Problem setup:

- $\gamma_1, \gamma_2, \gamma_3$ are orthonormal unit vectors (3×1 matrices); $\kappa > 0, \beta > 0$ are unconstrained.
- γ^T is the transpose of γ.
- v_0 is a 3×1 matrix. A is a 3×3 symmetric matrix. v_0 and A are functions of data and can be precomputed.

Parameters to be estimated: $\Theta = \{\kappa, \beta, \gamma_1, \gamma_2, \gamma_3\}$
The negative log-likelihood function that needs to be minimized is as follows:

$$L(\Theta) = \log c(\kappa, \beta) - \kappa \gamma_1^T v_0 - \beta \gamma_2^T A \gamma_2 + \beta \gamma_3^T A \gamma_3 \tag{0.1}$$

where

$$c(\kappa, \beta) = 2\pi \sum_{j=0}^{\infty} \frac{\Gamma(j + \frac{1}{2})}{\Gamma(j+1)} \beta^{2j} \left(\frac{\kappa}{2}\right)^{-2j-\frac{1}{2}} I_{2j+\frac{1}{2}}(\kappa) \tag{0.2}$$

$c(\kappa, \beta)$ is the normalization constant which is an infinite summation and is dependent on the Gamma function (Γ) and the modified Bessel function of the first kind $I_v(\kappa)$. It could be a bottleneck to efficiently implement this summation.

Equation 0.1 should be minimized subject to the following constraints:

$$\begin{aligned}
\gamma_1^T \gamma_1 &= 1 \\
\gamma_2^T \gamma_2 &= 1 \\
\gamma_3^T \gamma_3 &= 1 \\
\gamma_1^T \gamma_2 &= 0 \\
\gamma_1^T \gamma_3 &= 0 \\
\gamma_2^T \gamma_3 &= 0
\end{aligned}$$

As an example, I generated 1000 data points from a Kent distribution with the true values of the parameters as follows:

$$\begin{aligned}
\gamma_1^T &= (0.580, -0.388, 0.717) \\
\gamma_2^T &= (-0.753, -0.592, 0.289) \\
\gamma_3^T &= (0.312, -0.707, -0.635) \\
\kappa &= 100 \\
\beta &= 20
\end{aligned}$$

v_0 and A are precomputed (from the data generated) and are as follows:

$$v_0^T = (0.580, -0.385, 0.709)$$

and

$$A = \begin{pmatrix} 0.341 & -0.221 & 0.408 \\ -0.221 & 0.153 & -0.272 \\ 0.408 & -0.272 & 0.506 \end{pmatrix}$$

Using these values for v_0 and A, equation 0.1 needs to be minimized.