

Qualifying Paper 1

Variational Inference in Location-Scale Families: Exact Recovery of the Mean and Correlation Matrix

Professor: Charles C. Margossian

Isaac Rankin

October 24, 2025

Contents

1 Context	3
2 Contributions	3
2.1 Notation, definitions, and regularity conditions	3
2.2 Main contributions	4
3 Strengths and limitations	4
4 Discuss open questions and propose future research directions	5
5 Small project	6
5.1 Rotation symmetry and exact recovery of the mean	6
5.2 Discussion of properties	6
5.3 Proof of conjecture	7
5.3.1 The point of symmetry is indeed the mean	7
5.3.2 An optima is achieved when points of symmetry coincide	8
5.3.3 Concavity conditions ensure a unique minimiser	10
5.4 Simulations	11
Bibliography	15

1 Context

Variational inference (VI) concerns itself with approximating an unknown and typically intractable target distribution [1]. An optimal approximation is found by searching within a family of simpler distributions, and selecting the distribution that minimizes a notion of dissimilarity between the approximate distribution and the target distribution.

VI provides an alternative to Markov Chain Monte Carlo (MCMC). MCMC produces a sequence of samples which, after a sufficient number of samples, represent samples from the target distribution arbitrarily well [2]. VI on the other hand, is often much faster than MCMC [1], partly due to having restricted the search space of approximations to within the chosen variational family of distributions. VI's optimal approximation of the target distribution will be misspecified unless the target is within the family of approximations. Though not all features of the approximation match those of the target, the utility of the approximation lies in sharing properties of interest with the target. An open problem is: under which conditions does the optimal approximation found through VI share properties with the target distribution exactly?

The paper *Variational Inference in Location-Scale Families: Exact Recovery of the Mean and Correlation Matrix* examines the location-scale family of distributions – which is closed under linear transformation – and defines the optimal approximation from this family to the target distribution as that which minimizes the Kullback-Leibler (KL) divergence from the approximation to the target [3]. Thus, approximating the target distribution can be reframed as an optimization problem over the location and scale parameters. Margossian and Saul identify conditions for which VI is guaranteed to correctly identify the mean and correlation matrix of the target distribution. This is done through matching points of symmetry when the target distribution and the location-scale family share a certain symmetry.

2 Contributions

As the optimal approximate distribution found through VI is most often misspecified, there is a need to develop an understanding of which attributes can VI approximate, when this can be done exactly, and the severity of any errors. There are two directions this problem can be tackled:

- 1) VI finds an approximation for a feature of the target distribution being estimated, how can we bound the error of the approximation?
- 2) To achieve an error bound when estimating a feature of the target, which conditions on the target and the variational family must be met?

Margossian and Saul contributed to the second when the error bound is set to zero, finding sufficient conditions under which VI can exactly recover the mean and the correlation matrix of the target density. Given arbitrary target density and variational family, Margossian and Saul provide no guarantees on the error of approximation. This is a limitation of the paper, and an area of future research.

Their paper has two main results: one for sufficient conditions to recover the mean, and the other describing conditions sufficient for recovering the correlation matrix. Both results follow the same reasoning; when the target density and all members of the variational family share a symmetry, then the KL divergence will, under additional conditions, have a unique global minimiser when the point of symmetry of the approximation coincides with that of the target.

2.1 Notation, definitions, and regularity conditions

Before formally discussing the two main theorems in the paper, we first need to introduce notation, define terms, and impose conditions on the target and the variational family.

Denote the target density over \mathbb{R}^d as $p(z)$. If q_0 is a density over \mathbb{R}^d , then a location-scale family \mathcal{Q} can be

constructed as the set of all densities $q_{\nu,S}(z) = q_0(S^{-\frac{1}{2}}(z - \nu))|S|^{-\frac{1}{2}}$ where $z, \nu \in \mathbb{R}^d$, and $S \in \mathbb{R}^{d \times d}$ is a positive definite matrix.

- $e : \mathbb{R}^d \rightarrow \mathbb{R}$ is called even-symmetric about a point $\nu \in \mathbb{R}^d$ if $e(\nu - \xi) = e(\nu + \xi)$ for every $\xi \in \mathbb{R}^d$.
- $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is called odd-symmetric about a point $\nu \in \mathbb{R}^d$ if $-f(\nu - \xi) = f(\nu + \xi)$ for every $\xi \in \mathbb{R}^d$.
- $g : \mathbb{R}^d \rightarrow \mathbb{R}$ is called spherically-symmetric if $\|\xi_1\| = \|\xi_2\|$ implies $g(\xi_1) = g(\xi_2)$ where $\xi_1, \xi_2 \in \mathbb{R}^d$.
- $h(z) : \mathbb{R}^d \rightarrow \mathbb{R}$ is called elliptically-symmetric about $\nu \in \mathbb{R}^d$ if there exists a positive-definite scale matrix $M \in \mathbb{R}^{d \times d}$ such that $h(M^{-\frac{1}{2}}(z - \nu))$ is spherically-symmetric.

The Kullback Leibler divergence is defined as $\text{KL}(q \parallel p) = \int \log \frac{q(z)}{p(z)} q(z) dz$.

A location-scale family is called a location family if the scale parameter is fixed.

We seek to find the parameters ν, S to minimise $\text{KL}(q_{\nu,S} \parallel p)$. The objective function is an integral. Assume that the target density is differentiable allowing us to take its gradient. Also assume that the log-target density and every distribution in the variational family satisfy the conditions of the dominated convergence theorem. This allows taking the gradient of the objective function, an integral, and take the gradient inside the integral. To guarantee a unique global optima, suppose that the target density is log-concave on \mathbb{R}^d , and strictly log-concave on some open subset. This ensures that the target density is unimodal and that any stationary point of $\text{KL}(q_{\nu,S} \parallel p)$ is a unique minimiser.

2.2 Main contributions

Main result 1 (Exact recovery of the mean): If \mathcal{Q} is a location family, q_0 is even-symmetric about the origin, the regularity conditions described above are satisfied, the mean of p exists, and p is even-symmetric about μ , then $\text{KL}(q_{\nu} \parallel p)$ has a unique minimizer at $\nu = \mu$ and hence VI exactly recovers the mean of p .

This result leverages even symmetry. The proof of this result uses the fact that the point of symmetry of an even-symmetric density is its mean, the gradient of an even-symmetric function is odd-symmetric, multiplying an odd-symmetric function by an even-symmetric function results in an odd-symmetric function, and that the integral of an odd-symmetric function about the origin over a symmetric region is zero. The result also holds for location-scale families with general scale parameter. If the condition that $\log p$ is concave is relaxed, then there is a stationary point at $\nu = \mu$, which may correspond to a local or global optima, but not necessarily a global minima.

Main result 2 (Exact recovery of the correlation matrix): Suppose \mathcal{Q} is a location-scale family with q_0 spherically-symmetric, p is elliptically-symmetric about μ with scale matrix M , and that the regularity conditions described before hold. Then there is a unique minimizer of $\text{KL}(q_{\nu,S} \parallel p)$ which occurs when $\nu = \mu$ and $S = \gamma^2 M$ for some $\gamma \in \mathbb{R}$.

This result leverages elliptical symmetry. The proof can be thought of as a generalization of the proof of the first result, using the fact that spherical-symmetry implies even-symmetry. First, $\text{KL}(q_{\nu,S} \parallel p)$ is minimized with respect to the location parameter ν by setting to $\nu = \mu$ as before. The objective function is now a strictly convex function of $M^{-\frac{1}{2}}S^{\frac{1}{2}}$, where M is the scale matrix of the target density $p(z) = p_0(M^{-\frac{1}{2}}(z - \mu))|M|^{-\frac{1}{2}}$ since p is assumed to be elliptically-symmetric. This strictly convex function has a unique global minimizer at $S = \gamma^2 M$. As the correlation matrix is a ratio of entries from the scale matrix, the constant of proportionality γ^2 cancels, and VI exactly recover the correlation matrix of p .

3 Strengths and limitations

Before discussing theoretical limitation of the paper, I will make note of two minor issues that I identified.

First, there is an error in the paper. Equation 3 in Definition 1 has $q_{\nu,S}(z) = q_0(S^{-\frac{1}{2}}(z - \nu))|S|^{\frac{1}{2}}$ but it should be $q_{\nu,S}(z) = q_0(S^{-\frac{1}{2}}(z - \nu))|S|^{-\frac{1}{2}}$. Using Definition 1, the determinant of S would not cancel when performing a change of variables. In Equation 18 for the proof of Theorem 10, Margossian and Saul use the correct version of Definition 1.

Second, the GitHub page does not include the Stan file for the crescent distribution to reproduce figure 6.

It is true that there are open questions and that the methods in the paper do not apply to all situations. But Margossian and Saul do not make unfounded claims. The two theorems in the paper provide sufficient conditions under which VI can exactly recover the mean and the correlation matrix. These are not if and only if statements, but there is no claim that they are. The figures in the paper clearly contrast the behaviour of VI when the sufficient conditions are met, and when a condition is violated. Additionally, Margossian and Saul may be the first to investigate the abilities of VI to recover the correlation matrix. Recovering the correlation matrix when using location-scale families is a natural and clever idea. The correlation matrix is recoverable, unlike the scale or covariance matrix, as any constants of proportionality are cancelled.

The limitations of the paper become apparent when attempting to check whether the sufficient conditions are satisfied. Margossian and Saul provided an estimator for violations of even symmetry. To check whether a target is even-symmetric about μ , we must first estimate μ . If μ were known, then Theorem 8 would be of little use. We check if the target is even-symmetric about this estimate for μ . If it is approximately symmetric, and the regularity conditions are plausible, we can then estimate μ with VI. But we've already estimated μ , and a poor first estimate can lead to a poor even symmetry diagnostic, and in turn, a poor VI estimate of μ . What is needed is a bound on the approximation error for the mean in terms of a measurable characteristic of the target density.

To recover the correlation matrix, the target density must be elliptically symmetric about μ . A bootstrap test for spherical and elliptical symmetry exists, based on the fact that $z \in \mathbb{R}^d$ is spherically symmetric if and only if $\mathbb{E}[u^T z | v^T z] = 0$ for any orthogonal $u, v \in \mathbb{R}^d$ satisfying $u^T v = 0$ [4], with a transformation used to generalize the test for elliptical symmetry. This test can be applied to assess whether the target density is approximately elliptically symmetric, which then implies even symmetry. Again, it would be beneficial to bound the approximation error in terms of the degree of violation of the sufficient conditions.

One then wonders when a density is even or elliptically symmetric. Margossian and Saul describe situations for which the posterior in a Bayesian analysis may approximately exhibit these symmetries. When either the prior or likelihood dominates, it would be the dominating component that is primarily being approximated, and so this is not a particularly interesting Bayesian analysis. A symmetry in the posterior may also occur if the likelihood and prior share a common point of symmetry, though this is an optimistic scenario.

4 Discuss open questions and propose future research directions

There are three categories of future directions discussed, as well as an inspiring conjecture.

- 1) Using the correlation matrix found as a starting point for another algorithm.
- 2) Research symmetries in other variational families.
- 3) Bound the errors of VI in terms of violations of symmetries.

The conjecture is that for each symmetry of the target density, there is a corresponding feature that VI can recover. A Noether's theorem for VI.

The first is a reasonable continuation of this work. However, it is doubtful that the densities encountered in practice will satisfy the sufficient conditions required to exactly recover the correlation matrix. It would therefore be beneficial to bound the correlation matrix approximation error in terms of measurable violations of elliptical symmetry, which ties in with the third research direction. Or, find a broader class of targets for which VI can exactly recover their correlation matrices. The second direction is of great interest. I am interested in studying symmetries in exponential transformation models, which have a close connection to algebraic Lie groups [5]. The third is a natural avenue to explore, especially after observing Figures 1, 4, and 7. This is the bottleneck which limits the applicability of these sufficient conditions. How can we measure

symmetry, and how can that measurement be used to bound the approximation error?

The conjecture is of particular interest. It is this conjecture which inspired the small project in Section 5. Can an ‘asymmetric’ density be expressed as a ‘symmetric’ density under some generalized notion of symmetry? Under which conditions can a density be expressed as symmetric? If the target only exhibits a partial symmetry, a symmetry that holds for a subset of the domain, are there guarantees on which components can be recovered exactly and insights on how any errors behave? Is there an impossibility theorem to determine if a density cannot have a symmetry?

Borrowing from group theory, a ‘continuous symmetry’ is a topological group action that preserves structure [6]. That is, if $f : X \rightarrow Y$ and G is a group on X , then a subgroup $H \subseteq G$ is a symmetry of f if $f(h \cdot x) = f(x)$ for all $h \in H$ [6]. How can this definition of symmetry extend the work done by Margossian and Saul? Can this notion of symmetry allow for seemingly asymmetric densities to be framed as symmetric?

5 Small project

5.1 Rotation symmetry and exact recovery of the mean

My small project relaxes the even-symmetry sufficient condition to exactly recover the mean of the target density.

Let the function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be called n -rotationally symmetric about $\nu \in \mathbb{R}^d$ if $f(\nu + \xi) = f(\nu + R\xi)$ for every $\xi \in \mathbb{R}^d$ and a rotation matrix $R \in \mathbb{R}^{d \times d}$ where $R^n = I$ for some smallest $n \in \mathbb{N} \setminus \{0, 1\}$. Here R may be a proper rotation or an improper rotation, that is, a composition of a proper rotations and a reflection.

Regularity conditions: Assume that the target density is differentiable, and that the log-target density and every distribution in the variational family satisfy the conditions of the dominated convergence theorem. Also assume that the mean of p exists.

Conjecture: Suppose that R has no eigenvalues equal to 1, that is, for all $z \in \mathbb{R}^d \setminus 0$, $z \neq Rz$. If the target density p is n -rotationally symmetric about μ and \mathcal{Q} is a location family whose base distribution q_0 is spherically symmetric, then $\text{KL}(q_\nu \parallel p)$ has an optima at $\nu = \mu$. If, in addition, the target density is log-concave on \mathbb{R}^d and strictly log-concave on some open subset, then $\text{KL}(q_\nu \parallel p)$ has a unique global minimizer at $\nu = \mu$.

5.2 Discussion of properties

Notice that elliptical symmetry implies 2-rotational symmetry. However, if f is 3-rotationally symmetric, f need not be even-symmetric.

By definition, if f is n -rotationally symmetric about ν , then for any $\xi \in \mathbb{R}^d$,

$$f(\nu + \xi) = f(\nu + R\xi) = f(\nu + R^2\xi) = \cdots = f(\nu + R^n\xi) = f(\nu + \xi)$$

where $R^n = I$. If $R^m = -I$ for some $m \in \mathbb{N}$, then $R^{2m} = I$ and so $p(\mu + R^m\xi) = p(\mu + R^{2m}\xi) \implies p(\mu - \xi) = p(\mu + \xi)$, and we can immediately apply the even symmetry result of Margossian and Saul.

A spherical symmetry implies n -rotational symmetry for any $n \in \mathbb{N}$. Therefore an isotropic Gaussian location family is a universally applicable variational family as it matches any rotational symmetry.

For any rotation matrix $R \in \mathbb{R}^{d \times d}$, R is orthogonal and so $R^T = R^{-1}$. Therefore $1 = \det(I) = \det(RR^T) = \det(R)\det(R^T) = \det(R)\det(R) = \det(R)^2 \implies |\det(R)| = 1$.

A proper rotation matrix satisfies $\det(R) = 1$, while an improper rotation satisfies $\det(R) = -1$, corresponding to the composition of a proper rotation and a reflection [7, 8]. Preserving lengths of $\xi \in \mathbb{R}^d$ only requires $|\det(R)| = 1$.

Euler's rotation theorem states that in 3 dimensions, any sequence of proper rotations is equivalent to a single rotation about an axis [7]. R is a $d \times d$ matrix with real entries. By the fundamental theorem of algebra, R has d eigenvalues. When a polynomial has real coefficients, complex roots appear as conjugate pairs. Thus, if d is odd, R must have at least one real eigenvalue. As the norm of any eigenvalue of a rotation matrix is 1, if R has a single real eigenvalue then it must be either 1 or -1 . Therefore, we cannot have a situation as described by Euler. In general, for odd d we require an improper rotation so R has no eigenvalues equal to 1.

I have identified a direction of future work. I wish to investigate whether we can measure the violation of rotational symmetry by finding the nearest orthonormal matrix [9]. Thus, decompose a near-rotational symmetry into a true rotational component and a violation component. From here, attempt to bound the approximation error in terms of the violation component.

5.3 Proof of conjecture

5.3.1 The point of symmetry is indeed the mean

First, we show that the point of symmetry μ is indeed the mean of p .

Suppose that p is n^* -rotationally-symmetric about $\mu \in \mathbb{R}^d$ where $n^* = \arg \min_{n \in \mathbb{N} \setminus \{0,1\}} R^n = I$. That is $p(\mu + \xi) = p(\mu + R\xi)$ for any $\xi \in \mathbb{R}^d$ and rotation matrix $R \in \mathbb{R}^{d \times d}$ with no eigenvalues equal to 1.

The expected value of $z \sim p(z)$ is given by:

$$\mathbb{E}(z) = \int zp(z)dz$$

We will do change of variables two ways:

1) Let $z = \mu + y$ so $dz = dy$ and

$$\begin{aligned} \mathbb{E}(z) &= \int zp(z)dz \\ &= \int (\mu + y)p(\mu + y)dy && \text{Change of variables} \\ &= \mu \int p(\mu + y)dy + \int yp(\mu + y)dy && \text{Linearity} \\ &= \mu + \int yp(\mu + y)dy && p \text{ integrates to 1} \end{aligned}$$

2) Similarly, let $z = \mu + Rx$. R is a rotation matrix so the absolute value of the Jacobian for this transformation is 1 and hence $dz = dx$.

$$\begin{aligned} \mathbb{E}(z) &= \int zp(z)dz \\ &= \int (\mu + Rx)p(\mu + Rx)dx && \text{Change of variables} \\ &= \mu \int p(\mu + Rx)dx + R \int x p(\mu + Rx)dx && \text{Linearity} \\ &= \mu + R \int x p(\mu + Rx)dx && p \text{ integrates to 1} \\ &= \mu + R \int x p(\mu + x)dx && \text{Rotation assumption} \end{aligned}$$

By equating the two equations from separate change of variables we have:

$$\mathbb{E}(z) = \mu + \int yp(\mu + y)dy = \mu + R \int x p(\mu + x)dx$$

equivalently

$$\mathbb{E}(z) - \mu = \int yp(\mu + y)dy = R \int x p(\mu + x)dx$$

but this is the eigenvalue-eigenvector equation where $\vec{v} = \int yp(\mu + y)dy = \int x p(\mu + x)dx$. So $1\vec{v} = R\vec{v}$. But we assumed that R had no eigenvalues equal to 1. So $\vec{v} = \vec{0}$ is the only vector which satisfies $\vec{v} = R\vec{v}$.

Therefore $\mathbb{E}(z) - \mu = \vec{0}$ or $\mathbb{E}(z) = \mu$ and we have shown that indeed the point of symmetry of p is its mean.

Notice that if R does have an eigenvalue of 1, there exists an axis of symmetry, rather than a point of symmetry, and any point along that axis is a point of symmetry.

5.3.2 An optima is achieved when points of symmetry coincide

We want to show that $\nabla_\nu \text{KL}(q_\nu || p)|_{\nu=\mu} = 0$.

Suppose $q_0(\cdot)$ is spherically symmetric about 0. That is $q_0 : \mathbb{R}^d \rightarrow \mathbb{R}$ satisfies if $\|\xi_1\| = \|\xi_2\|$ then $q_0(\xi_1) = q_0(\xi_2)$ where $\xi_1, \xi_2 \in \mathbb{R}^d$. Let $q_\nu(z) = q_0(z - \nu)$ and \mathcal{Q} is the location family formed by allowing any $\nu \in \mathbb{R}^d$.

$$\begin{aligned} \text{KL}(q_\nu || p) &= \int \log \frac{q_\nu(z)}{p(z)} q_\nu(z) dz \\ &= \int [\log q_\nu(z) - \log p(z)] q_\nu(z) dz \\ &\quad \text{Let } \xi = z - \nu \text{ so } z = \xi + \nu \text{ and } dz = d\xi \\ &= \int [\log q_\nu(\xi + \nu) - \log p(\xi + \nu)] q_\nu(\xi + \nu) d\xi \\ &\quad \text{By definition 1, } q_\nu(z) = q_0(z - \nu) \\ &\quad \text{Substituting we have } q_\nu(z = \xi + \nu) = q_0((\xi + \nu) - \nu) = q_0(\xi) \\ &= \int [\log(q_0(\xi)) - \log p(\xi + \nu)] q_0(\xi) d\xi \\ &= \int \log(q_0(\xi)) q_0(\xi) d\xi - \int \log(p(\xi + \nu)) q_0(\xi) d\xi \\ &\quad \text{But } \mathcal{H}(q_0) = \int \log q_0(\xi) q_0(\xi) d\xi \text{ is the entropy of } q_0 \\ &= -\mathcal{H}(q_0) - \int \log(p(\xi + \nu)) q_0(\xi) d\xi \end{aligned}$$

Now, we take the gradient of $\text{KL}(q_\nu || p)$ with respect to ν .

$$\begin{aligned} \nabla_\nu \text{KL}(q_\nu || p) &= \nabla_\nu \left[-\mathcal{H}(q_0) - \int \log(p(\xi + \nu)) q_0(\xi) d\xi \right] \\ &= \nabla_\nu [-\mathcal{H}(q_0)] - \nabla_\nu \left[\int \log(p(\xi + \nu)) q_0(\xi) d\xi \right] \\ &= 0 - \nabla_\nu \left[\int \log(p(\xi + \nu)) q_0(\xi) d\xi \right] \end{aligned}$$

By the Dominated Convergence Theorem we have the following

$$= - \int \nabla_\nu [\log(p(\xi + \nu))] q_0(\xi) d\xi$$

We are using the notation that $\nabla_z f(z) \in \mathbb{R}^d$. Recall that the Jacobian \mathbf{J}_f of the vector f is given by $\mathbf{J}_f = (\nabla f)^T$. The chain rule states that if $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $g : \mathbb{R}^d \rightarrow \mathbb{R}$, then $\mathbf{J}_{g \circ f}(z) = \mathbf{J}_g(f(z))\mathbf{J}_f(z)$ where $z \in \mathbb{R}^d$ [10].

So we have

$$(\nabla_\nu \log p(\xi + \nu))^T = (\nabla_x \log p(x)|_{x=\xi+\nu})^T \mathbf{J}_\nu(\xi + \nu) = (\nabla_x \log p(x)|_{x=\xi+\nu})^T \mathbf{I} = (\nabla_x \log p(x)|_{x=\xi+\nu})^T$$

and

$$(\nabla_\xi \log p(\xi + \nu))^T = (\nabla_x \log p(x)|_{x=\xi+\nu})^T \mathbf{J}_\xi(\xi + \nu) = (\nabla_x \log p(x)|_{x=\xi+\nu})^T \mathbf{I} = (\nabla_x \log p(x)|_{x=\xi+\nu})^T$$

so

$$\nabla_\nu \log p(\xi + \nu) = \nabla_\xi \log p(\xi + \nu)$$

which could also be argued due the the symmetric roles of ξ and ν in $p(\cdot)$.

Therefore we have

$$\nabla_\nu \text{KL}(q_\nu \parallel p) = - \int \nabla_\xi [\log(p(\xi + \nu))] q_0(\xi) d\xi$$

If we can show that this is equal to zero for a certain $\nu = \nu^*$, then ν^* is a stationary point of $\text{KL}(q_\nu \parallel p)$.

Suppose $\nu = \mu$ where μ is the point of symmetry for p .

R is a rotation matrix so $R^T = R^{-1}$ and $|\det(R)| = 1$.

p is n -rotationally symmetric about μ , that is $p(\mu + \xi) = p(\mu + R\xi)$ for every $\xi \in \mathbb{R}^d$.

$q_0(\cdot)$ is spherically symmetric about 0.

Take any $\xi \in \mathbb{R}$ then $\|R\xi\|^2 = (R\xi)^T (R\xi) = \xi^T R^T R \xi = \xi^T R^{-1} R \xi = \xi^T I \xi = \xi^T \xi = \|\xi\|^2$.

Therefore, as $\|R\xi\| = \|\xi\|$, we have $q_0(\xi) = q_0(R\xi)$.

Let $c(\nu) = \int \nabla_\xi [\log(p(\xi + \nu))] q_0(\xi) d\xi$.

We want to show that $c(\mu) = 0$ which would imply that $\nabla_\nu \text{KL}(q_\nu \parallel p)|_{\nu=\mu} = 0$ and so μ is a stationary point.

Define Let $h(x) = \log(p(\mu + x)) : \mathbb{R}^d \rightarrow \mathbb{R}$.

So $h(Rx) = \log(p(\mu + Rx)) = \log(p(\mu + x)) = h(x)$ and thus h is rotational symmetric about the origin.

The Jacobian of h is denoted $\mathbf{J}(h) \in \mathbb{R}^{1 \times d} = (\nabla h)^T$ as $\nabla h \in \mathbb{R}^d$.

The chain rule gives that if $g(x) = f(Ax)$ for $x \in \mathbb{R}^d$ and $A \in \mathbb{R}^{d \times d}$, then $\mathbf{J}(g(x)) = \mathbf{J}(f(Ax)) A$.

Equivalently $\nabla g(x) = A^T \nabla f(Ax)$ as transpose reverses the order.

Matrix properties are confirmed in the *The Matrix Cookbook* [10].

Define $g(x) = h(Rx) = \log(p(\mu + Rx))$.

Then $\mathbf{J}(g(x)) = \mathbf{J}(h(Rx)) R$ which can be written as $(\nabla g(x))^T = (\nabla h(Rx))^T R$.

So $\nabla g(x) = R^T \nabla h(Rx)$.

But $h(x) = \log(p(\mu + x)) = \log(p(\mu + Rx)) = h(Rx)$.

So $g(x) = h(Rx) = h(x)$ and therefore we have that $g = h$ and $\nabla g = \nabla h$.

But we found that $\nabla g(x) = R^T \nabla h(Rx)$ so $\nabla h(x) = R^T \nabla h(Rx)$ or $R \nabla h(x) = \nabla h(Rx)$.

Therefore $R \nabla \log(p(\mu + x)) = \nabla \log(p(\mu + Rx))$.

We will use these properties along with the assumption that R has no eigenvalues of 1 to show that $c(\mu) = 0$

$$\begin{aligned}
c(\mu) &= \int \nabla_\xi [\log(p(\mu + \xi))] q_0(\xi) d\xi \\
&\quad \text{Let } \xi = Rz \text{ so } d\xi = |\det(R)| dz = dz \\
&= \int \nabla_z [\log(p(\mu + Rz))] q_0(Rz) dz \\
&\quad q_0 \text{ is spherically symmetric} \\
&= \int \nabla_z [\log(p(\mu + Rz))] q_0(z) dz \\
&\quad \text{Use } R \nabla_z \log(p(\mu + z)) = \nabla_z \log(p(\mu + Rz)) \\
&= \int R \nabla_z \log(p(\mu + z)) q_0(z) dz \\
&= R \int \nabla_z \log(p(\mu + z)) q_0(z) dz \\
&= R c(\mu)
\end{aligned}$$

So we have the eigenvalue equation $1 c(\mu) = R c(\mu)$. But we assumed that R had no eigenvalues equal to 1. So the only $c(\mu) \in \mathbb{R}^d$ that satisfies $c(\mu) = R c(\mu)$ is $c(\mu) = 0$.

Therefore $\nabla_\nu \text{KL}(q_\nu \parallel p)|_{\nu=\mu} = 0$ and μ is a stationary point.

5.3.3 Concavity conditions ensure a unique minimiser

It remains to show that μ is the unique minimizer.

We assumed that $\log p$ is concave on \mathbb{R}^d and strictly so on an open subset. We need to show that $\text{KL}(q_\nu \parallel p)$ is strictly convex as a function of ν . This will ensure that $\nu = \mu$ is the unique global minimizer.

We found that

$$\text{KL}(q_\nu \parallel p) = - \int \log p(\nu + \xi) q_0(\xi) d\xi := f(\nu)$$

so, as we can interchange differentiation and integration by the dominated converge assumption, the gradient of $f(\nu)$ is given by

$$\nabla_\nu f(\nu) = - \int q_0(\xi) \nabla_\nu \log p(\nu + \xi) d\xi$$

and the Hessian is

$$\nabla_\nu^2 f(\nu) = - \int q_0(\xi) \nabla_\nu^2 \log p(\nu + \xi) d\xi$$

Since we assumed that $\log p$ is concave, $\nabla^2 \log p(z) \preceq 0$, we can take any $w \in \mathbb{R}^d$:

$$w^T [\nabla_\nu^2 \log p(\nu + \xi)] w \leq 0$$

and because $q_0(\xi) \geq 0$ for all ξ

$$w^T \left(- \int q_0(\xi) \nabla_\nu^2 \log p(\nu + \xi) d\xi \right) w \geq 0$$

so $\nabla_\nu^2 f(\nu)$ is convex, the Hessian is positive semi-definite, and the point where $\nabla_\nu f(\nu) = 0$ is a mimimizer.

$$\nabla_\nu^2 f(\nu) = - \int q_0(\xi) \nabla_\nu^2 \log p(\nu + \xi) d\xi \succeq 0$$

If the subset of \mathbb{R}^d where $\log p$ is strictly concave has $q_0 > 0$, then this point is the unique minimizer.

5.4 Simulations

To demonstrate that when the target density is rotationally symmetric and the location variational family has a spherically symmetric base distribution, VI exactly recovers the mean, we consider three cases: 1) a 3-rotationally symmetric target density, 2) an 8-rotationally symmetric target density, and 3) a target with reflection symmetry where R has an eigenvalue equal to 1.

The first two examples support my (now proven) conjecture that if the target is rotationally symmetric, VI can exactly recover its mean. The first target density is chosen to demonstrate that we can relax the even-symmetry assumptions made by Margossian and Saul, and that VI can exactly recover the mean under more general rotational symmetries. The second example satisfies even symmetry, so the generalized symmetry sufficient condition is not required. The third example shows that, similar to the crescent distribution, reflective symmetry in 2 dimensions corresponds to R with eigenvalues -1 and 1 . We showed that when R has an eigenvalue equal to 1, then $\nabla_\nu \text{KL}(q_\nu \parallel p)$ evaluated at a point of symmetry is an eigenvector of R . The reason $\text{KL}(q_\nu \parallel p)$ is minimized at a particular ν remains unknown.

VI was implemented using RStan [11] in R. Code to reproduce the experiments is available on GitHub: <https://github.com/1saacRankin/rotationally-symmetric-VI>.

For each of the three target densities, we construct them stochastically in cylindrical coordinates, convert them to Cartesian coordinates, and then apply a translation.

Algorithm:

- 1) Choose a density for q , $q \sim f$
- 2) Specify a level curve function $r(\theta, q)$
- 3) Sample $\theta \sim \text{Unif}[0, 2\pi]$ and $q \sim f$
- 4) Compute $r(\theta, q)$
- 5) Find $z = (z_1, z_2) \in \mathbb{R}^2$ where $z_1 = r(\theta, q) \cos(\theta)$ and $z_2 = r(\theta, q) \sin(\theta)$
- 6) Repeat steps 3-5 n times to obtain a synthetic dataset of size n
- 7) Translate all points to (a, b) as such $z = (z_1 + a, z_2 + b)$ where $a, b \in \mathbb{R}$

Target density 1:

Choose $q \sim \text{Gam}(\text{shape} = 3, \text{scale} = 2)$, $r(\theta, q) = q \sin(3\theta)$, $n = 10000$, and $(a = 2, b = 1)$.

Estimate of the mean: (1.987309, 1.075028) True mean: (2, 1)

Target density 2:

Choose $q \sim |\mathcal{N}(\text{mean} = 0, \text{sd} = 2)|$, $r(\theta, q) = q |\cos(4\theta)|$, $n = 10000$, and $(a = 2, b = 0)$.

Estimate of the mean: (1.98091512, -0.02184659) True mean: (2, 0)

Target density 3:

Choose $q \sim \text{Gam}(\text{shape} = 10, \text{scale} = 0.5)$, $r(\theta, q) = q \sin(4\theta)(\cos(\theta) + \sin(\theta))$, $n = 10000$, and $(a = -1, b = 2)$.

Estimate of the mean: (-1.026192, 2.006354) True mean: (-1, 2)

For all three target densities, VI approximately recovers the mean. Target densities 1 and 2 do not have eigenvalues equal to 1, and so the recovery of the mean was expected. Target density 3 exhibits a reflection symmetry which corresponds to eigenvalues of -1 and 1 . It remains unclear why VI recovered the mean of target 3 but only managed to find an eigenvector in the also reflectively symmetric crescent distribution. Log-concavity is not formally assessed.

These results show that the sufficient conditions presented by Margossian and Saul can be relaxed. It is sufficient that the target density p exhibits rotational symmetry, a weaker condition than even symmetry.

Target density 1:

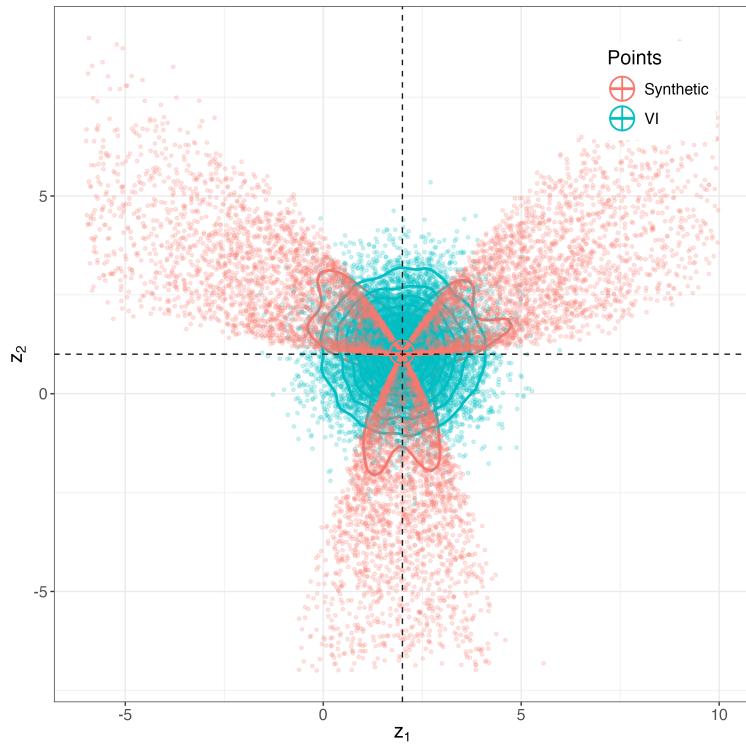


Figure 1: VI recovers the mean on a 3-rotationally symmetric target density

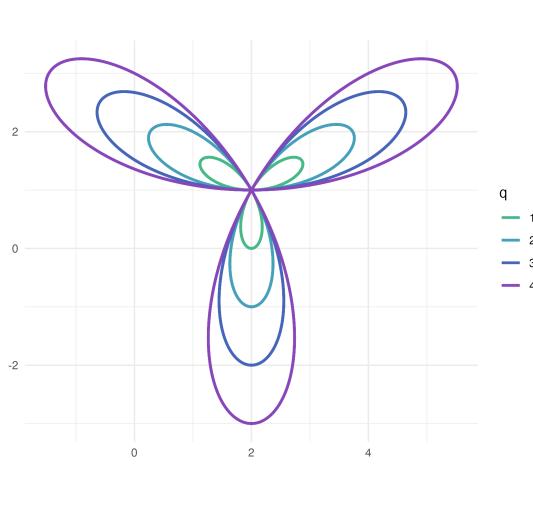


Figure 2: Level curves parametrized by q and θ

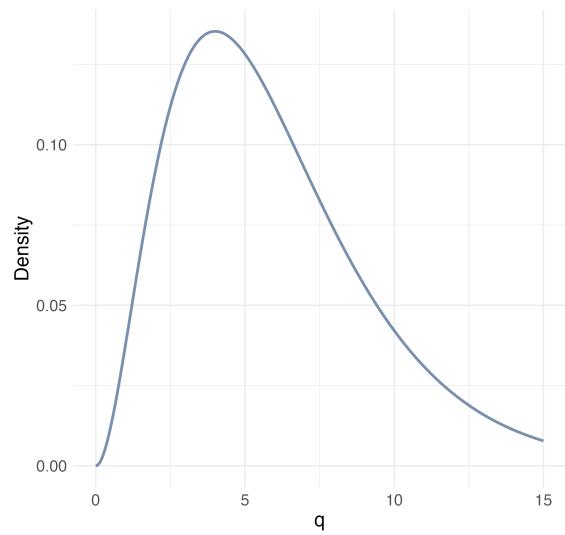


Figure 3: Density of q

Target density 2:

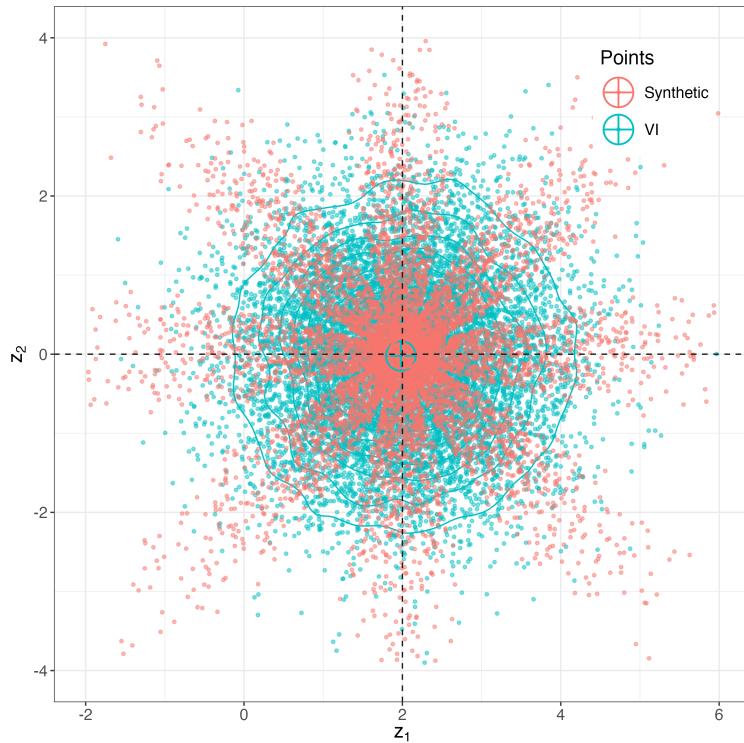


Figure 4: VI recovers the mean on an 8-rotationally symmetric target density

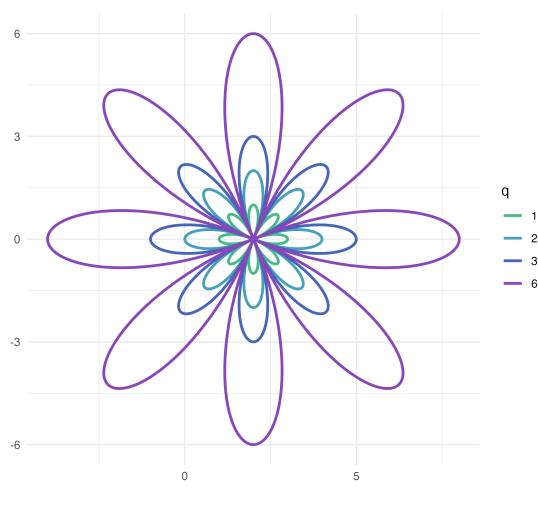


Figure 5: Level curves parametrized by q and θ

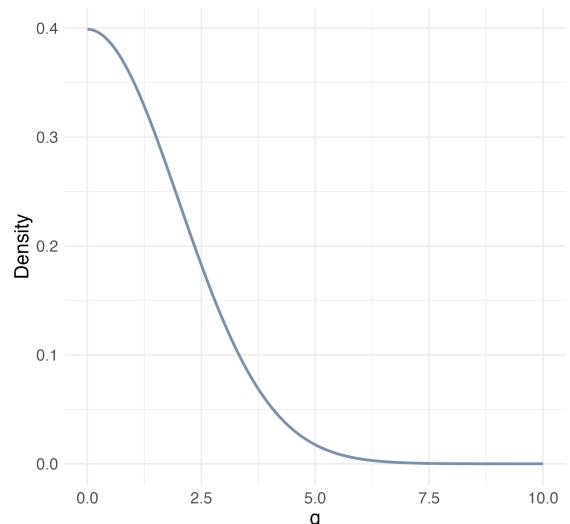


Figure 6: Density of q

Target density 3:

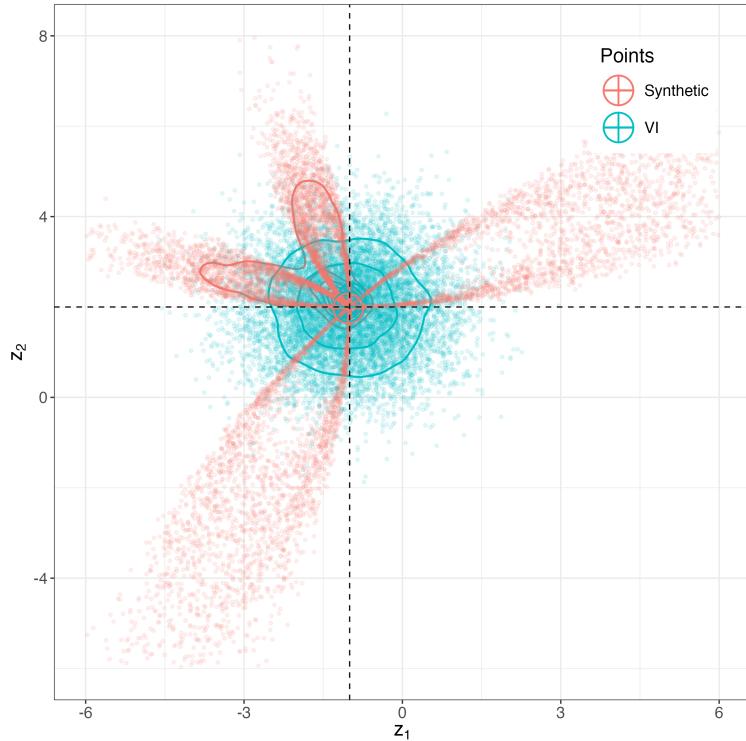


Figure 7: VI recovers the mean for a reflection-symmetric target

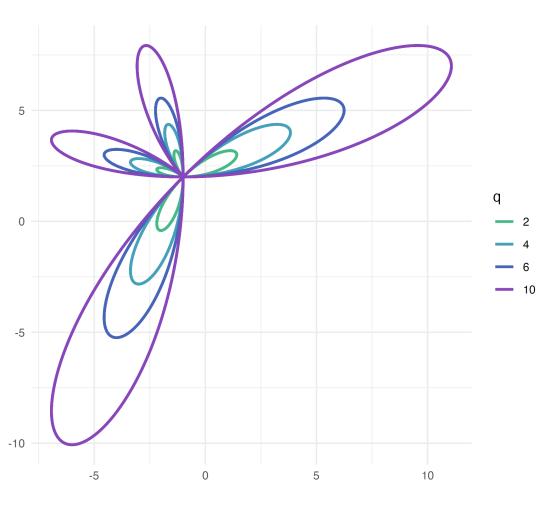


Figure 8: Level curves parametrized by q and θ

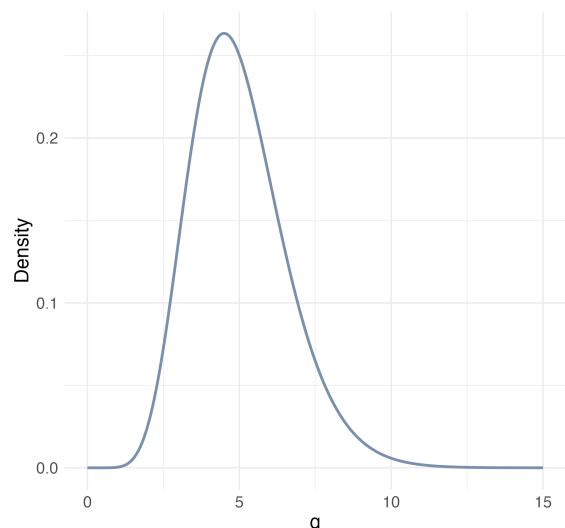


Figure 9: Density of q

Bibliography

- [1] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017.
- [2] Joshua S Speagle. A conceptual introduction to Markov chain Monte Carlo methods. *arXiv preprint arXiv:1909.12313*, 2019.
- [3] Charles C Margossian and Lawrence K Saul. Variational inference in location-scale families: Exact recovery of the mean and correlation matrix. *arXiv preprint arXiv:2410.11067*, 2024.
- [4] Isaia Albisetti, Fadoua Balabdaoui, and Hajo Holzmann. Testing for spherical and elliptical symmetry. *Journal of Multivariate Analysis*, 180:104667, 2020.
- [5] O Barndorff-Nielsen, Preben Blaesild, J Ledet Jensen, and B Jørgensen. Exponential transformation models. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 379(1776):41–65, 1982.
- [6] William H Barker and Roger Howe. *Continuous symmetry: from Euclid to Klein*. American Mathematical Soc., 2007.
- [7] Gabriele MT D’Eleuterio and Timothy D Barfoot. On the eigenstructure of rotations and poses: commonalities and peculiarities. *Proceedings of the Royal Society A*, 478(2266):20220080, 2022.
- [8] Adam Morawiec. *Orientations and rotations*. Springer, 2003.
- [9] Berthold KP Horn, Hugh M Hilden, and Shahriar Negahdaripour. Closed-form solution of absolute orientation using orthonormal matrices. *Journal of the Optical Society of America A*, 5(7):1127–1135, 1988.
- [10] Kaare Brandt Petersen, Michael Syskind Pedersen, et al. The matrix cookbook. *Technical University of Denmark*, 7(15):510, 2008.
- [11] Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76:1–32, 2017.