Supplement to
“Classification with the matrix-variate-t distribution”

by Geoffrey Z. Thompson, Ranjan Maitra, William Q. Meeker and Ashraf F. Bastawros

Iowa State University, Ames, Iowa, USA.

S-1 The EM algorithm for parameter estimation in the MxVt distribution

As mentioned in the paper, the MxVt distribution does not have closed-form ML estimators so we develop an EM algorithm by augmenting the data, in similar spirit as done for the vector-multivariate t-distribution Lange et al. (1989), and then present an ECME (Expectation/Conditional Maximization Either) algorithm (Liu and Rubin, 1994) to improve the speed of convergence of the EM algorithm. Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be independent realizations from the $t_{p,q}(\nu, \mathbf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Omega})$ density. Then each $\boldsymbol{X}_i$ can be augmented with latent Wishart-distributed weight matrices $\boldsymbol{S}_i$ as follows:

$$\begin{aligned}\boldsymbol{X}_i|\mathbf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \nu, \boldsymbol{S}_i &\sim \mathcal{N}_{p,q}(\mathbf{M}, \boldsymbol{S}_i^{-1}, \boldsymbol{\Omega})\\ \boldsymbol{S}_i|\mathbf{M}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}, \nu &\sim \mathcal{W}_p(\nu + p - 1, \boldsymbol{\Sigma}^{-1}), \quad \text{for } i = 1, 2, \ldots, n.\end{aligned} \tag{S-1}$$

To show the benefits of using the latent $\boldsymbol{S}_i$s, we first derive ML estimators with the complete data and then use that to derive an EM algorithm using only the observed data. We then modify the EM algorithm to its more efficient ECME derivative.

S-1.1 ML Estimation of parameters with complete data

Suppose that we have $(\boldsymbol{X}_i, \boldsymbol{S}_i), i = 1, 2, \ldots, n$ where each $\boldsymbol{S}_i \sim \mathcal{W}_p(\nu + p - 1, \boldsymbol{\Sigma}^{-1})$ and $\boldsymbol{X}_i \mid \boldsymbol{S}_i \sim \mathcal{N}_{p,q}(\mathbf{M}, \boldsymbol{S}^{-1}, \boldsymbol{\Omega})$ for each $i = 1, 2, \ldots, n$. Then the complete log-likelihood function ℓ_c of the parameters $(\mathbf{M}, \boldsymbol{\Omega})$ given the data $(\boldsymbol{X}_i, \boldsymbol{S}_i), i = 1, 2, \ldots, n$ can written as a sum of (conditional) MxVN log-likelihood functions ℓ_N and a sum of Wishart log-likelihood functions ℓ_W:

$$\ell_c(\mathbf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \nu; \boldsymbol{X}, \boldsymbol{S}) = \ell_N(\mathbf{M}, \boldsymbol{S}^{-1}, \boldsymbol{\Omega}; \boldsymbol{X} \mid \boldsymbol{S}) + \ell_W(\nu, \boldsymbol{\Sigma}; \boldsymbol{S})$$

From the definitions of the MxVN and Wishart distributions, we have, after ignoring additive constants,

$$\begin{aligned}\ell_N(\mathbf{M}, \boldsymbol{S}^{-1}, \boldsymbol{\Omega}; \boldsymbol{X} \mid \boldsymbol{S}) = &-\frac{np}{2}\log|\boldsymbol{\Omega}| + \frac{q}{2}\sum_{i=1}^n \log|\boldsymbol{S}_i|\\ &- \frac{1}{2}\mathrm{tr}\left[\sum_{i=1}^n \boldsymbol{S}_i\boldsymbol{X}_i\boldsymbol{\Omega}^{-1}\boldsymbol{X}_i^T + (\sum_{i=1}^n \boldsymbol{S}_i)\mathbf{M}\boldsymbol{\Omega}^{-1}\mathbf{M}^T - 2(\sum_{i=1}^n \boldsymbol{S}_i\boldsymbol{X}_i)\boldsymbol{\Omega}^{-1}\mathbf{M}^T\right]\end{aligned}$$

and

$$\begin{aligned}\ell_W(\nu, \boldsymbol{\Sigma}; \boldsymbol{S}) = &(\nu - 2)/2\sum_{i=1}^n \log|\boldsymbol{S}_i| - \sum_{i=1}^n \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{S}_i)/2 - n\nu p/2\log 2\\ &+ n(\nu + p - 1)/2\log|\boldsymbol{\Sigma}| - n\log\Gamma_p((\nu + p - 1)/2).\end{aligned}$$

To simplify computation of the ML estimators and their notation, we define the following complete data sufficient statistics for the parameters:

$$\boldsymbol{S}_{SX} = \sum_{i=1}^n \boldsymbol{S}_i\boldsymbol{X}_i; \quad \boldsymbol{S}_S = \sum_{i=1}^n \boldsymbol{S}_i; \quad \boldsymbol{S}_{XSX} = \sum_{i=1}^n \boldsymbol{X}_i^T\boldsymbol{S}_i\boldsymbol{X}_i \quad \boldsymbol{S}_{|S|} = \sum_{i=1}^n \log|\boldsymbol{S}_i|.$$

Taking derivatives of log-likelihoods yields the ML estimates:

$$\widehat{\mathbf{M}} = \left(\sum_{i=1}^{n} \boldsymbol{S}_i\right)^{-1} \sum_{i=1}^{n} \boldsymbol{S}_i \boldsymbol{X}_i = \boldsymbol{S}_S^{-1} \boldsymbol{S}_{SX},$$
$$\widehat{\boldsymbol{\Omega}} = \frac{1}{np} \sum_{i=1}^{n} (\boldsymbol{X}_i - \widehat{\mathbf{M}})^T \boldsymbol{S}_i (\boldsymbol{X}_i - \widehat{\mathbf{M}}) = \frac{1}{np} \left(\boldsymbol{S}_{XSX} - \boldsymbol{S}_{SX}^T \boldsymbol{S}_S^{-1} \boldsymbol{S}_{SX}\right),$$
$$\widehat{\boldsymbol{\Sigma}}^{-1} = \frac{1}{n(\nu + p - 1)} \sum_{i=1}^{n} \boldsymbol{S}_i = \frac{\boldsymbol{S}_S}{n(\nu + p - 1)}.$$

The ML estimate of ν can be obtained by finding the root of the equation:

$$n\psi_p((\nu + p - 1)/2) - (\boldsymbol{S}_{|S|} - np \log 2 + n \log |\boldsymbol{\Sigma}|) = 0$$

with $\psi_p(\cdot)$ the p-variate digamma function, defined as $\psi_p(x) = d \log \Gamma_p(x)/dx$. The ML estimate of ν may be obtained numerically by a one-dimensional search algorithm. We now use the development in this section in our EM algorithm for a sample from the MxVt distribution.

S-1.2 Estimating parameters from a MxVt sample

S-1.2.1 The EM algorithm

Let $\boldsymbol{X}_i, i = 1, 2, \ldots, n$ be independent identically distributed realizations from $t_{p,q}(\nu, \mathbf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Omega})$. As in the main article, we write $\boldsymbol{\Theta} \equiv \{\nu, \mathbf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}\}$. From the development in the introduction of this section, for each $i = 1, 2, \ldots, n$, let $\boldsymbol{S}_i$ be (unobserved) random matrices as per Equation (S-1) and Property 2. Then the expected complete log-likelihood function is

$$\begin{aligned} Q(\Theta; \Theta^{(t)}) = &-\frac{np}{2} \log |\boldsymbol{\Omega}| - \frac{n\nu p \log 2}{2} - n \log \Gamma_p\left(\frac{\nu + p - 1}{2}\right) + n\frac{\nu + p - 1}{2} \log |\boldsymbol{\Sigma}| \\ &+ \mathbb{E}_{\Theta^{(t)}} \left\{ \left[-\frac{1}{2} \mathrm{tr} \left(\sum_{i=1}^{n} \boldsymbol{S}_i \boldsymbol{X}_i \boldsymbol{\Omega}^{-1} \boldsymbol{X}_i^T + \sum_{i=1}^{n} \boldsymbol{S}_i \mathbf{M} \boldsymbol{\Omega}^{-1} \mathbf{M}^T - 2 \sum_{i=1}^{n} \boldsymbol{S}_i \boldsymbol{X}_i \boldsymbol{\Omega}^{-1} \mathbf{M}^T \right) \right.\right. \\ &\left.\left. + \frac{\nu - 2}{2} \sum_{i=1}^{n} \log |\boldsymbol{S}_i| - \frac{1}{2} \sum_{i=1}^{n} \mathrm{tr}(\boldsymbol{\Sigma} \boldsymbol{S}_i) + \frac{q}{2} \sum_{i=1}^{n} \log |\boldsymbol{S}_i| \middle| \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n \right] \right\}. \end{aligned} \tag{S-2}$$

E-step Using Property 2 and properties of the Wishart distribution, the expectation step (E-step) updates at the current value $\boldsymbol{\Theta}^{(t)}$ of $\boldsymbol{\Theta}$ are, by taking the expected values of the $\boldsymbol{S}_i$ given the current value of $\boldsymbol{\Theta}^{(t)}$:

$$\boldsymbol{S}_i^{(t+1)} \doteq \mathbb{E}_{\Theta^{(t)}}(\boldsymbol{S}_i | \boldsymbol{X}_i) = (\nu^{(t)} + p + q - 1)[(\boldsymbol{X}_i - \mathbf{M}^{(t)}) {\boldsymbol{\Omega}^{(t)}}^{-1} (\boldsymbol{X}_i - \mathbf{M}^{(t)})^T + \boldsymbol{\Sigma}^{(t)}]^{-1},$$
$$\mathbb{E}_{\Theta^{(t)}} \left(\log |\boldsymbol{S}_i| \big| \boldsymbol{X}_i\right) = \psi_p\left(\frac{\nu^{(t)} + p + q - 1}{2}\right) + p \log 2 + \log \left| \frac{\boldsymbol{S}_i^{(t+1)}}{\nu^{(t)} + p + q - 1} \right|,$$

with $\psi_p(\cdot)$ as the p-variate digamma function. Note that the updates for $\boldsymbol{S}_i^{(t+1)}$ exist by construction if the $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are positive definite. We define and store the expected sufficient statistics to reduce computational calculations and for notational convenience:

$$
\begin{aligned}
\boldsymbol{S}_S^{(t+1)} &\doteq \sum_{i=1}^{n} \boldsymbol{S}_i^{(t+1)}, \\
\boldsymbol{S}_{SX}^{(t+1)} &\doteq \sum_{i=1}^{n} \mathbb{E}_{\Theta^{(t)}}(\boldsymbol{S}_i \boldsymbol{X}_i | \boldsymbol{X}_i) = \sum_{i=1}^{n} \boldsymbol{S}_i^{(t+1)} \boldsymbol{X}_i, \\
\boldsymbol{S}_{XSX}^{(t+1)} &\doteq \sum_{i=1}^{n} \mathbb{E}_{\Theta^{(t)}}(\boldsymbol{X}_i^T \boldsymbol{S}_i \boldsymbol{X}_i | \boldsymbol{X}_i) = \sum_{i=1}^{n} \boldsymbol{X}_i^T \boldsymbol{S}_i^{(t+1)} \boldsymbol{X}_i, \\
\boldsymbol{S}_{|S|}^{(t+1)} &\doteq \mathbb{E}_{\Theta^{(t)}} \left[\sum_{i=1}^{n} \log |\boldsymbol{S}_i| \middle| \boldsymbol{X}_i \right],
\end{aligned}
$$

with the last expression needed only when we are also estimating ν. In that case, these statistics can be expressed with $(\nu^{(t)} + p + q - 1)$ factored out, and for convenience may be computed and stored that way when ν needs to be estimated. These quantities can be computed in $\boldsymbol{O}(npq^2) + \boldsymbol{O}(np^2q) + \boldsymbol{O}(np^3)$ flops.

Maximization step Based on the updated weight matrices $\boldsymbol{S}_i^{(t+1)}$ and statistics based on $\Theta^{(t)}$ and $\boldsymbol{X}$, we get the updates:

$$
\begin{aligned}
\widehat{\mathbf{M}} &= \left(\sum_{i=1}^{n} \boldsymbol{S}_i^{(t+1)} \right)^{-1} \sum_{i=1}^{n} \boldsymbol{S}_i \boldsymbol{X}_i = \boldsymbol{S}_S^{(t+1)^{-1}} \boldsymbol{S}_{SX}^{(t+1)}, \\
\widehat{\boldsymbol{\Omega}} &= \frac{1}{np} \sum_{i=1}^{n} (\boldsymbol{X}_i - \mathbf{M}^{(t)})^T \boldsymbol{S}_i^{(t+1)} (\boldsymbol{X}_i - \mathbf{M}^{(t)}) \\
&= \frac{1}{np} \left(\boldsymbol{S}_{XSX}^{(t+1)} - \boldsymbol{S}_{SX}^{(t+1)^T} \boldsymbol{S}_S^{(t+1)^{-1}} \boldsymbol{S}_{SX}^{(t+1)} \right), \\
\widehat{\boldsymbol{\Sigma}}^{-1} &= \frac{1}{n(\nu^{(t)} + p - 1)} \sum_{i=1}^{n} \boldsymbol{S}_i^{(t+1)} = \frac{\boldsymbol{S}_S^{(t+1)}}{n(\nu^{(t)} + p - 1)}.
\end{aligned}
$$

This can be computed in $\boldsymbol{O}(p^2q) + \boldsymbol{O}(pq^2)$ flops, which is negligible compared to the E-step computations. Again, treating the set of $\boldsymbol{S}_i^{(t+1)}$ as observed, the MLE of ν can be obtained:

$$
n \frac{d}{d\nu} \log \Gamma_p((\nu + p - 1)/2) - \frac{1}{2} (\boldsymbol{S}_{|S|} - np \log 2 + n \log |\widehat{\boldsymbol{\Sigma}}|) = 0.
$$

Defining $\kappa^{(t)} = \nu^{(t)} + p + q - 1$ for compactness:

$$
\begin{aligned}
0 &= n\psi_p((\nu + p - 1)/2) - \left(n\psi_p \left(\frac{\kappa^{(t)}}{2} \right) + \sum_{i=1}^{n} \log \left| \frac{\boldsymbol{S}_i^{(t+1)}}{\kappa^{(t)}} \right| - n \log \left| \frac{\boldsymbol{S}_S^{(t+1)}}{n(\nu^{(t)} + p - 1)} \right| \right) \\
&= \psi_p((\nu + p - 1)/2) - \left(\psi_p \left(\frac{\kappa^{(t)}}{2} \right) + \frac{1}{n} \sum_{i=1}^{n} \log \left| \boldsymbol{Z}_i^{(t+1)} \right| + p \log \frac{n(\nu^{(t)} + p - 1)}{\kappa^{(t)}} - \log \left| \boldsymbol{Z}_S^{(t+1)} \right| \right)
\end{aligned}
\tag{S-3}
$$

where $\boldsymbol{Z}_*$ is the appropriate $\boldsymbol{S}_*$ statistic with $(\nu^{(t)} + p + q - 1)$ factored out and ψ_p is the p-dimensional digamma function. This can be solved using a 1-dimensional search.

Since each $\boldsymbol{S}_i$ is positive definite by construction if the previous $\boldsymbol{\Sigma}^{(t)}$ and $\boldsymbol{\Omega}^{(t)}$ were positive definite, the updates $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\mathbf{M}}$ exist. The conditions for the positive definiteness of the update $\widehat{\boldsymbol{\Omega}}$ are less clear: it is the sum of matrices only guaranteed to be positive semi-definite and we do not have a proof of the necessary or sufficient sample size to guarantee the update is positive definite (a.s.) as required for the method. A solution for $\hat{\nu}$ is guaranteed to exist as long as $\widehat{\mathbf{\Sigma}}$ and $\widehat{\mathbf{\Omega}}$ exist and are positive definite.

ML Estimation with the Expectation/Conditional Maximization Either (ECME) algorithm First we note that, if ν is known, there is no need to partition the M-step into multiple constrained maximization steps. If ν is required to be estimated, there is no difference between a standard EM and a standard ECM (Expectation/Conditional Maximization) algorithm in this setting, since, as in the case of the multivariate t distribution, the complete data likelihood function factorizes into $\Theta_1 = (\mathbf{M}, \mathbf{\Sigma}, \mathbf{\Omega})$ and $\Theta_2 = (\nu)$. However, by partitioning it in this way, it is possible, similarly to the case of the multivariate t, to find a more efficient method of maximization. This is desirable because the M-step for ν can be slow. Here we present an ECME (Expectation/Conditional Maximization Either) algorithm that first maximizes the expected log-likelihood for $(\mathbf{M}, \mathbf{\Sigma}, \mathbf{\Omega})$ and then maximizes the actual log-likelihood over ν given the current values $(\mathbf{M}, \mathbf{\Sigma}, \mathbf{\Omega})$, similar to Liu and Rubin (1994).

Given $\Theta_1 = (\mathbf{M}, \mathbf{\Sigma}, \mathbf{\Omega})$, we can maximize for ν in Equation (1), yielding the set of equations provided in (2)

$$
\begin{aligned}
0 &= n\psi_p((\nu+p-1)/2) - \left\{ n\psi_p\left(\frac{\kappa}{2}\right) + \sum_{i=1}^{n} \log\left|\frac{\boldsymbol{S}_i^{(t+1)}}{\kappa}\right| - n\log\left|\frac{\boldsymbol{S}_S^{(t+1)}}{n(\nu+p-1)}\right| \right\} \\
&= \psi_p((\nu+p-1)/2) - \left\{ \psi_p\left(\frac{\kappa}{2}\right) + \frac{1}{n}\sum_{i=1}^{n} \log\left|\boldsymbol{Z}_i^{(t+1)}\right| + p\log\frac{n(\nu+p-1)}{\kappa} - \log\left|\boldsymbol{Z}_S^{(t+1)}\right| \right\}.
\end{aligned}
$$

The difference is that the solution for $\nu^{(t+1)}$ no longer depends on $\nu^{(t)}$, Solving this equation is slightly more computationally complex than solving Equation (S-3) (ν appears four times in the equation to be solved rather than once) but this converges in fewer total iterations. The ML estimating equation can be solved by a one-dimensional search, providing a ECME algorithm with the steps (as also provided in the main article):

1. **E-step:** Update $\boldsymbol{S}_i$ weights and statistics based on $\Theta^{(t)}$ and $\boldsymbol{X}$.
2. **CME-step:** Update $\Theta_1^{(t+1)} = (\mathbf{M}^{(t+1)}, \mathbf{\Sigma}^{(t+1)}, \mathbf{\Omega}^{(t+1)})$.
3. **CME-step:** Update $\Theta_2^{(t+1)} = \nu^{(t+1)}$ using the observed log-likelihood given the current values $(\mathbf{M}^{(t+1)}, \mathbf{\Sigma}^{(t+1)}, \mathbf{\Omega}^{(t+1)})$ by solving Equation (2).

Repeat these steps until convergence. Each iteration of this algorithm takes $\boldsymbol{O}(npq^2) + \boldsymbol{O}(np^2q) + \boldsymbol{O}(np^3)$ flops plus the number of iterations required by the second CME step.

S-1.2.2 Fitting with restrictions on the parameters

In some settings, restrictions on the parametrization of the center or scatter matrices are appropriate. In this section, we derive solutions in the cases of center matrices that are constant across rows, columns, or the entire matrix. In Roy and Khattree (2005) some results for restrictions on covariance matrices were derived and in this paper AR(1) covariance structures and compound symmetry (CS) variance structures were used; however, they were fit numerically as closed forms for the derivatives and determinants exist. Let $\mathbf{1}_{p,q}$ denote a $(p \times q)$ matrix consisting only of 1s. Then it can be shown that these are the appropriate M-step estimates for certain mean matrix constraints:

$$
\begin{aligned}
\mathbf{M} = \mathbf{1}_{p,q}\mu: &\qquad \widehat{\mathbf{M}} = \mathrm{tr}(\boldsymbol{S}_{SX}\widehat{\mathbf{\Omega}}^{-1}\mathbf{1}_{q,p})/\mathrm{tr}(\boldsymbol{S}_S\mathbf{1}_{p,q}\widehat{\mathbf{\Omega}}^{-1}\mathbf{1}_{q,p})\mathbf{1}_{p,q} \\
\mathbf{M} = \mathbf{1}_{p,1}\mu_{1,q}: &\qquad \widehat{\mathbf{M}} = \mathbf{1}_{p,p}\boldsymbol{S}_{SX}/(\mathbf{1}_{1,p}\boldsymbol{S}_S\mathbf{1}_{p,1}) \\
\mathbf{M} = \mu_{p,1}\mathbf{1}_{1,q}: &\qquad \widehat{\mathbf{M}} = \boldsymbol{S}_S^{-1}\boldsymbol{S}_{SX}\widehat{\mathbf{\Omega}}^{-1}\mathbf{1}_{q,q}/(\mathbf{1}_{1,q}\widehat{\mathbf{\Omega}}^{-1}\mathbf{1}_{q,1})
\end{aligned}
$$

which can be used to simplify the ECME algorithms further.

S-1.3 Performance Evaluations

S-1.3.1 Simulation Study

In the main paper, results pertaining to the recovery of the ν parameter were reported for a simulation study where 200 datasets were produced for $\nu = 5, 10, 20$ and $n = 35, 50, 100$ with a 0 mean matrix and identity scatter matrices. Here we report also the results for the recovery of the mean and covariance parameters. For $\boldsymbol{X} \sim t(\nu, \boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Omega})$, we have the result that $\mathrm{cov}(\mathrm{vec}(\boldsymbol{X})) = \boldsymbol{\Sigma} \otimes \boldsymbol{\Omega}/(\nu - 2)$. To compare all nine sets of simulations on the same scale, we correct each by the appropriate scaling factor such that each has an identity covariance matrix and then report the root mean square difference between the actual and fitted $\hat{\boldsymbol{M}}$ and $\hat{\boldsymbol{\Sigma}} \otimes \hat{\boldsymbol{\Omega}}$ in Figure S-1. The figures indicate performance improves as the sample size increases and indicates good recovery of the parameters in every case.

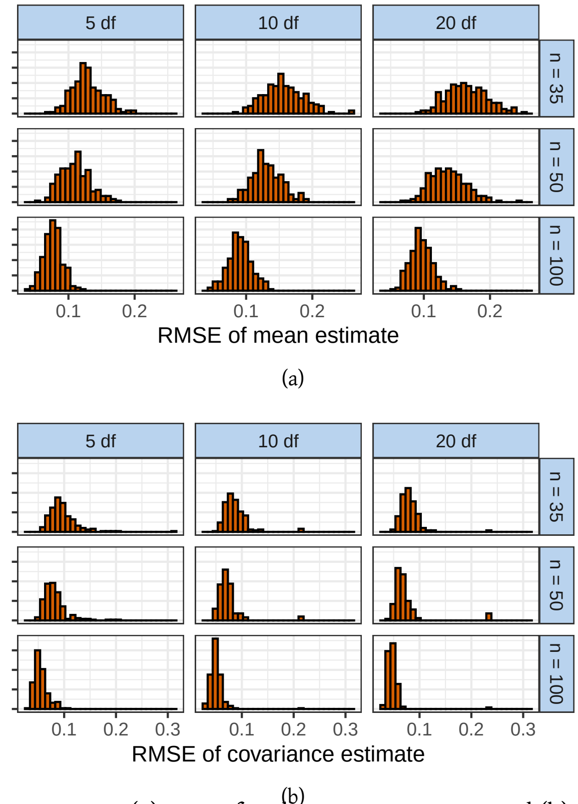

Figure S-1: (a) RMSE for the mean estimates and (b) RMSE for the covariance estimates for datasets of size $n = 35, 50, 100$ with true $\nu = 5, 10, 20$.

We provide a second simulation study to address concerns about model misspecification, namely, what happens when a matrix t distribution model is treated as a matrix normal or vice versa. Three datasets of size 100 with mean matrix 0 and parameters $\boldsymbol{\Sigma}$ a 5×5 AR(1) matrix with $\rho = 0.7$ and $\boldsymbol{\Omega}$ a draw from a standard Wishart distribution with $\nu = 10$ and dimension 8, with one dataset from a MxVt distribution with 6 degrees of freedom, one with 20 degrees of freedom, and one from a MxVN distribution.

In Figure S-2, we plot the log-likelihood, squared deviation from the mean, and the L^2 distance between the true and estimated covariance matrix. The top two rows indicate the results for the MxVt with $\nu = 6$ and 20 and the bottom indicates the results for the MxVN, fitted to a MxVN and to MxVt models with $\nu = 3, 4, \ldots, 100$. On the MxVt with $\nu = 6$ and 20, the MxVN performed poorly compared to the MxVt with ν near the true parameter values. On the MxVN, the MxVt performed poorly.

For all of the datasets, the MxVN has slightly worse recovery of the mean matrix than the MxVt distributions while the MxVN had estimates of the covariance matrix that were comparable to the best MxVt estimates. The L^2 norm of the covariance matrix was not accurate for low values of ν.

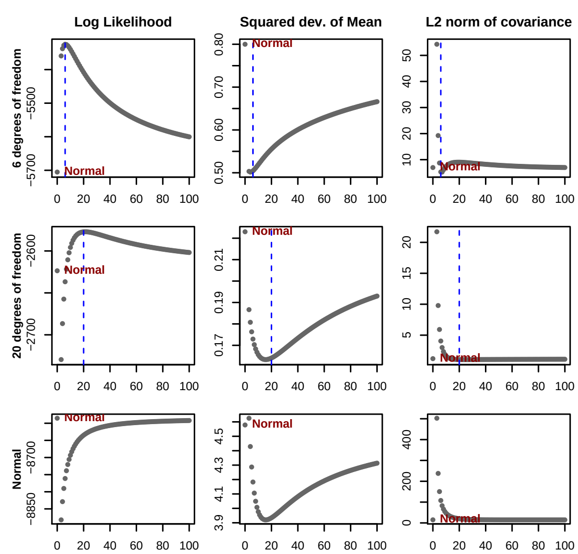

Figure S-2: The top rows contains results for a true $\nu = 6$ and 20 (the blue line) while the bottom row contains the results for a true matrix normal distribution.

The behavior here is suggestive of what occurs in the results when the method fails to converge. Simulations that fail to converge slowly increase likelihood as ν increases until either the maximum number of iterations or the upper bound of ν is reached. This scenario occurs more frequently when simulating from distributions with large ν

and small sample sizes or simulating from MxVN distributions with modest sample sizes (for larger sample sizes, even an MxVN will usually converge to some distribution with large ν). As Figure S-2 indicates, the likelihood surface is very flat for a true MxVN across values of ν. With a small sample size and ν not small, this may occur there was well.

S-1.3.2 Matching Fractured Surfaces

The knife surfaces were scanned using a standard non-contact 3D optical interferometer in corresponding regions, then the 2D Fourier frequencies were computed and compared. In Figure S-3, we illustrate one pair of corresponding images (out of 9) from one of the knife base-tip pairs (out of 38). On the left are a visualization of the output of the 3D optical interferometer for the two surfaces. Note that the images are presented as-is - they should fit together when one is flipped over. The blue depressed region on the top corresponds to the red elevated region on the bottom. On the right is a visualization of the 2D Fourier transform with the frequency ranges used for comparison highlighted - the two bands between the "low frequency" and "high frequency" region.

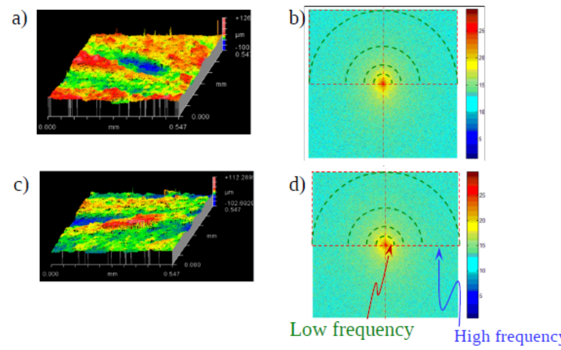

Figure S-3: Surface height 3D topographic maps for tip and base pair (a,c) and their corresponding 2D spectral analysis (b,d).

References

Lange, K. L., Little, R. J., and Taylor, J. M. (1989), "Robust statistical modeling using the t distribution," *Journal of the American Statistical Association*, 84, 881–896.

Liu, C. and Rubin, D. B. (1994), "The ECME Algorithm: A Simple Extension of EM and ECM with Faster Monotone Convergence," *Biometrika*, 81, 633–648. DOI: 10.1093/biomet/81.4.633.

Roy, A. and Khattree, R. (2005), "On discrimination and classification with multivariate repeated measures data," *Journal of Statistical Planning and Inference*, 134, 462 – 485. DOI: https://doi.org/10.1016/j.jspi.2004.04.012.