505.20002

[11] Terejanu, G., Singla, P., Singh, T., Scott, P.D.: Uncertainty propagation for nonlinear dynamic systems using Gaussian mixture models. Journal of Guidance, Control, and Dynamics **31**(6), 1623–1633 (2008) https://doi.org/10.2514/1.36247

[12] Huber, M.F.: Adaptive Gaussian Mixture filter based on statistical linearization. In: Proceedings of the 14th International Conference on Information Fusion (FUSION), pp. 1–8. IEEE, Chicago, IL (2011)

[13] Straka, O., Duník, J., Šimandl, M.: Gaussian sum unscented Kalman filter with adaptive scaling parameters. In: 14th International Conference on Information Fusion, pp. 1–8. IEEE, Chicago (2011)

[14] Vishwajeet, K., Singla, P.: Adaptive split/merge-based Gaussian mixture model approach for uncertainty propagation. Journal of Guidance, Control, and Dynamics **41**(3), 603–617 (2018) https://doi.org/10.2514/1.G002801

[15] Faubel, F., McDonough, J., Klakow, D.: The split and merge Unscented Gaussian mixture filter. IEEE Signal Processing Letters **16**(9), 786–789 (2009) https://doi.org/10.1109/lsp.2009.2024859

[16] Terejanu, G., Singla, P., Singh, T., Scott, P.D.: Adaptive Gaussian Sum filter for nonlinear Bayesian estimation. IEEE Transactions on Automatic Control **56**(9), 2151–2156 (2011) https://doi.org/10.1109/tac.2011.2141550

[17] Ho, N., Nguyen, X.: Identifiability and optimal rates of convergence for parameters of multiple types in finite mixtures. arXiv (2015). https://doi.org/10.48550/arXiv.1501.02497v1

[18] Rousseau, J., Mengersen, K.: Asymptotic Behaviour of the Posterior Distribution in Overfitted Mixture Models. Journal of the Royal Statistical Society Series B:

Statistical Methodology **73**(5), 689–710 (2011)

[19] Havre, Z., White, N., Rousseau, J., Mengersen, K.: Overfitting Bayesian Mixture Models with an Unknown Number of Components. PLOS One **10**(7) (2015)

[20] Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Series B: Statistical Methodology **39**(1), 1–22 (1977) https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

[21] Ma, J., Xu, L., Jordan, M.I.: Asymptotic Convergence Rate of the EM Algorithm for Gaussian Mixtures. Neural Computation **12**(12), 2881–2907 (2000)

[22] Zhao, R., Li, Y., Sun, Y.: Statistical Convergence of the EM Algorithm on Gaussian Mixture Models. arXiv (2018). https://doi.org/10.48550/arXiv.1810.04090

[23] Kwon, J., Caramanis, C.: The EM Algorithm gives Sample-Optimality for Learning Mixtures of Well-Separated Gaussians. In: Proceedings of Machine Learning Research, vol. 125, pp. 1–63 (2020)

[24] Springer, T., Urban, K.: Comparison of the EM algorithm and alternatives. Numerical Algorithms **67**, 335–364 (2014)

[25] Hasannasab, M., Hertrich, J., Laus, F., Steidl, G.: Alternatives to the EM algorithm for ML estimation of location, scatter matrix, and degree of freedom of the Student t distribution. Numerical Algorithms **87**, 77–118 (2021)

[26] Kyoya, S., Yamanishi, K.: Summarizing Finite Mixture Model with Overlapping Quantification. Entropy **23**(11) (2021)

[27] Hanselmann, A., Schrempf, O.C., Hanebeck, U.D.: Optimal parametric density estimation by minimizing an analytic distance measure. In: 2007 10th International Conference on Information Fusion, pp. 1–8. IEEE, Quebec (2007). https://doi.org/10.1109/icif.2007.4408100

[28] Musso, C., Oudjane, N.: L2-density Estimation Under Constraints. In: 2009 17th European Signal Processing Conference, pp. 2499–2503 (2009)

[29] Vittaldev, V., Russell, R.P.: Multidirectional Gaussian Mixture Models for Nonlinear Uncertainty Propagation. Computer Modeling in Engineering & Sciences **111**(1), 83–117 (2016)

[30] Mikhin, D., Xiourouppa, A.: A flexible and robust approach to univariate gaussian splitting using parameterized gaussian mixtures (2026) https://doi.org/10.21203/rs.3.rs-10658614/v1

[31] Zhang, Y., Pan, J., Li, L.K., Liu, W., Chen, Z., Liu, X., Wang, J.: On

the properties of Kullback-Leibler divergence between multivariate Gaussian distributions. In: Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S. (eds.) Advances in Neural Information Processing Systems 36 (NeurIPS 2023), vol. 36, pp. 58152–58165. Curran Associates, Inc., New Orleans (2023). https://proceedings.neurips.cc/paper_files/paper/2023/file/b5b4d92374323c53c24bbbc8ee0e715c-Paper-Conference.pdf

[32] Hoek, J., Elliott, R.J.: Mixtures of multivariate Gaussians. Stochastic Analysis and Applications **42**(4), 737–752 (2024) https://doi.org/10.1080/07362994.2024.2372605

[33] Jian, B., Vemuri, B.C.: A Robust Algorithm for Point Set Registration Using Mixture of Gaussians. In: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, vol. 2, pp. 1246–1251 (2005)

[34] Durrieu, J.-L., Thiran, J.-P., Kelly, F.: Lower and upper bounds for approximation of the Kullback-Leibler divergence between Gaussian mixture models. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4833–4836. IEEE, Kyoto (2012). https://doi.org/10.1109/icassp.2012.6289001

[35] Faubel, F., Klakow, D.: Further improvement of the adaptive level of detail transform: Splitting in direction of the nonlinearity. In: Proceedings of the 2010 European Signal Processing Conference (EUSIPCO 2010), pp. 850–854. IEEE, Aalborg, Denmark (2010)

[36] Leutnant, V., Krueger, A., Haeb-Umbach, R.: A versatile Gaussian splitting approach to non-linear state estimation and its application to noise-robust ASR. In: Interspeech 2011, pp. 1641–1644 (2011). https://doi.org/10.21437/Interspeech.2011-208

[37] Zanetti, R., Tuggle, K.: A novel Gaussian Mixture approximation for nonlinear estimation. In: 2018 21st International Conference on Information Fusion (FUSION), pp. 1100–1106. IEEE, Cambridge, UK (2018). https://doi.org/10.23919/icif.2018.8455485

[38] Agostini, D., Améndola, C., Ranestad, K.: Moment Identifiability of Homoscedastic Gaussian Mixtures. Foundations of Computational Mathematics **21**(3), 695–724 (2021)

[39] Huber, M.F., Bailey, T., Durrant-Whyte, H., Hanebeck, U.D.: On entropy approximation for Gaussian mixture random vectors. In: 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, pp. 181–188. IEEE, Seoul, Korea (South) (2008). https://doi.org/10.1109/mfi.2008.4648062

[40] DeMars, K.J., Bishop, R.H., Jah, M.K.: Entropy-based approach for uncertainty

propagation of nonlinear dynamical systems. Journal of Guidance, Control, and Dynamics **36**(4), 1047–1057 (2013) https://doi.org/10.2514/1.58987

[41] Xiourouppa, A.H., Mikhin, D., Humphries, M., Maclean, J.: Theoretical insights for bearings-only tracking in log-polar coordinates. IEEE Transactions on Aerospace and Electronic Systems **61**(4), 8534–8545 (2025)

[42] Abramowitz, M., Stegun, I.A. (eds.): Handbook of Mathematical Functions, Tenth edn. Dover books on mathematics. Dover Publications, New York, NY (1965)

[43] Graham, R.L., Knuth, D.E., Patashnik, O.: Concrete Mathematics, 2nd edn. Addison-Wesley, Upper Saddle River, NJ (2017)

[44] Mikhin, D., Xiourouppa, A.: Asymptotic analysis of parameterised univariate Gaussian splitting. arXiv (2026). https://doi.org/10.48550/arXiv.2606.04440

[45] OEIS Foundation Inc.: The On-Line Encyclopedia of Integer Sequences. Sequence A000045, published electronically at https://oeis.org (2026)

[46] OEIS Foundation Inc.: The On-Line Encyclopedia of Integer Sequences. Sequence A137565, published electronically at https://oeis.org (2026)

[47] Layman, J.W.: The Hankel transform and some of its properties. Journal of Integer Sequences **4**, 15 (2001)

[48] Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge (1985)

[49] Maz′ya, V., Schmidt, G.: On approximate approximations using Gaussian kernels. IMA Journal of Numerical Analysis **16**(1), 13–29 (1996) https://doi.org/10.1093/imanum/16.1.13

[50] Beatson, R.K., Light, W.A.: Quasi-interpolation in the absence of polynomial reproduction. In: Braess, D., Schumaker, L.L. (eds.) Numerical Methods in Approximation Theory vol. 9, pp. 21–39. Birkhäuser Basel, Basel, Switzerland (1992). https://doi.org/10.1007/978-3-0348-8619-2

[51] Baker, G.A., Graves-Morris, P.: Padé Approximants, 2nd edn. Encyclopedia of mathematics and its applications, vol. 59. Cambridge University Press, Cambridge (1996)

[52] Tsybakov, A.B.: Introduction to Nonparametric Estimation. Springer Series in Statistics. Springer, New York (2009)

[53] Tuggle, K.E.: Model selection for Gaussian mixture model filtering and sensor scheduling. Phd dissertation, The University of Texas at Austin, Austin, TX (August 2020). Available at https://repositories.lib.utexas.edu/

items/e73a80f9-6ff2-408d-958d-037c72226a50

[54] Klemelä, J.: Smoothing of Multivariate Data: Density Estimation and Visualization. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ (2009)