THE EFFICIENCIES OF THE SPATIAL MEDIAN AND SPATIAL SIGN COVARIANCE MATRIX FOR ELLIPTICALLY SYMMETRIC DISTRIBUTIONS

BY ANDREW F. MAGYAR

A dissertation submitted to the Graduate School—New Brunswick Rutgers, The State University of New Jersey in partial fulfillment of the requirements for the degree of Doctor of Philosophy Graduate Program in Statistics Written under the direction of David E. Tyler and approved by

> New Brunswick, New Jersey May, 2012

ABSTRACT OF THE DISSERTATION

The Efficiencies of the Spatial Median and Spatial Sign Covariance Matrix for Elliptically Symmetric Distributions

by Andrew F. Magyar Dissertation Director: David E. Tyler

The spatial median and spatial sign covariance matrix (SSCM) are popularly used robust alternatives for estimating the location vector and scatter matrix when outliers are present or it is believed the data arises from some distribution that is not multivariate normal. When the underlying distribution is an elliptical distribution, it has been observed that these estimators perform better under certain scatter structures. This dissertation is a detailed study of the efficiencies of the spatial median and the SSCM under the elliptical model, in particular the dependence of their efficiencies on the population scatter matrix. For the spatial median, it is shown this estimator is asymptotically most efficient compared to the MLE for the location vector when the population scatter matrix is proportional to the identity matrix. Furthermore, it is possible to construct an affinely equivariant version of the spatial median that is asymptotically more efficient than the spatial median. Asymptotic relative efficiencies of these two estimators are calculated to demonstrate how inefficient the spatial median can be as the underlying scatter structure becomes more elliptical. A simulation experiment is carried out to provide evidence of analogous result for finite samples. When the goal is estimating eigenprojection matrices, it is proven that under the elliptical model the eigenprojection estimates obtained from the Tyler matrix are asymptotically more efficient than those corresponding to the SSCM. Calculations of asymptotic relative efficiencies are presented to demonstrate the loss of efficiency in using eigenprojection estimates of the SSCM as opposed to the Tyler matrix, particularly when the scatter structure of the data is far from spherical. To assess the performance of these estimators in the finite sample setting, the notion of principal angles is used to define a means to compare eigenprojection estimators. Using this concept, simulations are implemented that support finite sample results similar to those for the asymptotic case. The implications of the above results are discussed, particularly in the application of principal component analysis. Future research directions are then proposed.

Acknowledgements

First and foremost, I would like to acknowledge my doctoral advisor, Dr. David E. Tyler, for mentoring me throughout my graduate studies. Not only has he introduced me to the field of robust statistics but also provided me with a topic that I have enjoyed researching these last few years. This dissertation could not have been completed without his guidance.

Secondly, I would like to acknowledge the entire Statistics Department at Rutgers University for giving me the chance to pursue graduate studies and embarking me on an interesting and rewarding career. My time here has enabled me to develop my love of mathematics and apply it to a field that I have grown to love equally. It was a privilege to study under such an outstanding faculty, each of whom has helped me in some capacity whether it be academically, intellectually or professionally.

I would also like to acknowledge Dr. Daniel Ocone from the Department of Mathematics as well as Dr. Richard McLean from the Department of Economics, both of whom I was fortunate to have had the opportunity to learn under their tutelage. Additionally, I would like to acknowledge my high-school mathematics teacher, Micheal O'Boyle, who sparked the fire of curiosity for mathematics and the desire to pursue it further.

Lastly, I would like to thank all my family and friends, both those from before graduate school and those I met during my time at Rutgers. I am truly blessed to have such a collection of people in my life. You have all provided me with a needed escape from work and a means to simply enjoy life in your company.

Dedication

This dissertation is dedicated to the two people whose love and support has helped me develop into the person I am today, my parents. To my mother, Dorothea, from whom I inherited the abilities necessary to pursue a career in the field of statistics and see this endeavor through. To my father, Rudolph, who always encouraged me to do what made me happy. While words alone cannot describe all that you have done for me, I have only these two to express my appreciation, Thank You. -Andy

Table of Contents

A	ostra	ct		ii
A	Acknowledgements			
De	Dedication			
\mathbf{Li}	st of	Figure	es	ix
1.	Pre	limina	ry Material	1
	1.1.	Introd	uction	1
	1.2.	Spheri	cally and Elliptically Symmetric Distributions	2
	1.3.	Multiv	variate Estimation and Equivariance	5
	1.4.	The G	oal of this Dissertation	6
2.	The	Spatia	al Median	8
	2.1.	Introd	uction	8
	2.2.	The S	patial Median	9
	2.3.	Theore	etical Results	11
		2.3.1.	Efficiency and Relative Efficiency of Location Estimators \ldots	11
		2.3.2.	The Asymptotic Efficiency of the Spatial Median	12
		2.3.3.	The Oracle Spatial Median and Affine Spatial Median	13
		2.3.4.	Asymptotic Efficiency Calculations	16
	2.4.	Finite	Sample Performance	21
		2.4.1.	Finite Sample Theory	21
		2.4.2.	Results	25
			Bivariate Distributions	25
			Trivariate Distributions	31

			10 Dimensional Distributions	37
	2.5.	Appen	ıdix	42
		2.5.1.	The Influence Function and Asymptotic Variance-Covariance Ma-	
			trix of M-estimators of Location	42
		2.5.2.	The Asymptotic Variance-Covariance of the Spatial Median	43
		2.5.3.	The Asymptotic Variance-Covariance of the Oracle Spatial Median	48
		2.5.4.	Proof of Theorem 3.3.1	51
		2.5.5.	Simulation Results	64
3.	The	Spati	al Sign Covariance Matrix	70
	3.1.	Introd	uction	70
	3.2.	Tyler's	s Scatter Matrix	72
	3.3.	Spatia	l Sign Covariance Matrix	73
	3.4.	Theore	etical Results	73
		3.4.1.	Optimality of the Tyler Matrix under the Elliptical Model	73
		3.4.2.	Asymptotic Calculations	77
	3.5.	Finite	Sample Performance	82
		3.5.1.	Finite Sample Theory	82
		3.5.2.	Results	87
	3.6.	Appen	ıdix	89
		3.6.1.	Asymptotic Distribution of Eigenprojections and Eigenvectors of	
			a Scatter Estimate	89
		3.6.2.	Tyler's Scatter Matrix	94
		3.6.3.	The Spatial Sign Covariance Matrix	101
		3.6.4.	Asymptotic Calculations	108
4.	Con	clusio	n	114
	4.1.	Implic	ations	114
	4.2.	Future	e Work	115
Re	References			

Vita	120
-----------------------	-----

List of Figures

2.1.	Asymptotic Relative Efficiencies of the Spatial Median to the Oracle	
	Spatial Median in \mathbb{R}^2	18
2.2.	Asymptotic Relative Efficiencies of the Spatial Median to the Oracle	
	Spatial Median in \mathbb{R}^3 : $\Lambda_0 = diag(1, 1, r^2)$	19
2.3.	Asymptotic Relative Efficiencies of the Spatial Median to the Oracle	
	Spatial Median in \mathbb{R}^3 : $\Lambda_0 = diag(1, r^2, r^2)$	20
2.4.	Asymptotic Relative Efficiencies of the Spatial Median to the Oracle	
	Spatial Median in \mathbb{R}^{10}	21
2.5.	Finite Sample Efficiencies of Affine Spatial Medians to the Oracle Spatial	
	Median for the Bivariate Spherical Normal	26
2.6.	Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial	
	Median for the Bivariate Normal	27
2.7.	Finite Sample Efficiencies of Affine Spatial Medians to the Oracle Spatial	
	Median for the Bivariate Spherical t_3	28
2.8.	Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial	
	Median for the Bivariate t_3	29
2.9.	Finite Sample Efficiencies of the Affine Spatial Median using Dümbgen's	
	matrix to the Oracle Spatial Median for the Bivariate Spherical Cauchy	
	and Slash	30
2.10	. Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial	
	Median for the Bivariate Cauchy	32
2.11	. Finite Sample Efficiencies of Affine Spatial Medians to the Oracle Spatial	
	Median for the Trivariate Spherical Normal and Cauchy	33

2.12	. Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial	
	Median for the Trivariate Normal and Cauchy - I	34
2.13	. Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial	
	Median for the Trivariate Normal and Cauchy - II	35
2.14	. Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial	
	Median for the Trivariate Normal and Cauchy - III	36
2.15	. Finite Sample Efficiencies of Affine Spatial Medians to the Oracle Spatial	
	Median for the Spherical Normal and Cauchy in \mathbb{R}^{10}	37
2.16	. Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial	
	Median for the Normal and Cauchy in \mathbb{R}^{10} - I \ldots	39
2.17	. Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial	
	Median for the Normal and Cauchy in \mathbb{R}^{10} - II $\ldots \ldots \ldots \ldots$	40
2.18	. Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial	
	Median for the Normal and Cauchy in \mathbb{R}^{10} - III	41
3.1.	Asymptotic Relative Efficiencies of the Eigenprojection Estimate of the	
	SSCM to the Corresponding Eigenprojection Estimate of the Tyler Ma-	
	trix in \mathbb{R}^2	79
3.2.	Asymptotic Relative Efficiencies of the Eigenprojection Estimate of the	
	SSCM to the Corresponding Eigenprojection Estimate of the Tyler Ma-	
	trix in \mathbb{R}^3	80
3.3.	Asymptotic Relative Efficiencies of the Eigenprojection Estimate of the	
	SSCM to the Corresponding Eigenprojection Estimate of the Tyler Ma-	
	trix in \mathbb{R}^{10}	81
3.4.	Finite Sample Relative Efficiencies of the Eigenprojection Estimate of	
	the SSCM to the Corresponding Eigenprojection Estimate of the Tyler	
	Matrix in \mathbb{R}^2	87
3.5.	Finite Sample Relative Efficiencies of the Eigenprojection Estimate of	
	the SSCM to the Corresponding Eigenprojection Estimate of the Tyler	
	Matrix in \mathbb{R}^3	88

3.6.	Finite Sample Relative Efficiencies of the Eigenprojection Estimate of	
	the SSCM to the Corresponding Eigenprojection Estimate of the Tyler	
	Matrix in \mathbb{R}^5	90

Chapter 1

Preliminary Material

1.1 Introduction

Many procedures in multivariate statistics make the assumption that the data arises as realizations from some multivariate normal distribution. A multivariate normal distribution is entirely specified by two parameters, its population mean vector and population covariance matrix, commonly symbolized by the Greek letters μ and Σ respectively. Most multivariate methods involve inference on one or both of these parameters with either one or both being assumed unknown. Thus it is often necessary to obtain estimates of these parameters from the observed data. The most popularly used estimators of μ and Σ are the sample mean vector and sample covariance matrix. Unfortunately, when the data deviates from the assumption of multivariate normality the reliability of these estimators, and any method that relies on them, is compromised. For instance, the presence of outliers in the data could have a significant impact on these estimates; extreme observations tend to pull the sample mean vector away from μ and towards themselves. Their effect on the sample covariance matrix is double in that they effect the sample mean vector used in calculating the sample covariance matrix as well as possibly biasing the estimated variation in their particular direction thus not giving an accurate estimate of Σ .

One attempt to address these short-comings is to implement methods from nonparametric statistics. Non-parametric procedures make minimal or no assumptions about the underlying distribution from which the data arises. Consequently, their performance does not depend on the underlying distribution hence will still produces reliable results. However, when the data does indeed come from a multivariate normal distribution, procedures involving estimation of μ and or Σ with the sample mean and sample covariance matrix often drastically out-perform any non-parametric procedure. Recall that in this situation, the sample mean vector and sample covariance matrix are functions of the complete, sufficient statistics thus contain the most information about the parameters μ and Σ . For this reason, alternative approaches have been taken to develop reliable procedures without sacrificing performance; robust statistics is one such instance. As opposed to assuming the data is from one particular distribution or discarding distributional assumptions altogether, robust statistics considers methods developed for a family of distributions. Because of this, these methods will produce reliable results and perform well for all distributions within the class they were developed for.

In robust multivariate statistics, considerable attention is given to the study of procedures under the assumption the data comes from some elliptically symmetric distribution (or elliptical distribution). This family of distributions is one generalization of the multivariate normal distribution. The characterization of an elliptical distribution involves specifying parameters that act as generalizations of μ and Σ ; these are the location vector and scatter matrix respectively and are represented with the same symbols. Similar to before, one or both of these parameters is assumed unknown and must be estimated from the data. Several estimators of location and scatter have been proposed to estimate the location vector and scatter matrix. Perhaps the most natural are the maximum likelihood estimators (MLE's) for μ and Σ assuming the data comes from a particular elliptical distribution. As a extension of these estimators, Maronna [23] developed M-estimators of location and scatter. Independently, Huber [14] developed an even broader class of estimators using arguments based on the concept of affine equivariance.

1.2 Spherically and Elliptically Symmetric Distributions

Elliptically symmetric distributions have played a central role in the development of robust multivariate statistics by serving as alternatives to the multivariate normal distribution with which to study the robustness properties of multivariate methods. Elliptical distributions arise by taking affine transformations of spherically symmetric distributions. A multivariate random vector, $\mathbf{z} \in \mathbb{R}^d$, is said to be spherically symmetric (or spherical) about the origin, $\mathbf{0}$, if $\mathbf{z} \sim \mathbf{Q}\mathbf{z}$ for any $d \times d$ orthogonal matrix \mathbf{Q} . The distribution of a spherically symmetric random variable will be denoted by G. If the measure induced by G is absolutely continuous with respect to Lebesgue measure in \mathbb{R}^d , then there exists a probability density function for the random vector \mathbf{z} of the form

$$C_{g,d}g\left(\|\mathbf{z}\|_{2}^{2}\right) = C_{g,d}g\left(\mathbf{z}^{t}\mathbf{z}\right)$$

where g is a fixed Lebesgue integrable function in \mathbb{R} and $C_{g,d}$ is a constant depending on both g and d that ensures that the expression is indeed a density (i.e. integrates to 1) in \mathbb{R}^d ; that is

$$C_{g,d} = \left(\int_{\mathbb{R}^d} g\left(\mathbf{z}^t \mathbf{z}\right) d\mathbf{z}\right)^{-1}$$

In this form, the motivation of the name spherically symmetric is obvious in that the contours of equal density are concentric hyper-spheres in \mathbb{R}^d . The vector **0** serves not only as the point of symmetry for the concentric hyper-spheres, but is also the mean vector if the distribution has finite first moments. If in addition the distribution has finite second moments, the covariance matrix of the distribution is proportional to the $d \times d$ identity matrix \mathbf{I}_d .

If the random vector $\mathbf{z} \in \mathbb{R}^d$ is spherically symmetric, then for any $d \times d$ nonsingular matrix, \mathbf{M} , and vector $\boldsymbol{\mu} \in \mathbb{R}^d$, the vector $\mathbf{x} = \mathbf{M}\mathbf{z} + \boldsymbol{\mu}$ is said to be elliptically distributed in *d*-dimensions with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma} = \mathbf{M}\mathbf{M}^t$. When a random vector $\mathbf{x} \in \mathbb{R}^d$ has an elliptical distribution it will be notated $\mathbf{x} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$; this distribution will be referred to as *F*. If the measure induced by *F* is absolutely continuous with respect to Lebesgue measure in \mathbb{R}^d , then \mathbf{x} has a density, denoted *f*, that is given by

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, g) = C_{g,d} det(\boldsymbol{\Sigma})^{-1/2} g\left(\|\mathbf{x} - \boldsymbol{\mu}\|_{\boldsymbol{\Sigma}^{-1}}^{2}\right)$$
$$= C_{g,d} det(\boldsymbol{\Sigma})^{-1/2} g\left((\mathbf{x} - \boldsymbol{\mu})^{t} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

where $\|\mathbf{x}\|_{\mathbf{A}} = \sqrt{\mathbf{x}^t \mathbf{A} \mathbf{x}}$. Analogously, the name elliptically symmetric originates from the fact the contours of equal density are concentric hyper-ellipsoids in \mathbb{R}^d , namely the hyper-ellipsoid given by the equation $(\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = c$. Similarly, the parameter $\boldsymbol{\mu}$ is the location vector and corresponds to the center of symmetry of the hyper-ellipsoids and also the mean vector when the distribution has finite first moments. The parameter $\boldsymbol{\Sigma}$ is referred to as the scatter matrix since it determines the spread and orientation of the concentric hyper-ellipsoids. In general, the distribution need not have finite second moments, but in cases where it does $\boldsymbol{\Sigma}$ is also referred to as the pseudocovariance matrix since the covariance matrix of the distribution is then proportional to $\boldsymbol{\Sigma}$. Note, in general the scatter matrix is not well-defined within the class of elliptical distributions. For a given elliptical distribution, the function g(s) can be replaced with the function $g_c(s) = c^d g(cs)$. It is possible to impose restrictions on the function g(s)to eliminate any ambiguities (such as take the constant c such that $g_c(s)$ has covariance matrix given by \mathbf{I}_d), however, will not be necessary for aims of this dissertation.

Necessarily, a scatter matrix is symmetric and positive definite. Recall from linear algebra that any symmetric, positive definite matrix has a spectral decomposition unique up to multiplication of the eigenvectors by ± 1 . For the matrix Σ , let $\lambda_1^2 \ge \lambda_2^2 \ge \cdots \ge \lambda_d^2$ denote its eigenvalues and $\mathbf{p}_1, \ldots, \mathbf{p}_d$ denote an orthonormal set of eigenvectors with \mathbf{p}_i belonging to the eigenspace corresponding to λ_i^2 . Let the eigenprojection associated with the eigenvalue λ_i^2 be denoted \mathbf{P}_i . Recall this is an projection matrix in to the eigenspace corresponding to the eigenvalue λ_i^2 , denoted \mathcal{P}_i . For repeated eigenvalues (i.e. an eigenvalue with algebraic multiplicity greater than 1), then the dimension of the corresponding eigenspace (the geometric multiplicity) could also be greater than 1. Suppose $\lambda_j^2 = \lambda_{j+1}^2 = \cdots = \lambda_{j+k-1}^2 = \lambda^2$. In this is instance the notation \mathbf{P}_{λ^2} and ${\cal P}_{\lambda^2}$ will be used to denote the corresponding eigenprojection and eigenspace associated with the eigenvalue λ^2 . Any pair of orthonormal vectors that span \mathcal{P}_{λ^2} could be taken for $\mathbf{p}_j, \ldots, \mathbf{p}_{j+k-1}$. If a given eigenvalue, say λ_{ι}^2 , has geometric multiplicity 1, then the eigenvector \mathbf{p}_{ι} is uniquely defined up to multiplication by ± 1 . Defining the matrices $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_d]$ and $\mathbf{\Lambda} = diag(\lambda_1^2, \dots, \lambda_d^2)$, then one can write $\mathbf{\Sigma} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^t$. Using the spectral decomposition, it is possible to define the unique, symmetric, positive definite

square root of Σ to be $\Sigma^{1/2} = \mathbf{P} \mathbf{\Lambda}^{1/2} \mathbf{P}^t$, where $\mathbf{\Lambda}^{1/2} = diag(|\lambda_1|, \dots, |\lambda_d|)$. There is a useful representation for any spherical distribution that will be utilized on several occasions in the proceeding chapters. Every spherical distribution in \mathbb{R}^d has a stochastic representation of the form $\mathbf{z} \sim R_G \mathbf{u}_d$ with R_G and \mathbf{u}_d independent. R_G is a nonnegative random variable referred to as the radial component of \mathbf{z} and \mathbf{u}_d is a random vector that is uniformly distributed on S_d , the unit hyper-sphere in \mathbb{R}^d . This implies that if $\mathbf{x} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$, then it has stochastic representation $\mathbf{x} \sim R_G \mathbf{M} \mathbf{u}_d + \boldsymbol{\mu}$, where \mathbf{M} is any matrix such that $\mathbf{M}\mathbf{M}^t = \boldsymbol{\Sigma}$. One such choice is $\mathbf{M} = \boldsymbol{\Sigma}^{1/2}$. If $\mathbf{x}_1, \dots, \mathbf{x}_n$ is an *i.i.d.* sample from some elliptical distribution $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$, this will be referred to as the elliptical model.

1.3 Multivariate Estimation and Equivariance

Intimately connected with elliptical distributions is the concept of affine equivariance. The location vector and scatter matrix for elliptical distributions are affinely equivariant; that is if $\mathbf{x} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$, then the random vector $\mathbf{x}^* = \mathbf{A}\mathbf{x} + \mathbf{b}$ will also be elliptically distributed with $\mathbf{x}^* \sim \mathcal{E}_d(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*; G)$ where $\boldsymbol{\mu}^* = \mathbf{A}\boldsymbol{\mu} + \mathbf{b}$ and $\boldsymbol{\Sigma}^* = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^t$ for any non-singular, $d \times d$ matrix \mathbf{A} and $\mathbf{b} \in \mathbb{R}^d$. Because of this property, when it is assumed that data arises from an elliptical distribution, it is natural to consider estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ that possess the property of affine equivariance. That is, if $\mathbf{x}_1, \ldots, \mathbf{x}_n$ yields estimates of the location vector and scatter matrix $\hat{\boldsymbol{\mu}}_n$ and $\hat{\boldsymbol{\Sigma}}_n$ respectively, then the estimators obtained for the transformed data $\mathbf{x}_i^* = \mathbf{A}\mathbf{x}_i + \mathbf{b}$, $i = 1, \ldots, n$, will be $\hat{\boldsymbol{\mu}}_n^* = \mathbf{A}\hat{\boldsymbol{\mu}}_n + \mathbf{b}$ and $\hat{\boldsymbol{\Sigma}}_n^* = \mathbf{A}\hat{\boldsymbol{\Sigma}}_n\mathbf{A}^t$.

M-estimators are examples of affinely equivariant estimators of location and scatter. However, in [23] the author showed the breakdown points of affinely equivariant M-estimators is at most 1/(d+1). Consequently, much work has been on the development of high breakdown affinely equivariant estimates such as the minimum volume ellipsoid estimate (MVE) and minimum covariance determinant estimate (MCD) [31], S-estimates ([9] & [18]), projection based estimates ([10], [25] & [37]), CM-estimates [16], MM-estimates ([32] & [37]), τ -estimates [19], and one-step versions of these estimates [20]. Unfortunately, high breakdown affinely equivariant estimators tend to be computationally intensive, especially for large d and n; current algorithms are only approximate and probabilistic in nature. Another pitfall of affinely equivariant estimators is that when n < d, any affinely equivariant estimate of location and scatter reduce to the sample mean vector and sample covariance matrix respectively, the latter being singular in this situation [38]. Consequently, these short-comings have led to the development of methods that discard the property of affine equivariance.

To the goal of the previous paragraph, one such approach has been to develop estimators that are only orthogonally equivariant. Estimators of the location vector and scatter matrix are said to be orthogonally equivariant if the data $\mathbf{x}_1, \ldots, \mathbf{x}_n$ yield estimates $\hat{\boldsymbol{\mu}}_n$ and $\hat{\boldsymbol{\Sigma}}_n$ respectively, then the estimates obtained for the transformed data $\mathbf{x}_i^* = \mathbf{Q}\mathbf{x}_i + \mathbf{b}, \ i = 1, \ldots, n$ will be $\hat{\boldsymbol{\mu}}_n^* = \mathbf{Q}\hat{\boldsymbol{\mu}}_n + \mathbf{b}$ and $\hat{\boldsymbol{\Sigma}}_n^* = \mathbf{Q}\hat{\boldsymbol{\Sigma}}_n\mathbf{Q}^t$ for any $d \times d$, orthogonal matrix \mathbf{Q} and $\mathbf{b} \in \mathbb{R}^d$. Orthogonal transformations are special cases of affine transformations, thus an analogous result holds for the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ under the class of elliptical distributions.

When considering the class of elliptical distributions, a benefit of using affinely equivariant estimators of location and scatter is that the form of the influence function can be derived by just considering the spherical case. Furthermore, the efficiencies of such estimators does not depend on either μ or Σ . These properties do not carry over to estimators that lack the property of affine equivariance. On the contrary, under elliptical models it has been observed that non-affinely equivariant estimators perform better under certain scatter structure than for others. Unfortunately, this fact is usually ignored when deciding which estimator to use. Using these estimators is rather Procrustean in that they are favoring certain scatter structures over others; to a degree this is letting the method determine the model.

1.4 The Goal of this Dissertation

As mentioned in section 1.4, the performance of non-affinely equivariant estimators under the elliptical model is dependent on Σ . One popular method for evaluating the performance of as estimator is to study the variability with which the estimator measure

the parameter of interest. While the nature of the parameter dictates the criterion of interest, it usually involves the variances or variance-covariance matrix of the estimators being studied. Naturally, the evaluation of the performance necessitates some sort of benchmark or alternative method that achieves the same goal, thus usually estimators are studied in reference to competing estimators. The evaluation of estimators via the prior paradigm is the basis of a notion called efficiency. This dissertation will study the efficiencies of two popularly used orthogonally equivariant estimators under elliptical models and their dependencies on Σ . The two considered are the estimators of location and scatter, the spatial median and the spatial sign covariance matrix (SSCM) respectively. These will be addressed in separate parts; the spatial median being discussed in Chapter II whereas the SSCM in Chapter III. The concept of efficiency for location and scatter estimates will be defined more explicitly in the subsequent sections. The efficiencies will be considered in both the asymptotic and finite sample cases. For both estimators, under elliptical models it will be shown that these estimators are asymptotically most efficient when the underlying scatter structure of the data is spherically symmetric. Simulation results will be presented in the finite sample case to support an analogous hypothesis. Technical details of proofs omitted from the body of the dissertation will be presented in two separate appendices at the ends of chapters 2 and 3. The first part of chapter 4 will discuss the implications of the above findings in robust Principal Component Analysis (PCA), a commonly used dimension reduction technique with broad applications ([7], [8], [12], [17] & [22]). The dissertation will conclude with a discussion of future research directions.

Chapter 2

The Spatial Median

2.1 Introduction

For estimating the location parameter of multivariate data, the spatial median is a commonly used robust alternative to the sample mean vector when it is believed that the data being analyzed either contains outlying observations or comes from a distribution that is not multivariate normal. Since the estimation of the spatial median does not require an estimate of the scatter matrix in its calculation, the spatial median has the added benefit that it exists even when the sample size is less than dimension thus making it a popular estimator of the location parameter for sparse data. However, unlike the sample mean vector, the spatial median is not affinely equivariant but only equivariant under translations, rescaling and orthogonal transformations. Because of this property, the spatial median is commonly used in orthogonally equivariant multivariate methods that require estimation of a location parameter as an intermediary such as principle component analysis (PCA). The reason the spatial median lacks the property of affine equivariance is that in its calculation it down-weighs observations in terms of their Euclidean distances as opposed to their Mahalanobis distance from the estimated center of the data. Thus in the the setting where the data is assumed to arise from some elliptical distribution, one might conjecture that the spatial median is less efficient in situations where the distribution is not spherical.

This chapter is broken down into the following sections. In section 2, the spatial median is discussed and it is shown that the estimating equation for it is simply the MLE for μ when the distribution for the elliptical model is a spherical Laplace (Double Exponential) distribution. In section 3, the main theoretical results are presented. The first subsection discusses the concept of relative efficiency and efficiency of a vector

estimator, both for finite samples and asymptotically. It is then shown that for the class of elliptically symmetric distributions, the spatial median is asymptotically most efficient when $\Sigma \propto \mathbf{I}_d$, that is the distribution is in fact spherically symmetric. In addition, it is possible for one to construct an affinely equivariant version of the spatial median that has the same asymptotic distribution as the spatial median at spherical symmetry but is asymptotically more efficient than the spatial median for all nonspherical elliptical distributions. Lastly, some calculations are presented to demonstrate the severity of the asymptotic inefficiencies. Section 4 carries out a simulation study of the efficiencies of the spatial median for finite samples. The first subsection contains theoretical results needed to carry out the simulations and describes how they were implemented. The second subsection contains the results and discussion.

2.2 The Spatial Median

Given a multivariate data set, $\mathbf{x}_1, \ldots, \mathbf{x}_n$, in \mathbb{R}^d , the spatial median is the vector $\tilde{\boldsymbol{\mu}}_{SM}$ that satisfies the following objective function

$$\widetilde{\boldsymbol{\mu}}_{SM} = \operatorname{argmin}_{\boldsymbol{\eta} \in \mathbb{R}^d} \sum_{i=1}^n \|\mathbf{x}_i - \boldsymbol{\eta}\|_2$$
(2.2.1)

Recall in the univariate case, d = 1, above reduces to an expression whose solution is given by the sample median of the dataset. Consequently, the spatial median can be thought of as one possible generalization of the median to the multivariate case. Perhaps the earliest reference of the spatial median was in [42]). In the literature, the spatial median is also referred to as the median centre or L_1 median since the minimization of the objective function involves minimizing the sum of the L_1 norms to the observations [13]. Brown [5] studied the asymptotic efficiency of the spatial median for the bivariate normal distribution as well as for standard multivariate normal distributions in dimensions greater than 2. The objective function in (2.2.1) has no explicit solution, however, it was shown to be uniquely defined when d > 2 ([15] & [27]). The spatial median is a special case of a monotonic M-estimate of multivariate location and thus can be computed via a simple re-iterated least squares (IRLS) algorithm [43]. In [39], the authors proposed a useful modification to improve the prior algorithm. A summary of the spatial median can be found in [28]. As mentioned in the introduction, the spatial median is not affinely equivariant, but only equivariant under orthogonal transformations, rescaling as well as translations. That is, for any $\mathbf{b} \in \mathbb{R}^d$, $c \in \mathbb{R}$ and $d \times d$ orthogonal matrix \mathbf{Q} , if the data is transformed as $\mathbf{x}_i^* = c\mathbf{Q}\mathbf{x}_i + \mathbf{b}$ for $i = 1, \ldots, n$, then the spatial median transforms as $\tilde{\boldsymbol{\mu}}_{SM}^* = c\mathbf{Q}\tilde{\boldsymbol{\mu}}_{SM} + \mathbf{b}$.

Referring back to the theory of maximum likelihood estimation, the sample median is the MLE of the location parameter when the data comes from a Laplace distribution. Analogously, the spatial median arises as the MLE when the data comes from an elliptical Laplace distribution with $\Sigma \propto \mathbf{I}_d$. Recall the density function of the multivariate Laplace distribution with location vector $\boldsymbol{\mu}$ and scatter matrix $\boldsymbol{\Sigma}$ is given by

$$f(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},g_L) = \frac{2\pi^{d/2}\Gamma(d)}{\Gamma(d/2)}det(\boldsymbol{\Sigma})^{-1/2}exp\left(-\sqrt{(\mathbf{x}-\boldsymbol{\mu})^t\,\boldsymbol{\Sigma}^{-1}\,(\mathbf{x}-\boldsymbol{\mu})}\right)$$

Given n observations, the likelihood function is then given by

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{X}, g_L) = \prod_{i=1}^n \frac{2\pi^{d/2} \Gamma(d)}{\Gamma(d/2)} det(\boldsymbol{\Sigma})^{-1/2} exp\left(-\sqrt{(\mathbf{x}_i - \boldsymbol{\mu})^t \, \boldsymbol{\Sigma}^{-1} \, (\mathbf{x}_i - \boldsymbol{\mu})}\right)$$
$$= \left(\frac{2\pi^{d/2} \Gamma(d)}{\Gamma(d/2)}\right)^n det(\boldsymbol{\Sigma})^{-1/2} exp\left(-\sum_{i=1}^n \sqrt{(\mathbf{x}_i - \boldsymbol{\mu})^t \, \boldsymbol{\Sigma}^{-1} \, (\mathbf{x}_i - \boldsymbol{\mu})}\right)$$

If Σ were known a priori, then maximizing the above likelihood entails maximizing the argument in the exponential, which is the same as minimizing the sum. Hence the the MLE for μ is

$$\widetilde{\mu}_{\Sigma} = argmin_{\eta \in \mathbb{R}^d} \sum_{i=1}^n \|\mathbf{x}_i - \eta\|_{\Sigma^{-1}}$$

If $\Sigma = \sigma^2 \mathbf{I}_d$, then the estimating equation for $\tilde{\mu}_{\sigma^2 \mathbf{I}_d}$ would yield the same solution as equation (2.2.1), that is $\tilde{\mu}_{\sigma^2 \mathbf{I}_d} = \tilde{\mu}_{SM}$. This characterization will be utilized in establishing the theoretical results to follow.

2.3 Theoretical Results

2.3.1 Efficiency and Relative Efficiency of Location Estimators

When sampling from elliptical distributions, it was conjectured in the introduction that the spatial median is asymptotically most efficient when Σ is proportional to the identity matrix. The spatial median is intended to give an estimate of the location vector, thus before proving the aforementioned result it is necessary to discuss the notions of efficiency and relative efficiency for vector estimators both in the finite sample case and asymptotically.

Let $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ be two different unbiased estimators of the vector parameter θ based on samples of size n. Let $\mathbf{V}_{\widehat{\theta}_n} = Var\left[\widehat{\theta}_n\right]$ and $\mathbf{V}_{\widetilde{\theta}_n} = Var\left[\widetilde{\theta}_n\right]$, assuming $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ have finite second moments. In this situation, comparing the estimators $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ reduces to a comparison of how they estimate linear combinations of the parameter being estimated, that is $\mathbf{a}^t \theta$ for some $\mathbf{a} \in \mathbb{R}^d$. The natural estimators for $\mathbf{a}^t \theta$ are $\mathbf{a}^t \widehat{\theta}_n$ and $\mathbf{a}^t \widetilde{\theta}_n$ with variances given by $Var\left[\mathbf{a}^t \widehat{\theta}_n\right] = \mathbf{a}^t \mathbf{V}_{\widehat{\theta}_n} \mathbf{a}$ and $Var\left[\mathbf{a}^t \widetilde{\theta}_n\right] = \mathbf{a}^t \mathbf{V}_{\widehat{\theta}_n} \mathbf{a}$ respectively. Since $\mathbf{a}^t \theta$ is a univariate parameter, convention is to focus on the ratio of variances, $\mathbf{a}^t \mathbf{V}_{\widehat{\theta}_n} \mathbf{a} / \mathbf{a}^t \mathbf{V}_{\widehat{\theta}_n} \mathbf{a}$. Comparison of $\widehat{\theta}_n$ to $\widetilde{\theta}_n$ involves locating the vector \mathbf{a} such that ratio $\mathbf{a}^t \mathbf{V}_{\widehat{\theta}_n} \mathbf{a} / \mathbf{a}^t \mathbf{V}_{\widehat{\theta}_n} \mathbf{a}$ is maximal or minimal. Results from linear algebra gives that the value of the ratio at its minimum/maximum is the same as the smallest/largest eigenvalue of the matrix $\mathbf{V}_{\widehat{\theta}_n}^{-1} \mathbf{V}_{\widehat{\theta}_n}$ with the vector \mathbf{a} giving the minimum/maximum being the corresponding eigenvector of the aforementioned matrix.

Comparing efficiencies in the asymptotic sense is analogous to the finite sample case. For the asymptotic case, assume $\hat{\theta}_n$ and $\tilde{\theta}_n$ are \sqrt{n} consistent, that is

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{n}-\boldsymbol{\theta}\right)\rightarrow_{\mathcal{D}}Norm_{d}\left(\boldsymbol{0},AV_{\widehat{\boldsymbol{\theta}}}\right),\ \sqrt{n}\left(\widetilde{\boldsymbol{\theta}}_{n}-\boldsymbol{\theta}\right)\rightarrow_{\mathcal{D}}Norm_{d}\left(\boldsymbol{0},AV_{\widetilde{\boldsymbol{\theta}}}\right)$$

where $AV_{\hat{\theta}}$ and $AV_{\hat{\theta}}$ are the asymptotic variance-covariance matrix of the estimators $\hat{\theta}_n$ and $\tilde{\theta}_n$ respectively. For any $\mathbf{a} \in \mathbb{R}^d$, it follows that

$$\sqrt{n} \left(\mathbf{a}^t \widehat{\boldsymbol{\theta}}_n - \mathbf{a}^t \boldsymbol{\theta} \right) \to_{\mathcal{D}} Norm \left(0, \mathbf{a}^t A V_{\widehat{\boldsymbol{\theta}}} \mathbf{a} \right), \ \sqrt{n} \left(\mathbf{a}^t \widetilde{\boldsymbol{\theta}}_n - \mathbf{a}^t \boldsymbol{\theta} \right) \to_{\mathcal{D}} Norm \left(0, \mathbf{a}^t A V_{\widetilde{\boldsymbol{\theta}}} \mathbf{a} \right)$$

Again the focus will be on the ratio $\mathbf{a}^t A V_{\widehat{\theta}} \mathbf{a} / \mathbf{a}^t A V_{\widehat{\theta}} \mathbf{a}$ for $\mathbf{a} \in \mathbb{R}^d$. This value of this ratio at its minimum/maximum equals the smallest/largest eigenvalue of the matrix $AV_{\widehat{\theta}}^{-1}AV_{\widehat{\theta}}$ with the vector \mathbf{a} being the corresponding eigenvectors of the aforementioned matrix.

Provided the distribution is such that the asymptotic version of the multivariate information inequality holds, the MLE of $\boldsymbol{\theta}$ is the best one can do, hence one is often interested in the asymptotic efficiency of an estimator relative to the MLE. The asymptotic variance-covariance matrix of the MLE of $\boldsymbol{\theta}$ is given by the inverse of the Fisher information matrix, denoted $\mathcal{I}(\boldsymbol{\theta})$. Thus one has $AV_{\hat{\boldsymbol{\theta}}} \geq \mathcal{I}^{-1}(\boldsymbol{\theta})$ where \geq refers to the usual partial ordering of symmetric matrices. Recall for two symmetric matrices, \mathbf{A} and $\mathbf{B}, \geq (>)$ is a partial ordering of these matrices such that $\mathbf{A} \geq \mathbf{B} \ (\mathbf{A} > \mathbf{B})$ if and only if $\mathbf{A} - \mathbf{B}$ is positive semi-definite (definite). The asymptotic efficiency of an estimator is defined to be

$$AE\left(\widehat{\boldsymbol{\theta}}\right) = max_{\mathbf{a}\in\mathbb{R}^{d}} \frac{\mathbf{a}^{t}\boldsymbol{\mathcal{I}}^{-1}\left(\boldsymbol{\theta}\right)\mathbf{a}}{\mathbf{a}^{t}AV_{\widehat{\boldsymbol{\theta}}}\mathbf{a}}$$

As mentioned above, the value of $AE\left(\widehat{\theta}\right)$ is given by the largest eigenvalue of the matrix $AV_{\widehat{\theta}}^{-1}\mathcal{I}^{-1}\left(\theta\right)$ with **a** being the corresponding eigenvector.

2.3.2 The Asymptotic Efficiency of the Spatial Median

Let $\mathbf{x} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$ be a random vector that is absolutely continuous with respect to Lebesgue measure. The Fisher information matrix for $\boldsymbol{\mu}$ for a fixed $\boldsymbol{\Sigma}$ and G is given by

$$\mathcal{I}(\boldsymbol{\mu}; \boldsymbol{\Sigma}, G) = \frac{1}{\alpha(G)} \boldsymbol{\Sigma}^{-1}$$

where $\alpha(G) = dE \left[u^2 \left(R_G^2 \right) \right]^{-1}$ is a scalar that depends only on G with u(s) = -2sg'(s)/g(s) [30]. Let $\hat{\mu}$ be any estimator of μ . By the asymptotic version of the information inequality one has that $AV_{\hat{\mu}} \geq \mathcal{I}^{-1}(\mu; \Sigma, G) = \alpha(G)\Sigma$. It is shown in section 2.5.2 of the appendix that the asymptotic variance-covariance matrix of the spatial median under the elliptical model is given below as

$$AV_{SM}(\mathbf{\Sigma}, G) = \beta(G) \mathbf{P} \mathcal{V}(\mathbf{\Lambda}) \mathbf{P}^{t}$$

where $\beta(G) = 1/E^2[1/R_G]$ and $\mathcal{V}(\mathbf{\Lambda}) = diag\left(\nu_1^2(\mathbf{\Lambda}), \dots, \nu_d^2(\mathbf{\Lambda})\right)$ with

$$\nu_i^2\left(\mathbf{\Lambda}\right) = \frac{E\left[\frac{\lambda_i^2 u_{d,i}^2}{\|\mathbf{\Lambda}^{1/2} \mathbf{u}_d\|_2^2}\right]}{E^2\left[\frac{\lambda_i^2 u_{d,i}^2}{\|\mathbf{\Lambda}^{1/2} \mathbf{u}_d\|_2^3} - \frac{1}{\|\mathbf{\Lambda}^{1/2} \mathbf{u}_d\|_2} \mathbf{I}_d\right]}$$

and $u_{d,i}$ representing the i^{th} component of the random unit vector \mathbf{u}_d .

For the family of elliptical distributions $\mathcal{E}_d(\mu, \Sigma; G)$ with fixed G further suppose that the following conditions are also satisfied

Conditions 3.2.1. (i) $E[1/R_G] > 0$ and (ii) g is bounded.

The following theorem formally states that for any family of elliptical distributions with fixed G that satisfies the condition given above, the asymptotic efficiency of the spatial median is uniquely maximized at a spherically symmetric member of this family,

Theorem 3.3.1. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represent an *i.i.d.* sample from $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$ that satisfies conditions 3.2.1. Then

$$AE\left(\widetilde{\boldsymbol{\mu}}_{SM};\sigma^{2}\mathbf{I}_{d},G\right) \geq AE\left(\widetilde{\boldsymbol{\mu}}_{SM};\boldsymbol{\Sigma},G\right)$$

with equality holding if and only if $\Sigma \propto I_d$.

Proof. See section 2.5.4 of the appendix.

2.3.3 The Oracle Spatial Median and Affine Spatial Median

As mentioned in the introduction to this chapter, the calculation of the spatial median involves down-weighing observations in terms of their Euclidean distances as opposed to their Mahalanobis distance from the estimated center of the data. Consequently, under the elliptical model one might surmise that an estimator that does the latter would work better in the case when $\Sigma \not\propto \mathbf{I}_d$. The following estimator does just the above. Consider the situation that Σ is known and not proportional to \mathbf{I}_d , define,

$$\widetilde{\mu}_{OSM} = \widetilde{\mu}_{\Sigma} = argmin_{\eta \in \mathbb{R}^d} \sum_{i=1}^n \|\mathbf{x}_i - \eta\|_{\Sigma^{-1}}$$

Note that this is the MLE for μ when sampling from an elliptical Laplace distribution when Σ is known; this follows from the likelihood function computed in section 2. Denote this estimator as the oracle spatial median since it requires knowing the scatter matrix of the data a priori. It is shown in section 2.5.3 of the appendix that the asymptotic variance-covariance matrix of the oracle spatial median is given by

$$AV_{\tilde{\boldsymbol{\mu}}_{OSM}}\left(\boldsymbol{\Sigma}, \boldsymbol{G}\right) = \frac{d}{\left(d-1\right)^2} \beta\left(\boldsymbol{G}\right) \boldsymbol{\Sigma}$$

As alluded to, one might suspect that for elliptical distributions, this estimator for μ is more efficient than the spatial median. This is stated in the following theorem,

Theorem 3.4.1. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represent an *i.i.d.* sample from $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$ that satisfies conditions 3.2.1. Then $AV_{\boldsymbol{\widetilde{\mu}}_{OSM}}(\boldsymbol{\Sigma}, G) < AV_{\boldsymbol{\widetilde{\mu}}_{SM}}(\boldsymbol{\Sigma}, G)$ unless $\boldsymbol{\Sigma} \propto \mathbf{I}_d$ in which case equality holds.

Proof. From the theory of maximum likelihood estimation, for the case when the particular distribution is the elliptical Laplace distribution, i.e. $G = G_L$, then $AV_{\tilde{\mu}_{OSM}}(\Sigma, G_L) < AV_{\tilde{\mu}_{SM}}(\Sigma, G_L)$ unless $\tilde{\mu}_{OSM}$ and $\tilde{\mu}_{SM}$ are asymptotically equivalent. However in the proof of Theorem 3.3.1, when $\Sigma \not\propto \mathbf{I}_d$ then $AV_{\tilde{\mu}_{SM}}(\Sigma, G) \not\ll \Sigma$ whereas $AV_{\tilde{\mu}_{OSM}}(\Sigma, G_L)$ is. Thus $\tilde{\mu}_{ASM}$ and $\tilde{\mu}_{SM}$ have different asymptotic variancecovariance matrices in this instance, thus strict inequality holds when $\Sigma \not\propto \mathbf{I}_d$. It follows that

$$AV_{\widetilde{\boldsymbol{\mu}}_{OSM}}\left(\boldsymbol{\Sigma}, G_{L}\right) = \frac{d}{\left(d-1\right)^{2}}\beta\left(G_{L}\right)\boldsymbol{\Sigma} < \beta\left(G_{L}\right)\mathbf{P}\boldsymbol{\mathcal{V}}\left(\boldsymbol{\Lambda}\right)\mathbf{P}^{t} = AV_{\widetilde{\boldsymbol{\mu}}_{SM}}\left(\boldsymbol{\Sigma}, G_{L}\right)$$

Canceling $\beta(G_L)$ from both sides gives then multiplying by $\beta(G)$ gives

$$AV_{\widetilde{\boldsymbol{\mu}}_{OSM}}\left(\boldsymbol{\Sigma}, \boldsymbol{G}\right) = \frac{d}{\left(d-1\right)^{2}}\beta\left(\boldsymbol{G}\right)\boldsymbol{\Sigma} < \beta\left(\boldsymbol{G}\right)\mathbf{P}\boldsymbol{\mathcal{V}}\left(\boldsymbol{\Lambda}\right)\mathbf{P}^{t} = AV_{\widetilde{\boldsymbol{\mu}}_{SM}}\left(\boldsymbol{\Sigma}, \boldsymbol{G}\right)$$

the desired result. When $\Sigma \propto \mathbf{I}_d$, then $\tilde{\mu}_{OSM} = \tilde{\mu}_{SM}$, thus the two estimators will have the same asymptotic variance-covariance matrix, hence equality holds. **QED**

In practice Σ is not known, however, can be estimated from the data. Let $\hat{\Sigma}_n$ be an affinely equivariant estimate of Σ based on a sample of size n. Replacing Σ with $\hat{\Sigma}_n$ in the definition of the oracle spatial median yields an affinely equivariant estimate of the location parameter, that is

$$\widetilde{\boldsymbol{\mu}}_{ASM} = \widetilde{\boldsymbol{\mu}}_{\widehat{\boldsymbol{\Sigma}}_n} = argmin_{\boldsymbol{\eta} \in \mathbb{R}^d} \sum_{i=1}^n \|\mathbf{x}_i - \boldsymbol{\eta}\|_{\widehat{\boldsymbol{\Sigma}}_n^{-1}}$$

Refer to this estimator as an affinely equivariant spatial median based on the scatter estimate $\hat{\Sigma}_n$, or simply an affine spatial median. If the elliptical distribution under consideration is such that

$$\widehat{\Sigma}_n = \Sigma + O_P\left(n^{1/2}\right) \tag{2.3.1}$$

then it follows by theorem 3 of [29] that $\sqrt{n} \left(\widetilde{\boldsymbol{\mu}}_{\widehat{\boldsymbol{\Sigma}}_n} - \widetilde{\boldsymbol{\mu}}_{OSM} \right) \rightarrow_P \mathbf{0}$, thus

$$\sqrt{n}\left(\widetilde{\boldsymbol{\mu}}_{\widehat{\boldsymbol{\Sigma}}_{n}}-\boldsymbol{\mu}\right) \rightarrow_{\mathcal{D}} Norm_{d}\left(\boldsymbol{0}, AV_{\widetilde{\boldsymbol{\mu}}_{\widehat{\boldsymbol{\Sigma}}}}\left(\boldsymbol{\Sigma}, G\right)\right)$$

where $AV_{\widetilde{\mu}_{\widehat{\Sigma}}}(\Sigma, G) = AV_{\widetilde{\mu}_{OSM}}(\Sigma, G).$

This leads to the following corollary to Theorem 3.4.1.

Corollary 3.4.1. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be a random sample from $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$ such that conditions 3.2.1 are satisfied. For any affinely equivariant estimate of scatter $\widehat{\boldsymbol{\Sigma}}_n$, such that $\widehat{\boldsymbol{\Sigma}}_n = \boldsymbol{\Sigma} + O_P(n^{1/2})$, it follows $AV_{\widehat{\boldsymbol{\mu}}_{\widehat{\boldsymbol{\Sigma}}}}(G) < AV_{\widehat{\boldsymbol{\mu}}_{SM}}(G, \boldsymbol{\Sigma})$ unless $\boldsymbol{\Sigma} \propto \mathbf{I}_d$ in which case equality holds. This corollary states that the spatial median is asymptotically inadmissible over the class of elliptical distributions. Note however for a particular estimator of scatter, this result holds true only when the conditions in equation (2.3.1) are met. For instance, if $\hat{\Sigma}_n$ is the sample covariance matrix, then this result holds only if the elliptical distribution has finite fourth moments. However, for a broad class of M-estimates of multivariate scatter defined in [23], the condition in equation (2.3.1) holds without any further assumptions on G.

2.3.4 Asymptotic Efficiency Calculations

In the previous section, it was mentioned that the oracle spatial median and affine spatial median are asymptotically equivalent provided that $\widehat{\Sigma}_n = \Sigma + O_P(n^{1/2})$. Furthermore, under the elliptical model it was shown that both of these were asymptotically more efficient at estimating the μ than the spatial median. To understand more precisely how inefficient the spatial median can be compared to either of the aforementioned estimators under the elliptical model, the asymptotic relative efficiency of the spatial median relative to the oracle spatial median will be computed under various dimensions and scatter structures. As will be seen, the values of the asymptotic efficiencies will not depend on the particular elliptical distribution.

The results in section 3.2 and 3.3 give,

$$AV_{\widetilde{\boldsymbol{\mu}}_{OSM}}\left(\boldsymbol{\Sigma},G\right) = \frac{d}{\left(d-1\right)^{2}}\beta\left(G\right)\boldsymbol{\Sigma} \text{ and } AV_{\widetilde{\boldsymbol{\mu}}_{SM}}\left(\boldsymbol{\Sigma},G\right) = \beta\left(G\right)\mathbf{P}\boldsymbol{\mathcal{V}}\left(\boldsymbol{\Lambda}\right)\mathbf{P}^{T}$$

with $\boldsymbol{\mathcal{V}}\left(\boldsymbol{\Lambda}\right)=diag\left(\nu_{1}^{2}\left(\boldsymbol{\Lambda}\right),\ldots,\nu_{d}^{2}\left(\boldsymbol{\Lambda}\right)\right)$ where

$$\nu_{i}^{2}\left(\mathbf{\Lambda}\right) = \frac{E\left[\frac{\lambda_{i}^{2}u_{d,i}^{2}}{\left\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\right\|_{2}^{2}}\right]}{E^{2}\left[\frac{\lambda_{i}^{2}u_{d,i}^{2}}{\left\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\right\|_{2}^{3}} - \frac{1}{\left\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\right\|_{2}}\mathbf{I}_{d}\right]}$$

Since $AV_{\tilde{\mu}_{SM}}(\Sigma, G)$ has the same eigenvectors as Σ , the asymptotic relative efficiency of $\tilde{\mu}_{SM}$ to $\tilde{\mu}_{OSM}$ reduces to a comparison of the eigenvalues of $AV_{\tilde{\mu}_{OSM}}(\Sigma, G)$ and $AV_{\tilde{\mu}_{SM}}(\Sigma, G)$ corresponding to the same eigenvectors. Thus without loss of generality one can reduce consideration to the simple case where $\mathbf{P} = \mathbf{I}_d$. As a consequence of this simplification there is a convenient interpretation of the aforementioned comparison of eigenvalues. For uncorrelated scatter structure, the eigenvalues of the asymptotic variance-covariance matrix correspond to the variances of the estimated components. Thus a comparison of the eigenvalues reduces to a comparison of the variability with which each estimator estimates the components of μ , i.e. the asymptotic relative efficiency for the components of the spatial median to the oracle spatial median. The asymptotic relative efficiency of each component can be computed and is equal to

$$ARE_{i}\left(\widetilde{\boldsymbol{\mu}}_{SM},\widetilde{\boldsymbol{\mu}}_{OSM};\boldsymbol{\Lambda},G\right) = \frac{\left[AV_{\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\Lambda}}}\left(\boldsymbol{\Lambda},G\right)\right]_{i}}{\left[AV_{\widetilde{\boldsymbol{\mu}}_{SM}}\left(\boldsymbol{\Lambda},G\right)\right]_{i}} = \frac{d}{\left(d-1\right)^{2}}\frac{\lambda_{i}^{2}}{\nu_{i}^{2}\left(\boldsymbol{\Lambda}\right)}$$

Note that for the expression $\nu_i^2(\mathbf{\Lambda})$, one has

$$\begin{split} \nu_{i}^{2}\left(c^{2}\mathbf{\Lambda}\right) &= \frac{E\left[\frac{c^{2}\lambda_{i}^{2}u_{d,i}^{2}}{\|c\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\|_{2}^{2}}\right]}{E^{2}\left[\frac{c^{2}\lambda_{i}^{2}u_{d,i}^{2}}{\|c\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\|_{2}^{3}} - \frac{1}{\|c\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\|_{2}}\mathbf{I}_{d}\right]} \\ &= \frac{E\left[\frac{c^{2}\lambda_{i}^{2}u_{d,i}^{2}}{c^{2}\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\|_{2}^{3}} - \frac{1}{c\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\|_{2}}\mathbf{I}_{d}\right]}{E^{2}\left[\frac{c^{2}\lambda_{i}^{2}u_{d,i}^{2}}{c^{3}\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\|_{2}^{3}} - \frac{1}{c\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\|_{2}}\mathbf{I}_{d}\right]} \\ &= \frac{E\left[\frac{\lambda_{i}^{2}u_{d,i}^{2}}{\left[\frac{\lambda_{i}^{2}u_{d,i}^{2}}{c\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\|_{2}^{3}} - \frac{1}{c\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\|_{2}}\mathbf{I}_{d}\right]}{E^{2}\left[\frac{\lambda_{i}^{2}u_{d,i}^{2}}{\left[\frac{\lambda_{i}^{2}u_{d,i}^{2}}{\left[\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\|_{2}^{3}} - \frac{1}{\left\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\|_{2}}\mathbf{I}_{d}\right]}\right]} \\ &= \frac{E\left[\frac{\lambda_{i}^{2}u_{d,i}^{2}}{\left[\frac{\lambda_{i}^{2}u_{d,i}^{2}}{\left\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\|_{2}^{3}} - \frac{1}{\left\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\|_{2}}\mathbf{I}_{d}\right]}\right]}{\frac{1}{c^{2}}E^{2}\left[\frac{\lambda_{i}^{2}u_{d,i}^{2}}{\left\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\|_{2}^{3}} - \frac{1}{\left\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\|_{2}}\mathbf{I}_{d}\right]} \\ &= c^{2}\nu_{i}^{2}\left(\mathbf{\Lambda}\right) \end{split}$$

thus

$$ARE_{i}\left(c^{2}\mathbf{\Lambda}\right) = \frac{d}{\left(d-1\right)^{2}} \frac{c^{2}\lambda_{i}^{2}}{c^{2}\nu_{i}^{2}\left(\mathbf{\Lambda}\right)} = \frac{d}{\left(d-1\right)^{2}} \frac{\lambda_{i}^{2}}{\nu_{i}^{2}\left(\mathbf{\Lambda}\right)} = ARE_{i}\left(\mathbf{\Lambda}\right)$$

this implies the size parameter does not matter in the calculation of the asymptotic relative efficiency. Define $r_i = \frac{\lambda_i}{\lambda_1}$, that is r_i is the ratio of the scale of the largest



Figure 2.1: Asymptotic Relative Efficiencies of the Spatial Median to the Oracle Spatial Median in \mathbb{R}^2

component to the scale of the i^{th} component. Consequently, when considering the asymptotic efficiency of each component one can take the case when the scale of the largest component is fixed; for simplicity assume it is equal to 1.

Consider the situation $\Sigma = \Lambda_0 = diag(\underbrace{\lambda^2, \ldots, \lambda^2}_{d_1}, \underbrace{r^2 \lambda^2, \ldots, r^2 \lambda^2}_{d-d_1})$ where $0 \le r \le 1$. Using the results in section 2.5.4 of the appendix, it follows from Lemma 5.2.3 that the efficiency of the any of the first d_1 components is given by

$$ARE_{d_1}(r) = \frac{{}_2F_1^2\left(\frac{1}{2}, \frac{d-d_1}{2}; \frac{d+2}{2}; 1-r^2\right)}{{}_2F_1\left(1, \frac{d-d_1}{2}; \frac{d+2}{2}; 1-r^2\right)}$$

The efficiency of the last $d - d_1$ components is given by

$$ARE_{d-d_1}(r) = \frac{{}_2F_1^2\left(\frac{1}{2}, \frac{d_1}{2}; \frac{d+2}{2}; 1-r^{-2}\right)}{{}_2F_1\left(1, \frac{d_1}{2}; \frac{d+2}{2}; 1-r^{-2}\right)}$$

where ${}_{2}F_{1}(a,b;c;k) = B^{-1}(b,c-b) \int_{0}^{1} x^{b-1} (1-x)^{c-b-1} (1-kx)^{-a} dx$ is the Gauss hypergeometric function. Starting with two dimensions, plotted in Figure 2.1 are the efficiencies of each component as a function of r.

For the component associated with the higher scale, the spatial median does not do



Figure 2.2: Asymptotic Relative Efficiencies of the Spatial Median to the Oracle Spatial Median in \mathbb{R}^3 : $\Lambda_0 = diag(1, 1, r^2)$

much worse than the oracle spatial median in estimating this component, even in the most extreme cases when Λ is nearly singular. However, for the component associated with the smaller scale, the asymptotic relative efficiency of the spatial median to the oracle spatial median is quite low when r is small indicating the drastic inferiority of the precision with which the spatial median estimates this component compared to the oracle spatial median.

For three dimensions, perhaps the most interesting scatter structures are those in which two of the λ 's are equal, that is $\lambda_1 = \lambda_2$ or $\lambda_2 = \lambda_3$. The first case corresponds to $r_2 = 1$, the efficiency for each of the components as a function of r is given in Figure 2.2,

For the second case, $r_2 = r_3 = r$. Presented in Figure 2.3 are the asymptotic efficiencies of the components as a function of r.

In both cases note that the efficiency for the larger components of the spatial median relative to the oracle spatial median is still relatively high even for nearly singular scatter structures. However for the smaller components as the scatter structure gets more singular, the relative efficiencies diminish drastically. Of the two situations, the scenario with which $\lambda_2 = \lambda_3$ is the most deleterious on the relative efficiencies of the



Figure 2.3: Asymptotic Relative Efficiencies of the Spatial Median to the Oracle Spatial Median in \mathbb{R}^3 : $\Lambda_0 = diag(1, r^2, r^2)$

components.

As the dimension increases, the number of possible scatter structures to consider greatly increases. The last dimension examined will be d = 10 under the following scatter structures,

In 2 and 3 dimensions, it was always that case the components associated with the larger scale had higher efficiencies. However, this is not true in general. In fact it is not only the scale that affects the efficiency of a given component, but also the number of components that have scales of similar magnitudes. Presented in Figure 2.4 are the efficiencies of the components for scatter structures given above.



Figure 2.4: Asymptotic Relative Efficiencies of the Spatial Median to the Oracle Spatial Median in \mathbb{R}^{10}

As the number of components with larger scales increases, the efficiencies of all the components, not just the ones with the larger scales, improves. This was also the case in three dimensions. Also of note, for scatter structures Λ_1 and Λ_2 , there are values of r in which the efficiencies are slightly higher for the components with the smaller scales, whereas for the rest of the scatter structures Λ_2 the efficiencies are always higher for the components with the larger scales.

2.4 Finite Sample Performance

2.4.1 Finite Sample Theory

For elliptical distributions it was proven that asymptotically, one can always find a more efficient estimator than the spatial median by using an affinely equivariant version of it. In section 3.5, exactly how asymptotically inefficient the spatial median is compared to the oracle spatial median (or affine spatial median) was considered under various dimensions and scatter structures. However, for finite samples, working out the exact distribution of the aforementioned estimators is intractable, thus one must resort to simulations in order to ascertain the efficiencies. For finite samples, there are two factors that must be considered when comparing the efficiency of the spatial median to an affinely equivariant version. The first is how the efficiency is affected by the fact one is estimating Σ with an affinely equivariant estimator of it. Asymptotically, this was shown not to matter provided the estimator of scatter converges in probability to Σ , however, for finite samples how one estimates Σ will be of consequence. The second consideration is how much efficiency is lost by sacrificing affine equivariance for only orthogonal equivariance and does this depend on the particular affine equivariant estimator of Σ . It will be shown that these two factors can be considered separately. To this end the following theorem is needed. The proof is relegated to section 2.5.5. of the appendix.

Theorem 4.1.1. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represent an i.i.d. sample from $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$. Let $\lambda_{\langle 1 \rangle}^2 > \lambda_{\langle 2 \rangle}^2 > \cdots > \lambda_{\langle m \rangle}^2$ be the distinct eigenvalues of $\boldsymbol{\Sigma}$ where $m \leq d$ is the number of mutually orthogonal eigenspaces of $\boldsymbol{\Sigma}$. For any orthogonally equivariant estimator of $\boldsymbol{\mu}$ based on a sample of size n, $\hat{\boldsymbol{\mu}}_n$, with variance-covariance matrix $\mathbf{V}_n(\hat{\boldsymbol{\mu}}) = Var[\hat{\boldsymbol{\mu}}_n]$, the following are true,

1) $\mathbf{V}_n(\widehat{\boldsymbol{\mu}})$ and $\boldsymbol{\Sigma}$ have the same eigenspaces; consequently they have the same eigenprojections and/or eigenvectors.

2) Let $\lambda_{1,n}^2 \geq \lambda_{2,n}^2 \geq \cdots \geq \lambda_{d,n}^2$, denote the eigenvalues of $\mathbf{V}_n(\widehat{\boldsymbol{\mu}})$. It follows the eigenspace associated with $\lambda_{i,n}^2$ is the same that is associated with λ_i^2 . Consequently, $\lambda_{i,n}^2 = \lambda_{j,n}^2$ if and only if $\lambda_i^2 = \lambda_j^2$.

Proof. See section 2.5.5 of the appendix.

The above theorem, coupled with affine equivariance arguments, implies that when sampling from an elliptical distribution, the variance-covariance matrix of the oracle spatial median, affine spatial median and spatial median have the following forms, provided they exist,

$$\mathbf{V}_{n}\left(\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\Sigma}}\right) = \alpha_{n}\left(G\right)\boldsymbol{\Sigma}, \quad \mathbf{V}_{n}\left(\widetilde{\boldsymbol{\mu}}_{\widehat{\boldsymbol{\Sigma}}}\right) = \beta_{\widehat{\boldsymbol{\Sigma}}_{n}}\left(G\right)\boldsymbol{\Sigma}, \quad \mathbf{V}_{n}\left(\widetilde{\boldsymbol{\mu}}_{SM}\right) = \mathbf{P}\boldsymbol{\mathcal{V}}_{n}\left(\boldsymbol{\Lambda};G\right)\mathbf{P}^{t}$$

with $\alpha_n(G)$ being a positive scalar that depends only on n and G whereas $\beta_{\widehat{\Sigma}_n}(G)$ depends on n, G and the choice of scatter estimator, $\widehat{\Sigma}_n$. The matrix $\mathcal{V}_n(\Lambda; G)$ is a diagonal and depends on n, G and Λ , that is $\mathcal{V}_n(\Lambda; G) = diag(\nu_{1,n}(\Lambda; G), \ldots, \nu_{d,n}(\Lambda; G))$.

As a consequence of the fact that the variance-covariance matrix of the oracle and affine spatial median are both proportional to Σ , the finite sample relative efficiency of $\tilde{\mu}_{\hat{\Sigma}_n}$ to $\tilde{\mu}_{\Sigma}$ reduces to a scalar quantity, namely $RE_n(\tilde{\mu}_{\hat{\Sigma}}, \tilde{\mu}_{\Sigma}) = \alpha_n(G) / \beta_{\hat{\Sigma}_n}(G)$. Furthermore, since the variance-covariance matrix of the spatial median has the same eigenvectors as Σ , the finite sample relative efficiency of $\tilde{\mu}_{SM}$ to either $\tilde{\mu}_{\hat{\Sigma}}$ to $\tilde{\mu}_{\Sigma}$ reduces to a comparison of eigenvalues of variance-covariance matrices corresponding to the same eigenvectors. Thus without loss of generality one can reduce consideration to the simple case where $\mathbf{P} = \mathbf{I}_d$. This simplification yields the same convenient interpretation as it did for comparing the asymptotic efficiencies in the previous section, namely the comparison of the eigenvalues reduces to a comparison of the variability with which each estimator estimates the components of μ , i.e. the finite sample relative efficiency for the components of the spatial median to an affine spatial median. This can be expressed as,

$$RE_{i,n}\left(\widetilde{\boldsymbol{\mu}}_{SM},\widetilde{\boldsymbol{\mu}}_{\widehat{\boldsymbol{\Sigma}}_{n}}\right) = \frac{\beta_{\widehat{\boldsymbol{\Sigma}}_{n}}\left(G\right)\lambda_{j}}{\nu_{j,n}\left(\boldsymbol{\Lambda};G\right)} = \frac{\alpha_{n}\left(G\right)\lambda_{i}}{\nu_{i,n}\left(\boldsymbol{\Lambda};G\right)} \times \frac{\beta_{\widehat{\boldsymbol{\Sigma}}_{n}}\left(G\right)}{\alpha_{n}\left(G\right)} = \frac{RE_{i,n}\left(\widetilde{\boldsymbol{\mu}}_{SM},\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\Sigma}}\right)}{RE_{n}\left(\widetilde{\boldsymbol{\mu}}_{\widehat{\boldsymbol{\Sigma}}},\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\Sigma}}\right)}$$
for $i = 1, \ldots, d$.

Note that $RE_{i,n}(\tilde{\mu}_{SM}, \tilde{\mu}_{\Sigma})$ only depends on n, Σ and G, but not the choice of estimator used for Σ . For a fixed n and G, this reflects how the relative efficiency of the j^{th} components is effected by the fact Σ is not proportional to the identity matrix. The term $RE_n(\tilde{\mu}_{\widehat{\Sigma}}, \tilde{\mu}_{\Sigma})$ depends only on $n, \hat{\Sigma}_n$ and G, thus can be viewed as a measure of how the relative efficiency of the components is affected by the choice of the scatter estimate, $\hat{\Sigma}_n$.

For the simulations, four elliptical distributions will be considered: the normal, Cauchy, t_3 and slash distributions; the reader is referred to [41]. The dimensions being considered are d = 2,3 and 10, with varying scatter structures for each. In two dimensions, the scatter structures considered are (1,2), (1,4), (1,8) and (1,16). For three dimensions the scatter structures considered are (3,4,5), (1,4,7), (1,4,4), (1,8,8), (1,1,4) and (1,1,8). Lastly for ten dimensions, the following scatter matrices were considered,

The choices of scatter estimates used for the affine spatial median were the sample covariance matrix and Dümbgen's scatter matrix, denoted \mathbf{C}_n and \mathbf{D}_n respectively. Dümbgen's matrix is defined to be the solution of the following estimating equation

$$\mathbf{D}_n = d \sum_{i \neq j} \frac{(\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^t}{(\mathbf{x}_i - \mathbf{x}_j)^t \mathbf{D}_n^{-1} (\mathbf{x}_i - \mathbf{x}_j)}$$

For further details on Dümbgen's scatter matrix, refer to [11].

To obtain an estimate of the value $\alpha_n(G)$ associated with the oracle spatial median, it is only necessary to consider the case $\Sigma = \mathbf{I}_d$ since $\mathbf{V}_n(\widetilde{\boldsymbol{\mu}}_{\Sigma}) \propto \Sigma$. For each dimension and distribution, 10,000 datasets of size n were generated under the spherical distribution $\mathcal{E}_d(\mathbf{0}, \mathbf{I}_d; G)$ and the oracle spatial median was calculated for each dataset. Using these 10,000 estimates, the variances of each component of the estimate were taken and then averaged to get an estimate of $\alpha_n(G)$. To get $\beta_{\widehat{\Sigma}_n}$, the same process can be applied to the affine spatial median.

For the spatial median, it was necessary to run simulations for each dimension, distribution and Σ , in order to obtain an estimate for $\mathcal{V}_n(\Lambda; G)$. Again, 10,000 datasets of size *n* were generated from the distribution $\mathcal{E}_d(\mathbf{0}, \Sigma; G)$ and the spatial median of the dataset was calculated for each dataset. The variances of the components of the spatial median were calculated to obtain estimates of $\nu_{i,n}(\Lambda; G)$. Recall, in situations where more than one of the diagonal elements of Λ were equal, the corresponding diagonal elements of $\mathcal{V}_n(\Lambda; G)$ must be equal by Theorem 4.1.1. Hence to get a better estimate of their common values, the average of the variances of the components for the repeated elements was taken.

2.4.2 Results

Bivariate Distributions

Bivariate Normal Distributions. The first distribution considered will be the bivariate normal distribution. Displayed in Figure 2.5 are the values of $RE_n(\tilde{\mu}_{\mathbf{C}}, \tilde{\mu}_{\mathbf{\Sigma}})$ and $RE_n(\tilde{\mu}_{\mathbf{D}}, \tilde{\mu}_{\mathbf{\Sigma}})$. Note that these correspond to the relative efficiency of the affine spatial medians to the spatial median under a spherical normal distribution, i.e. when $\mathbf{\Sigma} \propto \mathbf{I}_d$. The values of $RE_n(\tilde{\mu}_{\mathbf{\Sigma}}, \tilde{\mu}_{\mathbf{\Sigma}})$ are displayed in the vertical axis whereas the sample size, n, is displayed in the horizontal axis.

As can be seen from above, $RE_n(\tilde{\mu}_{\mathbf{C}}, \tilde{\mu}_{\mathbf{\Sigma}})$ is greater than 1 except for the case n = 4. Contrary to intuition, this suggests that in the situations where the data comes from a normal distribution in which Σ is known, it is better to estimate the location vector with an affinely equivariant version of the spatial median using the sample covariance matrix as an estimate of scatter rather than the known Σ . Note also that for n = 3, $RE_3(\tilde{\mu}_{\mathbf{D}}, \tilde{\mu}_{\Sigma})$ is greater than 1 as well, in fact it equals $RE_3(\tilde{\mu}_{\mathbf{C}}, \tilde{\mu}_{\Sigma})$. This follows from the fact that when the sample size is one greater than the dimension of the data (d = 2 and n = 3 in this case) the only affinely equivariant estimator of the location vector is in fact the sample mean vector [38]. For the normal, it is known that the sample mean is the MLE, hence explaining why the relative efficiencies of either affine spatial median to the oracle spatial median is greater than one when n = 3.

Presented in Figure 2.6 are the simulated values of $RE_{j,n}(\tilde{\mu}_{SM}, \tilde{\mu}_{\Sigma})$ for j = 1, 2, that is the finite sample relative efficiencies of the spatial median to the oracle spatial median for the two individual components when sampling from a bivariate normal distribution with covariance matrices diag(1, 2), diag(1, 4), diag(1, 8) and diag(1, 16).


Figure 2.5: Finite Sample Efficiencies of Affine Spatial Medians to the Oracle Spatial Median for the Bivariate Spherical Normal



Figure 2.6: Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial Median for the Bivariate Normal

Again, the horizontal axes represents the sample size and vertical axes is the relative efficiency of each component.

As the spread between the scales of the components gets larger, the relative efficiencies decrease, particularly for the component corresponding to the smaller scale. Also of note, for small sample sizes, the efficiencies for both components oscillates drastically before leveling off. This oscillating behavior is not only present in the bivariate normal case, but also for other distributions in differing dimensions and scatter structures and is attributable to the tendency of the spatial median to be one of the data points for odd sample sizes, see discussion [21]. In all cases as $n \to \infty$ the values of $RE_{j,n}(\tilde{\mu}_{SM}, \tilde{\mu}_{\Sigma})$ tend towards the values of $ARE_j(\tilde{\mu}_{SM}, \tilde{\mu}_{OSM}; \Lambda, G)$.



Figure 2.7: Finite Sample Efficiencies of Affine Spatial Medians to the Oracle Spatial Median for the Bivariate Spherical t_3

Bivariate elliptical t_3 distributions. The next distributions being considered are bivariate elliptical t_3 distributions. Like normal distributions, these distribution possess finite second moments, however, these distributions possess longer tails than normal distributions. Figure 2.7 presents the values of $RE_n(\tilde{\mu}_{\hat{\Sigma}}, \tilde{\mu}_{\Sigma})$ when using the sample covariance matrix and Dümbgen's matrix as estimates of scatter for the affine spatial median.

Unlike the normal, for the t_3 the relative efficiency of the affine spatial median to the oracle spatial median is less than 1, whether using either the the sample covariance matrix or Dümbgen's matrix as estimates of scatter. This agrees with the intuition that when constructing an affinely equivariant spatial median, it would be better to use the true population scatter matrix if it were known rather than some estimate of it. However, with the exception of the case when n = 3 which corresponds to the sample



Figure 2.8: Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial Median for the Bivariate t_3

mean as was previously mentioned, there is little loss in efficiency in estimating Σ . Additionally, it is better to use an affinely equivariant version of the spatial median using Dümbgen's matrix as the estimate of scatter rather than the sample covariance matrix.

Figure 2.8 presents the values of $RE_{j,n}(\tilde{\mu}_{SM}, \tilde{\mu}_{\Sigma})$ when sampling from a bivariate elliptical t_3 distribution. The same scatter matrices used for the normal cases are considered.

The same observations that were made about the normal case apply also to the t_3 distribution; namely, as the spread between the scales increases, the efficiencies decrease, particularly for the smaller component. Also, for small sample sizes the efficiencies of the components oscillate drastically before leveling off to the same asymptotic values



Figure 2.9: Finite Sample Efficiencies of the Affine Spatial Median using Dümbgen's matrix to the Oracle Spatial Median for the Bivariate Spherical Cauchy and Slash

as in the bivariate normal. The only noticeable difference being the bivariate normal has slightly higher efficiencies for smaller sample sizes.

Bivariate elliptical Cauchy and slash distributions. The last distributions considered are the bivariate elliptical Cauchy and slash distributions. These serve as examples of long-tailed distributions that do not possess finite first or second moments. Consequently, it is known that the sample covariance matrix does not converge for these distributions. Despite this fact, simulations were implemented to calculate the value of $RE_n(\tilde{\mu}_{\mathbf{C}}, \tilde{\mu}_{\mathbf{\Sigma}})$ and as to be expected were quite unstable. Consequently, only the values of $RE_n(\tilde{\mu}_{\mathbf{D}}, \tilde{\mu}_{\mathbf{\Sigma}})$ are presented. Figure 2.9 presents the results.

For both distributions, the relative efficiencies seem to approach the asymptotic value of 1, but considerably slower than the normal or t_3 distributions. Further note

that the smallest sample size presented in Figure 5 is n = 5. Simulations were implemented for the cases n = 3, 4, however the variances of the components for the estimators were highly variable. For n = 3, the affine spatial median corresponds to the sample mean, thus will not have have finite second moments if the underlying distribution of the data does not. The following theorem states when the spatial median, affine spatial median and oracle spatial median will lack finite second moments, the proof of which is relegated to the appendix.

Theorem 4.2.1. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represent an i.i.d. sample from $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$, with G absolutely continuous with respect to Lebesgue measure and not possessing finite second moments. For n = 3, the spatial median, oracle spatial median and any affinely equivariant spatial median do not possess finite second moments. This also holds for the case d = 2 and n = 4.

Proof. See section 2.5.5 of the appendix.

Depicted in Figure 2.10 are the simulated values of $RE_{j,n}(\tilde{\mu}_{SM}, \tilde{\mu}_{\Sigma})$ for j = 1, 2, for a Cauchy distribution with same scatter matrices considered in the normal and t_3 cases. The analogous plots for the slash distribution are almost identical to the Cauchy case and thus were omitted.

Again, the same trends noted for the normal and t_3 distributions are present; the only difference being that finite sample efficiencies being slightly lower in the Cauchy and slash distribution before leveling off to the same asymptotic values.

Trivariate Distributions

Like the bivariate case, there were similarities between the trivariate normal and t_3 as well as the trivariate Cauchy and slash. Consequently, only the results for the trivariate normal and Cauchy will be presented. First presented in Figure 2.11 are the values of $RE_n(\tilde{\mu}_{\hat{\Sigma}}, \tilde{\mu}_{\Sigma})$. For the normal, two versions of an affinely equivariant spatial median were considered, one using the sample covariance matrix as an estimator of scatter



Figure 2.10: Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial Median for the Bivariate Cauchy



Figure 2.11: Finite Sample Efficiencies of Affine Spatial Medians to the Oracle Spatial Median for the Trivariate Spherical Normal and Cauchy

matrix and one using Dümbgen's matrix. For the Cauchy distribution only Dümbgen's matrix was considered since the sample covariance matrix does not converge as noted in the bivariate case.

Like in the bivariate case, against intuition, the value of $RE_n(\tilde{\mu}_{\mathbf{C}}, \tilde{\mu}_{\mathbf{\Sigma}})$ is greater than 1, but now for all sample sizes considered. Though the affine spatial median using the sample covariance matrix has higher efficiency than the one using Dümbgen's matrix, in general not much is lost by estimating the population covariance matrix. This is not the case for small sample sizes from the Cauchy distribution where the efficiencies are quite low. Naturally the efficiency of the affine spatial median for the Cauchy is 0 for the case n = 4 since in this case, being one more than the dimension, the estimator reduces to the sample mean.



Figure 2.12: Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial Median for the Trivariate Normal and Cauchy - I

Now considering the affect differing scatter structures have one the efficiencies, plotted in Figure 2.12 are the simulated efficiencies for the normal and Cauchy distributions with scatter structures diag(3, 4, 5) and diag(1, 4, 7).

For both cases, the relative efficiency corresponding to the component with the smallest scale gets worse as the spread between the scales increases.

The other scatter structures of interest are the situations in which two of the scales are equal. Plotted in Figure 2.13 and 2.14 are the simulated relative efficiencies for the scatter matrices diag(1, 4, 4), diag(1, 8, 8), diag(1, 1, 4) and diag(1, 1, 8).

In both panels, as the spread between the scale of the components increases, the relative efficiencies for the components with the smaller scale diminishes. For the later two scatter structures, the relative efficiency corresponding to the larger component



Figure 2.13: Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial Median for the Trivariate Normal and Cauchy - II



Figure 2.14: Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial Median for the Trivariate Normal and Cauchy - III



Figure 2.15: Finite Sample Efficiencies of Affine Spatial Medians to the Oracle Spatial Median for the Spherical Normal and Cauchy in \mathbb{R}^{10}

also decreases as the spread between the scales of the components increases.

10 Dimensional Distributions

The last distributions considered will be elliptical distributions in 10 dimensions. Like in the bivariate and trivariate cases, there were stark similarities between the normal and t_3 distributions as well as between the Cauchy and slash distributions, thus as before only the results for the normal and Cauchy distribution will be considered. First considered is $RE_n(\tilde{\mu}_{\hat{\Sigma}}, \tilde{\mu}_{\Sigma})$. Again two versions of an affinely equivariant spatial median were considered, one using the sample covariance matrix as an estimator of scatter matrix and one using Dümbgen's matrix however, for the elliptical Cauchy distribution only Dümbgen's matrix was considered since the sample covariance matrix does not converge. Presented in Figure 2.15 are the simulated efficiencies. The results are similar to the trivariate case, $RE_n(\tilde{\mu}_{\mathbf{C}}, \tilde{\mu}_{\mathbf{\Sigma}}) > 1$ for all sample sizes, however not much is lost by estimating the scatter matrix with the exception of small sample size for the Cauchy distribution.

In 10-dimensions there is a greater variety of scatter structures to consider when determining how they affect the relative efficiency. The following were selected as a representative collection of interesting scatter matrices.

$$\begin{split} \mathbf{\Lambda}_1 &= diag(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_2 &= diag(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_3 &= diag(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_4 &= diag(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_5 &= diag(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_5 &= diag(1, 1, 1, 1, 1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_6 &= diag(1, 1, 1, 1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_6 &= diag(1, 1, 1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_6 &= diag(1, 1, 1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_6 &= diag(1, 1, 1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_6 &= diag(1, 1, 1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_6 &= diag(1, 1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_6 &= diag(1, 1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_6 &= diag(1, 1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_6 &= diag(1, 1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_6 &= diag(1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_6 &= diag(1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_6 &= diag(1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_6 &= diag(1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_6 &= diag(1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_6 &= diag(1, 1, 1, 1, 1, 1) \\ \mathbf{\Lambda}_6 &= diag(1, 1) \\ \mathbf{\Lambda}_6$$

Figure 2.16 presents the relative efficiencies for the normal and Cauchy distributions in 10 dimensions for scatter structures Λ_1 and Λ_2 .

In three dimensions, the components with the smaller scales had the smaller relative efficiencies; in 10 dimensions this is no longer the case as indicated by the plots corresponding to Λ_1 . The magnitudes of the relative efficiencies associated with the components not only appears to be related to the magnitude of the scales of the components, but also the number of components that have similar magnitudes. This is demonstrated in the plots corresponding to Λ_2 in which the relative efficiencies associated with the components with the larger scales are again the larger.

The other extreme is represented by scatter structures Λ_3 and Λ_4 . The simulated relative efficiencies for the normal and Cauchy distributions in 10 dimensions with these scatter structures are presented below in Figure 2.17.

Note that in general the relative efficiencies for scatter structures Λ_3 and Λ_4 are higher than for Λ_1 and Λ_2 . The situation where a few of the components scales are



Figure 2.16: Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial Median for the Normal and Cauchy in \mathbb{R}^{10} - I



Figure 2.17: Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial Median for the Normal and Cauchy in \mathbb{R}^{10} - II



Figure 2.18: Finite Sample Efficiencies of the Spatial Median to the Oracle Spatial Median for the Normal and Cauchy in \mathbb{R}^{10} - III

substantially larger than the rest is more detrimental to the relative efficiencies as opposed to the situation where a few of the components scales are smaller.

Lastly, scatter structures Λ_5 and Λ_6 represent intermediate situations to the prior. In Figure 2.17 are the relative efficiencies for these scatter structures,

In these situations, the components with the smallest relative efficiencies are again the ones with the smallest scales. However, overall the relative efficiencies are not as low as they are when a few of the components have larger scales.

2.5 Appendix

2.5.1 The Influence Function and Asymptotic Variance-Covariance Matrix of M-estimators of Location

The following is a review of material from chapter 10 of [24]. The general form of an M-estimator for the parameter θ is given by

$$\widehat{\boldsymbol{\theta}}_{n} = argmin_{\boldsymbol{\eta} \in \mathcal{S}_{\boldsymbol{\theta}}} \sum_{i=1}^{n} \rho\left(\mathbf{x}_{i}, \boldsymbol{\eta}\right)$$

where S_{θ} is the parameter space. When the parameter corresponds to the location vector of some elliptical distribution the function $\rho(\mathbf{x}, \boldsymbol{\eta})$ is of the form $\rho\left((\mathbf{x} - \boldsymbol{\eta})^t \widehat{\boldsymbol{\Sigma}}_n^{-1} (\mathbf{x} - \boldsymbol{\eta})\right)$, where $\widehat{\boldsymbol{\Sigma}}_n$ is some estimate of the scatter matrix. Additionally, the symbol $\boldsymbol{\mu}$ will be used instead of $\boldsymbol{\theta}$ as well as in all analogous definitions. One then has

$$\widehat{\boldsymbol{\mu}}_n = argmin_{\boldsymbol{\eta} \in \mathbb{R}^d} \sum_{i=1}^n \rho\left(\left(\mathbf{x}_i - \boldsymbol{\eta} \right)^t \widehat{\boldsymbol{\Sigma}}_n^{-1} \left(\mathbf{x}_i - \boldsymbol{\eta} \right) \right)$$

If $\rho(s)$ is differentiable, define $\psi(s) = \rho'(s)$. A necessary condition for the above equation to hold is

$$\sum_{i=1}^{n} \psi\left((\mathbf{x}_{i} - \widehat{\boldsymbol{\mu}}_{n})^{t} \, \widehat{\boldsymbol{\Sigma}}_{n}^{-1} \left(\mathbf{x}_{i} - \widehat{\boldsymbol{\mu}}_{n} \right) \right) \widehat{\boldsymbol{\Sigma}}_{n}^{-1} \left(\mathbf{x}_{i} - \widehat{\boldsymbol{\mu}}_{n} \right) = \mathbf{0}$$

Provided the estimator $\widehat{\Sigma}_n$ converges in probability to some nonsingular constant matrix $\widehat{\Sigma}_{\infty}(F)$, then the influence function of $\widehat{\mu}$ at **x** for the distribution F is defined as

$$IF_{\widehat{\boldsymbol{\mu}}}\left(\mathbf{x};F,\widehat{\boldsymbol{\Sigma}}\right) = -\mathbf{B}_{\psi}^{-1}\left(F,\widehat{\boldsymbol{\Sigma}}\right)\psi\left((\mathbf{x}-\widehat{\boldsymbol{\mu}}_{\infty}\left(F\right))^{t}\widehat{\boldsymbol{\Sigma}}_{\infty}^{-1}\left(F\right)\left(\mathbf{x}-\widehat{\boldsymbol{\mu}}_{\infty}\left(F\right)\right)\right)\left(\mathbf{x}-\widehat{\boldsymbol{\mu}}_{\infty}\left(F\right)\right)$$

where

$$\mathbf{B}_{\psi}\left(F,\widehat{\mathbf{\Sigma}}\right) = \frac{\partial}{\partial \boldsymbol{\eta}} E_F\left[\psi\left(\left(\mathbf{x}-\boldsymbol{\eta}\right)^t \widehat{\mathbf{\Sigma}}_{\infty}^{-1}\left(F\right)\left(\mathbf{x}-\boldsymbol{\eta}\right)\right)\left(\mathbf{x}-\boldsymbol{\eta}\right)\right]|_{\boldsymbol{\eta}=\widehat{\boldsymbol{\mu}}_{\infty}(F)}$$

If the conditions that allow for the interchanging of the expectation and derivative are satisfied then one has,

$$\mathbf{B}_{\psi}\left(F,\widehat{\mathbf{\Sigma}}\right) = E_{F}\left[\frac{\partial}{\partial \eta}\left\{\psi\left((\mathbf{x}-\eta)^{t}\,\widehat{\mathbf{\Sigma}}_{\infty}^{-1}\left(F\right)\left(\mathbf{x}-\eta\right)\right)\left(\mathbf{x}-\eta\right)\right\}|_{\eta=\widehat{\boldsymbol{\mu}}_{\infty}(F)}\right] = \\ = E_{F}\left[-2\psi'\left(\left(\mathbf{x}-\widehat{\boldsymbol{\mu}}_{\infty}\left(F\right)\right)^{t}\,\widehat{\mathbf{\Sigma}}_{\infty}^{-1}\left(F\right)\left(\mathbf{x}-\widehat{\boldsymbol{\mu}}_{\infty}\left(F\right)\right)\right)\widehat{\mathbf{\Sigma}}_{\infty}^{-1}\left(F\right)\left(\mathbf{x}-\widehat{\boldsymbol{\mu}}_{\infty}\left(F\right)\right)\left(\mathbf{x}-\widehat{\boldsymbol{\mu}}_{\infty}\left(F\right)\right)^{t}\right)\right.\\ \left.-\psi\left(\left(\mathbf{x}-\widehat{\boldsymbol{\mu}}_{\infty}\left(F\right)\right)^{t}\,\widehat{\mathbf{\Sigma}}_{\infty}^{-1}\left(F\right)\left(\mathbf{x}-\widehat{\boldsymbol{\mu}}_{\infty}\left(F\right)\right)\right)\mathbf{I}_{d}\right]$$

It follows the expression for the asymptotic covariance matrix of $\widehat{\boldsymbol{\mu}}$ as

$$\begin{aligned} AV_{\widehat{\boldsymbol{\mu}}}\left(F,\widehat{\boldsymbol{\Sigma}}\right) &= E_F\left[IF_{\widehat{\boldsymbol{\mu}}}\left(\widehat{\boldsymbol{\mu}}_{\infty}\left(F\right);F,\widehat{\boldsymbol{\Sigma}}\right)IF_{\widehat{\boldsymbol{\mu}}}^t\left(\widehat{\boldsymbol{\mu}}_{\infty}\left(F\right);F,\widehat{\boldsymbol{\Sigma}}\right)\right] \\ &= -\mathbf{B}_{\psi}^{-1}\left(F,\widehat{\boldsymbol{\Sigma}}\right)^{-1}\mathbf{A}_{\psi}\left(F,\widehat{\boldsymbol{\Sigma}}\right)\left(-\mathbf{B}_{\psi}^{-1}\left(F,\widehat{\boldsymbol{\Sigma}}\right)^{-1}\right)^t \end{aligned}$$

where

$$\mathbf{A}_{\psi}\left(F,\widehat{\mathbf{\Sigma}}\right) = E_F\left[\psi^2\left(\left(\mathbf{x} - \widehat{\boldsymbol{\mu}}_{\infty}\left(F\right)\right)^t \widehat{\mathbf{\Sigma}}_{\infty}^{-1}\left(\mathbf{x} - \widehat{\boldsymbol{\mu}}_{\infty}\left(F\right)\right)\right)\left(\mathbf{x} - \widehat{\boldsymbol{\mu}}_{\infty}\left(F\right)\right)\left(\mathbf{x} - \widehat{\boldsymbol{\mu}}_{\infty}\left(F\right)\right)^t\right]$$

2.5.2 The Asymptotic Variance-Covariance of the Spatial Median

Instead of focusing on just the spatial median, this section will prove results for a more general class of estimators for the location vector. Consider the family of elliptical power distributions which have density function given by

$$f_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = C_{p,d} det(\boldsymbol{\Sigma})^{-1/2} exp\left(-\left((\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)^p\right)$$

where p > 0 and $C_{p,d}$ is a constant that ensures the above expression in indeed a density (i.e. integrates to 1). Given *n* observations, the likelihood function is then given by

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{X}) = \prod_{i=1}^{n} C_{p,d} det (\boldsymbol{\Sigma})^{-1/2} exp\left(-\left((\mathbf{x}_{i} - \boldsymbol{\mu})^{t} \boldsymbol{\Sigma}^{-1} (\mathbf{x}_{i} - \boldsymbol{\mu})\right)^{p}\right)$$
$$= C_{p,d}^{n} det (\boldsymbol{\Sigma})^{-n/2} exp\left(-\sum_{i=1}^{n} \left((\mathbf{x}_{i} - \boldsymbol{\mu})^{t} \boldsymbol{\Sigma}^{-1} (\mathbf{x}_{i} - \boldsymbol{\mu})\right)^{p}\right)$$

If Σ were known a priori, then maximizing the above likelihood entails maximizing the argument in the exponential, which is the same as minimizing the sum. Hence the the maximum likelihood estimator μ is

$$\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\Sigma},p} = \operatorname{argmin}_{\boldsymbol{\eta} \in \mathbb{R}^d} \sum_{i=1}^n \|\mathbf{x}_i - \boldsymbol{\eta}\|_{\boldsymbol{\Sigma}^{-1}}^{2p}$$
(2.5.2.1)

If $\Sigma = \sigma^2 \mathbf{I}_d$, then the estimating equation for $\widetilde{\mu}_{\sigma^2 \mathbf{I}_d, p}$ would reduce to

$$\widetilde{\boldsymbol{\mu}}_{\sigma^{2}\mathbf{I}_{d},p} = \widetilde{\boldsymbol{\mu}}_{p} = \operatorname{argmin}_{\boldsymbol{\eta} \in \mathbb{R}^{d}} \sum_{i=1}^{n} \|\mathbf{x}_{i} - \boldsymbol{\eta}\|_{2}^{2p}$$
(2.5.2.2)

For this estimator, $\widehat{\Sigma}$ is fixed as \mathbf{I}_d and $\rho_p(x) = x^p$, thus giving $\psi_p(x) = px^{p-1}$. Since it is understood that $\widetilde{\mu}_p$ is referring to the M-estimator of location with those particular parameters, the redundancies in the notation utilized in the prior section will be suppressed with the subscript p indicating that expressions of interest pertain to $\widetilde{\mu}_p$. We thus have

$$\mathbf{A}_{p}(F) = E_{F}\left[\left\{p\left(\left(\mathbf{x}-\widetilde{\boldsymbol{\mu}}_{p,\infty}\right)^{t}\mathbf{I}_{d}^{-1}\left(\mathbf{x}-\widetilde{\boldsymbol{\mu}}_{p,\infty}\right)\right)^{p-1}\right\}^{2}\left(\mathbf{x}-\widetilde{\boldsymbol{\mu}}_{p,\infty}\right)\left(\mathbf{x}-\widetilde{\boldsymbol{\mu}}_{p,\infty}\right)^{t}\right]\\ = E_{F}\left[p^{2}\left\|\mathbf{x}-\widetilde{\boldsymbol{\mu}}_{p,\infty}\right\|_{2}^{4(p-1)}\left(\mathbf{x}-\widetilde{\boldsymbol{\mu}}_{p,\infty}\right)\left(\mathbf{x}-\widetilde{\boldsymbol{\mu}}_{p,\infty}\right)^{t}\right]$$

$$\mathbf{B}_{p}(F) = \frac{d}{d\boldsymbol{\eta}} E_{F} \left[p \left((\mathbf{x} - \boldsymbol{\eta})^{t} \mathbf{I}_{d}^{-1} (\mathbf{x} - \boldsymbol{\eta}) \right)^{p-1} (\mathbf{x} - \boldsymbol{\eta}) \right] |_{\boldsymbol{\eta} = \tilde{\boldsymbol{\mu}}_{p,\infty}}$$
$$= \frac{d}{d\boldsymbol{\eta}} E_{F} \left[p \| \mathbf{x} - \boldsymbol{\eta} \|_{2}^{2(p-1)} (\mathbf{x} - \boldsymbol{\eta}) \right] |_{\boldsymbol{\eta} = \tilde{\boldsymbol{\mu}}_{p,\infty}}$$

Provided the distribution F is well-behaved, one can swap the expectation and derivative in the expression to obtain,

$$\mathbf{B}_{p}\left(F\right) = E_{F}\left[\frac{d}{d\boldsymbol{\eta}}\left\{p\left\|\mathbf{x}-\boldsymbol{\eta}\right\|_{2}^{2\left(p-1\right)}\left(\mathbf{x}-\boldsymbol{\eta}\right)\right\}|_{\boldsymbol{\eta}=\widetilde{\boldsymbol{\mu}}_{p,\infty}}\right] =$$

$$= E_F \left[-2p\left(p-1\right) \left\| \mathbf{x} - \widetilde{\boldsymbol{\mu}}_{p,\infty} \right\|_2^{2(p-2)} \left(\mathbf{x} - \widetilde{\boldsymbol{\mu}}_{p,\infty} \right) \left(\mathbf{x} - \widetilde{\boldsymbol{\mu}}_{p,\infty} \right)^t - p \left\| \mathbf{x} - \widetilde{\boldsymbol{\mu}}_{p,\infty} \right\|_2^{2(p-1)} \mathbf{I}_d \right]$$

where obviously $\psi'_p(x) = p(p-1)x^{p-2}$. Before proceeding, the following lemma pertaining to orthogonally equivariant estimators of location for radially symmetric distributions is needed. A random vector, $\mathbf{x} \in \mathbb{R}^d$, is said to be radially symmetric about $\boldsymbol{\mu}$ if $(\mathbf{x} - \boldsymbol{\mu}) \sim -(\mathbf{x} - \boldsymbol{\mu})$.

Lemma 5.2.1. Let x_1, \ldots, x_n be a multivariate random sample from some multivariate distribution with location parameter μ and scatter matrix Σ that is radially symmetric about μ . Let $\hat{\mu}_n$ be any orthogonally equivariant estimator based on the sample of size n of location that converges in probability to a constant vector as $n \to \infty$. It follows that $\hat{\mu}$ is consistent for estimating μ , that is $\lim_{n\to\infty} \hat{\mu}_n = \mu$.

Proof. First consider the case $\boldsymbol{\mu} = \mathbf{0}$ and the scatter structure is uncorrelated (i.e. $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}$), denote this radially symmetric random vector as \mathbf{y} . Consider the orthogonal matrix $-\mathbf{I}_d$; it follows that $-\mathbf{I}_d\mathbf{y} = -\mathbf{y} \sim \mathbf{y}$. For notational convenience, let $\hat{\boldsymbol{\mu}}_{\infty}(\mathbf{y})$ denote the asymptotic value of $\hat{\boldsymbol{\mu}}$ when randomly sampling from the distribution of \mathbf{y} . Exploiting the fact that $-\mathbf{I}_d\mathbf{y}$ and \mathbf{y} have the same distribution it follows $\hat{\boldsymbol{\mu}}_{\infty}(-\mathbf{I}_d\mathbf{y}) = \hat{\boldsymbol{\mu}}_{\infty}(\mathbf{y})$. However, orthogonal equivariance implies $\hat{\boldsymbol{\mu}}_{\infty}(-\mathbf{I}_d\mathbf{y}) = -\mathbf{I}_d\hat{\boldsymbol{\mu}}_{\infty}(\mathbf{y}) = -\hat{\boldsymbol{\mu}}_{\infty}(\mathbf{y})$, thus $\hat{\boldsymbol{\mu}}_{\infty}(\mathbf{y}) = -\hat{\boldsymbol{\mu}}_{\infty}(\mathbf{y})$ implying $\hat{\boldsymbol{\mu}}_{\infty}(\mathbf{y}) = \mathbf{0}$. For the general case one can write $\mathbf{x} = \mathbf{P}\mathbf{y} + \boldsymbol{\mu}$ where \mathbf{P} is an orthogonal matrix and \mathbf{y} is distributed as before. It follows by orthogonal equivariance of $\hat{\boldsymbol{\mu}}$ that $\hat{\boldsymbol{\mu}}_{\infty}(\mathbf{x}) = \hat{\boldsymbol{\mu}}_{\infty}(\mathbf{P}\mathbf{y} + \boldsymbol{\mu}) = \mathbf{P}\hat{\boldsymbol{\mu}}_{\infty}(\mathbf{y}) + \hat{\boldsymbol{\mu}}_{\infty}(\boldsymbol{\mu}) = \mathbf{P}\mathbf{0} + \boldsymbol{\mu} = \boldsymbol{\mu}$. **QED**

Elliptically symmetric distributions are radially symmetric about their location parameter and $\tilde{\mu}_p$ is orthogonally equivariant so Lemma (5.2.1) applies. Hence for simplicity one can take $\mu = 0$. Using this gives,

$$\mathbf{A}_{p}(F) = E_{F}\left[\psi_{p}^{2}\left(\mathbf{x}^{t}\mathbf{x}\right)\mathbf{x}\mathbf{x}^{t}\right] = E_{F}\left[\left(p\|\mathbf{x}\|_{2}^{2(p-1)}\right)^{2}\mathbf{x}\mathbf{x}^{t}\right] = p^{2}E_{F}\left[\|\mathbf{x}\|_{2}^{4(p-1)}\mathbf{x}\mathbf{x}^{t}\right]$$

$$\begin{aligned} \mathbf{B}_{p}\left(F\right) &= E_{F}\left[-2\psi_{p}'\left(\mathbf{x}^{t}\mathbf{x}\right)\mathbf{x}\mathbf{x}^{t}-\psi_{p}\left(\mathbf{x}^{t}\mathbf{x}\right)\mathbf{I}_{d}\right] \\ &= E_{F}\left[-2p\left(p-1\right)\|\mathbf{x}\|_{2}^{2\left(p-2\right)}\mathbf{x}\mathbf{x}^{t}-p\|\mathbf{x}\|_{2}^{2\left(p-1\right)}\mathbf{I}_{d}\right] \\ &= -pE_{F}\left[2\left(p-1\right)\|\mathbf{x}\|_{2}^{2\left(p-2\right)}\mathbf{x}\mathbf{x}^{t}+\|\mathbf{x}\|_{2}^{2\left(p-1\right)}\mathbf{I}_{d}\right] \\ &= -pE_{F}\left[\|\mathbf{x}\|_{2}^{2\left(p-1\right)}\left(\mathbf{I}_{d}+2\left(p-1\right)\|\mathbf{x}\|_{2}^{-2}\mathbf{x}\mathbf{x}^{t}\right)\right] \end{aligned}$$

It is not difficult to see that $\mathbf{B}_{p}^{t}(F) = \mathbf{B}_{p}(F)$. When the random vector $\mathbf{x} \sim \mathcal{E}_{d}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$, the expressions for $AV_{p}(F)$ further simplifies. Using the stochastic representation $\mathbf{x} \sim R_{G} \boldsymbol{\Sigma}^{1/2} \mathbf{u}_{d}$, gives

$$E_{F} \left[\|\mathbf{x}\|_{2}^{4(p-1)} \mathbf{x} \mathbf{x}^{t} \right] = E_{F} \left[\left\| R_{G} \boldsymbol{\Sigma}^{1/2} \mathbf{u}_{d} \right\|_{2}^{4(p-1)} R_{G} \boldsymbol{\Sigma}^{1/2} \mathbf{u}_{d} \left(R_{G} \boldsymbol{\Sigma}^{1/2} \mathbf{u}_{d} \right)^{t} \right] \\ = E_{F} \left[R_{G}^{2(2p-1)} \left\| \boldsymbol{\Sigma}^{1/2} \mathbf{u}_{d} \right\|_{2}^{4(p-1)} \boldsymbol{\Sigma}^{1/2} \mathbf{u}_{d} \left(\boldsymbol{\Sigma}^{1/2} \mathbf{u}_{d} \right)^{t} \right] \\ = -p E_{F} \left[2 \left(p - 1 \right) \|\mathbf{x}\|_{2}^{2(p-2)} \mathbf{x} \mathbf{x}^{t} + \|\mathbf{x}\|_{2}^{2(p-1)} \mathbf{I}_{d} \right] \\ = E \left[R_{G}^{2(2p-1)} \right] E \left[\left\| \boldsymbol{\Sigma}^{1/2} \mathbf{u}_{d} \right\|_{2}^{4(p-1)} \boldsymbol{\Sigma}^{1/2} \mathbf{u}_{d} \left(\boldsymbol{\Sigma}^{1/2} \mathbf{u}_{d} \right)^{t} \right]$$

and

$$E_F\left[\|\mathbf{x}\|_2^{2(p-1)} \left(\mathbf{I}_d + 2(p-1) \|\mathbf{x}\|_2^{-2} \mathbf{x} \mathbf{x}^t\right)\right] =$$

$$= E_F \left[\left\| R_G \boldsymbol{\Sigma}^{1/2} \mathbf{u}_d \right\|_2^{2(p-1)} \left(\mathbf{I}_d + 2\left(p-1\right) \left\| R_G \boldsymbol{\Sigma}^{1/2} \mathbf{u}_d \right\|_2^{-2} R_G \boldsymbol{\Sigma}^{1/2} \mathbf{u}_d \left(R_G \boldsymbol{\Sigma}^{1/2} \mathbf{u}_d \right)^t \right) \right] =$$

$$= E_F \left[R_G^{2(p-1)} \left\| \mathbf{\Sigma}^{1/2} \mathbf{u}_d \right\|_2^{2(p-1)} \left(\mathbf{I}_d + 2(p-1) \left\| \mathbf{\Sigma}^{1/2} \mathbf{u}_d \right\|_2^{-2} \mathbf{\Sigma}^{1/2} \mathbf{u}_d \left(\mathbf{\Sigma}^{1/2} \mathbf{u}_d \right)^t \right) \right] =$$

$$= E\left[R_G^{2(p-1)}\right] E\left[\left\|\boldsymbol{\Sigma}^{1/2}\mathbf{u}_d\right\|_2^{2(p-1)} \left(\mathbf{I}_d + 2\left(p-1\right)\left\|\boldsymbol{\Sigma}^{1/2}\mathbf{u}_d\right\|_2^{-2}\boldsymbol{\Sigma}^{1/2}\mathbf{u}_d\left(\boldsymbol{\Sigma}^{1/2}\mathbf{u}_d\right)^t\right)\right]$$

Using the unique symmetric positive definite square root $\Sigma^{1/2} = \mathbf{P} \mathbf{\Lambda}^{1/2} \mathbf{P}^t$ and the facts $\mathbf{P}^t \mathbf{u}_d \sim \mathbf{u}_d$ and $\|\mathbf{P}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$ for any orthogonal matrix \mathbf{P} , it is not difficult to see that

$$E\left[\left\|\boldsymbol{\Sigma}^{1/2}\mathbf{u}_{d}\right\|_{2}^{4(p-1)}\boldsymbol{\Sigma}^{1/2}\mathbf{u}_{d}\left(\boldsymbol{\Sigma}^{1/2}\mathbf{u}_{d}\right)^{t}\right] = \mathbf{P}E\left[\left\|\boldsymbol{\Lambda}^{1/2}\mathbf{u}_{d}\right\|_{2}^{4(p-1)}\boldsymbol{\Lambda}^{1/2}\mathbf{u}_{d}\left(\boldsymbol{\Lambda}^{1/2}\mathbf{u}_{d}\right)^{t}\right]\mathbf{P}^{t}$$

and

$$E\left[\left\|\boldsymbol{\Sigma}^{1/2}\mathbf{u}_{d}\right\|_{2}^{2(p-1)}\left(\mathbf{I}_{d}+2\left(p-1\right)\left\|\boldsymbol{\Sigma}^{1/2}\mathbf{u}_{d}\right\|_{2}^{-2}\boldsymbol{\Sigma}^{1/2}\mathbf{u}_{d}\left(\boldsymbol{\Sigma}^{1/2}\mathbf{u}_{d}\right)^{t}\right)\right] =$$
$$=\mathbf{P}E\left[\left\|\boldsymbol{\Lambda}^{1/2}\mathbf{u}_{d}\right\|_{2}^{2(p-1)}\left(\mathbf{I}_{d}+2\left(p-1\right)\left\|\boldsymbol{\Lambda}^{1/2}\mathbf{u}_{d}\right\|_{2}^{-2}\boldsymbol{\Lambda}^{1/2}\mathbf{u}_{d}\left(\boldsymbol{\Lambda}^{1/2}\mathbf{u}_{d}\right)^{t}\right)\right]\mathbf{P}^{t}$$

Consequently, the asymptotic covariance matrix of $\tilde{\mu}_p$ under elliptical distributions will be denoted $AV_p(\Sigma, G)$ because it depends on them in the following manner

$$AV_{p}\left(\boldsymbol{\Sigma},G\right) = \beta_{p}\left(G\right)\mathbf{P}\boldsymbol{\mathcal{V}}_{p}\left(\boldsymbol{\Lambda}\right)\mathbf{P}^{t}$$

where $\beta_p(G) = E\left[R_G^{2(2p-1)}\right] / E^2\left[R_G^{2(p-1)}\right]$ and $\mathcal{V}_p(\Lambda) = \mathcal{B}_p^{-1}(\Lambda) \mathcal{A}_p(\Lambda) \mathcal{B}_p^{-1}(\Lambda)$ with

$$\mathcal{A}_{p}\left(\mathbf{\Lambda}\right) = E\left[\left\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\right\|_{2}^{4(p-1)}\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\left(\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\right)^{t}\right]$$

and

$$\boldsymbol{\mathcal{B}}_{p}\left(\boldsymbol{\Lambda}\right) = E\left[\left\|\boldsymbol{\Lambda}^{1/2}\mathbf{u}_{d}\right\|_{2}^{2(p-1)} \left(\mathbf{I}_{d} + 2\left(p-1\right)\left\|\boldsymbol{\Lambda}^{1/2}\mathbf{u}_{d}\right\|_{2}^{-2} \boldsymbol{\Lambda}^{1/2}\mathbf{u}_{d}\left(\boldsymbol{\Lambda}^{1/2}\mathbf{u}_{d}\right)^{t}\right)\right]$$

2.5.3 The Asymptotic Variance-Covariance of the Oracle Spatial Median

While the same machinery that was introduced in the prior two sections could be utilized to derive the asymptotic variance-covariance matrix of the oracle spatial median, an alternative approach is presented here. Again a generalization of the oracle spatial median will be considered, namely an oracle estimator of location obtained from the family of elliptical power distributions introduced in section 2.5.2. Recall for this distribution, the MLE for $\boldsymbol{\mu}$ for the case when $\boldsymbol{\Sigma}$ is known is given by the equation (2.5.2.1). Define the data matrix $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^t$ where the rows come from some elliptical distribution $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$. Consider the following transformation, $\mathbf{z}_i = \boldsymbol{\Sigma}^{-1/2} \mathbf{x}_i$ for $i = 1, \dots, n$. Since $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are affinely equivariant under the class of elliptical distribution, it follows the row of the data matrix $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)^t$ come from the elliptical distribution $\mathcal{E}_d(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}, \mathbf{I}_d; G)$. For $\tilde{\boldsymbol{\mu}}_{\boldsymbol{\Sigma},p}$ note that

$$\begin{split} \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\Sigma},p} &= \operatorname{argmin}_{\boldsymbol{\eta} \in \mathbb{R}^{d}} \sum_{i=1}^{n} \|\mathbf{x}_{i} - \boldsymbol{\eta}\|_{\boldsymbol{\Sigma}^{-1}}^{2p} \\ &= \operatorname{argmin}_{\boldsymbol{\eta} \in \mathbb{R}^{d}} \sum_{i=1}^{n} \left((\mathbf{x}_{i} - \boldsymbol{\eta})^{t} \, \boldsymbol{\Sigma}^{-1} (\mathbf{x}_{i} - \boldsymbol{\eta}) \right)^{p} \\ &= \operatorname{argmin}_{\boldsymbol{\eta} \in \mathbb{R}^{d}} \sum_{i=1}^{n} \left((\mathbf{x}_{i} - \boldsymbol{\eta})^{t} \, \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\Sigma}^{-1/2} (\mathbf{x}_{i} - \boldsymbol{\eta}) \right)^{p} \\ &= \operatorname{argmin}_{\boldsymbol{\eta} \in \mathbb{R}^{d}} \sum_{i=1}^{n} \left(\left(\boldsymbol{\Sigma}^{-1/2} \mathbf{x}_{i} - \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\eta} \right)^{t} \mathbf{I}_{d} \left(\boldsymbol{\Sigma}^{-1/2} \mathbf{x}_{i} - \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\eta} \right) \right)^{p} \\ &= \operatorname{argmin}_{\boldsymbol{\eta} \in \mathbb{R}^{d}} \sum_{i=1}^{n} \left(\left(\mathbf{z}_{i} - \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\eta} \right)^{t} \mathbf{I}_{d} \left(\mathbf{z}_{i} - \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\eta} \right) \right)^{p} \\ &= \operatorname{argmin}_{\boldsymbol{\eta} \in \mathbb{R}^{d}} \sum_{i=1}^{n} \left\| \mathbf{z}_{i} - \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\eta} \right\|_{\boldsymbol{\Sigma}^{-1}}^{2p} \end{split}$$

Define $\boldsymbol{\zeta} = \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\eta}$, thus $\boldsymbol{\eta} = \boldsymbol{\Sigma}^{1/2} \boldsymbol{\zeta}$.

$$\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\Sigma},p} = \operatorname{argmin}_{\boldsymbol{\Sigma}^{1/2}\boldsymbol{\zeta} \in \mathbb{R}^d} \sum_{i=1}^n \|\mathbf{z}_i - \boldsymbol{\zeta}\|_2^{2p}$$

Since the matrix $\Sigma^{1/2}$ is nonsingular, the transformation $\eta \to \Sigma^{1/2} \zeta$ is a bijection from \mathbb{R}^d to \mathbb{R}^d . Thus minimizing over η is the same as minimizing over $\zeta = \Sigma^{-1/2} \eta$. This gives,

$$\widetilde{\mu}_{\Sigma,p} = \Sigma^{1/2} argmin_{\boldsymbol{\zeta} \in \mathbb{R}^d} \sum_{i=1}^n \|\mathbf{z}_i - \boldsymbol{\zeta}\|_2^{2p}$$

Thus to ascertain $\tilde{\mu}_{\Sigma,p}$ one uses the inverse of the square root of the known scatter matrix to transform the data. One then finds $\tilde{\mu}_p$ for the transformed data, and then transforms it back using the square root of the known scatter matrix to arrive at $\tilde{\mu}_{\Sigma,p}$. This method is sometimes referred to as the transform-retransform method.

Not only does the previous fact provide a convenient computational method for finding $\tilde{\mu}_{\Sigma,p}$, but it also provides a quick way to ascertain both the finite sample and asymptotic variance-covariance matrix of $\tilde{\mu}_{\Sigma,p}$ under elliptical distributions. Let the estimate of $\tilde{\mu}_{\Sigma,p}$ based on the data matrix **X** be denoted $\tilde{\mu}_{\Sigma,p}(\mathbf{X})$, it was shown above that $\tilde{\mu}_{\Sigma,p}(\mathbf{X}) = \Sigma^{1/2} \tilde{\mu}_p(\mathbf{Z})$, thus

$$Var\left[\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\Sigma},p}\left(\mathbf{X}\right)\right] = Var\left[\boldsymbol{\Sigma}^{1/2}\widetilde{\boldsymbol{\mu}}_{p}\left(\mathbf{Z}\right)\right] = \boldsymbol{\Sigma}^{1/2}Var\left[\widetilde{\boldsymbol{\mu}}_{p}\left(\mathbf{Z}\right)\right]\left(\boldsymbol{\Sigma}^{1/2}\right)^{t}$$

This will also hold for the asymptotic covariance matrix,

$$AV_{\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\Sigma},p}}\left(\boldsymbol{\Sigma},G\right) = \boldsymbol{\Sigma}^{1/2}AV_{\widetilde{\boldsymbol{\mu}}_{p}}\left(\mathbf{I}_{d},G\right)\left(\boldsymbol{\Sigma}^{1/2}\right)^{t}$$

This gives the following formula

$$AV_{\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\Sigma},p}}\left(\boldsymbol{\Sigma},G\right) = \boldsymbol{\Sigma}^{1/2}\beta_{p}\left(G\right)\boldsymbol{\mathcal{V}}_{p}\left(\mathbf{I}_{d}\right)\boldsymbol{\Sigma}^{1/2}$$

with $\boldsymbol{\mathcal{V}}(\mathbf{I}_{d}) = \boldsymbol{\mathcal{B}}_{p}^{-1}(\mathbf{I}_{d}) \boldsymbol{\mathcal{A}}_{p}(\mathbf{I}_{d}) \boldsymbol{\mathcal{B}}_{p}^{-1}(\mathbf{I}_{d})$ where

$$\boldsymbol{\mathcal{A}}_{p}\left(\mathbf{I}_{d}\right) = E\left[\underbrace{\|\mathbf{u}_{d}\|_{2}^{4(p-1)}}_{=1}\mathbf{u}_{d}\mathbf{u}_{d}^{t}\right] = E\left[\mathbf{u}_{d}\mathbf{u}_{d}^{t}\right]$$

and

$$\boldsymbol{\mathcal{B}}_{p}\left(\mathbf{I}_{d}\right) = E\left[\underbrace{\left\|\mathbf{u}_{d}\right\|_{2}^{2\left(p-1\right)}}_{=1}\left(\mathbf{I}_{d}+2\left(p-1\right)\underbrace{\left\|\mathbf{u}_{d}\right\|_{2}^{-2}}_{=1}\mathbf{u}_{d}\mathbf{u}_{d}^{t}\right)\right] = E\left[\mathbf{I}_{d}+2\left(p-1\right)\mathbf{u}_{d}\mathbf{u}_{d}^{t}\right]$$

The above follow since $\|\mathbf{u}_d\|_2 = 1$. The matrices $\mathcal{A}_p(\mathbf{I}_d)$ and $\mathcal{B}_p(\mathbf{I}_d)$ are calculable. First it will be shown the matrix $E\left[\mathbf{u}_d\mathbf{u}_d^t\right]$ is proportional to \mathbf{I}_d . Define $\mathbf{u}_d^t = (u_{d,1}, \ldots, -u_{d,t}, \ldots, u_{d,d})$, where $u_{d,i}$ are the components of the unit random vector \mathbf{u}_d . It follows $\mathbf{u}_d^t \sim \mathbf{u}_d$. Thus for the matrix $E\left[\mathbf{u}_d\mathbf{u}_d^t\right]$, the $(\iota, i)^{th}$ component of this matrix is $E\left[u_{d,\iota}u_{d,i}\right]$. $\mathbf{u}_d^t \sim \mathbf{u}_d$ implies $u_{d,\iota}u_{d,i} \sim -u_{d,\iota}u_{d,i}$, thus $E\left[u_{d,\iota}u_{d,i}^t\right] = E\left[-u_{d,\iota}u_{d,i}\right] = -E\left[u_{d,\iota}u_{d,i}\right]$. Hence it must be the case $E\left[u_{d,\iota}u_{d,i}\right] = 0$. Since ι and i are arbitrary, it follows that every off-diagonal element of $E\left[u_{d,\iota}u_{d,i}\right]$ is 0. Furthermore, since $u_{d,i} \sim u_{d,j}$ is spherically symmetric, it follows $E\left[u_{d,\iota}u_{d,i}\right] = E\left[u_{d,j}u_{d,j}\right]$ for all i, j, thus $E\left[\mathbf{u}_d\mathbf{u}_d^t\right] \propto \mathbf{I}_d$. Write $E\left[\mathbf{u}_d\mathbf{u}_d^t\right] = \xi\mathbf{I}_d$; it is desired to get at the value of ξ . Taking the trace of $E\left[\mathbf{u}_d\left(\mathbf{u}_d\right)^t\right]$ gives,

$$tr\left(E\left[\mathbf{u}_{d}\mathbf{u}_{d}^{t}\right]\right) = E\left[tr\left(\mathbf{u}_{d}\mathbf{u}_{d}^{t}\right)\right]$$
$$= E\left[tr\left(\mathbf{u}_{d}^{t}\mathbf{u}_{d}\right)\right]$$
$$= 1$$

but

$$1 = tr\left(E\left[\mathbf{u}_{d}\mathbf{u}_{d}^{t}\right]\right) = tr\left(\xi\mathbf{I}_{d}\right) = \xi \times d$$

thus $\xi = 1/d$. This implies that $\mathcal{A}_p(\mathbf{I}_d) = \frac{1}{d}\mathbf{I}_d$ and $\mathcal{B}_p(\mathbf{I}_d) = \mathbf{I}_d + 2\frac{p-1}{d}\mathbf{I}_d = \frac{d+2(p-1)}{d}\mathbf{I}_d$. It then follows

$$\boldsymbol{\mathcal{V}}_{p}\left(\mathbf{I}_{d}\right) = \left(\frac{d+2\left(p-1\right)}{d}\mathbf{I}_{d}\right)^{-1}\frac{1}{d}\mathbf{I}_{d}\left(\frac{d+2\left(p-1\right)}{d}\mathbf{I}_{d}\right)^{-1} = \frac{d}{\left(d+2\left(p-1\right)\right)^{2}}\mathbf{I}_{d}$$

2.5.4 Proof of Theorem 3.3.1

Presented in this section are details for the proof of Theorem 3.3.1. To this end, several lemmas that are needed are first proved.

Lemma 5.2.2. Suppose $\mathbf{x} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Lambda}; G)$, let $AE_\iota(\lambda_1, \ldots, \lambda_d)$ denote the asymptotic efficiency associated with estimating the ι^{th} component of the location vector with $\tilde{\boldsymbol{\mu}}_p$. It follows that that only place where this multivariate function can have a stationary point is the when $\lambda_i = \lambda$ for $i \neq \iota$.

Proof. Without loss of generality one can consider the case $\mu = 0$. The asymptotic covariance matrix of the estimator $\tilde{\mu}_p$ when the scatter matrix is diagonal is given by

$$AV_{p}\left(\boldsymbol{\Lambda},G\right)=\beta_{p}\left(G\right)\boldsymbol{\mathcal{V}}_{p}\left(\boldsymbol{\Lambda}\right)$$

where $\beta_p(G)$ and $\mathcal{V}_p(\Lambda)$ are given in section 2.5.2. The only part of this expression that depends on the scatter matrix is $\mathcal{V}_p(\Lambda)$, which is a diagonal matrix. An explicit expression for the asymptotic efficiency of ι^{th} component of $\tilde{\mu}_p$ is given as

$$AE_{\iota}(\lambda_{1},\ldots,\lambda_{d}) = \frac{\alpha(G)}{\beta_{p}(G)} \frac{E^{2} \left[\left\| \mathbf{\Lambda}^{1/2} \mathbf{u}_{d} \right\|_{2}^{2(p-1)} + 2(p-1) \left\| \mathbf{\Lambda}^{1/2} \mathbf{u}_{d} \right\|_{2}^{2(p-2)} \lambda_{\iota}^{2} u_{d,\iota}^{2} \right]}{E \left[\left\| \mathbf{\Lambda}^{1/2} \mathbf{u}_{d} \right\|_{2}^{4(p-1)} \lambda_{\iota}^{2} u_{d,\iota}^{2} \right]} \lambda_{\iota}^{2}$$

For notational convenience, define the functions

$$C(\lambda_{1},...,\lambda_{d}) = E\left[\left\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\right\|_{2}^{2(p-1)}\right] = \int_{S_{d}} \frac{\Gamma(d/2)}{2\pi^{d/2}} \left(\lambda_{1}^{2}u_{d,1}^{2} + \dots + \lambda_{d}^{2}u_{d,d}^{2}\right)^{p-1} d\mathbf{u}_{d}$$

$$N_{\iota} (\lambda_1, \dots, \lambda_d) = E \left[\left\| \mathbf{\Lambda}^{1/2} \mathbf{u}_d \right\|_2^{2(p-2)} \lambda_{\iota}^2 u_{d,\iota}^2 \right]$$
$$= \int_{S_d} \frac{\Gamma (d/2)}{2\pi^{d/2}} \left(\lambda_1^2 u_{d,1}^2 + \dots + \lambda_d^2 u_{d,d}^2 \right)^{p-2} \lambda_{\iota}^2 u_{d,\iota}^2 d\mathbf{u}_d$$

$$D_{\iota}(\lambda_{1},...,\lambda_{d}) = E\left[\left\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\right\|_{2}^{4(p-1)}\lambda_{\iota}^{2}u_{d,\iota}^{2}\right]$$
$$= \int_{S_{d}}\frac{\Gamma(d/2)}{2\pi^{d/2}}\left(\lambda_{1}^{2}u_{d,1}^{2}+\cdots+\lambda_{d}^{2}u_{d,d}^{2}\right)^{2(p-1)}\lambda_{\iota}^{2}u_{d,\iota}^{2}d\mathbf{u}_{d}$$

The function $C(\lambda_1, \ldots, \lambda_d)$ is symmetric in all of its arguments, whereas the functions $N_{\iota}(\lambda_1, \ldots, \lambda_d)$ and $D_{\iota}(\lambda_1, \ldots, \lambda_d)$ are symmetric in every argument with the exceptions of λ_{ι} . Using these expressions one has

$$AE_{\iota}(\lambda_{1},\ldots,\lambda_{d}) = \frac{\alpha(G)}{\beta_{p}(G)} \frac{\left(C(\lambda_{1},\ldots,\lambda_{d})+2(p-1)N_{\iota}(\lambda_{1},\ldots,\lambda_{d})\right)^{2}}{D_{\iota}(\lambda_{1},\ldots,\lambda_{d})}\lambda_{\iota}^{2}$$

It follows that $AE_{\iota}(\lambda_1, \ldots, \lambda_d)$ is symmetric in every argument except λ_{ι} . It is desired to get the partial derivatives of $AE_{\iota}(\lambda_1, \ldots, \lambda_d)$; to this end the partial derivatives of the functions $N_{\iota}(\lambda_1, \ldots, \lambda_d)$, $C(\lambda_1, \ldots, \lambda_d)$ and $D_{\iota}(\lambda_1, \ldots, \lambda_d)$ are needed. One can verify the conditions necessary in order to interchange the partial derivative and integral are indeed satisfied with these functions. Suppressing the dependence on the λ 's one has

$$\frac{\partial C}{\partial \lambda_i} = 2(p-1) \int_{S_d} \frac{\Gamma(d/2)}{2\pi^{d/2}} \left(\lambda_1^2 u_{d,1}^2 + \dots + \lambda_d^2 u_{d,d}^2\right)^{p-2} \lambda_i u_{d,i}^2 d\mathbf{u}_d$$
$$= 2(p-1) \lambda_i E\left[\left\| \mathbf{\Lambda}^{1/2} \mathbf{u}_d \right\|_2^{2(p-2)} u_{d,i}^2 \right]$$

$$\begin{aligned} \frac{\partial N_{\iota}}{\partial \lambda_{i}} &= 2\left(p-2\right) \int_{S_{d}} \frac{\Gamma\left(d/2\right)}{2\pi^{d/2}} \left(\lambda_{1}^{2} u_{d,1}^{2} + \dots + \lambda_{d}^{2} u_{d,d}^{2}\right)^{p-3} \lambda_{i} u_{d,i}^{2} \lambda_{\iota}^{2} u_{d,\iota}^{2} d\mathbf{u}_{d} \\ &= 2\left(p-2\right) \lambda_{i} E\left[\left\|\mathbf{\Lambda}^{1/2} \mathbf{u}_{d}\right\|_{2}^{2\left(p-3\right)} u_{d,i}^{2} \lambda_{\iota}^{2} u_{d,\iota}^{2}\right] \end{aligned}$$

$$\frac{\partial D_{\iota}}{\partial \lambda_{i}} = 4 (p-1) \int_{S_{d}} \frac{\Gamma (d/2)}{2\pi^{d/2}} \left(\lambda_{1}^{2} u_{d,1}^{2} + \dots + \lambda_{d}^{2} u_{d,d}^{2} \right)^{2p-3} \lambda_{i} u_{d,i}^{2} \lambda_{\iota}^{2} u_{d,\iota}^{2} d\mathbf{u}_{d}
= 4 (p-1) \lambda_{i} E \left[\left\| \mathbf{\Lambda}^{1/2} \mathbf{u}_{d} \right\|_{2}^{2(2p-3)} u_{d,i}^{2} \lambda_{\iota}^{2} u_{d,\iota}^{2} \right]$$

The partial $\frac{\partial C}{\partial \lambda_i}$ is symmetric in every argument except λ_i whereas $\frac{\partial N_{\iota}}{\partial \lambda_i}$ and $\frac{\partial D_{\iota}}{\partial \lambda_i}$ are symmetric in every argument except λ_i and λ_{ι} . Using these gives

$$\frac{\partial AE_{\iota}}{\partial \lambda_{i}} = \frac{\alpha \left(G\right)}{\beta_{p}\left(G\right)} \frac{2D_{\iota}\left(C+2\left(p-1\right)N_{\iota}\right)\left(\frac{\partial C}{\partial \lambda_{i}}+2\left(p-1\right)\frac{\partial N_{\iota}}{\partial \lambda_{i}}\right)-\left(C+2\left(p-1\right)N_{\iota}\right)^{2}\frac{\partial D_{\iota}}{\partial \lambda_{i}}}{D_{\iota}^{2}}\lambda_{\iota}^{2} \\ = \frac{\alpha \left(G\right)}{\beta_{p}\left(G\right)}\lambda_{\iota}^{2}\frac{\left(C+2\left(p-1\right)N_{\iota}\right)}{D_{\iota}^{2}}\left[2D_{\iota}\left(\frac{\partial C}{\partial \lambda_{i}}+2\left(p-1\right)\frac{\partial N_{\iota}}{\partial \lambda_{i}}\right)-\left(C+2\left(p-1\right)N_{\iota}\right)\frac{\partial D_{\iota}}{\partial \lambda_{i}}\right]$$

which is consequently symmetric in every argument except λ_i and λ_i . Write the bracketed terms as $4(p-1)\lambda_i F_{i,i}(\lambda_1,\ldots,\lambda_d)$ where

$$\begin{split} F_{\iota,i} &= E\left[\left\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\right\|_{2}^{4(p-1)}\lambda_{\iota}^{2}u_{d,\iota}^{2}\right]\left(E\left[\left\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\right\|_{2}^{2(p-2)}u_{d,i}^{2}\right] + 2\left(p-2\right)E\left[\left\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\right\|_{2}^{2(p-3)}u_{d,i}^{2}\lambda_{\iota}^{2}u_{d,\iota}^{2}\right]\right) \\ &- E\left[\left\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\right\|_{2}^{2(2p-3)}u_{d,i}^{2}\lambda_{\iota}^{2}u_{d,\iota}^{2}\right]\left(E\left[\left\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\right\|_{2}^{2(p-1)}\right] + 2\left(p-1\right)E\left[\left\|\mathbf{\Lambda}^{1/2}\mathbf{u}_{d}\right\|_{2}^{2(p-2)}\lambda_{\iota}^{2}u_{d,\iota}^{2}\right]\right) \end{split}$$

Without loss of generality, take the case $\iota = d$. Then one has,

$$\frac{\partial AE_{d}}{\partial \lambda_{i}} = 4\left(p-1\right)\frac{\alpha\left(G\right)}{\beta_{p}\left(G\right)}\lambda_{i}\lambda_{d}^{2}\frac{\left(C+2\left(p-1\right)N_{d}\right)}{D_{d}^{2}}F_{d,i}$$

Let $(1), \ldots, (d-1)$ be a permutation of the indices $1, \ldots, d-1$. Because of the symmetries of the function AE_d , it follows for $c \neq 0$, if $\frac{\partial}{\partial \lambda_i} AE_d \left(\lambda_1^*, \ldots, \lambda_{d-1}^*, \lambda_d^*\right) = c$, then $\frac{\partial}{\partial \lambda_{(i)}} AE_d \left(\lambda_{(1)}^*, \ldots, \lambda_{(d-1)}^*, \lambda_d^*\right) = c$. Canceling out the expression $4(p-1) \frac{\alpha(G)}{\beta_p(G)}$ because it does not depend on the λ 's gives

$$\lambda_i^* \lambda_d^{*2} \frac{\left(C\left(\lambda_1^*, \dots, \lambda_{d-1}^*, \lambda_d^*\right) - 2\left(p-1\right) N_d\left(\lambda_1^*, \dots, \lambda_{d-1}^*, \lambda_d^*\right)\right)}{D_d^2\left(\lambda_1^*, \dots, \lambda_{d-1}^*, \lambda_d^*\right)} F_i\left(\lambda_1^*, \dots, \lambda_{d-1}^*, \lambda_d^*\right) = c$$

$$\lambda_{(i)}^{*}\lambda_{d}^{*2} \frac{\left(C\left(\lambda_{(1)}^{*}, \dots, \lambda_{(d-1)}^{*}, \lambda_{d}^{*}\right) - 2\left(p-1\right)D_{d}\left(\lambda_{(1)}^{*}, \dots, \lambda_{(d-1)}^{*}, \lambda_{d}^{*}\right)\right)}{D_{d}^{2}\left(\lambda_{(1)}^{*}, \dots, \lambda_{(d-1)}^{*}, \lambda_{d}^{*}\right)}F_{(i)}\left(\lambda_{(1)}^{*}, \dots, \lambda_{(d-1)}^{*}, \lambda_{d}^{*}\right) = c$$

Note that

$$\lambda_d^{*2} \frac{\left(C\left(\lambda_1^*, \dots, \lambda_{d-1}^*, \lambda_d^*\right) - 2\left(p-1\right) D_d\left(\lambda_1^*, \dots, \lambda_{d-1}^*, \lambda_d^*\right)\right)}{D_d^2\left(\lambda_1^*, \dots, \lambda_{d-1}^*, \lambda_d^*\right)}$$
$$= \lambda_d^{*2} \frac{\left(C\left(\lambda_{(1)}^*, \dots, \lambda_{(d-1)}^*, \lambda_d^*\right) - 2\left(p-1\right) D_d\left(\lambda_{(1)}^*, \dots, \lambda_{(d-1)}^*, \lambda_d^*\right)\right)}{D_d^2\left(\lambda_{(1)}^*, \dots, \lambda_{(d-1)}^*, \lambda_d^*\right)}$$

and

$$F_{d,i}\left(\lambda_1^*,\ldots,\lambda_{d-1}^*,\lambda_d^*\right) = F_{d,(i)}\left(\lambda_{(1)}^*,\ldots,\lambda_{(d-1)}^*,\lambda_d^*\right)$$

The first equality follows since all the arguments are symmetric in every argument except λ_d . For the second note that $F_{d,i}$ is symmetric in every argument except λ_i and λ_d . However $F_{d,i}$ and $F_{d,(i)}$ involve derivatives with respect to λ_i and $\lambda_{(i)}$ respectively, thus the two terms are equal. Upon canceling terms, it follows that $\frac{\partial}{\partial \lambda_i} AE_d \left(\lambda_1^*, \ldots, \lambda_{d-1}^*, \lambda_d^*\right) = c = \frac{\partial}{\partial \lambda_{(i)}} AE_d \left(\lambda_{(1)}^*, \ldots, \lambda_{(d-1)}^*, \lambda_d^*\right)$ implies $\lambda_i = \lambda_{(i)}$.

The case for c = 0 needs more care since one cannot cancel the parts of the expression that equal 0. However, a slight modification will give the desired result. Consider a sequence $\lambda_i^*(n)$ such that $\lim_{n\to\infty} \lambda_i^*(n) = \lambda_i^*$. We thus have

$$\lim_{n \to \infty} \frac{\partial}{\partial \lambda_i} A E_d \left(\lambda_1^*, \dots, \lambda_i^* \left(n \right), \dots \lambda_{d-1}^*, \lambda_d^* \right) = 0$$

At each point along this sequence it was shown

$$\frac{\partial}{\partial\lambda_{i}}AE_{d}\left(\lambda_{1}^{*},\ldots,\lambda_{i}^{*}\left(n\right),\ldots\lambda_{d-1}^{*},\lambda_{d}^{*}\right)=\frac{\partial}{\partial\lambda_{\left(i\right)}}AE_{d}\left(\lambda_{\left(1\right)}^{*},\ldots,\lambda_{\left(i\right)}^{*}\left(n\right),\ldots\lambda_{\left(d-1\right)}^{*},\lambda_{d}^{*}\right)$$

By the previous arguments we have $\lambda_i^*(n) = \lambda_{(i)}^*(n)$ for every n. However the functions λ_d , $(C - 2(p-1)N_d)/D_d^2$, λ_i and $F_{d,i}$ are all continuous functions of the λ 's. Thus if

and

 $\lim_{n\to\infty}\lambda_i^*(n) \neq \lim_{n\to\infty}\lambda_{(i)}^*(n)$, it would contradict the continuity of one of the above functions. Hence it must be the case $\lambda_i^* = \lambda_{(i)}^*$. The main result follows by repeating the above arguments. **QED**

Lemma 5.2.3 Let $\mathbf{x} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Lambda}; G)$ where $\boldsymbol{\Lambda} = diag\left(\underbrace{\kappa^2 \lambda^2, \cdots, \kappa^2 \lambda^2}_{d_1}, \underbrace{\lambda^2, \cdots, \lambda^2}_{d-d_1}\right)$. Under this scatter structure, the ARE's of $\widetilde{\boldsymbol{\mu}}_p$ to $\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\Lambda},p}$ for estimating the components of $\boldsymbol{\mu}$ is given by

$$ARE_{d_1}(\kappa) = \frac{{}_2F_1^2\left(-\left(p-1\right), \frac{d-d_1}{2}; \frac{d+2}{2}; 1-\kappa^{-2}\right)}{{}_2F_1\left(-2\left(p-1\right), \frac{d-d_1}{2}; \frac{d+2}{2}; 1-\kappa^{-2}\right)}$$

for the first d_1 components and

$$ARE_{d-d_{1}}(\kappa) = \frac{{}_{2}F_{1}^{2}\left(-\left(p-1\right), \frac{d_{1}}{2}; \frac{d+2}{2}; 1-\kappa^{2}\right)}{{}_{2}F_{1}\left(-2\left(p-1\right), \frac{d_{1}}{2}; \frac{d+2}{2}; 1-\kappa^{2}\right)}$$

for the last $d - d_1$ components where

 $_{2}F_{1}(a,b;c;k) = B^{-1}(b,c-b)\int_{0}^{1} x^{b-1} (1-x)^{c-b-1} (1-kx)^{-a} dx$ is the Gauss hypergeometric function.

Proof. Recall the expression for the ARE of $\tilde{\mu}_p$ to the $\tilde{\mu}_{\Lambda,p}$ for estimating the i^{th} component of μ is given by

$$ARE_{i}\left(\widetilde{\boldsymbol{\mu}}_{p},\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\Lambda},p};\boldsymbol{\Lambda},G\right) = \frac{\left[AV_{\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\Lambda},p}}\left(G\right)\right]_{i}}{\left[AV_{\widetilde{\boldsymbol{\mu}}_{p}}\left(G,\boldsymbol{\Lambda}\right)\right]_{i}} = \frac{d}{\left(d+2\left(p-1\right)\right)^{2}}\frac{\lambda_{i}^{2}}{\nu_{p,i}^{2}\left(\boldsymbol{\Lambda}\right)}$$

where

$$\nu_{p,i}^{2}(\mathbf{\Lambda}) = \frac{E\left[\left\|\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{u}_{d}\right\|_{2}^{4(p-1)}\lambda_{i}^{2}u_{d,i}^{2}\right]}{E^{2}\left[\left\|\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{u}_{d}\right\|_{2}^{2(p-1)}\left(1+2(p-1)\left\|\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{u}_{d}\right\|_{2}^{-2}\lambda_{i}^{2}u_{d,i}^{2}\right)\right]}$$
$$= \frac{E\left[\left\|\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{u}_{d}\right\|_{2}^{4(p-1)}\lambda_{i}^{2}u_{d,i}^{2}\right]}{E^{2}\left[\left\|\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{u}_{d}\right\|_{2}^{2(p-1)}+2(p-1)\left\|\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{u}_{d}\right\|_{2}^{2(p-2)}\lambda_{i}^{2}u_{d,i}^{2}\right]}$$

Define $U_1 = u_{d,1}^2 + \cdots + u_{d,d_1}^2$, $U_2 = u_{d,d_1+1}^2 + \cdots + u_{d,d}^2$ and $U = \kappa^2 U_1 + U_2$. The ARE of any of the last $(d - d_1)$ components of $\tilde{\mu}_p$ to $\tilde{\mu}_{\Lambda,p}$ can be written as

$$ARE_{d-d_{1}}(\kappa) = \frac{d}{(d+2(p-1))^{2}} \frac{E^{2} \left[\left(\lambda^{2} U \right)^{p-1} + 2(p-1) \left(\lambda^{2} U \right)^{p-2} \lambda^{2} u_{d,d}^{2} \right]}{E \left[\left(\lambda^{2} U \right)^{2(p-1)} \lambda^{2} u_{d,d}^{2} \right]} \lambda^{2}$$
$$= \frac{d}{(d+2(p-1))^{2}} \frac{E^{2} \left[U^{p-1} + 2(p-1) U^{p-2} u_{d,d}^{2} \right]}{E \left[U^{2(p-1)} u_{d,d}^{2} \right]}$$

Note that $U_2 = 1 - U_1$. By exchangeability,

$$E\left[u_{d,d}^{2}h\left(U\right)\right] = \frac{1}{d-d_{1}}\sum_{j=d_{1}+1}^{d} E\left[u_{d,j}^{2}h\left(U\right)\right] = \frac{1}{d-d_{1}}E\left[U_{2}h\left(U\right)\right] = \frac{1}{d-d_{1}}E\left[(1-U_{1})h\left(U\right)\right]$$

Also $U = 1 - kU_1$ where $k = 1 - \kappa^2$. Substituting these identities above gives

$$ARE_{d-d_1}(\kappa) = \frac{d}{(d+2(p-1))^2} \frac{E^2 \left[(1-kU_1)^{p-1} + 2\frac{p-1}{d-d_1} (1-U_1) (1-kU_1)^{p-2} \right]}{\frac{1}{d-d_1} E \left[(1-U_1) (1-kU_1)^{2(p-1)} \right]}$$

Since the vector \mathbf{u}_d is uniformly distributed on S_d , it follows that

 $U_1 \sim Beta(d_1/2, (d-d_1)/2)$ and $U_2 \sim Beta((d-d_1)/2, d_1/2)$. Thus for the expectations one has

$$E\left[(1-U_{1})^{s}\left(1-kU_{1}\right)^{-t}\right] = \frac{1}{B\left(\frac{d_{1}}{2}, \frac{d-d_{1}}{2}\right)} \int_{0}^{1} x^{\frac{d_{1}}{2}-1} (1-x)^{\frac{d-d_{1}}{2}+s-1} (1-kx)^{-t} dx$$
$$= \frac{1}{B\left(\frac{d_{1}}{2}, \frac{d-d_{1}}{2}\right)} \int_{0}^{1} x^{\frac{d_{1}}{2}-1} (1-x)^{\frac{d+2s}{2}-\frac{d_{1}}{2}-1} (1-kx)^{-t} dx$$
$$= \frac{B\left(\frac{d_{1}}{2}, \frac{d+2s}{2}-\frac{d_{1}}{2}\right)}{B\left(\frac{d_{1}}{2}, \frac{d-d_{1}}{2}\right)} {}_{2}F_{1}\left(t, \frac{d_{1}}{2}; \frac{d+2s}{2}; k\right)$$

where $B(a, b) = \Gamma(a) \Gamma(b) / \Gamma(a + b)$ is the Beta function. The integral representation of the Gauss hypergeometric function is valid for $\Re(c) > \Re(b) > 0$ so the last equality is justified. These facts then give

$$E\left[(1-kU_{1})^{p-1}\right] = {}_{2}F_{1}\left(-(p-1), \frac{d_{1}}{2}; \frac{d}{2}; k\right)$$

$$E\left[(1-U_{1})(1-kU_{1})^{p-2}\right] = \frac{B\left(\frac{d_{1}}{2}, \frac{d+2}{2} - \frac{d_{1}}{2}\right)}{B\left(\frac{d_{1}}{2}, \frac{d-d_{1}}{2}\right)} {}_{2}F_{1}\left(-(p-2), \frac{d_{1}}{2}; \frac{d+2}{2}; k\right)$$

$$= \frac{\Gamma\left(\frac{d_{1}}{2}\right)\Gamma\left(\frac{d-d_{1}+2}{2}\right)\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d+2}{2}\right)\Gamma\left(\frac{d-d_{1}}{2}\right)} {}_{2}F_{1}\left(-(p-2), \frac{d_{1}}{2}; \frac{d+2}{2}; k\right)$$

$$= \frac{\frac{d-d_{1}}{2}\Gamma\left(\frac{d-d_{1}}{2}\right)\Gamma\left(\frac{d}{2}\right)}{\frac{d}{2}\Gamma\left(\frac{d}{2}\right)\Gamma\left(\frac{d-d_{1}}{2}\right)} {}_{2}F_{1}\left(-(p-2), \frac{d_{1}}{2}; \frac{d+2}{2}; k\right)$$

$$\begin{split} E\left[\left(1-U_{1}\right)\left(1-kU_{1}\right)^{2(p-1)}\right] &= \frac{B\left(\frac{d_{1}}{2},\frac{d+2}{2}-\frac{d_{1}}{2}\right)}{B\left(\frac{d_{1}}{2},\frac{d-d_{1}}{2}\right)}{2}{}_{2}F_{1}\left(-2\left(p-1\right),\frac{d_{1}}{2};\frac{d+2}{2};k\right)\\ &= \frac{\Gamma\left(\frac{d_{1}}{2}\right)\Gamma\left(\frac{d-d_{1}+2}{2}\right)\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d+2}{2}\right)\Gamma\left(\frac{d-d_{1}}{2}\right)^{2}}{}_{2}F_{1}\left(-2\left(p-1\right),\frac{d_{1}}{2};\frac{d+2}{2};k\right)\\ &= \frac{\frac{d-d_{1}}{2}\Gamma\left(\frac{d-d_{1}}{2}\right)\Gamma\left(\frac{d}{2}\right)}{\frac{d}{2}\Gamma\left(\frac{d}{2}\right)\Gamma\left(\frac{d-d_{1}}{2}\right)^{2}}{}_{2}F_{1}\left(-2\left(p-1\right),\frac{d_{1}}{2};\frac{d+2}{2};k\right)\\ &= \frac{d-d_{1}}{d}{}_{2}F_{1}\left(-2\left(p-1\right),\frac{d_{1}}{2};\frac{d+2}{2};k\right)\end{split}$$

This then gives,

$$ARE_{d-d_{1}}(\kappa) = \frac{d}{(d+2(p-1))^{2}} \frac{\left(2F_{1}\left(-(p-1), \frac{d_{1}}{2}; \frac{d}{2}; k\right) + 2\frac{p-1}{d} F_{1}\left(-(p-2), \frac{d_{1}}{2}; \frac{d+2}{2}; k\right)\right)^{2}}{\frac{1}{d^{2}}F_{1}\left(-2(p-1), \frac{d_{1}}{2}; \frac{d+2}{2}; k\right)}$$

For the numerator, identity 15.2.17 from [1] $_2F_1(a,b;c-1;k) - \frac{a}{c-1^2}F_1(a+1,b;c;k) = \frac{c-a-1}{c-1^2}F_1(a,b;c;k)$ with a = -(p-1), $b = \frac{d_1}{2}$ and $c = \frac{d+2}{2}$ yields,

$$\begin{aligned} ARE_{d-d_1}\left(\kappa\right) &= \frac{d}{\left(d+2\left(p-1\right)\right)^2} \frac{\left(\frac{d+2\left(p-1\right)}{d} 2F_1\left(-\left(p-1\right), \frac{d_1}{2}; \frac{d+2}{2}; k\right)\right)^2}{\frac{1}{d^2}F_1\left(-2\left(p-1\right), \frac{d_1}{2}; \frac{d+2}{2}; k\right)} \\ &= \frac{2F_1^2\left(-\left(p-1\right), \frac{d_1}{2}; \frac{d+2}{2}; k\right)}{2F_1\left(-2\left(p-1\right), \frac{d_1}{2}; \frac{d+2}{2}; k\right)} \end{aligned}$$

Substituting for k gives the desired result.

To get the expression for $ARE_{d_1}(\kappa)$ note that the efficiencies of the components would be the same whether the scatter matrix is Λ or $\Lambda^* = diag\left(\underbrace{\lambda^2, \cdots, \lambda^2}_{d_1}, \underbrace{\frac{\lambda^2}{\kappa^2}, \cdots, \frac{\lambda^2}{\kappa^2}}_{d-d_1}\right)$. Interchanging the roles of d_1 and $(d-d_1)$, set $d'_1 = d - d_1$. Substituting κ^{-1} and d'_1 into the expression above gives the desired expression.

$$\begin{aligned} ARE_{d-d_{1}'}\left(\kappa^{-1}\right) &= \frac{{}_{2}F_{1}^{2}\left(-\left(p-1\right),\frac{d_{1}'}{2};\frac{d+2}{2};1-\kappa^{-2}\right)}{{}_{2}F_{1}\left(-2\left(p-1\right),\frac{d_{1}'}{2};\frac{d+2}{2};1-\kappa^{-2}\right)} \\ &= \frac{{}_{2}F_{1}^{2}\left(-\left(p-1\right),\frac{d-d_{1}}{2};\frac{d+2}{2};1-\kappa^{-2}\right)}{{}_{2}F_{1}\left(-2\left(p-1\right),\frac{d-d_{1}}{2};\frac{d+2}{2};1-\kappa^{-2}\right)}\end{aligned}$$

Lemma 5.2.4 For $1 - \frac{d_1}{2} , the function$

$$f_p(x) = \frac{{}_2F_1^2\left(-\left(p-1\right), \frac{d-d_1}{2}; \frac{d+2}{2}; x\right)}{{}_2F_1\left(-2\left(p-1\right), \frac{d-d_1}{2}; \frac{d+2}{2}; x\right)} = \frac{{}_2F_1^2\left(1-p, \frac{d-d_1}{2}; \frac{d+2}{2}; x\right)}{{}_2F_1\left(2\left(1-p\right), \frac{d-d_1}{2}; \frac{d+2}{2}; x\right)}$$

is uniquely maximized at x = 0.

Proof. Note that $f_p(0) = 1$. Taking derivatives of $f_p(x)$, it can be shown that $f'_p(0) = 0$ and $f''_p(0) < 0$, thus x = 0 is indeed a local maximum, however, it is is desired to prove it is a global max. Break this down into two cases, $x \in [0, 1)$ and $x \le 0$.

Case i) $x \in [0, 1)$ The Gauss hypergeometric function is the analytic continuation to the whole complex plane of the Gauss hypergeometric series, thus the the value of the function and the series correspond on the series' radius of convergence in the complex plane, which is $|z| \leq 1$ when $\Re(c - a - b) > 0$ as is the case when $1 - \frac{d_1}{2} . Thus$ $for <math>x \in [0, 1)$, $f_p(x)$ can be written in terms of infinite series,

$$f_p(x) = \frac{{}_2F_1^2\left(1-p,\frac{d-d_1}{2};\frac{d+2}{2};x\right)}{{}_2F_1\left(2\left(1-p\right),\frac{d-d_1}{2};\frac{d+2}{2};x\right)} = \frac{\left(\sum_{n=0}^{\infty} \frac{\left(\frac{d-d_1}{2}\right)_n (1-p)_n}{\left(\frac{d+2}{2}\right)_n n!} x^n\right)^2}{\sum_{n=0}^{\infty} \frac{\left(\frac{d-d_1}{2}\right)_n (2(1-p))_n}{\left(\frac{d+2}{2}\right)_n n!} x^n}$$

where $(x)_n = x (x+1) (x+2) \cdots (x+n-1) = \frac{\Gamma(x+n)}{\Gamma(x)}$ is the Pochhammer symbol.

If it can be shown that the coefficients of the terms of the power series in the denominator are greater or equal to those in the numerator, than it follows that the function $f_p(x)$ is monotonically decreasing as x goes from 0 to 1. However, the numerator is problematic in that it is squared, thus to get at its power series expansion, it is necessary to take the Cauchy product.

Within the radius of convergence, one is guaranteed that series obtained by taking the Cauchy product in the numerator converges to the value of the original series being squared. Recall the Cauchy product of two series is given by,

$$\left(\sum_{n=0}^{\infty} a_n\right) \left(\sum_{n=0}^{\infty} b_n\right) = \sum_{n=0}^{\infty} \sum_{j=0}^{n} a_j b_{n-j}$$

Doing this for the numerator gives

$$\left(\sum_{n=0}^{\infty} \frac{\left(\frac{d-d_1}{2}\right)_n (1-p)_n}{\left(\frac{d+2}{2}\right)_n n!} x^n\right)^2 = \sum_{n=0}^{\infty} \sum_{j=0}^{n} \frac{\left(\frac{d-d_1}{2}\right)_j (1-p)_j}{\left(\frac{d+2}{2}\right)_j j!} x^j \frac{\left(\frac{d-d_1}{2}\right)_{n-j} (1-p)_{n-j}}{\left(\frac{d+2}{2}\right)_{n-j} (n-j)!} x^{n-j}$$

Thus for the coefficients it is desired to show

$$\sum_{j=0}^{n} \frac{\left(\frac{d-d_1}{2}\right)_j (1-p)_j}{\left(\frac{d+2}{2}\right)_j j!} \frac{\left(\frac{d-d_1}{2}\right)_{n-j} (1-p)_{n-j}}{\left(\frac{d+2}{2}\right)_{n-j} (n-j)!} \le \frac{\left(\frac{d-d_1}{2}\right)_n}{\left(\frac{d+2}{2}\right)_n} \frac{(2(1-p))_n}{n!}$$

For n = 0, 1, equality holds. Consider the function

$$h(x) = \frac{\frac{d-d_1}{2} + x}{\frac{d+2}{2} + x}$$

Clearly $\lim_{x\to\infty} h(x) = 1$. Computing its derivative gives

$$h'(x) = \frac{\left(\frac{d+2}{2} + x\right) - \left(\frac{d-d_1}{2} + x\right)}{\left(\frac{d+2}{2} + x\right)^2} = \frac{\frac{d_1+2}{2}}{\left(\frac{d+2}{2} + x\right)^2} > 0$$

for all x. First, it will be shown

$$\frac{\left(\frac{d-d_1}{2}\right)_i}{\left(\frac{d+2}{2}\right)_i}\frac{\left(\frac{d-d_1}{2}\right)_{n-i}}{\left(\frac{d+2}{2}\right)_{n-i}} \le \frac{\left(\frac{d-d_1}{2}\right)_n}{\left(\frac{d+2}{2}\right)_n}$$

Note the on the left hand side, expressions with the subscript i or (n - i) is the product of i or (n - i) terms, thus there are a total of n terms in both the numerator and the denominator, the same is true for the right hand side. Obviously, when i = 0 or i = n, there is equality. However, for all other i the inequality is strict. To see this first consider the case i = 1. This gives,

$$\begin{aligned} \frac{\frac{d-d_1}{2}}{\frac{d+2}{2}} \times \frac{\left(\frac{d-d_1}{2}\right)_{n-1}}{\left(\frac{d+2}{2}\right)_{n-1}} &= h\left(0\right) \frac{\left(\frac{d-d_1}{2}\right)_{n-1}}{\left(\frac{d+2}{2}\right)_{n-1}} \\ &< h\left(n-1\right) \frac{\left(\frac{d-d_1}{2}\right)_{n-1}}{\left(\frac{d+2}{2}\right)_{n-1}} \\ &= \frac{\frac{d-d_1}{2} + n - 1}{\frac{d-d_1}{2} + n - 1} \times \frac{\left(\frac{d-d_1}{2}\right)_{n-1}}{\left(\frac{d+2}{2}\right)_{n-1}} \\ &= \frac{\left(\frac{d-d_1}{2}\right)_n}{\left(\frac{d+2}{2}\right)_n} \end{aligned}$$

The argument is similar for all other i, just continue to pull out terms. This leaves

$$\sum_{j=0}^{n} \frac{\left(\frac{d-d_1}{2}\right)_j (1-p)_j}{\left(\frac{d+2}{2}\right)_j j!} \frac{\left(\frac{d-d_1}{2}\right)_{n-j} (1-p)_{n-j}}{\left(\frac{d+2}{2}\right)_{n-j} (n-j)!} \le \frac{\left(\frac{d-d_1}{2}\right)_n}{\left(\frac{d+2}{2}\right)_n} \sum_{j=0}^{n} \frac{(1-p)_j}{j!} \frac{(1-p)_{n-j}}{(n-j)!}$$

Focus on the sum on the right hand side; rewrite it as

$$\sum_{j=0}^{n} \frac{(1-p)_{j}}{j!} \frac{(1-p)_{n-j}}{(n-j)!} = \sum_{j=0}^{n} \frac{n!}{n!} \frac{\Gamma(1-p+j)}{\Gamma(1-p)j!} \frac{\Gamma(1-p+n-j)}{\Gamma(1-p)(n-j)!} =$$
$$= \sum_{j=0}^{n} {}_{n}C_{j} \frac{1}{n!} \frac{\Gamma(1-p+j)}{\Gamma(1-p)} \frac{\Gamma(1-p+n-j)}{\Gamma(1-p)} =$$

$$=\sum_{j=0}^{n} {}_{n}C_{j}\frac{\Gamma\left(1\right)}{\Gamma\left(n+1\right)}\frac{\Gamma\left(n+2-2p\right)}{\Gamma\left(n+2-2p\right)}\frac{\Gamma\left(2-2p\right)}{\Gamma\left(2-2p\right)}\frac{\Gamma\left(1-p+j\right)}{\Gamma\left(1-p\right)}\frac{\Gamma\left(1-p+n-j\right)}{\Gamma\left(1-p\right)}=$$

$$=\frac{\Gamma\left(n+2-2p\right)}{n!\Gamma\left(2-2p\right)}\sum_{j=0}^{n}{}_{n}C_{j}\frac{B\left(j+1-p,n-j+1-p\right)}{B\left(1-p,1-p\right)}=\frac{\Gamma\left(n+2-2p\right)}{n!\Gamma\left(2-2p\right)}=\frac{(2\left(1-p\right))_{n}}{n!}$$

where ${}_{n}C_{j} = \frac{n!}{j!(n-j)!}$. The terms in the sum give the probability a Beta-Binomial random variable with parameters a = 1 - p and b = 1 - p is equal to j. Thus summing from 0 to n will give 1. We thus have

$$\sum_{j=0}^{n} \frac{\left(\frac{d-d_{1}}{2}\right)_{j} (1-p)_{j}}{\left(\frac{d+2}{2}\right)_{j} j!} \frac{\left(\frac{d-d_{1}}{2}\right)_{n-j} (1-p)_{n-j}}{\left(\frac{d+2}{2}\right)_{n-j} (n-j)!} \leq \frac{\left(\frac{d-d_{1}}{2}\right)_{n}}{\left(\frac{d+2}{2}\right)_{n}} \sum_{j=0}^{n} \frac{(1-p)_{j}}{j!} \frac{(1-p)_{n-j}}{(n-j)!}$$
$$= \frac{\left(\frac{d-d_{1}}{2}\right)_{n}}{\left(\frac{d+2}{2}\right)_{n}} \frac{(2(1-p))_{n}}{n!}$$

The last inequality follows from the fact that when $0 , <math>1 \le 2(1-p) < 2$, hence $\frac{(2(1-p))_n}{n!} \ge 1$.

Case ii) $x \leq 0$ Using the identity $_2F_1(a,b;c;x) = (1-x)^{-a} {}_2F_1\left(a,c-b;c;\frac{x}{x-1}\right)$ yields

$$\frac{{}_{2}F_{1}^{2}\left(1-p,\frac{d-d_{1}}{2};\frac{d+2}{2};x\right)}{{}_{2}F_{1}\left(2\left(1-p\right),\frac{d-d_{1}}{2};\frac{d+2}{2};x\right)} = \frac{\left(\left(1-x\right)^{-\left(1-p\right)}{}_{2}F_{1}\left(1-p,\frac{d+2}{2}-\frac{d-d_{1}}{2};\frac{d+2}{2};\frac{x}{x-1}\right)\right)^{2}}{\left(1-x\right)^{-2\left(1-p\right)}{}_{2}F_{1}\left(2\left(1-p\right),\frac{d+2}{2}-\frac{d-d_{1}}{2};\frac{d+2}{2};\frac{x}{x-1}\right)}$$
$$= \frac{{}_{2}F_{1}^{2}\left(1-p,\frac{d_{1}+2}{2};\frac{d+2}{2};\frac{x}{x-1}\right)}{{}_{2}F_{1}\left(2\left(1-p\right),\frac{d_{1}+2}{2};\frac{d+2}{2};\frac{x}{x-1}\right)}$$
Let $y = \frac{x}{x-1}$, as x goes from $-\infty$ to 0, y goes from 0 to 1. Define the function

$$f(y) = \frac{{}_{2}F_{1}^{2}\left(1-p,\frac{d_{1}+2}{2};\frac{d+2}{2};y\right)}{{}_{2}F_{1}\left(2\left(1-p\right),\frac{d_{1}+2}{2};\frac{d+2}{2};y\right)}$$

The fact this is maximized at y = 0 is shown in the same manner as case (i) with the exception that for this case

$$h(y) = \frac{\frac{d_1+2}{2}+y}{\frac{d+2}{2}+y}$$

Clearly $\lim_{y\to\infty} h(y) = 1$. Computing its derivative gives

$$h'(y) = \frac{\left(\frac{d+2}{2} + y\right) - \left(\frac{d_{1}+2}{2} + y\right)}{\left(\frac{d+2}{2} + y\right)^2} = \frac{\frac{d-d_1}{2}}{\left(\frac{d-d_1}{2} + y\right)^2} > 0$$

for all y. **QED**

The necessary machinery has been established to prove Theorem 3.3.1, however, a similar theorem will be proved for the estimator $\tilde{\mu}_p$ in which the spatial median is a special case $\left(\tilde{\mu}_{1/2} = \tilde{\mu}_{SM}\right)$.

Theorem 5.2.1. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represent an *i.i.d.* sample from $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$ with the conditions necessary for the information inequality to hold. Then

$$AE\left(\widetilde{\boldsymbol{\mu}}_{p};\sigma^{2}\mathbf{I}_{d},G\right) \geq AE\left(\widetilde{\boldsymbol{\mu}}_{p};\boldsymbol{\Sigma},G\right)$$

Furthermore, for $1 - \frac{d_1}{2} , equality holds if and only if <math>\Sigma \propto \mathbf{I}_d$.

Proof. Focus on the ratio

$$\begin{aligned} \frac{AE\left(\widetilde{\boldsymbol{\mu}}_{p};\sigma^{2}\mathbf{I}_{d},G\right)}{AE\left(\widetilde{\boldsymbol{\mu}}_{p};\boldsymbol{\Sigma},G\right)} &= \min_{\mathbf{a}\in\mathbb{R}^{d}}\frac{\alpha\left(G\right)\mathbf{a}^{t}\sigma^{2}\mathbf{I}_{d}\mathbf{a}}{\beta_{p}\left(G\right)\mathbf{a}^{t}\boldsymbol{\mathcal{V}}_{p}\left(\sigma^{2}\mathbf{I}_{d}\right)\mathbf{a}}/\min_{\mathbf{a}\in\mathbb{R}^{d}}\frac{\alpha\left(G\right)\mathbf{a}^{t}\boldsymbol{\Sigma}\mathbf{a}}{\beta_{p}\left(G\right)\mathbf{a}^{t}\mathbf{P}\boldsymbol{\mathcal{V}}_{p}\left(\Lambda\right)\mathbf{P}^{t}\mathbf{a}} \\ &= \min_{\mathbf{a}\in\mathbb{R}^{d}}\frac{\mathbf{a}^{t}\sigma^{2}\mathbf{I}_{d}\mathbf{a}}{\mathbf{a}^{t}\boldsymbol{\mathcal{V}}_{p}\left(\sigma^{2}\mathbf{I}_{d}\right)\mathbf{a}}/\min_{\mathbf{a}\in\mathbb{R}^{d}}\frac{\mathbf{a}^{t}\boldsymbol{\Sigma}\mathbf{a}}{\mathbf{a}^{t}\mathbf{P}\boldsymbol{\mathcal{V}}_{p}\left(\Lambda\right)\mathbf{P}^{t}\mathbf{a}} \end{aligned}$$

since it does not depend on the particular radial distribution, one can consider a particular family of distribution. For convenience, pick the family of elliptical power distributions mentioned in section 2.5.2. When $\Sigma = \sigma^2 \mathbf{I}_d$, $\tilde{\mu}_p$ corresponds to the MLE of μ , thus $AE\left(\tilde{\mu}_p; \sigma^2 \mathbf{I}_d, G_p\right) = 1$ and the inequality is established since $AE\left(\tilde{\mu}_p; \Sigma, G_p\right) \leq 1$. One than proceeds just as in the proof of theorem 3.4.1 to get the general case.

For uniqueness in the case $1 - \frac{d_1}{2} , note that one can write,$

$$AE\left(\widetilde{\boldsymbol{\mu}}_{p};\boldsymbol{\Sigma},G\right) = min_{\mathbf{a}\in\mathbb{R}^{d}}\frac{\mathbf{a}^{t}AV_{MLE}\left(\boldsymbol{\Sigma};G\right)\mathbf{a}}{\mathbf{a}^{t}AV_{\widetilde{\boldsymbol{\mu}}_{r}}\left(\boldsymbol{\Sigma};G\right)\mathbf{a}}$$

Above corresponds to the largest eigenvalue of the matrix $AV_{\tilde{\mu}_p}^{-1}(\Sigma; G) AV_{MLE}(\Sigma; G)$. Writing out these matrices explicitly, one has

$$AV_{\widetilde{\boldsymbol{\mu}}_{p}}^{-1}(\boldsymbol{\Sigma}; G) AV_{MLE}(\boldsymbol{\Sigma}; G) = \left(\beta_{p}(G) \mathbf{P}\boldsymbol{\mathcal{V}}_{p}(\boldsymbol{\Lambda}) \mathbf{P}^{t}\right)^{-1} \left(\alpha(G) \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{t}\right)$$
$$= \frac{\alpha(G)}{\beta_{p}(G)} \mathbf{P}\boldsymbol{\mathcal{V}}_{p}^{-1}(\boldsymbol{\Lambda}) \boldsymbol{\Lambda}\mathbf{P}^{t}$$

As in the previous paragraph, the scatter structure for which the asymptotic efficiency is maximized does not depend on the particular elliptical distribution, G. For reasons that will be apparent as the proof proceeds, multiply the expression on the right by $\frac{d}{(d+2(p-1))^2}/\frac{d}{(d+2(p-1))^2}$. This gives

$$AV_{\widetilde{\boldsymbol{\mu}}_{p}}^{-1}\left(\boldsymbol{\Sigma};G\right)AV_{MLE}\left(\boldsymbol{\Sigma};G\right) = \frac{\alpha\left(G\right)}{\beta_{p}\left(G\right)}\frac{\left(d+2\left(p-1\right)\right)^{2}}{d}\left\{\frac{d}{\left(d+2\left(p-1\right)\right)^{2}}\mathbf{P}\boldsymbol{\mathcal{V}}_{p}^{-1}\left(\boldsymbol{\Lambda}\right)\boldsymbol{\Lambda}\mathbf{P}^{t}\right\}$$

The expression in the braces is the matrix $AV_{\tilde{\mu}_p}^{-1}(\Sigma; G_p) AV_{\tilde{\mu}_{\Sigma,p}}(\Sigma; G_p)$, hence if it can be shown that this expression is uniquely maximized under spherical symmetry, the result will hold for the asymptotic efficiency as well. Furthermore, since the matrices $\mathbf{P}\mathcal{V}_p^{-1}(\mathbf{\Lambda}) \mathbf{\Lambda} \mathbf{P}^t$ and $\mathcal{V}_p^{-1}(\mathbf{\Lambda}) \mathbf{\Lambda}$, have the same eigenvalues, one need only consider the latter simplified case when $\mathbf{P} = \mathbf{I}_d$. The matrix $\mathcal{V}_p^{-1}(\mathbf{\Lambda}) \mathbf{\Lambda}$ is indeed a diagonal matrix, hence its eigenvalues are simply the diagonal components. To this end focus on one of the diagonal components, say the ι^{th} diagonal component. This corresponds to the asymptotic efficiency of the ι^{th} component of the estimator $\tilde{\mu}_p$ to $\tilde{\mu}_{\mathbf{\Lambda},p}$, it is given by,

$$AE_{\iota}\left(\tilde{\mu}_{p},\tilde{\mu}_{\Lambda,p};\Lambda,G_{p}\right) = \frac{dE^{2}\left[\left\|\Lambda^{\frac{1}{2}}\mathbf{u}_{d}\right\|_{2}^{2(p-1)} + 2(p-1)\left\|\Lambda^{\frac{1}{2}}\mathbf{u}_{d}\right\|_{2}^{2(p-2)}\lambda_{\iota}^{2}u_{d,\iota}^{2}\right]}{(d(p+2(p-1)))^{2}E\left[\left\|\Lambda^{\frac{1}{2}}\mathbf{u}_{d}\right\|_{2}^{4(p-1)}\lambda_{\iota}^{2}u_{d,\iota}^{2}\right]}\lambda_{\iota}^{2}$$

In lemma 5.2.2 it is shown that as a multivariate function of $\lambda_1, \ldots, \lambda_d$, the only place where the gradient of the function $AE_{\iota}\left(\tilde{\mu}_p, \tilde{\mu}_{\Lambda,p}; \Lambda, G_p\right)$ can be **0** is when $\lambda_i = \lambda$ for $i \neq \iota$. This is a special case of situation considered in Lemma 5.2.3, where $d_1 = 1$, thus there is a closed form expression for the values of $AE_{\iota}\left(\tilde{\mu}_p, \tilde{\mu}_{\Lambda,p}; \Lambda, G_p\right)$ in terms of Gauss hypergeometric functions in this case. Lemma 5.2.4 however shows that this function is uniquely maximized when $\kappa = 1$, which corresponds to the case of spherical symmetry, thus proving the results. **QED**

Note that one can also obtain similar results to Theorem 2.3.4.1 and Corollary 3.4.1 for oracle and affine versions of the estimator $\tilde{\mu}_p$.

2.5.5 Simulation Results

Below are the proofs of theorems states in section 2.4.1.

Theorem 4.1.1. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represent an i.i.d. sample from $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$. Let $\lambda_{\langle 1 \rangle}^2 > \lambda_{\langle 2 \rangle}^2 > \cdots > \lambda_{\langle m \rangle}^2$ be the distinct eigenvalues of $\boldsymbol{\Sigma}$ where $m \leq d$ is the number of mutually orthogonal eigenspaces of $\boldsymbol{\Sigma}$. For any orthogonally equivariant estimator of $\boldsymbol{\mu}$ based on a sample of size n, $\hat{\boldsymbol{\mu}}_n$, with variance-covariance matrix $\mathbf{V}_n(\hat{\boldsymbol{\mu}}) = Var[\hat{\boldsymbol{\mu}}_n]$, the following are true,

1) $\mathbf{V}_n(\widehat{\boldsymbol{\mu}})$ and $\boldsymbol{\Sigma}$ have the same eigenspaces; consequently they have the same eigenprojections and/or eigenvectors.

2) Let $\lambda_{1,n}^2 \geq \lambda_{2,n}^2 \geq \cdots \geq \lambda_{d,n}^2$, denote the eigenvalues of $\mathbf{V}_n(\widehat{\boldsymbol{\mu}})$. It follows the eigenspace associated with $\lambda_{i,n}^2$ is the same that is associated with λ_i^2 . Consequently, $\lambda_{i,n}^2 = \lambda_{j,n}^2$ if and only if $\lambda_i^2 = \lambda_j^2$.

Proof. Without loss of generality, assume $\boldsymbol{\mu} = \boldsymbol{0}$. Start with the case $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}$. First it will be shown that the $\mathbf{V}_n(\widehat{\boldsymbol{\mu}})$ is a diagonal matrix. Partition $\mathbf{V}_n(\widehat{\boldsymbol{\mu}})$ into

$$\mathbf{V}_{n}\left(\widehat{\boldsymbol{\mu}}\right) = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1d} \\ v_{21} & v_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ v_{d1} & \cdots & \cdots & v_{dd} \end{pmatrix} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{v}_{1} & \mathbf{V}_{12} \\ \mathbf{v}_{1}^{t} & v_{jj} & \mathbf{v}_{2}^{t} \\ \mathbf{V}_{12}^{t} & \mathbf{v}_{2} & \mathbf{V}_{22} \end{pmatrix}$$

Consider the following diagonal orthogonal matrix

$$\mathbf{Q}_{\iota} = \left(egin{array}{ccc} \mathbf{I}_{\iota-1} & & & \ & -1 & & \ & & \mathbf{I}_{d-\iota} \end{array}
ight)$$

Because of orthogonal equivariance, if one uses the estimator, $\hat{\boldsymbol{\mu}}_n$, for the data $\mathbf{x}_i \in \mathbb{R}^d$, $i = 1, \dots, n$, then the corresponding estimator for the data $\mathbf{Q}_i \mathbf{x}_i \in \mathbb{R}^d$, $i = 1, \dots, n$, would be $\mathbf{Q}_i \hat{\boldsymbol{\mu}}_n$. Additionally, $Var\left[\mathbf{Q}_i \hat{\boldsymbol{\mu}}_n\right] = \mathbf{Q}_i Var\left[\hat{\boldsymbol{\mu}}_n\right] \mathbf{Q}_i^t = \mathbf{Q}_i \mathbf{V}_n\left(\hat{\boldsymbol{\mu}}\right) \mathbf{Q}_i^t$. Since $\boldsymbol{\mu} = \mathbf{0}$ and the components of \mathbf{x}_i are uncorrelated it follows $\mathbf{Q}_i \mathbf{x}_i \sim \mathbf{x}_i$, thus $\mathbf{Q}_i \hat{\boldsymbol{\mu}}_n \sim \hat{\boldsymbol{\mu}}_n$. This implies $\mathbf{Q}_i \mathbf{V}_n\left(\hat{\boldsymbol{\mu}}\right) \mathbf{Q}_i^t = \mathbf{V}_n\left(\hat{\boldsymbol{\mu}}\right)$. However,

$$\mathbf{Q}_{\iota}\mathbf{V}_{n}\left(\widehat{\boldsymbol{\mu}}\right)\mathbf{Q}_{\iota}^{t} = \left(\begin{array}{cccc} \mathbf{V}_{11} & -\mathbf{v}_{1} & \mathbf{V}_{12} \\ -\mathbf{v}_{1}^{t} & v_{jj} & -\mathbf{v}_{2}^{t} \\ \mathbf{V}_{12}^{t} & -\mathbf{v}_{2} & \mathbf{V}_{22} \end{array}\right)$$

Thus $\mathbf{v}_1 = -\mathbf{v}_1$ and $\mathbf{v}_2 = -\mathbf{v}_2$ or $\mathbf{v}_1 = \mathbf{0}$ and $\mathbf{v}_2 = \mathbf{0}$, i.e the ι^{th} component of the estimator $\hat{\boldsymbol{\mu}}_n$ is uncorrelated with the rest of the components. Doing this for every $\iota \in \{1, \ldots, d\}$, it can be shown that every off-diagonal component is 0. Hence $\mathbf{V}_n(\hat{\boldsymbol{\mu}})$ is a diagonal matrix, thus the canonical unit vectors $\mathbf{e}_{\iota}, \iota \in \{1, \ldots, d\}$ can be taken as an eigenbasis.

To prove (2), suppose that diagonal elements of Λ (i.e. eigenvalues) λ_j and λ_k are equal. Let $\mathbf{K}_{j,k}$ be the permutation matrix that swaps the j^{th} and k^{th} component of any vector in \mathbb{R}^d ; $\mathbf{K}_{j,k}$ is an orthogonal matrix. It follows that $\mathbf{K}_{j,k}\mathbf{x}_i \sim \mathbf{x}_i$ since the j^{th} and k^{th} component of \mathbf{x}_i both have location parameter 0, are uncorrelated but have the same

scale. Consequently, $\mathbf{K}_{j,k}\widehat{\boldsymbol{\mu}}_n \sim \widehat{\boldsymbol{\mu}}_n$ and as in part (1) we have $\mathbf{K}_{j,k}\mathbf{V}_n(\widehat{\boldsymbol{\mu}})\mathbf{K}_{j,k}^t = \mathbf{V}_n(\widehat{\boldsymbol{\mu}})$. Since, $\mathbf{V}_n(\widehat{\boldsymbol{\mu}})$ is diagonal, pre-multiplying by $\mathbf{K}_{j,k}$ and post-multiplying by $\mathbf{K}_{j,k}^t$ simply swaps the j^{th} and k^{th} diagonal elements. In order for equality to hold, it must be that the j^{th} and k^{th} diagonal elements are equal, thus proving the eigenvalues are equal.

The fact that $\mathbf{V}_n(\hat{\boldsymbol{\mu}})$ and $\boldsymbol{\Lambda}$ can be represented by the same eigenbasis and have repeated eigenvalues at the corresponding indices implies (1), i.e. that $\mathbf{V}_n(\hat{\boldsymbol{\mu}})$ and $\boldsymbol{\Lambda}$ have the same eigenspaces, thus the same eigenprojections and/or eigenvectors.

For the general case, using the spectral decomposition, $\Sigma = \mathbf{P} \mathbf{\Lambda} \mathbf{P}^t$, one can write $\mathbf{\Lambda} = \mathbf{P}^t \Sigma \mathbf{P}$. Suppose $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n \sim \mathcal{E}_d(\mathbf{0}, \mathbf{\Lambda}; G)$, it was shown the results of the theorem hold for $\mathbf{V}_n(\hat{\boldsymbol{\mu}})$ where $\hat{\boldsymbol{\mu}}_n$ is the estimator using the data \mathbf{y}_i , $i = 1, \dots, n$. Let $\mathbf{x}_i = \mathbf{P} \mathbf{y}_i$, $i = 1, \dots, n$; it follows $\mathbf{x}_i \sim \mathcal{E}_d(\mathbf{0}, \Sigma; G)$. $\mathbf{\Lambda}$ and Σ not only have the same eigenvalues, but also the structure of the eigenspaces of Σ can be obtained from $\mathbf{\Lambda}$ by rotating the aforementioned eigenspaces by the orthogonal matrix \mathbf{P} . By orthogonal equivariance, the estimator using the transformed data is $\mathbf{P} \hat{\boldsymbol{\mu}}_n$. Thus, $Var[\mathbf{P} \hat{\boldsymbol{\mu}}_n] = \mathbf{P} Var[\hat{\boldsymbol{\mu}}_n] \mathbf{P}^t = \mathbf{P} \mathbf{V}_n(\hat{\boldsymbol{\mu}}) \mathbf{P}^t$, thus the covariance matrix has the same eigenvalues as the matrix $\mathbf{V}_n(\hat{\boldsymbol{\mu}})$. Furthermore the structure of the eigenspaces by the orthogonal matrix \mathbf{P} . By orthogonal matrix \mathbf{P} , thus implying the general result. **QED**

We have the following corollary.

Corollary 5.5.1 Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represent an i.i.d. sample from $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$. Let $\lambda_{\langle 1 \rangle}^2 > \lambda_{\langle 2 \rangle}^2 > \cdots > \lambda_{\langle m \rangle}^2$ be the distinct eigenvalues of $\boldsymbol{\Sigma}$ where $m \leq d$ is the number of mutually orthogonal eigenspaces of $\boldsymbol{\Sigma}$. For any orthogonally equivariant estimator of $\boldsymbol{\mu}, \, \boldsymbol{\mu}$, with asymptotic variance-covariance matrix $AV_{\boldsymbol{\mu}}$, the following are true,

1) $AV_{\hat{\mu}}$ and Σ have the same eigenspaces; consequently they have the same eigenprojections and/or eigenvectors.

2) Let $\lambda_{1,n}^2 \geq \lambda_{2,n}^2 \geq \cdots \geq \lambda_{d,n}^2$, denote the eigenvalues of $AV_{\widehat{\mu}}$. It follows the eigenspace associated with $\lambda_{i,n}^2$ is the same that is associated with λ_i^2 . Consequently,

$$\lambda_{i,n}^2 = \lambda_{j,n}^2$$
 if and only if $\lambda_i^2 = \lambda_j^2$.

Proof. The proof follows the same way as that of theorem 4.1.1.

Theorem 4.2.1. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represent an i.i.d. sample from $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$, with G absolutely continuous with respect to Lebesgue measure and not possessing finite second moments. For n = 3, the spatial median, oracle spatial median and any affinely equivariant spatial median do not possess finite second moments. This also holds for the case d = 2 and n = 4.

Proof. First consider the case where the sample size is 3, denote the points $\mathbf{x}_1, \mathbf{x}_2$ and \mathbf{x}_3 . Since it is assumed the underlying probability density function is absolutely continuous with respect to Lebesgue measure, the probability they are collinear is 0. Thus For any dimension greater than or equal to 2, the 3 points will always be coplanar. Consequently, they will form a triangle in some 2-dimensional affine subspace. In this instance the point that corresponds to the spatial median is a long solved problem in geometry and is called the Fermat point of the triangle. The Fermat point of a triangle is found as follows. If the largest angle in the triangle is less than $2\pi/3$ radians, than the Fermat point corresponds to the unique interior point, denoted \mathbf{x}^* , such that if one were to draw a line from this point to every vertex, the angle between each line is $2\pi/3$. If one of the angles of the triangle is larger than $2\pi/3$, than the Fermat point is the vertex of that angle.

Define the following events on the sample space.

- E_1 = the angle formed at vertex $\mathbf{x}_1 \ge 2\pi/3$
- E_2 = the angle formed at vertex $\mathbf{x}_2 \ge 2\pi/3$
- E_3 = the angle formed at vertex $\mathbf{x}_3 \ge 2\pi/3$
- E_4 = the angle formed at every vertex is $< 2\pi/3$

These four events form a partition of the sample space (excluding the case of co-linearity which has probability 0). The spatial median is given by the following random vector,

$$\widetilde{\boldsymbol{\mu}}_{SM} = I_{E_1} \mathbf{x}_1 + I_{E_2} \mathbf{x}_2 + I_{E_3} \mathbf{x}_3 + I_{E_4} \mathbf{x}^*$$

where I_A is the indicator function for the event A (i.e. equal to 1 if event A occurs and 0 else). The covariance matrix of this random vector if it existed, would be given by the following expression

$$\mathbf{V}_{3}\left(\widetilde{\boldsymbol{\mu}}_{SM}\right) = Var\left[I_{E_{1}}\mathbf{x}_{1}\right] + Var\left[I_{E_{2}}\mathbf{x}_{2}\right] + Var\left[I_{E_{3}}\mathbf{x}_{3}\right] + Var\left[I_{E_{4}}\mathbf{x}^{*}\right] = 3\boldsymbol{\Sigma} + \boldsymbol{\Sigma}^{*}$$

The covariance matrices $Cov \left[I_{E_i} \mathbf{x}_i, I_{E_j} \mathbf{x}_j\right]$ and $Cov \left[I_{E_i} \mathbf{x}_i, I_{E_4} \mathbf{x}^*\right]$ reduce to the zero matrix because the events $E_i, i = 1, 2, 3, 4$ are all mutually exclusive. Suppose $\mathbf{V}_3(\widetilde{\boldsymbol{\mu}}_{SM})$ had finite second moments. Then for every non-zero vector, $\mathbf{a} \in \mathbb{R}^d$, it must be the case that

$$\mathbf{a}^{t}\mathbf{V}_{3}\left(\widetilde{\boldsymbol{\mu}}_{SM}\right)\mathbf{a} = \mathbf{a}^{t}\left(3\boldsymbol{\Sigma} + \boldsymbol{\Sigma}^{*}\right)\mathbf{a} = 3\mathbf{a}^{t}\boldsymbol{\Sigma}\mathbf{a} + \mathbf{a}^{t}\boldsymbol{\Sigma}^{*}\mathbf{a} < \infty$$

This would necessitate $\mathbf{a}^t \Sigma \mathbf{a} < \infty$, contradicting the fact that \mathbf{x} has finite second moments.

Now take the case the dimension is 2 and there are 4 sample points, $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ and \mathbf{x}_4 . In this situation there are only two possible ways the points can be arranged (again since the probability density function is continuous, the probability they are collinear is 0), the four points could either form a quadrilateral, or one of the points could be in the convex hull of the other three (which consequently form a triangle). The problem of finding the spatial median for four coplanar points is a solved geometric problem. If the four points form a quadrilateral, construct the two lines formed by connecting opposite vertices. These two lines will meet at unique point in the interior of the quadrilateral denoted \mathbf{x}^* . This point is the spatial median of the four points. If one of the points lies in the convex hull of the other three, the spatial median corresponds to this point in the convex hull.

Define the following events on the sample space.

 $E_1 = \mathbf{x}_1$ is in the convex hull formed by the other three points $E_2 = \mathbf{x}_2$ is in the convex hull formed by the other three points $E_3 = \mathbf{x}_3$ is in the convex hull formed by the other three points $E_4 = \mathbf{x}_4$ is in the convex hull formed by the other three points $E_5 =$ the four points form a quadrilateral

These five events form a partition of the sample space (excluding the case of collinearity which has probability 0). The spatial median is now given by the following random vector,

$$\widetilde{\mu}_{SM} = I_{E_1} \mathbf{x}_1 + I_{E_2} \mathbf{x}_2 + I_{E_3} \mathbf{x}_3 + I_{E_4} \mathbf{x}_4 + I_{E_5} \mathbf{x}^*$$

Arguing similarly, its covariance matrix, if it existed, is given by,

$$\mathbf{V}_4\left(\widetilde{\boldsymbol{\mu}}_{SM}\right) = 4\boldsymbol{\Sigma} + \boldsymbol{\Sigma}^*$$

Suppose $\mathbf{V}_4(\widetilde{\boldsymbol{\mu}}_{SM})$ had finite second moments, it would then follow that $\mathbf{a}^t \boldsymbol{\Sigma} \mathbf{a} < \infty$, a contradiction.

To get the result for the oracle spatial median or affine spatial median, recall that one can calculate there via the transform-retransform method. Thus in both cases it would reduce to first finding the spatial median for the data $\mathbf{z}_i = \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{x}_i$. Applying the same arguments used above, one could show that the spatial median for the \mathbf{z}_i will not have a finite second moments, thus nor will the oracle spatial median or affine spatial median.

Chapter 3

The Spatial Sign Covariance Matrix

3.1 Introduction

One reason multivariate procedures are implemented is to understand the relationships between several quantitative variables of interest on the same experimental unit. For the case the multivariate random vector arises from some elliptical distribution, the scatter matrix, Σ , is the parameter that dictates the variability and relationships between the components of the vector. Consequently, many methods involve obtaining estimates of this parameter or characteristics of this parameter. Perhaps the most commonly used estimator for Σ is the sample covariance matrix. Despite its popularity, its limitations are just as infamous. In particular, the sample covariance matrix is sensitive to outliers; a few non-typical observations can lead to an estimate that is in reality not close to Σ . This fact is particularly problematic when the data arises from distributions that have wider tails than the multivariate Gaussian distribution, a situation where extreme observations are more likely. Consequently, much research has focused on developing estimators that still provide suitable estimates of Σ despite the fact outliers are present.

One estimator proposed as an alternative to the sample covariance matrix is Tyler's scatter matrix [35]. Tyler's matrix is an example of an affinely equivariant M-estimate of scatter. While affinely equivariant M-estimators of scatter address the problem of not being affected by outliers, i.e. are robust estimators of Σ , this does not hold for every situation. As mentioned in the introduction, when $n \leq d+1$, it can be shown that any affinely equivariant estimator of scatter is indeed the sample covariance matrix [38]. To handle the case when $n \leq d+1$, scatter estimators that sacrifice affine equivariance have been proposed; one such instance is the the spatial sign covariance matrix (SSCM). The SSCM possesses the more restrictive property of orthogonal equivariance. In its calculation, the SSCM down-weighs observations based on their Euclidean distance from the estimated center of data. On the other hand, Tyler's matrix down-weighs observations based on their Mahalanobis distance from the estimated center of data. If the data arises from some elliptical distribution, the previous fact implies the Tyler matrix is down-weighing observations in accordance with the likelihood that observations will occur. Thus one might conjecture that under the elliptical model with $\Sigma \not\propto \mathbf{I}_d$ Tyler's scatter matrix outperforms the SSCM.

Many procedures are not interested in Σ but rather its eigenprojections and eigenvectors, PCA being an example, thus estimation of these is the primary goal. Under the elliptical model, one might surmise that if Tyler's matrix outperforms the SSCM in terms of estimating Σ , then an eigenprojection/eigenvector of the Tyler matrix would outperform the corresponding eigenprojection/eigenvector of the SSCM as an estimate of the corresponding eigenprojection/eigenvector of Σ . This chapter of the dissertation focuses on comparing Tyler's matrix to the SSCM when the goal is estimating eigenprojections and eigenvectors of Σ under the elliptical model. Section 2 discusses Tyler's matrix and its significance as the MLE for Σ when the data follows an angular central Gaussian distribution. In section 3, the SSCM is reviewed in more detail. Section 4 presents the main theoretic results regarding Tyler's matrix and SSCM under the elliptical model, namely the superiority of the Tyler matrix to the SSCM when the goal is estimation of eigenprojections or eigenvectors. In order to quantify exactly how superior Tyler's matrix is compared to the SSCM under the elliptical model for estimating eigenprojections a specific, yet informative, scatter structure is considered that admits a non-arbitrary definition of asymptotic relative efficiencies for these estimators. Calculation of these asymptotic relative efficiencies under this scenario are then presented. This is all included in section 5. In section 6, simulations are implemented to compare the performance of the eigenprojection matrices based on Tyler's matrix and the SSCM under the elliptical model. To do so, a measure of relative efficiency is defined based on principal angles; the first part of section 6 reviews this concept. Discussion of how the simulations were implemented and the results are then presented.

3.2 Tyler's Scatter Matrix

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be a multivariate sample. In [35], the author proposed using the solution to the following implicit equation as an estimate of the scatter matrix of the the data

$$\widehat{\boldsymbol{\mathcal{T}}}_{\widehat{\boldsymbol{\mu}}_n} = \frac{d}{n} \sum_{i=1}^n \frac{(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_n) (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_n)^t}{(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_n)^t \, \widehat{\boldsymbol{\mathcal{T}}}_{\widehat{\boldsymbol{\mu}}_n}^{-1} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_n)}$$

where $\hat{\mu}_n$ is some estimator of location. The proof of the existence of the estimate involves showing it is the limiting point of a specific algorithm, thus providing a means to obtain the estimate. For the case the data comes from some elliptical distribution $\mathcal{E}_d(\mu, \Sigma; G)$ and $\hat{\mu}$ is a consistent estimator of μ , the asymptotic distribution of the estimator $\hat{\mathcal{T}}_{\hat{\mu}}$ does not depend on the particular elliptical distribution (i.e. it does not depend on G). In addition, for the case that μ is known (i.e. $\hat{\mu}_n = \mu$) the finite sample distribution of the Tyler matrix also does not depend on G. For this special case, the notation will be $\hat{\mathcal{T}}_n$, where it is understood that the estimator of location is the known location vector, μ . Tyler's matrix is the most robust estimate of the scatter matrix for elliptical distributions in that it minimizes the maximum asymptotic variance.

An important characterization of Tyler's matrix is that it is the MLE for Σ when the data comes from an angular central Gaussian distribution [36]. The probability density function of the angular central Gaussian distribution is given by,

$$f(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\delta_1) = \frac{\Gamma(d/2)}{2\pi^{d/2}} \left(\mathbf{x}-\boldsymbol{\mu}\right)^t \boldsymbol{\Sigma}^{-1} \left(\mathbf{x}-\boldsymbol{\mu}\right) \mathbb{I}_{\left\{\|\mathbf{x}\|_2=1\right\}}$$

where $\mathbb{I}_{\{A\}}$ is the indicator function for the set A and $\hat{\mu}_n$ is some estimator of location. Being a probability distribution on S_d , in its stochastic representation, the radial component assigns all the mass to the value 1, this distribution is the Dirac delta measure at x = 1. Thus for a random variable following an angular Gaussian distribution, one will write $\mathbf{x} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \delta_1)$.

3.3 Spatial Sign Covariance Matrix

Given a multivariate sample $\mathbf{x}_1, \ldots, \mathbf{x}_n$, the spatial sign covariance matrix, or SSCM, is given by

$$\widehat{\boldsymbol{\mathcal{S}}}_{\widehat{\boldsymbol{\mu}}_n} = \frac{1}{n} \sum_{i=1}^n \frac{\left(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_n\right) \left(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_n\right)^t}{\left(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_n\right)^t \left(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_n\right)}$$

where $\hat{\mu}_n$ is some estimator of location. Of note is that $Tr\left(\hat{\boldsymbol{S}}_{\hat{\mu}_n}\right) = 1$. Similar to Tyler's matrix, under the elliptical model, the asymptotic distribution of the SSCM does not depend on G provided $\hat{\mu}$ is a consistent estimate of μ . Additionally, if μ is known, then the finite sample distribution also does not depend on G; for this situation the notation $\hat{\boldsymbol{S}}_n$ will be utilized. Unlike Tyler's matrix, this estimate of scatter has a closed form solution. This estimate of scatter was first introduced in [3], though the name spatial sign covariance matrix was coined by Visuri in [40]. Since the asymptotic distribution of the SSCM under the elliptical model does not depend on $\hat{\mu}$ provided $\hat{\mu}$ is a consistent estimate of μ , one can assume that μ is known. For this situation let $E\left[\hat{\boldsymbol{S}}_n\right] = \boldsymbol{\Xi}$, notice this is the same for all n. In general, $\boldsymbol{\Xi} \not\propto \boldsymbol{\Sigma}$, thus the SSCM is a biased estimate of $\boldsymbol{\Sigma}$ both for finite samples and asymptotically. In particular, it can be shown the eigenvalues of the SSCM are biased estimates of the eigenvalues of $\boldsymbol{\Sigma}$.

3.4 Theoretical Results

3.4.1 Optimality of the Tyler Matrix under the Elliptical Model

As mentioned in the introduction, since the SSCM is only orthogonally equivariant, it is speculated that for the elliptical model, Tyler's matrix would be a better estimator of Σ . For many multivariate statistical applications, the focus is not estimating Σ but rather its eigenprojections/eigenvectors. As mentioned, the SSCM is a biased estimator of Σ , however it is shown in section 3.6.3 the appendix that $\Xi = \mathbf{P} \Delta \mathbf{P}^t$ where Δ is a diagonal matrix. Consequently, the eigenprojections of the SSCM have been proposed in applications that require estimating the eigenprojections/eigenvectors of Σ . Most notably, [17] and [22] suggested its use in PCA. Since the Tyler matrix is a superior estimator of Σ than the SSCM under the elliptical model, one might suspect that eigenprojections and eigenvectors of the Tyler matrix would be better estimators than the corresponding ones of the the SSCM.

Before stating this formally in a theorem, the following notation must be established. Recall, for the scatter matrix Σ , the eigenprojection associated with the eigenvalue λ_i^2 will be denoted \mathbf{P}_i . For the matrix $\mathbf{\Xi}$, let $\delta_1^2 \geq \cdots \geq \delta_d^2$ denote its eigenvalues. It is shown in the Appendix that \mathbf{P}_i will also be the eigenprojection associated with the eigenvalue δ_i^2 . It is also shown, the eigenvalues λ_i^2 and δ_i^2 will have the same geometric multiplicities, denote the geometric multiplicities as d_1, \ldots, d_d . For any subset of indices $\mathcal{N} = \{h, h+1, \dots, h+p-1\}$, the total eigenprojection of either Σ or Ξ associated with the eigenvalues associated with these indices is $\mathbf{P}_{\mathcal{N}} = \sum_{i=h}^{h+p-1} \frac{1}{d_i} \mathbf{P}_i$. For the scatter estimate $\hat{\mathcal{T}}_{\hat{\mu}}$, let $\hat{\mathbf{T}}_i$ be the eigenprojection associated with its i^{th} largest eigenvalue. For multivariate distributions absolutely continuous with respect to Lebesgue measure in \mathbb{R}^d , the eigenvalues of $\widehat{\mathcal{T}}_{\widehat{\mu}}$ will be distinct with probability 1. It follows that $\widehat{\mathbf{T}}_{\mathcal{N}} =$ $\sum_{i=h}^{h+p-1} \widehat{\mathbf{T}}_i$ is an estimator of the eigenprojection $\mathbf{P}_{\mathcal{N}}$. Similarly, for the scatter estimate $\widehat{\boldsymbol{\mathcal{S}}}_{\widehat{\boldsymbol{\mu}}}$, let $\widehat{\mathbf{S}}_i$ be the eigenprojection associated with its i^{th} largest eigenvalue. Again, for multivariate distributions absolutely continuous with respect to Lebesgue measure in \mathbb{R}^d , the eigenvalues of $\widehat{S}_{\widehat{\mu}}$ will be distinct with probability 1. It follows that $\widehat{S}_{\mathcal{N}} =$ $\sum_{i=h}^{h+p-1} \widehat{\mathbf{S}}_i$ is an estimator of the eigenprojection $\mathbf{P}_{\mathcal{N}}$.

Much literature has been devoted to the distributional properties of the eigenvalues and eigenvectors of a scatter estimator. The results presented in this part utilize methods introduced in [33]. In his dissertation, perturbation methods were used to derive Taylor series expansions of projection matrices into the eigenspaces of a matrix. In [34], the author implements these methods to derive the asymptotic distributions of projection matrices into the eigenspaces of scatter estimates. A review of these methods is relegated to section 3.6.1 of the appendix. These methods are used to prove subsequent theorems pertaining to estimation of the eigenprojections and eigenvectors of a scatter matrix estimate. The results of [34] imply

$$\sqrt{n}\left(\operatorname{\mathbf{vec}}\left(\widehat{\mathbf{T}}_{\mathcal{N}}-\mathbf{P}_{\mathcal{N}}\right)\right) \rightarrow_{\mathcal{D}} Norm_{d^{2}}\left(\mathbf{0},AV_{\widehat{\mathbf{T}}_{\mathcal{N}}}\left(\mathbf{\Sigma}\right)\right)$$

and

$$\sqrt{n}\left(\operatorname{vec}\left(\widehat{\mathbf{S}}_{\mathcal{N}}-\mathbf{P}_{\mathcal{N}}\right)\right) \rightarrow_{\mathcal{D}} Norm_{d^{2}}\left(\mathbf{0}, AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}\left(\mathbf{\Sigma}\right)\right)$$

where $AV_{\widehat{\mathbf{T}}_{\mathcal{N}}}(\mathbf{\Sigma})$ and $AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}(\mathbf{\Sigma})$ are the asymptotic variance-covariance matrices of the estimators $\widehat{\mathbf{T}}_{\mathcal{N}}$ and $\widehat{\mathbf{S}}_{\mathcal{N}}$ respectively. Note that these depend on the underlying scatter matrix of the data, $\mathbf{\Sigma}$. However, an important caveat is that the aforementioned asymptotic variance-covariance matrices are not of full rank. The form of these matrices is described in the Appendix. Before the main result is proven, the following lemma is needed the proof of which is relegated to section 3.6.2 and 3.6.3 of the appendix.

Lemma 4.1.1. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represent an i.i.d. sample from $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$ and let $\widehat{\boldsymbol{\mu}}$ be an asymptotically consistent estimator of $\boldsymbol{\mu}$. It then follows that $AV_{\widehat{\boldsymbol{\mathcal{S}}}_{\widehat{\boldsymbol{\mu}}}}(\sigma^2 \boldsymbol{I}_d) \propto AV_{\widehat{\boldsymbol{\mathcal{T}}}_{\widehat{\boldsymbol{\mu}}}}(\sigma^2 \boldsymbol{I}_d)$.

Proof. See section 3.6.3. of the appendix.

Theorem 4.1.1. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represent an *i.i.d.* sample from $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$ and let $\hat{\boldsymbol{\mu}}$ be an asymptotically consistent estimator of $\boldsymbol{\mu}$. For $\widehat{\mathbf{T}}_{\mathcal{N}}$ and $\widehat{\mathbf{S}}_{\mathcal{N}}$, estimators of $\mathbf{P}_{\mathcal{N}}$, the total eigenprojection of $\boldsymbol{\Sigma}$ associated with the eigenvalues indexed by \mathcal{N} , it follows

$$AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}(\mathbf{\Sigma}) \ge AV_{\widehat{\mathbf{T}}_{\mathcal{N}}}(\mathbf{\Sigma})$$

Furthermore, as $\Sigma \to \sigma^2 \mathbf{I}_d$, the above expression approaches equality.

Proof. Since the asymptotic distribution of either estimator does not depend on G, one can choose a convenient elliptical family to work with. In particular choose the family of angular central Gaussian distributions, i.e. $G = \delta_1$. Recall that in this

situation, the Tyler matrix is the MLE, thus it attains the Cramer-Rao lower bound. Further note that the eigenprojections of a matrix are attainable as functions of the matrix of interest, i.e. Σ in our case. By the invariance properties of the MLE, it follows that $\hat{\mathbf{T}}_{\mathcal{N}}$ is the MLE for $\mathbf{P}_{\mathcal{N}}$ under the angular central Gaussian model. Thus the matrix $AV_{\hat{\mathbf{T}}_{\mathcal{N}}}(\Sigma)$ is less than or equal to the asymptotic variance-covariance matrix of any other asymptotically unbiased estimator of $\mathbf{P}_{\mathcal{N}}$. The eigenprojection matrices of the SSCM are asymptotically unbiased estimators of $\mathbf{P}_{\mathcal{N}}$, thus proving the inequality. By Lemma 4.1.1., as $\Sigma \to \sigma^2 \mathbf{I}_d$, it follows $AV_{\hat{\mathbf{S}}_{\hat{\mu}}}(\Sigma)$ approaches a matrix that is proportional to $AV_{\hat{\mathcal{T}}_{\hat{\mu}}}(\Sigma)$. This implies the matrices $AV_{\hat{\mathbf{T}}_{\mathcal{N}}}(\Sigma)$ and $AV_{\hat{\mathbf{S}}_{\mathcal{N}}}(\Sigma)$ approach equality (See section 3.6.3 of the appendix for details). **QED**

When sampling from elliptical distributions, the above theorem implies that asymptotically it is better to use Tyler's matrix to obtain estimates of eigenprojections as opposed to the SSCM. In the special case $\lambda_i^2 \neq \lambda_i^2$, for $i \neq \iota$, the geometric multiplicity associated with the eigenvalue λ_i^2 is 1, thus there is a unique eigenvector (unique up to multiplication by ± 1) associated with the eigenvalue λ_i^2 . In this paradigm, one is not so interested in the eigenprojection associated with λ_i^2 as they are the eigenvector itself. To this end, let $\mathbf{p}_1, \ldots, \mathbf{p}_d$ be a set of eigenvectors for $\boldsymbol{\Sigma}$. Let $\hat{\mathbf{t}}_i$ and $\hat{\mathbf{s}}_i$, $i = 1, \ldots, d$, be a set of eigenvectors for the scatter estimates $\hat{\mathcal{T}}_{\hat{\mu}}$ and $\hat{\boldsymbol{S}}_{\hat{\mu}}$ respectively. Again, utilizing results from [34] one has

$$\sqrt{n}\left(\widehat{\mathbf{t}}_{\iota}-\mathbf{p}_{\iota}\right) \rightarrow_{\mathcal{D}} Norm_{d}\left(\mathbf{0}, AV_{\widehat{\mathbf{t}}_{\iota}}\left(\mathbf{\Sigma}\right)\right)$$

and

$$\sqrt{n} \left(\widehat{\mathbf{s}}_{\iota} - \mathbf{p}_{\iota} \right) \rightarrow_{\mathcal{D}} Norm_d \left(\mathbf{0}, AV_{\widehat{\mathbf{s}}_{\iota}} \left(\mathbf{\Sigma} \right) \right)$$

where $AV_{\hat{\mathbf{t}}_{\iota}}(\boldsymbol{\Sigma})$ and $AV_{\hat{\mathbf{s}}_{\iota}}(\boldsymbol{\Sigma})$ are the asymptotic variance-covariance matrices of the estimators $\hat{\mathbf{t}}_{\iota}$ and $\hat{\mathbf{s}}_{\iota}$ respectively. Again note that these depend on the underlying scatter matrix of the data, $\boldsymbol{\Sigma}$. It can be shown the rank of the aforementioned asymptotic variance-covariance matrices is d-1, thus the limiting distributions are in fact degenerate distributions that lie in a d-1 dimensional subspace. The following theorem pertaining the the asymptotic distribution of eigenvector estimates based on the scatter estimates $\hat{\mathcal{T}}_{\hat{\mu}}$ and $\hat{\mathcal{S}}_{\hat{\mu}}$ is presented.

Theorem 4.1.2. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represent an *i.i.d.* sample from $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$ and let $\hat{\boldsymbol{\mu}}$ be an asymptotically consistent estimator of $\boldsymbol{\mu}$. If $\lambda_{\iota}^2 \neq \lambda_i^2$, for $i \neq \iota$ it follows

$$AV_{\widehat{\mathbf{s}}_{\iota}}(\mathbf{\Sigma}) \geq AV_{\widehat{\mathbf{t}}_{\iota}}(\mathbf{\Sigma})$$

Equality holds when the vector under consideration for the quadratic forms is \mathbf{p}_{ι} . Furthermore, as $\mathbf{\Sigma} \to \sigma^2 \mathbf{I}_d$, the above expression approaches equality for any vector considered in the quadratic forms.

Proof. This is proved in an analogous manner to Theorem 4.1.1. The modification is the fact that the function of Σ being estimated is no longer the eigenprojection \mathbf{P}_{ι} , but rather the eigenvector \mathbf{p}_{ι} . **QED**

3.4.2 Asymptotic Calculations

The results of the previous section showed that under the elliptical model, when the goal is estimating the eigenprojections/eigenvectors of Σ the corresponding eigenprojections/eigenvectors of Tyler's matrix have a smaller asymptotic variance than those of the SSCM. Furthermore, it was shown that for a given eigenprojection/eigenvector of interest, when the underlying elliptical distribution approaches spherical symmetry, these two estimators tend to the same estimator. To get a better sense of how much better the eigenprojection estimator $\widehat{\mathbf{T}}_{\mathcal{N}}$ is compared to $\widehat{\mathbf{S}}_{\mathcal{N}}$, the matrices $AV_{\widehat{\mathbf{T}}_{\mathcal{N}}}(\Sigma)$ and $AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}(\Sigma)$ will be computed for specific scatter structures. Unfortunately, there is not a general consensus as to how to compare scatter estimators or their eigenprojection matrices and/or eigenvectors. Fortunately, the specific scatter structures considered admit a non-arbitrary means to compare them. The scatter structures considered are

of the form,

$$\Sigma = \Lambda_0 = diag\left(\underbrace{\lambda^2, \dots, \lambda^2}_{d_1}, \underbrace{r^2\lambda^2, \dots, r^2\lambda^2}_{d-d_1}\right)$$

for $0 \leq r \leq 1$. Note that this scatter structure has at most two eigenvalues; the eigenvalue λ^2 corresponds to the eigenspace $\mathbf{E}_{D_1} = \sum_{i=1}^{d_1} \mathbf{E}_i$ and the eigenvalue $r^2 \lambda^2$ corresponds to the eigenspace $\mathbf{E}_{D_1}^{\perp} = \mathbf{I}_d - \mathbf{E}_{D_1} = \sum_{i=d_1+1}^{d} \mathbf{E}_i$ where $\mathbf{E}_i = \mathbf{e}_i \mathbf{e}_i^t$ with \mathbf{e}_i being the i^{th} canonical unit vector. Small values of r correspond to the situation where the majority of the variation of the data lies in a d_1 dimensional subspace of \mathbb{R}^d . Let $\widehat{\mathbf{T}}_{\mathcal{N}}$ and $\widehat{\mathbf{S}}_{\mathcal{N}}$ be estimators of \mathbf{E}_{D_1} . Under this scatter structure, it is shown in the appendix that

$$AV_{\widehat{\mathbf{T}}_{\mathcal{N}}}\left(\mathbf{\Lambda}_{0}\right) = \sum_{i=1}^{d_{1}} \sum_{j=d_{1}+1}^{d} \alpha_{\mathcal{T}_{d,d_{1}}}\left(\mathbf{e}_{i}\mathbf{e}_{i}^{t} \otimes \mathbf{e}_{j}\mathbf{e}_{j}^{t} + \mathbf{e}_{i}\mathbf{e}_{j}^{t} \otimes \mathbf{e}_{j}\mathbf{e}_{i}^{t} + \mathbf{e}_{j}\mathbf{e}_{i}^{t} \otimes \mathbf{e}_{i}\mathbf{e}_{i}^{t} + \mathbf{e}_{j}\mathbf{e}_{j}^{t} \otimes \mathbf{e}_{i}\mathbf{e}_{i}^{t}\right)$$

$$AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}\left(\mathbf{\Lambda}_{0}\right) = \sum_{i=1}^{d_{1}} \sum_{j=d_{1}+1}^{d} \alpha_{\mathbf{S}_{d,d_{1}}} \left(\mathbf{e}_{i}\mathbf{e}_{i}^{t} \otimes \mathbf{e}_{j}\mathbf{e}_{j}^{t} + \mathbf{e}_{i}\mathbf{e}_{j}^{t} \otimes \mathbf{e}_{j}\mathbf{e}_{i}^{t} + \mathbf{e}_{j}\mathbf{e}_{i}^{t} \otimes \mathbf{e}_{i}\mathbf{e}_{i}^{t} + \mathbf{e}_{j}\mathbf{e}_{j}^{t} \otimes \mathbf{e}_{i}\mathbf{e}_{i}^{t}\right)$$

where

$$\alpha \tau_{d,d_1} = \frac{d+2}{d} \frac{r^2}{(1-r^2)^2}$$

and

$$\alpha_{\boldsymbol{\mathcal{S}}_{d,d_1}} = \frac{\frac{d}{d+2}r^2 \,_2F_1\left(2,\frac{d_1+2}{2};\frac{d+4}{2};1-r^2\right)}{\left({}_2F_1\left(1,\frac{d_1+2}{2};\frac{d+2}{2};1-r^2\right) - r^2 \,_2F_1\left(1,\frac{d_1}{2};\frac{d+2}{2};1-r^2\right)\right)^2}$$

These above forms imply $AV_{\widehat{\mathbf{T}}_{\mathcal{N}}}(\mathbf{\Lambda}_0)$ and $AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}(\mathbf{\Lambda}_0)$ only differ by a constant, thus a natural method of comparing the two is to compare these constants. In particular, the asymptotic relative efficiency of $\widehat{\mathbf{S}}_{\mathcal{N}}$ to $\widehat{\mathbf{T}}_{\mathcal{N}}$ will reduce to the ratio,



Figure 3.1: Asymptotic Relative Efficiencies of the Eigenprojection Estimate of the SSCM to the Corresponding Eigenprojection Estimate of the Tyler Matrix in \mathbb{R}^2

$$ARE_{d,d_{1}}\left(\widehat{\mathbf{S}}_{\mathcal{N}},\widehat{\mathbf{T}}_{\mathcal{N}},r\right) = \frac{\frac{d+2}{d}\frac{r^{2}}{(1-r^{2})^{2}}}{\frac{d}{d+2}r^{2}{}_{2}F_{1}\left(2,\frac{d_{1}+2}{2};\frac{d+4}{2};1-r^{2}\right)}}{\left({}_{2}F_{1}\left(1,\frac{d_{1}+2}{2};\frac{d+2}{2};1-r^{2}\right)-r^{2}{}_{2}F_{1}\left(1,\frac{d_{1}+2}{2};\frac{d+4}{2};1-r^{2}\right)}\right)^{2}}$$
$$= \frac{2F_{1}^{2}\left(1,\frac{d_{1}+2}{2};\frac{d+4}{2};1-r^{2}\right)}{{}_{2}F_{1}\left(2,\frac{d_{1}+2}{2};\frac{d+4}{2};1-r^{2}\right)}$$

The last equality is shown in section 3.7 of the appendix.

As mentioned, $\widehat{\mathbf{T}}_{\mathcal{N}}$ and $\widehat{\mathbf{S}}_{\mathcal{N}}$ are estimators of \mathbf{E}_{D_1} ; it follows that $(\mathbf{I}_d - \widehat{\mathbf{T}}_{\mathcal{N}})$ and $(\mathbf{I}_d - \widehat{\mathbf{S}}_{\mathcal{N}})$ are estimators of \mathbf{E}_{D_1} . Consequently, since the estimators of \mathbf{E}_{D_1} and $\mathbf{E}_{D_1}^{\perp}$, only differ by the constant matrix \mathbf{I}_d , it will be that $AV_{\widehat{\mathbf{T}}_{\mathcal{N}}}(\mathbf{\Lambda}_0) = AV_{\mathbf{I}_d - \widehat{\mathbf{T}}_{\mathcal{N}}}(\mathbf{\Lambda}_0)$ and $AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}(\mathbf{\Lambda}_0) = AV_{\mathbf{I}_d - \widehat{\mathbf{S}}_{\mathcal{N}}}(\mathbf{\Lambda}_0)$, thus one need only consider the asymptotic efficiency for one of the eigenprojections.

Plotted in Figure 3.1 is $ARE_{d,d_1}\left(\widehat{\mathbf{S}}_{\mathcal{N}}, \widehat{\mathbf{T}}_{\mathcal{N}}, r\right)$ in two dimensions. The asymptotic relative efficiency is quite low when r is close to 0. This indicates the drastic inferiority of the SSCM to the Tyler matrix in estimating \mathbf{E}_{D_1} for nearly singular scatter structures.

In three dimensions, the two situations to consider are $\lambda_1 = \lambda_2$ and $\lambda_2 = \lambda_3$, i.e. $d_1 = 2$ and $d_1 = 1$ respectively. Presented in Figure 3.2 on the following page are the



Figure 3.2: Asymptotic Relative Efficiencies of the Eigenprojection Estimate of the SSCM to the Corresponding Eigenprojection Estimate of the Tyler Matrix in \mathbb{R}^3

asymptotic relative efficiencies under both scatter structures as a function of r. Just as in two dimensions, the asymptotic relative efficiency of \mathbf{E}_{D_1} is low when r is close to 0. Of the two scatter structures considered, the one that is more deleterious in terms of asymptotic efficiencies is $\mathbf{\Lambda}_0 = diag(1, r^2, r^2)$.

In order to investigate how the dimension of the eigenspace of the eigenvalue of interest affects the asymptotic relative efficiency, the following scatter structures will be considered in 10 dimensions.

Plotted in Figure 3.3 are the asymptotic relative efficiencies for the above scatter structures. The larger the dimension of the eigenspace associated with the larger eigenvalue,



Figure 3.3: Asymptotic Relative Efficiencies of the Eigenprojection Estimate of the SSCM to the Corresponding Eigenprojection Estimate of the Tyler Matrix in \mathbb{R}^{10}

the higher the asymptotic relative efficiency. This indicates not much is lost by using the corresponding eigenprojection estimate of the SSCM, even under nearly singular scatter structures. However, as the dimension of the eigenspace associated with the larger eigenvalue, the asymptotic relative efficiencies steadily decreases, in some instances being quite poor for nearly singular scatter structures.

3.5 Finite Sample Performance

3.5.1 Finite Sample Theory

The results of the previous section showed that asymptotically, the inefficiency of the SSCM compared to to the Tyler's matrix for estimating the eigenprojections of the scatter matrix Λ_0 can be quite severe under the elliptical model when Λ_0 is far from singular. However, for practical purposes, it is desired to understand how severe the inefficiency can be in the finite sample setting. Working out the finite sample distributions of the aforementioned scatter estimates, let alone the distribution of their eigenprojections and eigenvectors, is quite formidable. Furthermore, in the finite sample setting, for a given distribution there are two factors to consider when studying the relative efficiency of the SSCM to Tyler's matrix: the first being the effect of the scatter structure of the data and the second being the particular estimator used for the location vector.

Fortunately, under the elliptical model in the situation where the location vector is known a priori, the only factor one need consider is the scatter structure of the data. As mentioned in the sections discussing these estimators, when the known location vector is used in the construction of these estimators, the finite sample distributions of these estimators does not depend on the particular elliptical distribution (i.e. does not depend on G) hence nor will the finite sample distributions of their eigenprojections or eigenvectors. Unfortunately, the exact form of these distributions is not known, so one must still resort to simulations to ascertain their behavior. However the previous fact dictates that one only need to implement simulations for one distribution. While it is hardly ever the case that one will know the location vector a priori, considering this is beyond the scope of this dissertation. Consequently, only the effect of the scatter structures of the data will be considered in these simulations, leaving the point of estimating the location parameter as a future endeavor.

As mentioned in the last section, there is not a general consensus on how best to compare scatter estimates, let alone eigenprojection estimates. Fortunately, by considering a simplified (yet still informative) scatter structure, it was shown that the asymptotic variance-covariance matrices of the eigenprojection estimators obtained from Tyler's matrix and the SSCM differed only by a constant under the elliptical model, thus reducing the problem of comparing the two to a single number. This aforementioned scatter structure was when there were only two principal component spaces associated with Λ . Unfortunately, the finite sample variance-covariance matrices of these aforementioned estimators do not possess the same form as their asymptotic counter-parts, even when considering the same scatter structures under the elliptical model. While the form of the finite sample variance-covariance matrices for both the aforementioned estimators can be worked out under the elliptical model using symmetry and equivariance arguments, a different means will used to compare them. The means to compare them will be based on the idea of principal (canonical) angles, thus this concept will be discussed.

The notation used in the proceeding paragraphs will be consistent with that utilized in [26]. Principal angles can be used to describe how far apart one linear subspace is from another. Let L and M be linear subspaces of \mathbb{R}^n with $\dim(L) = l \leq \dim(M) = m$. The principal angles between L and M,

$$0 \le \theta_1 \le \theta_2 \le \dots \le \theta_l \le \frac{\pi}{2}$$

are given by

$$\cos\theta_i = \frac{\langle \mathbf{x}_i, \mathbf{y}_i \rangle}{\|\mathbf{x}_i\| \|\mathbf{y}_i\|} = max \left\{ \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} : \ \mathbf{x} \in L, \ \mathbf{x} \perp \mathbf{x}_k, \ \mathbf{y} \in M, \ \mathbf{y} \perp \mathbf{y}_k, \ k = 1, \dots, i-1 \right\}$$

(see [2]). It follows from the above definition, when the two subspaces coincide (i.e. L = M) then the principal angles are necessarily 0. As a simple example; suppose

it is desired to compare two linear subspaces of dimension 1. In \mathbb{R}^d , linear subspaces of dimension 1 can be visualized as lines through the origin, thus the principal angle between the two linear subspaces is simply the minimal angle between these two vectors. Applying this concept for the problem at hand, recall that an eigenprojections and estimators of them simply are projection matrices into subspaces spanned by their columns. Thus, the concept of principal angles can be used to describe how far an eigenprojection estimate is from the eigenprojection it is meant to estimate by comparing the subspace they both project into respectively. In [4], the authors present a means to compute the principal angles between two linear subspaces in the following lemma,

Lemma 1 Let the columns of $Q_L \in \mathbb{R}^{n \times l}$ and $Q_M \in \mathbb{R}^{n \times m}$ be orthonormal basis for L and M respectively, and let,

$$\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_l \ge 0$$

be the singular values of $Q_M^T Q_L$, then

$$\cos \theta_i = \sigma_i \ i = 1, \ldots, l,$$

and

$$\sigma_1 = \cdots = \sigma_k$$
 iff $\dim (L \cap M) = k$.

Also useful is the following theorem from [26]

Theorem 3 The non-zero principal angles between L, M (are) equal to the non-zero principal angles between L^{\perp} , M^{\perp} .

The above two results are helpful if one wishes to consider the same scatter structures for finite sample simulations that were considered in the asymptotic calculations in section 3.4.2, that is scatter structures where there are only two principal component spaces. The results imply that when the the dimension of the eigenspace being estimated is 1 or d-1, then there will only be 1 non-zero principal angle. (Note that when d= 2 or 3, the subspace can only have dimensions, 1 or 1, 2 respectively.). For the simulations, recall the finite sample distribution of the Tyler matrix and SSCM do not depend on the underlying elliptical distribution provided μ is known, thus the only parameters that must be considered are the sample size and the scatter structure of the elliptical model. Simulations were undertaken in dimensions d = 2, 3 and 5. In two dimensions, the scatter structures considered are diag(2,1), diag(16,1) and diag(128,1). In three dimensions, the scatter structures considered are diag(2,1,1), diag(16,1,1), diag(128,1,1), diag(2,2,1), diag(16,16,1) and diag(128,128,1). For five dimensions, the following scatter structures are considered,

$$\begin{split} \mathbf{\Lambda}_1 &= diag(2,1,1,1,1) & \mathbf{\Lambda}_7 &= diag(2,2,2,1,1) \\ \mathbf{\Lambda}_2 &= diag(16,1,1,1,1) & \mathbf{\Lambda}_8 &= diag(16,16,16,1,1) \\ \mathbf{\Lambda}_3 &= diag(128,1,1,1,1) & \mathbf{\Lambda}_9 &= diag(128,128,128,1,1) \\ \mathbf{\Lambda}_4 &= diag(2,2,1,1,1) & \mathbf{\Lambda}_{10} &= diag(2,2,2,2,1) \\ \mathbf{\Lambda}_5 &= diag(16,16,1,1,1) & \mathbf{\Lambda}_{11} &= diag(16,16,16,16,1) \\ \mathbf{\Lambda}_6 &= diag(128,128,1,1,1) & \mathbf{\Lambda}_{12} &= diag(128,128,128,128,1) \end{split}$$

For the Tyler matrix, because it is affinely equivariant it was only necessary to implement simulations for the case $\Sigma = \mathbf{I}_d$. For a fixed sample size and dimension, 10,000 datasets were generated from an angular central Gaussian distribution. The Tyler matrix was calculated for each dataset, denote this $\hat{\mathcal{T}}_n$. By affine equivariance, the Tyler matrices for uncorrelated scatter structures other than spherical symmetry were obtainable by transforming the matrices via $\Lambda^{1/2} \hat{\mathcal{T}}_n \Lambda^{1/2}$. Once transformed, by Lemma 1 in order to obtain the principal angles it is necessary to obtain an orthonormal basis for the eigenspaces of interest. Recall the two eigenspaces of interest are the subspace the matrix $\hat{\mathbf{T}}_{D_1}$ projects into and the subspace spanned by the canonical unit vector $\mathbf{e}_1, \ldots, \mathbf{e}_{d_1}$, denote them $\widehat{\mathbb{T}}_{D_1}$, and \mathbb{E}_{D_1} respectively. The spectral decomposition of the matrix $\mathbf{\Lambda}^{1/2} \widehat{\boldsymbol{\mathcal{T}}}_n \mathbf{\Lambda}^{1/2}$ was obtained to get the eigenvectors $\widehat{\mathbf{t}}_1, \ldots, \widehat{\mathbf{t}}_{d_1}$, which serve as an orthonormal basis of $\widehat{\mathbb{T}}_{D_1}$. An orthonormal basis for \mathbb{E}_{D_1} is simply the canonical unit vector $\mathbf{e}_1, \ldots, \mathbf{e}_{d_1}$. Define the matrices $Q_{\widehat{\mathbb{T}}_{D_1}} = (\widehat{\mathbf{t}}_1, \ldots, \widehat{\mathbf{t}}_{d_1})$ and $Q_{\mathbb{E}_{D_1}} = (\mathbf{e}_1, \ldots, \mathbf{e}_{d_1})$. Denote the principal angles between the subspaces $\widehat{\mathbb{T}}_{D_1}$ and \mathbb{E}_{D_1} as $\tau_1, \ldots, \tau_{d_1}$. As described in Lemma 1, the singular values of the matrix $Q_{\widehat{\mathbb{T}}_{D_1}}^t Q_{\mathbb{E}_{D_1}}$ will be $cos\tau_1, \ldots, cos\tau_{d_1}$. The arccosine of these singular values were then taken to obtain the necessary principal angles.

For the SSCM, since it is only orthogonally equivariant, it was necessary to implement simulations for each sample size, dimension and scatter structure considered. For fixed values of the prior mentioned parameters, 10,000 datasets were generated and the SSCM was calculated for each, denoted as $\hat{\boldsymbol{S}}_n$. Similarly, the two eigenspaces of interest are the subspace the matrix $\hat{\boldsymbol{S}}_{D_1}$ projects into, denoted $\hat{\boldsymbol{S}}_{D_1}$, and \mathbb{E}_{D_1} . The spectral decomposition of the matrix $\hat{\boldsymbol{S}}_n$ was obtained to get the eigenvectors $\hat{\boldsymbol{s}}_1, \ldots, \hat{\boldsymbol{s}}_{d_1}$, which serve as an orthonormal basis of $\hat{\boldsymbol{S}}_{D_1}$. Define the matrices $Q_{\hat{\boldsymbol{S}}_{D_1}} = (\hat{\boldsymbol{s}}_1, \ldots, \hat{\boldsymbol{s}}_{d_1})$ and $Q_{\mathbb{E}_{D_1}}$ as before. Denote the principal angles between the subspaces $\hat{\boldsymbol{S}}_{D_1}$ and \mathbb{E}_{D_1} as $\boldsymbol{\varsigma}_1, \ldots, \boldsymbol{\varsigma}_{d_1}$. As described in Lemma 1, the singular values of the matrix $Q_{\hat{\boldsymbol{S}}_{D_1}}^t Q_{\mathbb{E}_{D_1}}$ will be $cos \boldsymbol{\varsigma}_1, \ldots, cos \boldsymbol{\varsigma}_{d_1}$. The arccosine of these singular values were then taken to obtain the necessary principal angles.

Note that the principal angles will have expectation 0. This is artifact of the way principal angles are defined, that is as positive angles; consequently, there expectation will be positive. This fact motivates a method to compare the eigenprojection estimates obtained via the Tyler matrix and SSCM. Intuitively, the smaller the expected values of the principal angles, the closer the eigenprojection estimator is to estimating the true eigenprojection. Consequently, the sum of the expected values of the principal angles was used a measure to compare different eigenprojection estimates, the ratio of these sums for the two different estimators being the value of interest. That is define

$$RE_n\left[\widehat{\mathbf{S}}_{D_1}, \widehat{\mathbf{T}}_{D_1}; \mathbf{\Lambda}\right] = \frac{E\left[\tau_1 + \dots + \tau_{d_1}\right]}{E\left[\varsigma_1 + \dots + \varsigma_{d_1}\right]}$$



Figure 3.4: Finite Sample Relative Efficiencies of the Eigenprojection Estimate of the SSCM to the Corresponding Eigenprojection Estimate of the Tyler Matrix in \mathbb{R}^2

3.5.2 Results

In two dimensions, the only possibility is $d_1 = 1$, thus an exhaustive study of the relative efficiency of the eigenprojection estimate obtained from the SSCM to that of the Tyler matrix is possible. In this situation, τ_1 and ς_1 correspond to the angle between the eigenvectors $\hat{\mathbf{t}}_1$ and \mathbf{e}_1 or $\hat{\mathbf{s}}_1$ and \mathbf{e}_1 respectively (the convention for $\hat{\mathbf{t}}_1$ and $\hat{\mathbf{s}}_1$ is that their first component is positive). The values of the relative efficiencies for various scatter structures are given in Figure 3.4.

When $\Lambda = diag(2, 1)$, the relative efficiency of the principal angle estimates is nearly 1, indicating the the SSCM estimates the eigenprojection matrix associated with the eigenvalue 2 almost as well as the Tyler matrix. However, for scatter structures that are more elliptical, as sample size increases, the relative efficiency decreases before leveling off. For the case $\Lambda = diag(16, 1)$, the efficiency is around 0.85 for sample sizes as small as 9. Interestingly, the sample size for which the relative efficiencies are the worst is n = 2, however, this will not be the case for all elliptical distribution. Recall in this case, the Tyler matrix turns out to be equal to the sample covariance matrix. Unlike the Tyler matrix or SSCM, the sample covariance matrix does depend on the radial component of the elliptical distribution (i.e. does depend on R_G).

In three dimensions for scatter structures which have only two principal component spaces, there are two situations to consider. The first situation is when the eigenspace associated with the larger scale has dimension 1; the second being when the eigenspace





Figure 3.5: Finite Sample Relative Efficiencies of the Eigenprojection Estimate of the SSCM to the Corresponding Eigenprojection Estimate of the Tyler Matrix in \mathbb{R}^3

associated with the larger scale has dimension 2. In either situation, Lemma 1 implies that there will be only one non-zero canonical angle. The values of the relative efficiencies are presented in Figure 3.5.

For both cases, the relative efficiencies when the underlying scatter structure is close to spherical symmetry ($\Lambda = diag(2, 1, 1)$ and $\Lambda = diag(2, 2, 1)$) are close to 1 indicating the SSCM is performing nearly as well the Tyler matrix in estimating the eigenprojection matrix associated with the eigenvalue 2. As the scatter structure move away from spherical symmetry, the relative efficiencies decrease in both situations. At n = 15, the relative efficiencies for the scatter structures $\Lambda = diag(16, 1, 1)$ and $\Lambda = diag(16, 16, 1)$ are 0.86 and 0.9 respectively. Of the two situation, the one that is more deleterious towards accurately estimating the eigenprojection matrices is when the dimension of the eigenspace associated with the larger scale is 1. Like in two dimensions, the case when the number of samples equals the dimension, n = 3, does not follow the general trend. Similarly, this is attributable to the fact the Tyler matrix equals the sample covariance matrix in this case; since the later depends on the radial component of the elliptical distribution, these values do not hold for all elliptical distributions.

The last dimension considered permits more interesting scatter structures to consider, even when there are only two principal component spaces. In five dimensions, the eigenspace associated with the larger scale can have dimension $d_1 = 1, 2, 3$ or 4. Figure 3.6 displays the plots relative efficiencies in five dimensions for the scatter matrices in consideration.

The same trends present in two and three dimensions are also present in five dimensions, namely as the scatter matrix becomes more elliptical the relative efficiencies decrease. Fixing the values of the larger scale, the relative efficiency increases as the dimension of the eigenspace associated with the larger scale increases. Thus, like in three dimensions, the situation where the relative efficiency is worst is when the eigenspace associated with the largest scale is one dimensional. Also similar is the fact that for n = 5, the relative efficiency does not follow the general trend, the reason being the same as it was in two and three dimensions.

3.6 Appendix

3.6.1 Asymptotic Distribution of Eigenprojections and Eigenvectors of a Scatter Estimate

Two theorems from [33] will be needed. For authenticity the notation is consistent with that in the original presentation.

Let \mathbf{M} be a $p \times p$ matrix which is symmetric in the metric of the positive definite symmetric matrix $\mathbf{\Gamma}$, i.e. $\mathbf{\Gamma}\mathbf{M}$ is symmetric. Denote the eigenvalues of \mathbf{M} by $\lambda_1 \geq \lambda_2 \geq$ $\cdots \geq \lambda_p$. Define \mathbf{M}_n be a sequence of estimates of \mathbf{M} such that \mathbf{M}_n is symmetric in the metric of the positive definite symmetric matrix $\mathbf{\Gamma}_n$. The following assumption are needed

- 1) $\Gamma_n \to \Gamma$ in probability.
- 2) $a_n(\mathbf{M}_n \mathbf{M}) \rightarrow_d \mathbf{N}$ where a_n is an increasing sequence of positive numbers





Figure 3.6: Finite Sample Relative Efficiencies of the Eigenprojection Estimate of the SSCM to the Corresponding Eigenprojection Estimate of the Tyler Matrix in \mathbb{R}^5

such that $a_n \to \infty$ as $n \to \infty$, and $\text{vec}(\mathbf{N})$ is multivariate Normal with mean **0** and covariance matrix Σ_0 .

In the original exposition, the author considers the following subset of eigenvalues of \mathbf{M} , $\{\lambda_i, \lambda_{i+1}, \ldots, \lambda_{i+m-1}\}$. Note that, this includes situations where some, or all, of the eigenvalues are equal. Consequently, define ω to be the distinct eigenvalues in the set $\{\lambda_i, \lambda_{i+1}, \ldots, \lambda_{i+m-1}\}$. One could use the corresponding eigenvalues of \mathbf{M}_n as estimates of the eigenvalues of \mathbf{M} ; thus define $\hat{\omega} = \{\hat{\lambda}_i, \hat{\lambda}_{i+1}, \ldots, \hat{\lambda}_{i+m-1}\}$. For the eigenvalue λ of \mathbf{M} , let \mathbf{P}_{λ} be the eigenprojection of \mathbf{M} associated with λ . Similarly, for λ an eigenvalues of \mathbf{M}_n (not necessarily equal to the one in the previous sentence), let $\widehat{\mathbf{P}}_{\lambda}$ be defined similarly. Lastly, define $\mathbf{P}_0 = \sum_{\lambda \in \omega} \mathbf{P}_{\lambda}$ and $\widehat{\mathbf{P}}_0 = \sum_{\lambda \in \hat{\omega}} \widehat{\mathbf{P}}_{\lambda}$ the total eigenprojections of \mathbf{M} and \mathbf{M}_n associated with the eigenvalues in ω and $\hat{\omega}$ respectively. The first order Taylor approximation of $\widehat{\mathbf{P}}_0$ about \mathbf{P}_0 is given in the following lemma from the paper.

Lemma 4.1. Let $d_0 = \min \{\lambda_{i-1} - \lambda_i, \lambda_{i+m-1} - \lambda_{i+m}\}$, and $d_1 = (\lambda_i - \lambda_{i+m-1})$. Also define the norm $\|\mathbf{B}\| = [\max \text{ eigenvalue} (\mathbf{\Gamma}^{-1}\mathbf{B}'\mathbf{\Gamma}\mathbf{B})]^{1/2}$. If $\|\mathbf{B}_n - \mathbf{B}\| \leq d_0/2$ then

$$\widehat{\mathbf{P}}_{0} = \mathbf{P}_{0} - \sum_{\lambda \in \omega} \left[\mathbf{P}_{\lambda} \left(\mathbf{M}_{n} - \mathbf{M} \right) \left(\mathbf{M} - \lambda \mathbf{I}_{d} \right)^{+} + \left(\mathbf{M} - \lambda \mathbf{I}_{d} \right)^{+} \left(\mathbf{M}_{n} - \mathbf{M} \right) \mathbf{P}_{\lambda} \right] + \mathbf{E}_{n}$$
where $\|\mathbf{E}_{n}\| \leq \left(1 + d_{1}/d_{0} \right) \left(2 \|\mathbf{M}_{n} - \mathbf{M}\| / d_{0} \right)^{2} \left(1 - 2 \|\mathbf{M}_{n} - \mathbf{M}\| / d_{0} \right)^{-1}$

This lemma is useful in that it gives the asymptotic distribution of the eigenprojection $\hat{\mathbf{P}}_0$, namely

$$a_{n}\mathbf{vec}\left(\widehat{\mathbf{P}}_{0}-\mathbf{P}_{0}\right)\rightarrow_{d}\mathbf{vec}\left(\mathbf{N}_{0}\right)=-\sum_{\lambda\in\omega}\left[\left(\mathbf{M}-\lambda\mathbf{I}_{d}\right)^{+}\otimes\mathbf{P}_{\lambda}+\mathbf{P}_{\lambda}^{\prime}\otimes\left(\mathbf{M}-\lambda\mathbf{I}_{d}\right)^{+}\right]\mathbf{vec}\left(\mathbf{N}\right)$$

where $\mathbf{vec}(\mathbf{N}_0)$ has a multivariate Normal distribution with mean **0** and covariance matrix $\mathbf{C}'_{\omega} \boldsymbol{\Sigma}_0 \mathbf{C}_{\omega}$ with

$$\mathbf{C}_{\omega}' = -\sum_{\lambda \in \omega} \left[(\mathbf{M} - \lambda \mathbf{I}_d)^+ \otimes \mathbf{P}_\lambda + \mathbf{P}_\lambda' \otimes (\mathbf{M} - \lambda \mathbf{I}_d)^+
ight]$$

The above expression can be simplified. Since projection matrices are symmetric, the transposes are superfluous and thus will be omitted. Recall one can write a symmetric positive-definite matrix in terms of its eigenvalues and eigenprojections via the formula $\mathbf{M} = \sum \mu \mathbf{P}_{\mu}$. Thus one has

$$\mathbf{M} - \lambda \mathbf{I}_d = \sum_{\mu \neq \lambda} \mu \mathbf{P}_{\mu} + \lambda \mathbf{P}_{\lambda} - \lambda \mathbf{I}_d = \sum_{\mu \neq \lambda} \mu \mathbf{P}_{\mu} - \lambda \left(\mathbf{I}_d - \mathbf{P}_{\lambda} \right)$$

Assuming **M** is of full rank, it follows

$$\sum_{\mu} \mathbf{P}_{\mu} = \sum_{\mu \neq \lambda} \mathbf{P}_{\mu} + \mathbf{P}_{\lambda} = \mathbf{I}_{d}$$

This then gives

$$\mathbf{M} - \lambda \mathbf{I}_d = \sum_{\mu \neq \lambda} \mu \mathbf{P}_{\mu} - \lambda \sum_{\mu \neq \lambda} \mathbf{P}_{\mu} = \sum_{\mu \neq \lambda} (\mu - \lambda) \mathbf{P}_{\mu}$$

Note that the matrix $(\mathbf{M} - \lambda \mathbf{I}_d)$ has eigenvalues $(\mu - \lambda)$ where $\mu \neq \lambda$. It can be easily shown that a generalized inverse of $(\mathbf{M} - \lambda \mathbf{I}_d)$ is of the form

$$(\mathbf{M} - \lambda \mathbf{I}_d)^+ = \sum_{\mu \neq \lambda} (\mu - \lambda)^{-1} \mathbf{P}_\mu$$

Using this fact gives

$$\mathbf{C}'_{\omega} = -\sum_{\lambda \in \omega} \left[\left(\sum_{\mu \neq \lambda} (\mu - \lambda)^{-1} \mathbf{P}_{\mu} \right) \otimes \mathbf{P}_{\lambda} + \mathbf{P}_{\lambda} \otimes \left(\sum_{\mu \neq \lambda} (\mu - \lambda)^{-1} \mathbf{P}_{\mu} \right) \right]$$
$$= \sum_{\lambda \in \omega} \left[\left(\sum_{\mu \neq \lambda} (\lambda - \mu)^{-1} \mathbf{P}_{\mu} \right) \otimes \mathbf{P}_{\lambda} + \mathbf{P}_{\lambda} \otimes \left(\sum_{\mu \neq \lambda} (\lambda - \mu)^{-1} \mathbf{P}_{\mu} \right) \right]$$

For any $\lambda, \mu \in \omega$, the matrices $(\lambda - \mu)^{-1} [\mathbf{P}_{\mu} \otimes \mathbf{P}_{\lambda} + \mathbf{P}_{\lambda} \otimes \mathbf{P}_{\mu}]$ and $(\mu - \lambda)^{-1} [\mathbf{P}_{\lambda} \otimes \mathbf{P}_{\mu} + \mathbf{P}_{\mu} \otimes \mathbf{P}_{\lambda}]$ will appear in the double sum. However, since $(\lambda - \mu)^{-1} = -(\mu - \lambda)^{-1}$, these terms will cancel with each other. Utilizing this fact yields

$$\mathbf{C}'_{\omega} = \sum_{\lambda \in \omega} \sum_{\mu \notin \omega} (\lambda - \mu)^{-1} \left[\mathbf{P}_{\mu} \otimes \mathbf{P}_{\lambda} + \mathbf{P}_{\lambda} \otimes \mathbf{P}_{\mu} \right]$$

As an aside, note that the above matrix is symmetric, so the transpose may also be disregarded. For the λ_i that are represented in ω , let \mathcal{N} be the collection of these indices (i.e. the subscripts). It will be shown that

$$\mathbf{C}_{\omega} = \sum_{i \in \mathcal{N}} \sum_{j \notin \mathcal{N}} \left(\lambda_i - \lambda_j\right)^{-1} \left[\mathbf{p}_j \mathbf{p}_j^t \otimes \mathbf{p}_i \mathbf{p}_i^t + \mathbf{p}_i \mathbf{p}_i^t \otimes \mathbf{p}_j \mathbf{p}_j^t \right]$$

where $\mathbf{p}_1, \ldots, \mathbf{p}_d$ is an orthonormal set of eigenvectors for the matrix \mathbf{M} . To see this, suppose that the eigenvalues $\lambda_m = \cdots = \lambda_{m+k-1} = \lambda$. Let $\mathbf{p}_m, \ldots, \mathbf{p}_{m+k-1}$ be a set of orthonormal eigenvectors corresponding to the eigenvalue λ . The eigenprojection matrix associated with the eigenvalue λ , denoted \mathbf{P}_{λ} , is $\mathbf{P}_{\lambda} = \mathbf{p}_m \mathbf{p}_m^t + \cdots + \mathbf{p}_{m+k-1} \mathbf{p}_{m+k-1}^t$. Suppose for a different, mutually exclusive set of eigenvalues $\lambda_n = \cdots = \lambda_{n+l-1} = \mu$. The same result holds for eigenprojection matrix \mathbf{P}_{μ} . Suppose that $\lambda \in \omega$ and $\mu \notin \omega$, thus the indices $m, \ldots, (m+k-1)$ are in \mathcal{N} whereas $n, \ldots, (n+l-1)$ are not. The claim is

$$(\lambda - \mu)^{-1} \left[\mathbf{P}_{\mu} \otimes \mathbf{P}_{\lambda} + \mathbf{P}_{\lambda} \otimes \mathbf{P}_{\mu} \right] = \sum_{i=m}^{m+k-1} \sum_{j=n}^{n+l-1} (\lambda_i - \lambda_j)^{-1} \left[\mathbf{p}_j \mathbf{p}_j^t \otimes \mathbf{p}_i \mathbf{p}_i^t + \mathbf{p}_i \mathbf{p}_i^t \otimes \mathbf{p}_j \mathbf{p}_j^t \right]$$

For the eigenvalues corresponding to the index of summation, one has $\lambda_i = \lambda$ and $\lambda_j = \mu$, thus they can be factored out of the sums. Consequently, one need only focus attention on showing the sum is equal to the sum of the two Kronecker products on the left hand side. Since the Kronecker product is distributive, one has

$$\sum_{i=m}^{m+k-1}\sum_{j=n}^{n+l-1}\left[\mathbf{p}_{j}\mathbf{p}_{j}^{t}\otimes\mathbf{p}_{i}\mathbf{p}_{i}^{t}+\mathbf{p}_{i}\mathbf{p}_{i}^{t}\otimes\mathbf{p}_{j}\mathbf{p}_{j}^{t}\right]=\sum_{i=m}^{m+k-1}\left[\left(\sum_{j=n}^{n+l-1}\mathbf{p}_{j}\mathbf{p}_{j}^{t}\right)\otimes\mathbf{p}_{i}\mathbf{p}_{i}^{t}+\mathbf{p}_{i}\mathbf{p}_{i}^{t}\otimes\left(\sum_{j=n}^{n+l-1}\mathbf{p}_{j}\mathbf{p}_{j}^{t}\right)\right]$$

The sums within the brackets equal \mathbf{P}_{μ} . Applying this trick again gives

$$\sum_{i=m}^{m+k-1} \left[\mathbf{P}_{\mu} \otimes \mathbf{p}_{i} \mathbf{p}_{i}^{t} + \mathbf{p}_{i} \mathbf{p}_{i}^{t} \otimes \mathbf{P}_{\mu} \right] = \mathbf{P}_{\mu} \otimes \left(\sum_{i=m}^{m+k-1} \mathbf{p}_{i} \mathbf{p}_{i}^{t} \right) + \left(\sum_{i=m}^{m+k-1} \mathbf{p}_{i} \mathbf{p}_{i}^{t} \right) \otimes \mathbf{P}_{\mu}$$
$$= \mathbf{P}_{\mu} \otimes \mathbf{P}_{\lambda} + \mathbf{P}_{\lambda} \otimes \mathbf{P}_{\mu}$$

thus showing the result.

Relating these theorems back to the problem at hand. The matrix **M** that is to be estimated will be either the scatter matrix of the data, Σ or the matrix Ξ . The estimators being used, \mathbf{M}_n , for Σ or Ξ are either Tyler's matrix or the SSCM respectively, based on samples of size n; thus $a_n = \sqrt{n}$. The matrices, Γ and Γ_n with which the aforementioned matrices are to be symmetric in the metric of is simply \mathbf{I}_d in both cases. Consequently, $\Sigma_0 = AV_{\widehat{T}}(\Sigma)$ or $\Sigma_0 = AV_{\widehat{S}}(\Sigma)$.

3.6.2 Tyler's Scatter Matrix

Existence and uniqueness of $\widehat{\mathcal{T}}_{\widehat{\mu}_n}$ as well as the its properties were discussed in [35]. Of interest for this section is the asymptotic distribution of Tyler's matrix. A technicality that must be addressed proceeding is the issue that the Tyler matrix is only asymptotically consistent for Σ up to proportionality. Hence to properly treat the asymptotic distribution of the Tyler matrix, both the Tyler matrix and Σ must be standardized. In [35], the author does so by standardizing both matrices so that each has trace equal to d. To this end define $\mathbf{M}^{\circ} = d\mathbf{M}/Trace(\mathbf{M})$; it then follows the matrix \mathbf{M}° will have trace equal to d. Using this standardization it was shown that the asymptotic distribution of the standardized Tyler matrix under the elliptical model is given by

$$\sqrt{n}\left(\operatorname{\mathbf{vec}}\left(\widehat{\boldsymbol{\mathcal{T}}}_{\widehat{\boldsymbol{\mu}}_{n}}^{\circ}-\boldsymbol{\Sigma}^{\circ}\right)\right)\rightarrow_{\mathcal{D}} Norm_{d^{2}}\left(\boldsymbol{0},AV_{\widehat{\boldsymbol{\mathcal{T}}}^{\circ}}\left(\boldsymbol{\Sigma}^{\circ}\right)\right)$$

where $\widehat{\mu}$ is some consistent estimate of μ and

$$\begin{aligned} AV_{\widehat{\boldsymbol{\tau}}^{\circ}}\left(\boldsymbol{\Sigma}^{\circ}\right) &= \frac{d+2}{d}\left(\mathbf{I}_{d^{2}} + \mathbf{K}_{d,d}\right)\left(\boldsymbol{\Sigma}^{\circ} \otimes \boldsymbol{\Sigma}^{\circ}\right) - \frac{2}{d}\frac{d+2}{d}\mathbf{vec}\left(\boldsymbol{\Sigma}^{\circ}\right)\mathbf{vec}^{t}\left(\boldsymbol{\Sigma}^{\circ}\right) \\ &= \left(\frac{d}{Trace\left(\boldsymbol{\Sigma}\right)}\right)^{2}\frac{d+2}{d}\left(\left(\mathbf{I}_{d^{2}} + \mathbf{K}_{d,d}\right)\left(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}\right) - \frac{2}{d}\mathbf{vec}\left(\boldsymbol{\Sigma}\right)\mathbf{vec}^{t}\left(\boldsymbol{\Sigma}\right)\right) \end{aligned}$$

Using the spectral decomposition of $\pmb{\Sigma}$ reduces the above to

$$\begin{aligned} AV_{\widehat{\boldsymbol{\tau}}^{\circ}}\left(\boldsymbol{\Sigma}^{\circ}\right) &= \left(\frac{d}{Trace\left(\boldsymbol{\Sigma}\right)}\right)^{2} \frac{d+2}{d} \left(\mathbf{P} \otimes \mathbf{P}\right) \left(\left(\mathbf{I}_{d^{2}} + \mathbf{K}_{d,d}\right) \left(\boldsymbol{\Lambda} \otimes \boldsymbol{\Lambda}\right) - \frac{2}{d} \mathbf{vec}\left(\boldsymbol{\Lambda}\right) \mathbf{vec}^{t}\left(\boldsymbol{\Lambda}\right)\right) \left(\mathbf{P}^{t} \otimes \mathbf{P}^{t}\right) \\ &= \left(\mathbf{P} \otimes \mathbf{P}\right) AV_{\widehat{\boldsymbol{\tau}}}\left(\boldsymbol{\Lambda}^{\circ}\right) \left(\mathbf{P}^{t} \otimes \mathbf{P}^{t}\right) \end{aligned}$$

Where the second inequality utilizes the fact $Trace(\Sigma) = Trace(\Lambda)$. Furthermore, the results of the previous section yield

$$\sqrt{n}\left(\operatorname{\mathbf{vec}}\left(\widehat{\mathbf{T}}_{\mathcal{N}}-\mathbf{P}_{\mathcal{N}}\right)\right) \rightarrow_{\mathcal{D}} Norm_{d^{2}}\left(\mathbf{0},AV_{\widehat{\mathbf{T}}_{\mathcal{N}}}\left(\mathbf{\Sigma}\right)\right)$$

where

$$AV_{\widehat{\mathbf{T}}_{\mathcal{N}}}\left(\mathbf{\Sigma}\right) = \mathbf{C}_{\mathcal{N}}^{t}\left(\mathbf{\Sigma}^{\circ}\right)AV_{\widehat{\mathcal{T}}}\left(\mathbf{\Sigma}^{\circ}\right)\mathbf{C}_{\mathcal{N}}\left(\mathbf{\Sigma}^{\circ}\right)$$

Note that

$$\begin{aligned} \mathbf{C}_{\mathcal{N}}(\mathbf{\Sigma}^{\circ}) &= \sum_{i \in \mathcal{N}} \sum_{j \notin \mathcal{N}} \left(\frac{d}{Trace(\mathbf{\Sigma})} \lambda_i^2 - \frac{d}{Trace(\mathbf{\Sigma})} \lambda_j^2 \right)^{-1} \left[\mathbf{p}_j \mathbf{p}_j^t \otimes \mathbf{p}_i \mathbf{p}_i^t + \mathbf{p}_i \mathbf{p}_i^t \otimes \mathbf{p}_j \mathbf{p}_j^t \right] \\ &= \frac{Trace(\mathbf{\Sigma})}{d} \sum_{i \in \mathcal{N}} \sum_{j \notin \mathcal{N}} \left(\lambda_i^2 - \lambda_j^2 \right)^{-1} \left[\mathbf{p}_j \mathbf{p}_j^t \otimes \mathbf{p}_i \mathbf{p}_i^t + \mathbf{p}_i \mathbf{p}_i^t \otimes \mathbf{p}_j \mathbf{p}_j^t \right] \\ &= \frac{Trace(\mathbf{\Sigma})}{d} \mathbf{C}_{\mathcal{N}}(\mathbf{\Sigma}) \end{aligned}$$

Thus in the expression for $AV_{\widehat{\mathbf{T}}_{\mathcal{N}}}(\mathbf{\Sigma})$, the term $Trace(\mathbf{\Sigma})/d$ cancels out indicating that the asymptotic distribution of the eigenprojection estimate does not depend the

scaling of Σ or the estimator $\hat{\mathcal{T}}$. Thus the superscript \circ is superfluous when discussing eigenprojection and will be omitted in the exposition (as was purposely done above).

Recall one can write a symmetric positive-definite matrix in terms of its eigenvalues and eigenvectors via the formula $\Sigma = \sum_{i=1}^{d} \lambda_i^2 \mathbf{p}_i \mathbf{p}_i^t$. Thus for any $d \times d$ orthogonal matrix \mathbf{Q} , if the random vector \mathbf{x} has scatter matrix Σ , then the random vector $\mathbf{Q}\mathbf{x}$ will have scatter matrix $\mathbf{Q}\Sigma\mathbf{Q}^t$. The transformed scatter matrix can be written as $\Sigma =$ $\sum_{i=1}^{d} \lambda_i^2 \mathbf{Q} \mathbf{p}_i \mathbf{p}_i^t \mathbf{Q}^t$. This implies that eigenvectors and consequently, the eigenprojection matrices of scatter matrices are orthogonally equivariant (note that they are not affinely equivariant). The same is true for the eigenprojection matrices of scatter matrices that are estimates. In addition, for the transformed data one has the following

$$\begin{aligned} \mathbf{C}_{\mathcal{N}} \left(\mathbf{Q} \mathbf{\Sigma} \mathbf{Q}^{t} \right) &= \sum_{i \in \mathcal{N}} \sum_{j \notin \mathcal{N}} \left(\lambda_{i}^{2} - \lambda_{j}^{2} \right)^{-1} \left[\mathbf{Q} \mathbf{p}_{j} \mathbf{p}_{j}^{t} \mathbf{Q}^{t} \otimes \mathbf{Q} \mathbf{p}_{i} \mathbf{p}_{i}^{t} \mathbf{Q}^{t} + \mathbf{Q} \mathbf{p}_{i} \mathbf{p}_{i}^{t} \mathbf{Q}^{t} \otimes \mathbf{Q} \mathbf{p}_{j} \mathbf{p}_{j}^{t} \mathbf{Q}^{t} \right] \\ &= \sum_{i \in \mathcal{N}} \sum_{j \notin \mathcal{N}} \left(\lambda_{i}^{2} - \lambda_{j}^{2} \right)^{-1} \left(\mathbf{Q} \otimes \mathbf{Q} \right) \left[\mathbf{p}_{j} \mathbf{p}_{j}^{t} \otimes \mathbf{p}_{i} \mathbf{p}_{i}^{t} + \mathbf{p}_{i} \mathbf{p}_{i}^{t} \otimes \mathbf{p}_{j} \mathbf{p}_{j}^{t} \right] \left(\mathbf{Q}^{t} \otimes \mathbf{Q}^{t} \right) \\ &= \left(\mathbf{Q} \otimes \mathbf{Q} \right) \mathbf{C}_{\mathcal{N}} \left(\mathbf{\Sigma} \right) \left(\mathbf{Q}^{t} \otimes \mathbf{Q}^{t} \right) \end{aligned}$$

Specifically, above implies $\mathbf{C}_{\mathcal{N}}(\mathbf{\Sigma}) = (\mathbf{P} \otimes \mathbf{P}) \mathbf{C}_{\mathcal{N}}(\mathbf{\Lambda}) (\mathbf{P}^t \otimes \mathbf{P}^t)$. Hence the asymptotic variance-covariance matrix of the eigenprojection estimate $\widehat{\mathbf{T}}_{\mathcal{N}}$ is

$$\begin{aligned} AV_{\widehat{\mathbf{T}}_{\mathcal{N}}}\left(\mathbf{\Sigma}\right) &= \left(\mathbf{P}\otimes\mathbf{P}\right)\mathbf{C}_{\mathcal{N}}^{t}\left(\mathbf{\Lambda}\right)\left(\mathbf{P}^{t}\otimes\mathbf{P}^{t}\right)\left(\mathbf{P}\otimes\mathbf{P}\right)AV_{\widehat{\mathcal{T}}}\left(\mathbf{\Lambda}\right)\left(\mathbf{P}^{t}\otimes\mathbf{P}^{t}\right)\left(\mathbf{P}\otimes\mathbf{P}\right)\mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\right)\left(\mathbf{P}^{t}\otimes\mathbf{P}^{t}\right) \\ &= \left(\mathbf{P}\otimes\mathbf{P}\right)\mathbf{C}_{\mathcal{N}}^{t}\left(\mathbf{\Lambda}\right)AV_{\widehat{\mathcal{T}}}\left(\mathbf{\Lambda}\right)\mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\right)\left(\mathbf{P}^{t}\otimes\mathbf{P}^{t}\right) \\ &= \left(\mathbf{P}\otimes\mathbf{P}\right)AV_{\mathcal{N}}\left(\mathbf{\Lambda}\right)\left(\mathbf{P}^{t}\otimes\mathbf{P}^{t}\right) \end{aligned}$$

Consequently, one can examine a simplified case when considering the asymptotic distribution of an eigenprojection estimate based on a scatter estimate; in particular, take the case $\mathbf{P} = \mathbf{I}_d$. In this situation, the eigenprojection matrices can be written in terms of the matrices $\mathbf{E}_i = \mathbf{e}_i \mathbf{e}_i^t$; that is for a given eigenvalue of $\mathbf{\Lambda}$, the corresponding eigenprojection matrix can be written by summing the matrices \mathbf{E}_i across the indices i corresponding to that eigenvalue. These facts imply

$$\mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\right) = \sum_{i \in \mathcal{N}} \sum_{j \notin \mathcal{N}} \left(\lambda_{i}^{2} - \lambda_{j}^{2}\right)^{-1} \left(\mathbf{e}_{i} \mathbf{e}_{i}^{t} \otimes \mathbf{e}_{j} \mathbf{e}_{j}^{t} + \mathbf{e}_{j} \mathbf{e}_{j}^{t} \otimes \mathbf{e}_{i} \mathbf{e}_{i}^{t}\right)$$

Note that matrices of the form $\mathbf{e}_i \mathbf{e}_i^t \otimes \mathbf{e}_j \mathbf{e}_j^t$ for $i, j \in \mathcal{N}$ or $i, j \notin \mathcal{N}$ will not be represented in the double summation. Further note $\mathbf{C}_{\mathcal{N}}^t(\mathbf{\Lambda}) = \mathbf{C}_{\mathcal{N}}(\mathbf{\Lambda})$. With the above representation, one can readily describe the form of the matrix $AV_{\widehat{\mathbf{T}}_{\mathcal{N}}}(\mathbf{\Lambda})$. One has

$$\begin{aligned} AV_{\widehat{\mathbf{T}}_{\mathcal{N}}}\left(\mathbf{\Lambda}\right) &= \mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\right) \left(\frac{d+2}{d} \left(\mathbf{I}_{d^{2}} + \mathbf{K}_{d,d}\right) \left(\mathbf{\Lambda} \otimes \mathbf{\Lambda}\right) - \frac{2}{d} \frac{d+2}{d} \mathbf{vec}\left(\mathbf{\Lambda}\right) \mathbf{vec}^{t}\left(\mathbf{\Lambda}\right)\right) \mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\right) \\ &= \frac{d+2}{d} \mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\right) \left(\mathbf{I}_{d^{2}} + \mathbf{K}_{d,d}\right) \left(\mathbf{\Lambda} \otimes \mathbf{\Lambda}\right) \mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\right) - \frac{2}{d} \frac{d+2}{d} \mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\right) \mathbf{vec}\left(\mathbf{\Lambda}\right) \mathbf{vec}^{t}\left(\mathbf{\Lambda}\right) \mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\right) \\ &= \frac{d+2}{d} \left(\mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\right) \left(\mathbf{\Lambda} \otimes \mathbf{\Lambda}\right) \mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\right) + \mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\right) \mathbf{K}_{d,d}\left(\mathbf{\Lambda} \otimes \mathbf{\Lambda}\right) \mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\right) - \frac{2}{d} \mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\right) \mathbf{vec}\left(\mathbf{\Lambda}\right) \mathbf{vec}^{t}\left(\mathbf{\Lambda}\right) \mathbf{c}_{\mathcal{N}}\left(\mathbf{\Lambda}\right) \right) \end{aligned}$$

Consider the matrix products separately. Writing out $\mathbf{C}_{\mathcal{N}}(\Lambda) (\Lambda \otimes \Lambda) \mathbf{C}_{\mathcal{N}}(\Lambda)$ explicitly gives

$$\left(\sum_{i\in\mathcal{N}}\sum_{j\notin\mathcal{N}}\frac{\mathbf{e}_{i}\mathbf{e}_{i}^{t}\otimes\mathbf{e}_{j}\mathbf{e}_{j}^{t}+\mathbf{e}_{j}\mathbf{e}_{j}^{t}\otimes\mathbf{e}_{i}\mathbf{e}_{i}^{t}}{\left(\lambda_{i}^{2}-\lambda_{j}^{2}\right)}\right)\left(\mathbf{\Lambda}\otimes\mathbf{\Lambda}\right)\left(\sum_{m\in\mathcal{N}}\sum_{n\notin\mathcal{N}}\frac{\mathbf{e}_{m}\mathbf{e}_{m}^{t}\otimes\mathbf{e}_{n}\mathbf{e}_{n}^{t}+\mathbf{e}_{n}\mathbf{e}_{n}^{t}\otimes\mathbf{e}_{m}\mathbf{e}_{m}^{t}}{\left(\lambda_{m}^{2}-\lambda_{n}^{2}\right)}\right)$$

There will be four different cross-product terms to consider. One such pair will be of the form

$$\frac{\left(\mathbf{e}_{i}\mathbf{e}_{i}^{t}\otimes\mathbf{e}_{j}\mathbf{e}_{j}^{t}\right)\left(\mathbf{\Lambda}\otimes\mathbf{\Lambda}\right)\left(\mathbf{e}_{m}\mathbf{e}_{m}^{t}\otimes\mathbf{e}_{n}\mathbf{e}_{n}^{t}\right)}{\left(\lambda_{i}^{2}-\lambda_{j}^{2}\right)\left(\lambda_{m}^{2}-\lambda_{n}^{2}\right)}=\frac{\left(\mathbf{e}_{i}\mathbf{e}_{i}^{t}\mathbf{\Lambda}\mathbf{e}_{m}\mathbf{e}_{m}^{t}\right)\otimes\left(\mathbf{e}_{j}\mathbf{e}_{j}^{t}\mathbf{\Lambda}\mathbf{e}_{n}\mathbf{e}_{n}^{t}\right)}{\left(\lambda_{i}^{2}-\lambda_{j}^{2}\right)\left(\lambda_{m}^{2}-\lambda_{n}^{2}\right)}$$

Since Λ is diagonal, this matrix will be a **0** matrix unless i = m and j = n, in which case it reduces to $\frac{\lambda_i^2 \lambda_j^2}{(\lambda_i^2 - \lambda_j^2)^2} \mathbf{e}_i \mathbf{e}_i^t \otimes \mathbf{e}_j \mathbf{e}_j^t$. Thus the contributions from these cross-product terms can be written as

$$\sum_{i \in \mathcal{N}} \sum_{j \notin \mathcal{N}} \frac{\lambda_i^2 \lambda_j^2}{\left(\lambda_i^2 - \lambda_j^2\right)^2} \mathbf{e}_i \mathbf{e}_i^t \otimes \mathbf{e}_j \mathbf{e}_j^t$$

Corresponding to the cross-product terms,
$$\frac{\left(\mathbf{e}_{j}\mathbf{e}_{j}^{t}\otimes\mathbf{e}_{i}\mathbf{e}_{i}^{t}\right)\left(\mathbf{\Lambda}\otimes\mathbf{\Lambda}\right)\left(\mathbf{e}_{n}\mathbf{e}_{n}^{t}\otimes\mathbf{e}_{m}\mathbf{e}_{m}^{t}\right)}{\left(\lambda_{i}^{2}-\lambda_{j}^{2}\right)\left(\lambda_{m}^{2}-\lambda_{n}^{2}\right)}$$

one can show similarly that the contributions from these cross-product terms can be written as

$$\sum_{i \in \mathcal{N}} \sum_{j \notin \mathcal{N}} \frac{\lambda_j^2 \lambda_i^2}{\left(\lambda_i^2 - \lambda_j^2\right)^2} \mathbf{e}_j \mathbf{e}_j^t \otimes \mathbf{e}_i \mathbf{e}_i^t = \sum_{i \in \mathcal{N}} \sum_{j \notin \mathcal{N}} \frac{\lambda_i^2 \lambda_j^2}{\left(\lambda_i^2 - \lambda_j^2\right)^2} \mathbf{e}_j \mathbf{e}_j^t \otimes \mathbf{e}_i \mathbf{e}_i^t$$

For the cross-product terms

$$\frac{\left(\mathbf{e}_{i}\mathbf{e}_{i}^{t}\otimes\mathbf{e}_{j}\mathbf{e}_{j}^{t}\right)\left(\mathbf{\Lambda}\otimes\mathbf{\Lambda}\right)\left(\mathbf{e}_{n}\mathbf{e}_{n}^{t}\otimes\mathbf{e}_{m}\mathbf{e}_{m}^{t}\right)}{\left(\lambda_{i}^{2}-\lambda_{j}^{2}\right)\left(\lambda_{m}^{2}-\lambda_{n}^{2}\right)}=\frac{\left(\mathbf{e}_{i}\mathbf{e}_{i}^{t}\mathbf{\Lambda}\mathbf{e}_{n}\mathbf{e}_{n}^{t}\right)\otimes\left(\mathbf{e}_{j}\mathbf{e}_{j}^{t}\mathbf{\Lambda}\mathbf{e}_{m}\mathbf{e}_{m}^{t}\right)}{\left(\lambda_{i}^{2}-\lambda_{j}^{2}\right)\left(\lambda_{n}^{2}-\lambda_{m}^{2}\right)}$$

Again, by the diagonal nature of Λ , these will reduce to a **0** matrix unless i = n and j = m. This is impossible since $i, m \in \mathcal{N}$ but $j, n \notin \mathcal{N}$, thus the contributions from these terms will be a **0** matrix. By the same argument, the contributions from the terms

$$\frac{\left(\mathbf{e}_{j}\mathbf{e}_{j}^{t}\otimes\mathbf{e}_{i}\mathbf{e}_{i}^{t}\right)\left(\mathbf{\Lambda}\otimes\mathbf{\Lambda}\right)\left(\mathbf{e}_{m}\mathbf{e}_{m}^{t}\otimes\mathbf{e}_{n}\mathbf{e}_{n}^{t}\right)}{\left(\lambda_{i}^{2}-\lambda_{j}^{2}\right)\left(\lambda_{m}^{2}-\lambda_{n}^{2}\right)}=\frac{\left(\mathbf{e}_{j}\mathbf{e}_{j}^{t}\mathbf{\Lambda}\mathbf{e}_{m}\mathbf{e}_{m}^{t}\right)\otimes\left(\mathbf{e}_{i}\mathbf{e}_{i}^{t}\mathbf{\Lambda}\mathbf{e}_{n}\mathbf{e}_{n}^{t}\right)}{\left(\lambda_{i}^{2}-\lambda_{j}^{2}\right)\left(\lambda_{m}^{2}-\lambda_{n}^{2}\right)}$$

will also be 0. This leaves

$$\mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\otimes\mathbf{\Lambda}
ight)\mathbf{C}_{\mathcal{N}} = \sum_{i\in\mathcal{N}}\sum_{j\notin\mathcal{N}}rac{\lambda_{i}^{2}\lambda_{j}^{2}}{\left(\lambda_{i}^{2}-\lambda_{j}^{2}
ight)^{2}}\left(\mathbf{e}_{i}\mathbf{e}_{i}^{t}\otimes\mathbf{e}_{j}\mathbf{e}_{j}^{t}+\mathbf{e}_{j}\mathbf{e}_{j}^{t}\otimes\mathbf{e}_{i}\mathbf{e}_{i}^{t}
ight)$$

Next consider the matrix product $\mathbf{C}_{\mathcal{N}}(\mathbf{\Lambda}) \mathbf{K}_{d,d}(\mathbf{\Lambda} \otimes \mathbf{\Lambda}) \mathbf{C}_{\mathcal{N}}(\mathbf{\Lambda})$. Writing this out explicitely yields

$$\sum_{i \in \mathcal{N}} \sum_{j \notin \mathcal{N}} \frac{\mathbf{e}_i \mathbf{e}_i^t \otimes \mathbf{e}_j \mathbf{e}_j^t + \mathbf{e}_j \mathbf{e}_j^t \otimes \mathbf{e}_i \mathbf{e}_i^t}{\left(\lambda_i^2 - \lambda_j^2\right)} \sum_{k=1}^d \sum_{l=1}^d \mathbf{e}_k \mathbf{e}_l^t \otimes \mathbf{e}_l \mathbf{e}_k^t \left(\mathbf{\Lambda} \otimes \mathbf{\Lambda}\right) \sum_{m \in \mathcal{N}} \sum_{n \notin \mathcal{N}} \frac{\mathbf{e}_m \mathbf{e}_m^t \otimes \mathbf{e}_n \mathbf{e}_n^t + \mathbf{e}_n \mathbf{e}_n^t \otimes \mathbf{e}_m \mathbf{e}_m^t}{\left(\lambda_m^2 - \lambda_n^2\right)}$$

Again, there will be four different cross product terms to address. First consider

$$\frac{\left(\mathbf{e}_{i}\mathbf{e}_{i}^{t}\otimes\mathbf{e}_{j}\mathbf{e}_{j}^{t}\right)\left(\mathbf{e}_{k}\mathbf{e}_{l}^{t}\otimes\mathbf{e}_{l}\mathbf{e}_{k}^{t}\right)\left(\mathbf{\Lambda}\otimes\mathbf{\Lambda}\right)\left(\mathbf{e}_{m}\mathbf{e}_{m}^{t}\otimes\mathbf{e}_{n}\mathbf{e}_{n}^{t}\right)}{\left(\lambda_{i}^{2}-\lambda_{j}^{2}\right)\left(\lambda_{m}^{2}-\lambda_{n}^{2}\right)}=\frac{\left(\mathbf{e}_{i}\mathbf{e}_{i}^{t}\mathbf{e}_{k}\mathbf{e}_{l}^{t}\mathbf{\Lambda}\mathbf{e}_{m}\mathbf{e}_{m}^{t}\right)\otimes\left(\mathbf{e}_{j}\mathbf{e}_{j}^{t}\mathbf{e}_{l}\mathbf{e}_{k}^{t}\mathbf{\Lambda}\mathbf{e}_{n}\mathbf{e}_{n}^{t}\right)}{\left(\lambda_{i}^{2}-\lambda_{j}^{2}\right)\left(\lambda_{m}^{2}-\lambda_{n}^{2}\right)}$$

Since Λ is diagonal, the matrix $\mathbf{e}_i \mathbf{e}_i^t \mathbf{e}_k \mathbf{e}_l^t \Lambda \mathbf{e}_m \mathbf{e}_m^t$ will be a **0** matrix unless k = i, l = m. Similarly, the matrix $\mathbf{e}_j \mathbf{e}_j^t \mathbf{e}_l \mathbf{e}_k^t \Lambda \mathbf{e}_n \mathbf{e}_n^t$ will be a **0** matrix unless l = j, k = n. These two conditions holding concordantly imply $k \in \mathcal{N}, l \in \mathcal{N}$ and $k \notin \mathcal{N}, l \notin \mathcal{N}$, which is impossible. Thus the contributions from these cross-product terms will be a **0** matrix. A similar argument can be applied to show that the contributions from the cross-product terms

$$\frac{\left(\mathbf{e}_{j}\mathbf{e}_{j}^{t}\otimes\mathbf{e}_{i}\mathbf{e}_{i}^{t}\right)\left(\mathbf{e}_{k}\mathbf{e}_{l}^{t}\otimes\mathbf{e}_{l}\mathbf{e}_{k}^{t}\right)\left(\mathbf{\Lambda}\otimes\mathbf{\Lambda}\right)\left(\mathbf{e}_{n}\mathbf{e}_{n}^{t}\otimes\mathbf{e}_{m}\mathbf{e}_{m}^{t}\right)}{\left(\lambda_{i}^{2}-\lambda_{j}^{2}\right)\left(\lambda_{m}^{2}-\lambda_{n}^{2}\right)}$$

will also result in a **0** matrix.

For cross-product terms of the form

$$\frac{\left(\mathbf{e}_{i}\mathbf{e}_{i}^{t}\otimes\mathbf{e}_{j}\mathbf{e}_{j}^{t}\right)\left(\mathbf{e}_{k}\mathbf{e}_{l}^{t}\otimes\mathbf{e}_{l}\mathbf{e}_{k}^{t}\right)\left(\mathbf{\Lambda}\otimes\mathbf{\Lambda}\right)\left(\mathbf{e}_{n}\mathbf{e}_{n}^{t}\otimes\mathbf{e}_{m}\mathbf{e}_{m}^{t}\right)}{\left(\lambda_{i}^{2}-\lambda_{j}^{2}\right)\left(\lambda_{m}^{2}-\lambda_{n}^{2}\right)} = \frac{\left(\mathbf{e}_{i}\mathbf{e}_{i}^{t}\mathbf{e}_{k}\mathbf{e}_{l}^{t}\mathbf{\Lambda}\mathbf{e}_{n}\mathbf{e}_{n}^{t}\right)\otimes\left(\mathbf{e}_{j}\mathbf{e}_{j}^{t}\mathbf{e}_{l}\mathbf{e}_{k}^{t}\mathbf{\Lambda}\mathbf{e}_{m}\mathbf{e}_{m}^{t}\right)}{\left(\lambda_{i}^{2}-\lambda_{j}^{2}\right)\left(\lambda_{m}^{2}-\lambda_{n}^{2}\right)}$$

The fact $\mathbf{\Lambda}$ is diagonal implies the matrix $\mathbf{e}_i \mathbf{e}_i^t \mathbf{e}_k \mathbf{e}_l^t \mathbf{\Lambda} \mathbf{e}_n \mathbf{e}_n^t$ will be a **0** matrix unless k = i, l = n. Similarly, the matrix $\mathbf{e}_j \mathbf{e}_j^t \mathbf{e}_l \mathbf{e}_k^t \mathbf{\Lambda} \mathbf{e}_m \mathbf{e}_m^t$ will be a **0** matrix unless l = j, k = m. These two conditions holding concordantly imply $k \in \mathcal{N}, l \notin \mathcal{N}$. The terms where the previous conditions hold reduce to $\frac{\lambda_k^2 \lambda_l^2}{(\lambda_k^2 - \lambda_l^2)^2} \mathbf{e}_k \mathbf{e}_l^t \otimes \mathbf{e}_l \mathbf{e}_k^t$, thus the contributions from these terms can be written as

$$\sum_{k \in \mathcal{N}} \sum_{l \notin \mathcal{N}} \frac{\lambda_k^2 \lambda_l^2}{\left(\lambda_k^2 - \lambda_l^2\right)^2} \mathbf{e}_k \mathbf{e}_l^t \otimes \mathbf{e}_l \mathbf{e}_k^t$$

Lastly, for the cross-product terms of the form

$$\frac{\left(\mathbf{e}_{j}\mathbf{e}_{j}^{t}\otimes\mathbf{e}_{i}\mathbf{e}_{i}^{t}\right)\left(\mathbf{e}_{k}\mathbf{e}_{l}^{t}\otimes\mathbf{e}_{l}\mathbf{e}_{k}^{t}\right)\left(\mathbf{\Lambda}\otimes\mathbf{\Lambda}\right)\left(\mathbf{e}_{m}\mathbf{e}_{m}^{t}\otimes\mathbf{e}_{n}\mathbf{e}_{n}^{t}\right)}{\left(\lambda_{i}^{2}-\lambda_{j}^{2}\right)\left(\lambda_{m}^{2}-\lambda_{n}^{2}\right)}=\frac{\left(\mathbf{e}_{j}\mathbf{e}_{j}^{t}\mathbf{e}_{k}\mathbf{e}_{l}^{t}\mathbf{\Lambda}\mathbf{e}_{m}\mathbf{e}_{m}^{t}\right)\otimes\left(\mathbf{e}_{i}\mathbf{e}_{i}^{t}\mathbf{e}_{l}\mathbf{e}_{k}^{t}\mathbf{\Lambda}\mathbf{e}_{n}\mathbf{e}_{n}^{t}\right)}{\left(\lambda_{i}^{2}-\lambda_{j}^{2}\right)\left(\lambda_{m}^{2}-\lambda_{n}^{2}\right)}$$

Since Λ is diagonal, it follows the matrix $\mathbf{e}_{j}\mathbf{e}_{j}^{t}\mathbf{e}_{k}\mathbf{e}_{l}^{t}\Lambda\mathbf{e}_{m}\mathbf{e}_{m}^{t}$ will be a **0** matrix unless k = j, l = m. Similarly, the matrix $\mathbf{e}_{i}\mathbf{e}_{i}^{t}\mathbf{e}_{l}\mathbf{e}_{k}^{t}\Lambda\mathbf{e}_{n}\mathbf{e}_{n}^{t}$ will be a **0** matrix unless l = i, k = n. These two conditions holding concordantly imply $k \notin \mathcal{N}, l \in \mathcal{N}$. The terms where the previous conditions hold reduce to $\frac{\lambda_{l}^{2}\lambda_{k}^{2}}{(\lambda_{l}^{2}-\lambda_{k}^{2})^{2}}\mathbf{e}_{k}\mathbf{e}_{l}^{t}\otimes\mathbf{e}_{l}\mathbf{e}_{k}^{t}$, thus the contributions from these terms can be written as

$$\sum_{k \notin \mathcal{N}} \sum_{l \in \mathcal{N}} \frac{\lambda_l^2 \lambda_k^2}{\left(\lambda_l^2 - \lambda_k^2\right)^2} \mathbf{e}_k \mathbf{e}_l^t \otimes \mathbf{e}_l \mathbf{e}_k^t = \sum_{l \in \mathcal{N}} \sum_{k \notin \mathcal{N}} \frac{\lambda_l^2 \lambda_k^2}{\left(\lambda_l^2 - \lambda_k^2\right)^2} \mathbf{e}_k \mathbf{e}_l^t \otimes \mathbf{e}_l \mathbf{e}_k^t$$
$$= \sum_{k' \in \mathcal{N}} \sum_{l' \notin \mathcal{N}} \frac{\lambda_{k'}^2 \lambda_{l'}^2}{\left(\lambda_{k'}^2 - \lambda_{l'}^2\right)^2} \mathbf{e}_{l'} \mathbf{e}_{k'}^t \otimes \mathbf{e}_{k'} \mathbf{e}_{l'}^t$$

Where the last equality follows by defining the new indices, l' = k and k' = l. This then leaves

$$\mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\right)\mathbf{K}_{d,d}\left(\mathbf{\Lambda}\otimes\mathbf{\Lambda}\right)\mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\right) = \sum_{k\in\mathcal{N}}\sum_{l\notin\mathcal{N}}\frac{\lambda_{k}^{2}\lambda_{l}^{2}}{\left(\lambda_{k}^{2}-\lambda_{l}^{2}\right)^{2}}\left(\mathbf{e}_{k}\mathbf{e}_{l}^{t}\otimes\mathbf{e}_{l}\mathbf{e}_{k}^{t}+\mathbf{e}_{l}\mathbf{e}_{k}^{t}\otimes\mathbf{e}_{k}\mathbf{e}_{l}^{t}\right)$$

Finally consider the expression

$$\begin{split} \mathbf{C}_{\mathcal{N}}\left(\mathbf{\Lambda}\right)\mathbf{vec}\left(\mathbf{\Lambda}\right) &= \left(\sum_{i\in\mathcal{N}}\sum_{j\notin\mathcal{N}}\frac{\mathbf{e}_{i}\mathbf{e}_{i}^{t}\otimes\mathbf{e}_{j}\mathbf{e}_{j}^{t}+\mathbf{e}_{j}\mathbf{e}_{j}^{t}\otimes\mathbf{e}_{i}\mathbf{e}_{i}^{t}}{\lambda_{i}^{2}-\lambda_{j}^{2}}\right)\mathbf{vec}\left(\mathbf{\Lambda}\right) \\ &= \sum_{i\in\mathcal{N}}\sum_{j\notin\mathcal{N}}\frac{\mathbf{vec}\left(\mathbf{e}_{j}\mathbf{e}_{j}^{t}\mathbf{\Lambda}\mathbf{e}_{i}\mathbf{e}_{i}^{t}\right)+\mathbf{vec}\left(\mathbf{e}_{i}\mathbf{e}_{i}^{t}\mathbf{\Lambda}\mathbf{e}_{j}\mathbf{e}_{j}^{t}\right)}{\lambda_{i}^{2}-\lambda_{j}^{2}}\end{split}$$

Again, the fact that Λ is diagonal implies this expression will be a zero matrix unless i = j; however, the latter never occurs since i and j are in separate indexing sets. Consequently, $\mathbf{C}_{\mathcal{N}}^{t}(\Lambda) \operatorname{vec}(\Lambda) = \mathbf{0}$. These facts imply that

$$AV_{\widehat{\mathbf{T}}_{\mathcal{N}}}\left(\mathbf{\Lambda}\right) = \frac{d+2}{d} \sum_{i \in \mathcal{N}} \sum_{j \notin \mathcal{N}} \frac{\lambda_i^2 \lambda_j^2}{\left(\lambda_i^2 - \lambda_j^2\right)^2} \left(\mathbf{e}_i \mathbf{e}_i^t \otimes \mathbf{e}_j \mathbf{e}_j^t + \mathbf{e}_i \mathbf{e}_j^t \otimes \mathbf{e}_j \mathbf{e}_i^t + \mathbf{e}_j \mathbf{e}_i^t \otimes \mathbf{e}_i \mathbf{e}_j^t + \mathbf{e}_j \mathbf{e}_i^t \otimes \mathbf{e}_i \mathbf{e}_i^t\right)$$

3.6.3 The Spatial Sign Covariance Matrix

Recall, the spatial sign covariance matrix (SSCM) is computed as follows

$$\widehat{\boldsymbol{\mathcal{S}}}_{\widehat{\boldsymbol{\mu}}_n} = \frac{1}{n} \sum_{i=1}^n \frac{\left(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_n\right) \left(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_n\right)^t}{\left(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_n\right)^t \left(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_n\right)}$$

where $\hat{\mu}_n$ is some estimate of location. Under the elliptical model, provided that $\hat{\mu}$ is a consistent estimate of μ , the asymptotic distribution of $\hat{S}_{\hat{\mu}_n}$ will not depend on $\hat{\mu}$. Consequently, one can assume that that location parameter is known and for simplicity can take it to be the origin.

For notational convenience define the following random matrix,

$$oldsymbol{\mathcal{X}} = rac{\mathbf{x}\mathbf{x}^t}{\mathbf{x}^t\mathbf{x}}$$

Note that $\hat{\boldsymbol{S}}_n$ is the average of *n* observations with the same distribution as $\boldsymbol{\mathcal{X}}$. Provided the distribution of the random matrix $\boldsymbol{\mathcal{X}}$ has finite second moments, the Strong Law of Large Numbers implies

$$lim_{n\to\infty}\widehat{\boldsymbol{\mathcal{S}}}_n = E\left[\boldsymbol{\mathcal{X}}\right] = \boldsymbol{\Xi}$$

In general, $\Xi \neq \Sigma$. Furthermore, by the central limit theorem one has

$$\sqrt{n}\mathbf{vec}\left(\widehat{\boldsymbol{\mathcal{S}}}_{n}-\boldsymbol{\Xi}\right)\rightarrow_{\mathcal{D}}Norm_{d^{2}}\left(\boldsymbol{0},AV_{\widehat{\boldsymbol{\mathcal{S}}}}\left(\boldsymbol{\mathcal{X}}\right)\right)$$

where

$$AV_{\widehat{\mathbf{S}}}(\mathbf{\mathcal{X}}) = Var\left[\operatorname{vec}(\mathbf{\mathcal{X}})\right] = E\left[\operatorname{vec}(\mathbf{\mathcal{X}} - \mathbf{\Xi})\operatorname{vec}^{t}(\mathbf{\mathcal{X}} - \mathbf{\Xi})\right]$$

It is desired to get at the form of the matrices and Ξ and $AV_{\widehat{\mathbf{S}}}(\mathbf{X})$ when $\mathbf{x} \sim \mathcal{E}_d(\mathbf{0}, \Sigma; G)$. In this situation, the distribution of random matrix \mathbf{X} is the same for all distributions G; conveniently, this distribution has finite second moments (the entries of the matrix are all bounded by 1) thus the above results hold. By orthogonal equivariance, one need only consider the case $\mathbf{y} \sim \mathcal{E}_d(\mathbf{0}, \mathbf{\Lambda}; G)$ (thus $\mathbf{x} \sim \mathbf{P}\mathbf{y}$). To see this note that

$$oldsymbol{\mathcal{X}} = rac{\mathbf{P}\mathbf{y}\left(\mathbf{P}\mathbf{y}
ight)^t}{\left(\mathbf{P}\mathbf{y}
ight)^t\mathbf{P}\mathbf{y}} = rac{\mathbf{P}\mathbf{y}\mathbf{y}^t\mathbf{P}^t}{\mathbf{y}\mathbf{P}^t\mathbf{P}\mathbf{y}} = \mathbf{P}rac{\mathbf{y}\mathbf{y}^t}{\mathbf{y}^t\mathbf{y}}\mathbf{P}^t = \mathbf{P}oldsymbol{\mathcal{Y}}\mathbf{P}^t$$

where

$$oldsymbol{\mathcal{Y}} = rac{\mathbf{y}\mathbf{y}^t}{\mathbf{y}^t\mathbf{y}}$$

It then follows

$$\Xi = E\left[\mathbf{P}\mathbf{\mathcal{Y}}\mathbf{P}^t\right] = \mathbf{P}E\left[\mathbf{\mathcal{Y}}\right]\mathbf{P}^t = \mathbf{P}\mathbf{\Delta}\mathbf{P}^t$$

where $\mathbf{\Delta} = E[\mathbf{\mathcal{Y}}]$. It will be shown that $\mathbf{\Delta}$ is a diagonal matrix. Recall $\mathbf{y} \in \mathbb{R}^d$, hence it is of the form $\mathbf{y} = (y_1, \dots, y_d)^t$ (y_i represent the *d* components of \mathbf{y} , not the *n* observations in the sample). Define

$$\gamma_{ij} = \frac{y_i y_j}{\mathbf{y}^t \mathbf{y}} = \frac{y_i y_j}{y_1^2 + \dots + y_d^2}$$

thus $\boldsymbol{\mathcal{Y}}$ is a matrix where the (i, j)th entry is γ_{ij} . One can write it as

$$oldsymbol{\mathcal{Y}} = \sum_{i=1}^d \sum_{j=1}^d \gamma_{ij} \left(\mathbf{e}_i \otimes \mathbf{e}_j^t
ight)$$

Hence $\mathbf{\Delta} = E\left[\{\gamma_{ij}\}_{i,j=1}^{d}\right] = \{E\left[\gamma_{ij}\right]\}_{i,j=1}^{d}$. It will now be shown that $E\left[\gamma_{ij}\right] = 0$ for $i \neq j$. Define another random vector $\mathbf{y}' = (y_1, y_2, \dots, -y_i, \dots, y_d)$, since $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Lambda}$ is diagonal it follows $\mathbf{y} \sim \mathbf{y}'$. Thus

$$\gamma_{ij} = rac{y_i y_j}{\mathbf{y}^t \mathbf{y}} \sim -\gamma_{ij} = rac{-y_i y_j}{\mathbf{y}^t \mathbf{y}}$$

 So

$$E[\gamma_{ij}] = E[-\gamma_{ij}] = E\left[\frac{-y_i y_j}{\mathbf{y}^t \mathbf{y}}\right] = -E\left[\frac{y_i y_j}{\mathbf{y}^t \mathbf{y}}\right] = -E[\gamma_{ij}]$$

Thus implying $E[\gamma_{ij}] = 0$ for $i \neq j$. This kind of argument will be used again, however the details omitted.

Under the elliptical model, the asymptotic variance-covariance matrix for the SSCM will only depend on the scatter structure, thus the notation $AV_{\hat{s}}(\Sigma)$ is utilized. It follows

$$\begin{aligned} AV_{\widehat{\boldsymbol{\mathcal{S}}}}\left(\boldsymbol{\Sigma}\right) &= E\left[\operatorname{\mathbf{vec}}\left(\mathbf{P}\boldsymbol{\mathcal{Y}}\mathbf{P}^{t} - \mathbf{P}\boldsymbol{\Delta}\mathbf{P}^{t}\right)\operatorname{\mathbf{vec}}^{t}\left(\mathbf{P}\boldsymbol{\mathcal{Y}}\mathbf{P}^{t} - \mathbf{P}\boldsymbol{\Delta}\mathbf{P}^{t}\right)\right] \\ &= E\left[\operatorname{\mathbf{vec}}\left(\mathbf{P}\left(\boldsymbol{\mathcal{Y}}-\boldsymbol{\Delta}\right)\mathbf{P}^{t}\right)\operatorname{\mathbf{vec}}^{t}\left(\mathbf{P}\left(\boldsymbol{\mathcal{Y}}-\boldsymbol{\Delta}\right)\mathbf{P}^{t}\right)\right] \\ &= E\left[\left(\mathbf{P}\otimes\mathbf{P}\right)\operatorname{\mathbf{vec}}\left(\boldsymbol{\mathcal{Y}}-\boldsymbol{\Delta}\right)\left(\left(\mathbf{P}\otimes\mathbf{P}\right)\operatorname{\mathbf{vec}}\left(\boldsymbol{\mathcal{Y}}-\boldsymbol{\Delta}\right)\right)^{t}\right] \\ &= \left(\mathbf{P}\otimes\mathbf{P}\right)E\left[\operatorname{\mathbf{vec}}\left(\boldsymbol{\mathcal{Y}}-\boldsymbol{\Delta}\right)\operatorname{\mathbf{vec}}^{t}\left(\boldsymbol{\mathcal{Y}}-\boldsymbol{\Delta}\right)\right]\left(\mathbf{P}^{t}\otimes\mathbf{P}^{t}\right) \\ &= \left(\mathbf{P}\otimes\mathbf{P}\right)\mathcal{H}\left(\mathbf{P}^{t}\otimes\mathbf{P}^{t}\right)\end{aligned}$$

where

$$\mathcal{H} = E\left[\operatorname{vec}\left(\mathcal{Y} - \boldsymbol{\Delta}\right)\operatorname{vec}^{t}\left(\mathcal{Y} - \boldsymbol{\Delta}\right)\right] = E\left[\left(\operatorname{vec}\left(\mathcal{Y}\right) - \operatorname{vec}\left(\boldsymbol{\Delta}\right)\right)\left(\operatorname{vec}\left(\mathcal{Y}\right) - \operatorname{vec}\left(\boldsymbol{\Delta}\right)\right)^{t}\right]$$

$$=E\left[\operatorname{vec}\left(\boldsymbol{\mathcal{Y}}\right)\operatorname{vec}^{t}\left(\boldsymbol{\mathcal{Y}}\right)-\operatorname{vec}\left(\boldsymbol{\mathcal{Y}}\right)\operatorname{vec}^{t}\left(\boldsymbol{\Delta}\right)-\operatorname{vec}\left(\boldsymbol{\Delta}\right)\operatorname{vec}^{t}\left(\boldsymbol{\mathcal{Y}}\right)+\operatorname{vec}\left(\boldsymbol{\Delta}\right)\operatorname{vec}^{t}\left(\boldsymbol{\Delta}\right)\right]$$

Since $E[\mathbf{vec}(\boldsymbol{\mathcal{Y}})] = \mathbf{vec}(\boldsymbol{\Delta})$, which is a vector of constants, above simplifies to

$$\mathcal{H} = E\left[\operatorname{vec}\left(\mathcal{Y}\right)\operatorname{vec}^{t}\left(\mathcal{Y}\right)\right] - \operatorname{vec}\left(\boldsymbol{\Delta}\right)\operatorname{vec}^{t}\left(\boldsymbol{\Delta}\right)$$

Write

$$\mathbf{vec}\left(\boldsymbol{\mathcal{Y}}\right) = \sum_{i=1}^{d} \sum_{j=1}^{d} \gamma_{ij} \mathbf{vec}\left(\mathbf{e}_{i} \otimes \mathbf{e}_{j}^{t}\right) = \sum_{i=1}^{d} \sum_{j=1}^{d} \gamma_{ij} \mathbf{vec}\left(\mathbf{e}_{i} \mathbf{e}_{j}^{t}\right)$$

thus

$$\mathbf{vec}\left(\boldsymbol{\mathcal{Y}}\right)\mathbf{vec}^{t}\left(\boldsymbol{\mathcal{Y}}\right) = \left(\sum_{i=1}^{d}\sum_{j=1}^{d}\gamma_{ij}\mathbf{vec}\left(\mathbf{e}_{i}\mathbf{e}_{j}^{t}\right)\right)\left(\sum_{k=1}^{d}\sum_{l=1}^{d}\gamma_{kl}\mathbf{vec}\left(\mathbf{e}_{k}\mathbf{e}_{l}^{t}\right)\right)^{t}$$
$$= \sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d}\sum_{l=1}^{d}\gamma_{ij}\gamma_{kl}\mathbf{vec}\left(\mathbf{e}_{i}\mathbf{e}_{j}^{t}\right)\mathbf{vec}^{t}\left(\mathbf{e}_{k}\mathbf{e}_{l}^{t}\right)$$

This gives

$$E\left[\operatorname{\mathbf{vec}}\left(\boldsymbol{\mathcal{Y}}\right)\operatorname{\mathbf{vec}}^{t}\left(\boldsymbol{\mathcal{Y}}\right)\right] = \sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d}\sum_{l=1}^{d}E\left[\gamma_{ij}\gamma_{kl}\right]\operatorname{\mathbf{vec}}\left(\mathbf{e}_{i}\mathbf{e}_{j}^{t}\right)\operatorname{\mathbf{vec}}^{t}\left(\mathbf{e}_{k}\mathbf{e}_{l}^{t}\right)$$

Repeating arguments similar to the one employed to show Δ was diagonal, it can be shown that the following terms are zero.

$$E[\gamma_{ij}\gamma_{ii}], E[\gamma_{ij}\gamma_{jj}], E[\gamma_{ij}\gamma_{kk}], E[\gamma_{ij}\gamma_{kl}], E[\gamma_{ij}\gamma_{jk}]$$
$$E[\gamma_{ii}\gamma_{ij}], E[\gamma_{ii}\gamma_{ji}], E[\gamma_{ii}\gamma_{kl}], E[\gamma_{ij}\gamma_{ki}]$$

Thus, the only expectations that are nonnegative are

$$E[\gamma_{ii}\gamma_{ii}], E[\gamma_{ij}\gamma_{ij}], E[\gamma_{ii}\gamma_{jj}], E[\gamma_{ij}\gamma_{ji}]$$

which corresponding to i = j = k = l, i = k & j = l, i = j & k = l and i = l & j = krespectively.

It can be shown that the matrix $AV_{\widehat{\mathbf{S}}}(\mathbf{\Sigma})$ has a similar structure to $AV_{\widehat{\mathbf{T}}}(\mathbf{\Sigma})$; namely, the elements of the matrices will be non-zero/zero in the same locations. First consider the situation $\mathbf{\Sigma} = \mathbf{\Lambda}$. Recall the form of $AV_{\widehat{\mathbf{T}}}(\mathbf{\Lambda})$ in this situation is

$$AV_{\widehat{\boldsymbol{\mathcal{T}}}}\left(\boldsymbol{\Lambda}\right) = \frac{d+2}{d} \left(\mathbf{I}_{d^{2}} + \mathbf{K}_{d,d}\right) \left(\boldsymbol{\Lambda} \otimes \boldsymbol{\Lambda}\right) + \frac{2}{d} \frac{d+2}{d} \mathbf{vec}\left(\boldsymbol{\Lambda}\right) \mathbf{vec}^{t}\left(\boldsymbol{\Lambda}\right)$$

The matrix $(\mathbf{\Lambda} \otimes \mathbf{\Lambda})$ has non-zero/zero elements at the same locations as the matrix $(\mathbf{I}_d \otimes \mathbf{I}_d) = \mathbf{I}_{d^2}$. Similarly, the matrices $\mathbf{vec}(\mathbf{\Lambda}) \mathbf{vec}^t(\mathbf{\Lambda})$ and $\mathbf{vec}(\mathbf{\Delta}) \mathbf{vec}^t(\mathbf{\Delta})$ have non-zero/zero elements at the same locations as the matrix $\mathbf{vec}(\mathbf{I}_d) \mathbf{vec}^t(\mathbf{I}_d)$. Hence one can focus on the non-zero/zero terms of the matrices \mathbf{I}_{d^2} , $\mathbf{K}_{d,d}$ and $\mathbf{vec}(\mathbf{I}_d) \mathbf{vec}^t(\mathbf{I}_d)$. Writing these out in terms of the unit canonical vectors, one has

$$\mathbf{I}_{d^2} = \mathbf{I}_d \otimes \mathbf{I}_d = \left(\sum_{i=1}^d \mathbf{e}_i \mathbf{e}_i^t\right) \otimes \left(\sum_{j=1}^d \mathbf{e}_j \mathbf{e}_j^t\right) = \sum_{i=1}^d \sum_{j=1}^d \mathbf{e}_i \mathbf{e}_i^t \otimes \mathbf{e}_j \mathbf{e}_j^t =$$
$$= \sum_{i=1}^d \sum_{j=1}^d \left(\mathbf{e}_i \otimes \mathbf{e}_j\right) \left(\mathbf{e}_i^t \otimes \mathbf{e}_j^t\right) = \sum_{i=1}^d \sum_{j=1}^d \left(\mathbf{e}_i \otimes \mathbf{e}_j\right) \left(\mathbf{e}_i \otimes \mathbf{e}_j\right)^t = \sum_{i=1}^d \sum_{j=1}^d \operatorname{vec}\left(\mathbf{e}_j \mathbf{e}_i^t\right) \operatorname{vec}^t\left(\mathbf{e}_j \mathbf{e}_i^t\right)$$

Thus \mathbf{I}_{d^2} is non-zero at the locations corresponding to the terms $E[\gamma_{ii}\gamma_{ii}]$ and $E[\gamma_{ij}\gamma_{ij}]$.

$$\mathbf{K}_{d,d} = \sum_{i=1}^{d} \sum_{j=1}^{d} \mathbf{e}_{i} \mathbf{e}_{j}^{t} \otimes \mathbf{e}_{j} \mathbf{e}_{i}^{t} = \sum_{i=1}^{d} \sum_{j=1}^{d} (\mathbf{e}_{i} \otimes \mathbf{e}_{j}) (\mathbf{e}_{j}^{t} \otimes \mathbf{e}_{i}^{t}) =$$
$$= \sum_{i=1}^{d} \sum_{j=1}^{d} (\mathbf{e}_{i} \otimes \mathbf{e}_{j}) (\mathbf{e}_{j} \otimes \mathbf{e}_{i})^{t} = \sum_{i=1}^{d} \sum_{j=1}^{d} \mathbf{vec} (\mathbf{e}_{j} \mathbf{e}_{i}^{t}) \mathbf{vec}^{t} (\mathbf{e}_{i} \mathbf{e}_{j}^{t})$$

Thus $\mathbf{K}_{d,d}$ is non-zero at the locations corresponding to the terms $E[\gamma_{ii}\gamma_{ii}]$ and $E[\gamma_{ij}\gamma_{ji}]$. Lastly,

$$\mathbf{vec}\left(\mathbf{I}_{d}
ight)\mathbf{vec}^{t}\left(\mathbf{I}_{d}
ight) = \mathbf{vec}\left(\sum_{i=1}^{d}\mathbf{e}_{i}\mathbf{e}_{i}^{t}
ight)\mathbf{vec}^{t}\left(\sum_{j=1}^{d}\mathbf{e}_{j}\mathbf{e}_{j}^{t}
ight) =$$

$$= \left(\sum_{i=1}^{d} \operatorname{\mathbf{vec}}\left(\mathbf{e}_{i} \mathbf{e}_{i}^{t}\right)\right) \left(\sum_{j=1}^{d} \operatorname{\mathbf{vec}}^{t}\left(\mathbf{e}_{j} \mathbf{e}_{j}^{t}\right)\right) = \sum_{i=1}^{d} \sum_{j=1}^{d} \operatorname{\mathbf{vec}}\left(\mathbf{e}_{i} \mathbf{e}_{i}^{t}\right) \operatorname{\mathbf{vec}}^{t}\left(\mathbf{e}_{j} \mathbf{e}_{j}^{t}\right)$$

Thus $\operatorname{vec}(\mathbf{I}_d) \operatorname{vec}^t(\mathbf{I}_d)$ is non-zero at the locations corresponding to the terms $E[\gamma_{ii}\gamma_{ii}]$, $E[\gamma_{ii}\gamma_{jj}]$ and the contributions from the term $-\operatorname{vec}(\boldsymbol{\Delta}) \operatorname{vec}^t(\boldsymbol{\Delta})$. The general case follows from the fact that the asymptotic variance-covariance matrices for both cases are equal to $AV_{\widehat{\mathcal{T}}}(\boldsymbol{\Sigma}) = (\mathbf{P} \otimes \mathbf{P}) AV_{\widehat{\mathcal{T}}}(\boldsymbol{\Lambda}) (\mathbf{P}^t \otimes \mathbf{P}^t)$

and $AV_{\widehat{\boldsymbol{\mathcal{S}}}}(\boldsymbol{\Sigma}) = (\mathbf{P} \otimes \mathbf{P}) AV_{\widehat{\boldsymbol{\mathcal{S}}}}(\boldsymbol{\Lambda}) (\mathbf{P}^t \otimes \mathbf{P}^t)$ respectively. In fact, when $\boldsymbol{\Sigma} \propto \mathbf{I}_d$, it can be shown that $AV_{\widehat{\boldsymbol{\mathcal{S}}}}(\sigma^2 \mathbf{I}_d)$ and $AV_{\widehat{\boldsymbol{\mathcal{T}}}}(\sigma^2 \mathbf{I}_d)$ are proportional. This is stated formerly in the following lemma

Lemma 4.1.1. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ represent an i.i.d. sample from $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; G)$ and let $\widehat{\boldsymbol{\mu}}$ be an asymptotically consistent estimator of $\boldsymbol{\mu}$. It then follows that $AV_{\widehat{\boldsymbol{S}}_{\widehat{\boldsymbol{\mu}}}}(\sigma^2 \boldsymbol{I}_d) \propto AV_{\widehat{\boldsymbol{T}}_{\widehat{\boldsymbol{\mu}}}}(\sigma^2 \boldsymbol{I}_d)$.

Proof. Under spherical symmetry, the results of [34] Theorem 1 dictate that the form of the covariance matrix of any orthogonally equivariant estimate of scatter is $\sigma_1 (\mathbf{I}_{d^2} + \mathbf{K}_{d,d}) (\mathbf{I}_d \otimes \mathbf{I}_d) + \sigma_2 \mathbf{vec} (\boldsymbol{\Sigma}) \mathbf{vec}^t (\boldsymbol{\Sigma})$, where $\sigma_1 \geq 0$ and $\sigma_2 \geq -2\sigma_1/d$. The form of this matrix implies that the off-diagonal elements of the scatter estimate are uncorrelated with each other as well as uncorrelated with the diagonal elements. Additionally, the off-diagonal elements each have variance σ_1 , whereas the diagonal elements

have variances $2\sigma_1 + \sigma_2$. The covariance between any two diagonal elements is σ_2 . Both Tyler's matrix and the SSCM are orthogonally equivariant, thus the prior results hold for both estimates. Recall the values of these constants for the Tyler matrix are given by

$$\sigma_{\boldsymbol{\mathcal{T}}_1} = rac{d+2}{d} ext{ and } \sigma_{\boldsymbol{\mathcal{T}}_2} = -rac{d+2}{d}rac{2}{d} = -\sigma_{\boldsymbol{\mathcal{T}}_1}rac{2}{d}$$

To show the desired result for the SSCM, one must show $\sigma_{\mathbf{S}_1} = c (d+2)/d$ and $\sigma_{\mathbf{S}_2} = -c\sigma_{\mathbf{T}_1}(2/d)$ for some c > 0. This is equivalent to showing $\sigma_{\mathbf{S}_1}/\sigma_{\mathbf{S}_2} = -d/2$. Without loss of generality assume $\boldsymbol{\mu} = \mathbf{0}$, for $\mathbf{z} \sim \mathcal{E}_d(\mathbf{0}, \sigma^2 \mathbf{I}_d; G)$ define

$$oldsymbol{\mathcal{Z}} = rac{\mathbf{z} \mathbf{z}^t}{\mathbf{z}^t \mathbf{z}}$$

Define

$$\zeta_{ij} = \frac{z_i z_j}{\mathbf{z}^t \mathbf{z}} = \frac{z_i z_j}{z_1^2 + \dots + z_d^2}$$

that is ζ_{ij} the ij^{th} element of the random matrix $\boldsymbol{\mathcal{Z}}$. Since \mathbf{I}_d is diagonal it follows that $E[\boldsymbol{\mathcal{Z}}]$ will be as well. Further more, the fact $\zeta_{ii}, i = 1, \ldots, d$, are identically distributed implies that is proportional to the identity matrix. To find this constant of proportionality, note that $Trace(\boldsymbol{\mathcal{Z}}) = 1$, thus it follows $E[\boldsymbol{\mathcal{Z}}] = \frac{1}{d}\mathbf{I}_d$. The previous notation and fact give

$$\sigma_{S_1} = E\left[\zeta_{ij}^2\right] = E\left[\frac{z_i^2 z_j^2}{\left(z_1^2 + \dots + z_d^2\right)^2}\right]$$
$$\sigma_{S_2} = E\left[\left(\zeta_{ii} - \frac{1}{d}\right)\left(\zeta_{jj} - \frac{1}{d}\right)\right] = E\left[\zeta_{ii}\zeta_{jj}\right] - \frac{1}{d}E\left[\zeta_{ii}\right] - \frac{1}{d}E\left[\zeta_{jj}\right] + \frac{1}{d^2} =$$
$$= E\left[\frac{z_i^2 z_j^2}{\left(z_1^2 + \dots + z_d^2\right)^2}\right] - \frac{1}{d^2} = \sigma_{S_1} - \frac{1}{d^2}$$

To proceed further, the value $\sigma_{\mathcal{S}_1}$ must be calculated. Using the stochastic representation of the random vector \mathbf{z} , it can be be shown that the distribution of $z_i^2 z_j^2 / (z_1^2 + \cdots + z_d^2)^2$ will not depend on the radial component of the elliptical random vector, that is it does not depend on G. This reduces to $\sigma_{\mathcal{S}_1}$.

$$\sigma_{\boldsymbol{\mathcal{S}}_1} = E\left[\frac{u_i^2 u_j^2}{\left(u_1^2 + \dots + u_d^2\right)^2}\right] = E\left[u_i^2 u_j^2\right]$$

where $\mathbf{u} = (u_1, \dots, u_d)^t$ is a unit vector uniformly distributed on the unit sphere in \mathbb{R}^d . Define $U_1 = u_i^2$ and $U_2 = 1 - u_i^2$. By exchangeability it follows,

$$E\left[u_i^2 u_j^2\right] = \frac{1}{d-1} \sum_{j \neq i} E\left[u_i^2 u_j^2\right] = \frac{1}{d-1} E\left[U_1 U_2\right] = \frac{1}{d-1} E\left[U_1 \left(1 - U_1\right)\right]$$

It is known that $U_1 \sim Beta(1/2, (d-1)/2)$, thus

$$E[U_1(1-U_1)] = E[U_1 - U_1^2] = E[U_1] - Var[U_1] - E^2[U_1] =$$
$$= \frac{1}{d} - \frac{(d-1)/4}{(d/2)^2(d+2)/2} - \frac{1}{d^2} = \frac{1}{d}\frac{d-1}{d+2}$$

Therefore $\sigma_{\boldsymbol{\mathcal{S}}_1} = 1/d (d+2)$ and $\sigma_{\boldsymbol{\mathcal{S}}_2} = 1/d (d+2) - 1/d^2 = -2/d^2 (d+2)$. Hence $\sigma_{\boldsymbol{\mathcal{S}}_1}/\sigma_{\boldsymbol{\mathcal{S}}_2} = -d/2$, thus proving the result.**QED**

Just as with Tyler's scatter matrix, it will be the case that

$$\sqrt{n}\left(\widehat{\mathbf{S}}_{\mathcal{N}}-\mathbf{P}_{\mathcal{N}}\right) \rightarrow_{\mathcal{D}} Norm_{d^{2}}\left(\mathbf{0},AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}\left(\mathbf{\Sigma}\right)\right)$$

where

$$AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}(\mathbf{\Sigma}) = \mathbf{C}_{\mathcal{N}}^{t}(\mathbf{\Xi}) AV_{\widehat{\mathbf{S}}}(\mathbf{\Sigma}) \mathbf{C}_{\mathcal{N}}(\mathbf{\Xi})$$

The fact that $AV_{\widehat{\mathbf{S}}}(\mathbf{\Sigma})$ will be zero/non-zero in the same locations as the matrix $AV_{\widehat{\mathbf{T}}}(\mathbf{\Sigma})$ implies the same about the matrices $AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}(\mathbf{\Sigma})$ and $AV_{\widehat{\mathbf{T}}_{\mathcal{N}}}(\mathbf{\Sigma})$. This fact further implies $AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}(\mathbf{\Sigma}) = (\mathbf{P} \otimes \mathbf{P}) AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}(\mathbf{\Lambda}) (\mathbf{P}^t \otimes \mathbf{P}^t)$, thus it is desired to get the exact from of the matrix $AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}(\mathbf{\Lambda})$. In this situation it follows

$$AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}\left(\mathbf{\Lambda}\right) = \mathbf{C}_{\mathcal{N}}^{t}\left(\mathbf{\Delta}\right) \left(E\left[\mathbf{vec}\left(\boldsymbol{\mathcal{Y}}\right)\mathbf{vec}^{t}\left(\boldsymbol{\mathcal{Y}}\right)\right] - \mathbf{vec}\left(\mathbf{\Delta}\right)\mathbf{vec}^{t}\left(\mathbf{\Delta}\right)\right) \mathbf{C}_{\mathcal{N}}\left(\mathbf{\Delta}\right)$$

However, there will be some simplification. Note that

$$\mathbf{C}_{\mathcal{N}}^{t}\left(\boldsymbol{\Delta}\right)\mathbf{vec}\left(\boldsymbol{\Delta}\right) = \left\{\sum_{i\in\mathcal{N}}\sum_{j\notin\mathcal{N}}\frac{\mathbf{e}_{i}\mathbf{e}_{i}^{t}\otimes\mathbf{e}_{j}\mathbf{e}_{j}^{t} + \mathbf{e}_{j}\mathbf{e}_{j}^{t}\otimes\mathbf{e}_{i}\mathbf{e}_{i}^{t}}{\delta_{i}^{2} - \delta_{j}^{2}}\right\}^{t}\mathbf{vec}\left(\boldsymbol{\Delta}\right)$$

The expression above involves terms of the form

$$\left(\mathbf{e}_{i}\mathbf{e}_{i}^{t}\otimes\mathbf{e}_{j}\mathbf{e}_{j}^{t}\right)\mathbf{vec}\left(\mathbf{\Delta}\right)+\left(\mathbf{e}_{j}\mathbf{e}_{j}^{t}\otimes\mathbf{e}_{i}\mathbf{e}_{i}^{t}\right)\mathbf{vec}\left(\mathbf{\Delta}\right)=\mathbf{vec}\left(\mathbf{e}_{j}\mathbf{e}_{j}^{t}\mathbf{\Delta}\mathbf{e}_{i}\mathbf{e}_{i}^{t}\right)+\mathbf{vec}\left(\mathbf{e}_{i}\mathbf{e}_{i}^{t}\mathbf{\Delta}\mathbf{e}_{j}\mathbf{e}_{j}^{t}\right)$$

The term $\mathbf{e}_k^t \Delta \mathbf{e}_l$ is the $k l^{th}$ element of the matrix Δ which is necessarily 0 since it is diagonal. This implies $\mathbf{C}_{\mathcal{N}}^t(\Delta) \operatorname{vec}(\Delta) = \mathbf{0}$. This leaves,

$$\begin{aligned} AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}\left(\mathbf{\Lambda}\right) &= \mathbf{C}_{\mathcal{N}}^{t}\left(\mathbf{\Delta}\right) E\left[\mathbf{vec}\left(\mathbf{\mathcal{Y}}\right)\mathbf{vec}^{t}\left(\mathbf{\mathcal{Y}}\right)\right] \mathbf{C}_{\mathcal{N}}\left(\mathbf{\Delta}\right) \\ &= \mathbf{C}_{\mathcal{N}}^{t}\left(\mathbf{\Delta}\right) \left\{\sum_{i=1}^{d} \sum_{j=1}^{d} \sum_{k=1}^{d} \sum_{l=1}^{d} E\left[\gamma_{ij}\gamma_{kl}\right]\mathbf{vec}\left(\mathbf{e}_{i}\mathbf{e}_{j}^{t}\right)\mathbf{vec}^{t}\left(\mathbf{e}_{k}\mathbf{e}_{l}^{t}\right)\right\} \mathbf{C}_{\mathcal{N}}\left(\mathbf{\Delta}\right) \end{aligned}$$

The fact that $AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}(\mathbf{\Sigma})$ and $AV_{\widehat{\mathbf{T}}_{\mathcal{N}}}(\mathbf{\Sigma})$ have non-zero/zero elements at the same location necessitates that

$$AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}\left(\mathbf{\Lambda}\right) = \sum_{i \in \mathcal{N}} \sum_{j \notin \mathcal{N}} \frac{E\left[\gamma_{ij}\gamma_{ij}\right]}{\left(\delta_{i}^{2} - \delta_{j}^{2}\right)^{2}} \left(\mathbf{e}_{i}\mathbf{e}_{i}^{t} \otimes \mathbf{e}_{j}\mathbf{e}_{j}^{t} + \mathbf{e}_{i}\mathbf{e}_{j}^{t} \otimes \mathbf{e}_{j}\mathbf{e}_{i}^{t} + \mathbf{e}_{j}\mathbf{e}_{i}^{t} \otimes \mathbf{e}_{i}\mathbf{e}_{j}^{t} + \mathbf{e}_{j}\mathbf{e}_{j}^{t} \otimes \mathbf{e}_{i}\mathbf{e}_{i}^{t}\right)$$

Since, $\mathbf{y} \sim \mathcal{E}_d(\mathbf{0}, \mathbf{\Lambda}, G)$, the distribution of \mathcal{Y} does not depend on G, consequently nor will γ_{ij} . Using the stochastic representation of the random vector \mathbf{Y} , it can be shown,

$$\delta_{\iota}^{2} = E\left[\gamma_{\iota\iota}\right] = E\left[\frac{\lambda_{\iota}^{2}u_{\iota}^{2}}{\sum_{i=1}^{d}\lambda_{i}^{2}u_{i}^{2}}\right]$$
$$E\left[\gamma_{ij}\gamma_{ij}\right] = E\left[\frac{\lambda_{i}^{2}\lambda_{j}^{2}u_{i}^{2}u_{j}^{2}}{\left(\sum_{k=1}^{d}\lambda_{k}^{2}u_{k}^{2}\right)\left(\sum_{l=1}^{d}\lambda_{l}^{2}u_{l}^{2}\right)}\right]$$

3.6.4 Asymptotic Calculations

Results stated in section 3.4.1 without proof are presented in this section. First the asymptotic covariance matrices of the estimators $\widehat{\mathbf{T}}_{\mathcal{N}}$ and $\widehat{\mathbf{S}}_{\mathcal{N}}$ will be calculated under the elliptical model for the scatter matrix

$$\mathbf{\Lambda}_0 = diag\left(\underbrace{\lambda^2, \dots, \lambda^2}_{d_1}, \underbrace{r^2 \lambda^2, \dots, r^2 \lambda^2}_{d-d_1}\right)$$

In this instance, $\mathbf{P}_{\mathcal{N}} = \sum_{i=1}^{d_1} \mathbf{E}_i = E_{D_1}$. For $\widehat{\mathbf{T}}_{\mathcal{N}}$ it easily follows,

$$AV_{\widehat{\mathbf{T}}_{\mathcal{N}}}(\mathbf{\Lambda}_{0}) = \sum_{i=1}^{d_{1}} \sum_{j=d_{1}+1}^{d} \frac{d+2}{d} \frac{r^{2}\lambda^{2}\lambda^{2}}{(\lambda^{2}-r^{2}\lambda^{2})^{2}} \left(\mathbf{e}_{i}\mathbf{e}_{i}^{t} \otimes \mathbf{e}_{j}\mathbf{e}_{j}^{t} + \mathbf{e}_{i}\mathbf{e}_{j}^{t} \otimes \mathbf{e}_{j}\mathbf{e}_{i}^{t} + \mathbf{e}_{j}\mathbf{e}_{i}^{t} \otimes \mathbf{e}_{i}\mathbf{e}_{j}^{t} + \mathbf{e}_{j}\mathbf{e}_{i}^{t} \otimes \mathbf{e}_{i}\mathbf{e}_{i}^{t}\right)$$
$$= \sum_{i=1}^{d_{1}} \sum_{j=d_{1}+1}^{d} \frac{d+2}{d} \frac{r^{2}}{(1-r^{2})^{2}} \left(\mathbf{e}_{i}\mathbf{e}_{i}^{t} \otimes \mathbf{e}_{j}\mathbf{e}_{j}^{t} + \mathbf{e}_{i}\mathbf{e}_{j}^{t} \otimes \mathbf{e}_{j}\mathbf{e}_{i}^{t} + \mathbf{e}_{j}\mathbf{e}_{i}^{t} \otimes \mathbf{e}_{i}\mathbf{e}_{j}^{t} + \mathbf{e}_{j}\mathbf{e}_{i}^{t} \otimes \mathbf{e}_{i}\mathbf{e}_{i}^{t}\right)$$

For the SSCM, it was shown in the previous section that

$$AV_{\widehat{\mathbf{S}}_{\mathcal{N}}}\left(\mathbf{\Lambda}\right) = \sum_{i=1}^{d_{1}} \sum_{j=d_{1}+1}^{d} \frac{E\left[\gamma_{ij}\gamma_{ij}\right]}{\left(\delta_{i}^{2}-\delta_{j}^{2}\right)^{2}} \left(\mathbf{e}_{i}\mathbf{e}_{i}^{t}\otimes\mathbf{e}_{j}\mathbf{e}_{j}^{t} + \mathbf{e}_{i}\mathbf{e}_{j}^{t}\otimes\mathbf{e}_{j}\mathbf{e}_{i}^{t} + \mathbf{e}_{j}\mathbf{e}_{i}^{t}\otimes\mathbf{e}_{i}\mathbf{e}_{j}^{t} + \mathbf{e}_{j}\mathbf{e}_{i}^{t}\otimes\mathbf{e}_{i}\mathbf{e}_{i}^{t}\right)$$

where

$$\begin{split} \delta_i^2 &= E\left[\frac{\lambda^2 u_i^2}{\lambda^2 u_1^2 + \dots + \lambda^2 u_{d_1}^2 + r^2 \lambda^2 u_{d_1+1}^2 + \dots + r^2 \lambda^2 u_d^2}\right] \\ &= E\left[\frac{u_i^2}{u_1^2 + \dots + u_{d_1}^2 + r^2 u_{d_1+1}^2 + \dots + r^2 u_d^2}\right] \end{split}$$

$$\begin{split} \delta_j^2 &= E\left[\frac{r^2\lambda^2 u_j^2}{\lambda^2 u_1^2 + \dots + \lambda^2 u_{d_1}^2 + r^2\lambda^2 u_{d_1+1}^2 + \dots + r^2\lambda^2 u_d^2}\right] \\ &= r^2 E\left[\frac{u_j^2}{u_1^2 + \dots + u_{d_1}^2 + r^2 u_{d_1+1}^2 + \dots + r^2 u_d^2}\right] \end{split}$$

$$E[\gamma_{ij}\gamma_{ij}] = E\left[\frac{\lambda^2 u_i^2 r^2 \lambda^2 u_j^2}{\left(\lambda^2 u_1^2 + \dots + \lambda^2 u_{d_1}^2 + r^2 \lambda^2 u_{d_1+1}^2 + \dots + r^2 \lambda^2 u_d^2\right)^2}\right]$$
$$= r^2 E\left[\frac{u_i^2 u_j^2}{\left(u_1^2 + \dots + u_{d_1}^2 + r^2 u_{d_1+1}^2 + \dots + r^2 u_d^2\right)^2}\right]$$

for all $i = 1, ..., d_1$ and $j = d_1 + 1, ..., d$. Fortunately, for this scatter structure, the values of δ_{ι}^2 and $E[\gamma_{ij}\gamma_{ij}]$ can be expressed in terms of Gauss hypergeometric functions. Similar to the proof of Lemma 4.1.1, define the variables $U_1 = u_1^2 + \cdots + u_{d_1}^2$, $U_2 = u_{d_1+1}^2 + \cdots + u_d^2$, consequently $U_1 = 1 - U_2$. Further define and $U = U_1 + r^2 U_2 = 1 - kU_2$ where $k = 1 - r^2$. Exploiting this notation and exchangeability, one can write

$$\delta_i^2 = \frac{1}{d_1} E\left[U_1 \left(1 - kU_2 \right)^{-1} \right] = \frac{1}{d_1} E\left[\left(1 - U_2 \right) \left(1 - kU_2 \right)^{-1} \right]$$
$$\delta_j^2 = \frac{r^2}{d - d_1} E\left[U_2 \left(1 - kU_2 \right)^{-1} \right] = \frac{1 - k}{d - d_1} E\left[U_2 \left(1 - kU_2 \right)^{-1} \right]$$

Also, by exchangeability it follows

$$E\left[\gamma_{ij}\gamma_{ij}\right] = \frac{r^2}{d_1\left(d-d_1\right)}E\left[U_1U_2\left(1-kU_2\right)^{-2}\right] = \frac{1-k}{d_1\left(d-d_1\right)}E\left[U_2\left(1-U_2\right)\left(1-kU_2\right)^{-2}\right]$$

Using the fact, $U_2 \sim Beta\left(\frac{d-d_1}{2}, \frac{d_1}{2}\right)$ one can write,

$$E\left[U_{2}^{r}\left(1-U_{2}\right)^{s}\left(1-kU_{2}\right)^{-t}\right] = \frac{1}{B\left(\frac{d-d_{1}}{2},\frac{d_{1}}{2}\right)} \int_{0}^{1} x^{\frac{d-d_{1}}{2}+r-1} \left(1-x\right)^{\frac{d_{1}}{2}+s-1} \left(1-kx\right)^{-t} dx$$
$$= \frac{1}{B\left(\frac{d-d_{1}}{2},\frac{d_{1}}{2}\right)} \int_{0}^{1} x^{\frac{d-d_{1}+2r}{2}-1} \left(1-x\right)^{\frac{d+2s+2r}{2}-\frac{d-d_{1}+2r}{2}-1} \left(1-kx\right)^{-t} dx$$
$$= \frac{B\left(\frac{d-d_{1}+2r}{2},\frac{d+2s+2r}{2}-\frac{d-d_{1}+2r}{2}\right)}{B\left(\frac{d-d_{1}}{2},\frac{d_{1}}{2}\right)} {}_{2}F_{1}\left(t,\frac{d-d_{1}+2r}{2};\frac{d+2s+2r}{2};k\right)$$

where $B(a,b) = \Gamma(a) \Gamma(b) / \Gamma(a+b)$ is the Beta function and ${}_{2}F_{1}(a,b;c;k) = B^{-1}(b,c-b) \int_{0}^{1} x^{b-1} (1-x)^{c-b-1} (1-kx)^{-a} dx$ is the Gauss hypergeometric function. The integral representation of the Gauss hypergeometric function is valid for $\Re(c) > \Re(b) > 0$. These facts then give

$$E\left[U_{2}\left(1-kU_{2}\right)^{-1}\right] = \frac{B\left(\frac{d-d_{1}+2}{2}, \frac{d+2}{2} - \frac{d-d_{1}+2}{2}\right)}{B\left(\frac{d-d_{1}}{2}, \frac{d_{1}}{2}\right)} {_{2}F_{1}\left(1, \frac{d-d_{1}+2}{2}; \frac{d+2}{2}; k\right)}$$

$$= \frac{\Gamma\left(\frac{d-d_{1}+2}{2}\right)\Gamma\left(\frac{d_{1}}{2}\right)}{\Gamma\left(\frac{d+2}{2}\right)} \frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d-d_{1}}{2}\right)^{2}} F_{1}\left(1, \frac{d-d_{1}+2}{2}; \frac{d+2}{2}; k\right)$$

$$= \frac{\frac{d-d_{1}}{2}\Gamma\left(\frac{d-d_{1}}{2}\right)\Gamma\left(\frac{d}{2}\right)}{\frac{d}{2}\Gamma\left(\frac{d}{2}\right)} \frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d-d_{1}}{2}\right)\Gamma\left(\frac{d-d_{1}}{2}\right)^{2}} F_{1}\left(1, \frac{d-d_{1}+2}{2}; \frac{d+2}{2}; k\right)$$

$$= \frac{d-d_{1}}{d} {}_{2}F_{1}\left(1, \frac{d-d_{1}+2}{2}; \frac{d+2}{2}; k\right)$$

$$E\left[(1-U_{2})(1-kU_{2})^{-1}\right] = \frac{B\left(\frac{d-d_{1}}{2},\frac{d+2}{2}-\frac{d-d_{1}}{2}\right)}{B\left(\frac{d-d_{1}}{2},\frac{d_{1}}{2}\right)} {_{2}}F_{1}\left(1,\frac{d-d_{1}}{2};\frac{d+2}{2};k\right)$$

$$= \frac{\Gamma\left(\frac{d-d_{1}}{2}\right)\Gamma\left(\frac{d_{1}+2}{2}\right)}{\Gamma\left(\frac{d+2}{2}\right)} \frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d-d_{1}}{2}\right)\Gamma\left(\frac{d_{1}}{2}\right)} {_{2}}F_{1}\left(1,\frac{d-d_{1}}{2};\frac{d+2}{2};k\right)$$

$$= \frac{\Gamma\left(\frac{d-d_{1}}{2}\right)\frac{d_{1}}{2}\Gamma\left(\frac{d_{1}}{2}\right)}{\frac{d_{1}}{2}\Gamma\left(\frac{d_{2}}{2}\right)} \frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d-d_{1}}{2}\right)\Gamma\left(\frac{d_{1}}{2}\right)} {_{2}}F_{1}\left(1,\frac{d-d_{1}}{2};\frac{d+2}{2};k\right)$$

$$= \frac{d_{1}}{d} {}_{2}F_{1}\left(1,\frac{d-d_{1}}{2};\frac{d+2}{2};k\right)$$

$$E\left[U_{2}\left(1-U_{2}\right)\left(1-kU_{2}\right)^{-2}\right] = \frac{B\left(\frac{d-d_{1}+2}{2},\frac{d+2+2}{2}-\frac{d-d_{1}+2}{2}\right)}{B\left(\frac{d-d_{1}}{2},\frac{d_{1}}{2}\right)}{2}F_{1}\left(2,\frac{d-d_{1}+2}{2};\frac{d+2+2}{2};k\right)$$

$$= \frac{\Gamma\left(\frac{d-d_{1}+2}{2}\right)\Gamma\left(\frac{d_{1}+2}{2}\right)}{\Gamma\left(\frac{d+4}{2}\right)}\frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d-d_{1}}{2}\right)\Gamma\left(\frac{d_{1}}{2}\right)}{2}F_{1}\left(2,\frac{d-d_{1}+2}{2};\frac{d+4}{2};k\right)$$

$$= \frac{\frac{d-d_{1}}{2}\Gamma\left(\frac{d-d_{1}}{2}\right)\frac{d_{1}}{2}\Gamma\left(\frac{d}{2}\right)}{\frac{d+2}{2}\frac{d}{2}\Gamma\left(\frac{d}{2}\right)}\frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d-d_{1}}{2}\right)\Gamma\left(\frac{d_{1}}{2}\right)}{2}F_{1}\left(2,\frac{d-d_{1}+2}{2};\frac{d+4}{2};k\right)$$

$$= \frac{d_{1}\left(d-d_{1}\right)}{d\left(d+2\right)}2F_{1}\left(2,\frac{d-d_{1}+2}{2};\frac{d+4}{2};k\right)$$

Using these gives,

$$\delta_i^2 = \frac{1}{d_1} \frac{d_1}{d_2} F_1\left(1, \frac{d-d_1}{2}; \frac{d+2}{2}; k\right) = \frac{1}{d^2} F_1\left(1, \frac{d-d_1}{2}; \frac{d+2}{2}; k\right)$$

$$\delta_j^2 = \frac{1-k}{d-d_1} \frac{d-d_1}{d} {}_2F_1\left(1, \frac{d-d_1+2}{2}; \frac{d+2}{2}; k\right) = \frac{1-k}{d} {}_2F_1\left(1, \frac{d-d_1+2}{2}; \frac{d+2}{2}; k\right)$$

$$E[\gamma_{ij}\gamma_{ij}] = \frac{1-k}{d_1(d-d_1)} \frac{d_1(d-d_1)}{d(d+2)} {}_2F_1\left(2, \frac{d-d_1+2}{2}; \frac{d+4}{2}; k\right)$$
$$= \frac{1-k}{d(d+2)} {}_2F_1\left(2, \frac{d-d_1+2}{2}; \frac{d+4}{2}; k\right)$$

Thus,

$$\begin{aligned} \frac{E\left[\gamma_{ij}\gamma_{ij}\right]}{\left(\delta_{i}^{2}-\delta_{j}^{2}\right)^{2}} &= \frac{\frac{1-k}{d(d+2)^{2}}F_{1}\left(2,\frac{d-d_{1}+2}{2};\frac{d+4}{2};k\right)}{\left(\frac{1}{d^{2}}F_{1}\left(1,\frac{d-d_{1}}{2};\frac{d+2}{2};k\right)-\frac{1-k}{d^{2}}F_{1}\left(1,\frac{d-d_{1}+2}{2};\frac{d+2}{2};k\right)\right)^{2}} \\ &= \frac{\frac{d}{d+2}F_{1}\left(1-k\right)\left(2,\frac{d-d_{1}+2}{2};\frac{d+4}{2};k\right)}{\left(2F_{1}\left(1,\frac{d-d_{1}}{2};\frac{d+2}{2};k\right)-(1-k)_{2}F_{1}\left(1,\frac{d-d_{1}+2}{2};\frac{d+2}{2};k\right)\right)^{2}}\end{aligned}$$

This then gives

$$\begin{aligned} ARE_{d,d_1}\left(\widehat{\mathbf{S}}_{\mathcal{N}}, \widehat{\mathbf{T}}_{\mathcal{N}}, r\right) &= \frac{\frac{d+2}{d} \frac{r^2}{(1-r^2)^2}}{\frac{d+2}{d+2}(1-k)_2 F_1\left(2, \frac{d-d_1+2}{2}; \frac{d+4}{2}; k\right)}{\left(2F_1\left(1, \frac{d-d_1}{2}; \frac{d+2}{2}; k\right) - (1-k)_2 F_1\left(1, \frac{d-d_1+2}{2}; \frac{d+2}{2}; k\right)\right)^2} \\ &= \frac{\left(\frac{d+2}{d}\right)^2 \frac{1-k}{k^2}}{\left(1-k)_2 F_1\left(2, \frac{d-d_1+2}{2}; \frac{d+4}{2}; k\right)\right)}{\left(2F_1\left(1, \frac{d-d_1}{2}; \frac{d+2}{2}; k\right) - (1-k)_2 F_1\left(1, \frac{d-d_1+2}{2}; \frac{d+2}{2}; k\right)\right)^2} \\ &= \frac{\left(\frac{d+2}{d}\right)^2}{\frac{2F_1\left(2, \frac{d-d_1+2}{2}; \frac{d+4}{2}; k\right)}{\left(\frac{1}{k^2} F_1\left(1, \frac{d-d_1}{2}; \frac{d+2}{2}; k\right) - \frac{1-k}{k^2} F_1\left(1, \frac{d-d_1+2}{2}; \frac{d+2}{2}; k\right)\right)^2} \end{aligned}$$

Expression in the denominator of the denominator, identity 15.2.20 from [1], $\frac{1-z}{z} {}_2F_1(a, b; c; z) - \frac{1}{z} {}_2F_1(a - 1, b; c; z) + \frac{c-b}{c} {}_2F_1(a, b; c + 1; z) = 0$, where $z = k, a = \frac{d-d_1+2}{2}, b = 1$ and $c = \frac{d+2}{2}$, yields the final result

$$ARE_{d,d_1}\left(\widehat{\mathbf{S}}_{\mathcal{N}}, \widehat{\mathbf{T}}_{\mathcal{N}}, r\right) = \frac{\left(\frac{d+2}{d}\right)^2}{\frac{2F_1\left(2, \frac{d-d_1+2}{2}; \frac{d+4}{2}; k\right)}{\left(\frac{d}{d+2^2}F_1\left(1, \frac{d_1+2}{2}; \frac{d+4}{2}; 1-r^2\right)\right)}} = \frac{2F_1^2\left(1, \frac{d_1+2}{2}; \frac{d+4}{2}; 1-r^2\right)}{2F_1\left(2, \frac{d_1+2}{2}; \frac{d+4}{2}; 1-r^2\right)}$$

Chapter 4

Conclusion

4.1 Implications

As shown for both the spatial median and SSCM, when sampling from an elliptical distribution, these estimators are at their best in terms of efficiency when the underlying scatter structure of the elliptical distribution is in fact spherical. This situation is the least interesting scenario for multivariate problems. For example, PCA is most useful is when the majority of the variation in the data lies in some d_1 dimensional subspace with $d_1 < d$. If it is assumed that data arises from some elliptical model, then this scenario implies that Σ is far from spherical. A popular approach to robust PCA is to ascertain the eigendecomposition of some robust estimator of the scatter matrix, the computation of which usually involves estimating the location vector, usually with some robust estimator as well. As a procedure, PCA is orthogonally equivariant. Consequently, the estimators used for the prior approach need only possess this property as well. Being orthogonally equivariant, the spatial median and SSCM are commonly used in robust versions of PCA. However, these estimators are at their best is when PCA is least useful, spherical symmetric data admits nearly no reduction in dimensionality. Where PCA yields the greatest amount of dimension reduction is when these estimators are at their worst in terms of efficiency. Unfortunately, the fact the performance of orthogonally equivariant estimators depends on the underlying scatter structure of the data is hardly ever considered in practice. Utilizing an orthogonally equivariant estimator without considering the underlying scatter structure is rather Procrustean in that it does not take into account the scatter structure of the data, but rather down-weighs observations based on Euclidean distances.

4.2 Future Work

Based on the results for the spatial median and SSCM, one might hypothesize that the efficiencies of other orthogonally equivariant estimators of location and scatter are maximized under the elliptical model when the underlying scatter structure of the data is spherically symmetric. Interestingly, this is not the case; a counter-example is with Huber's skipped mean for the bivariate Gaussian distribution. This fact motivates a future research direction; a further exploration of which scatter structures are favorable to other orthogonally equivariant estimators under the elliptical model. Included with the direction is a study of the properties of the estimator that dictate what scatter structures are favorable. This direction would provide insight to researchers in other fields when deciding upon which orthogonally equivariant estimators to use for their data.

A second research direction is the development and study of methods that are improvements of orthogonally equivariant procedures in the sense their efficiencies are not as sensitive to the underlying scatter structure of the data as the former estimators. To this end the author proposes studying hybrid estimators of the form,

$$\widehat{\boldsymbol{\mu}} = \operatorname{argmin}_{\boldsymbol{\theta} \in \mathbb{R}^d} \sum_{i=1}^n u_1 \left(\left(\mathbf{x}_i - \boldsymbol{\theta} \right)^t \widehat{\boldsymbol{\Sigma}}_*^{-1} \left(\mathbf{x}_i - \boldsymbol{\theta} \right) \right)$$
(4.1)

$$\widehat{\boldsymbol{\Sigma}} = \operatorname{argmin}_{\mathbf{V}\in\mathcal{P}_d} \sum_{i=1}^n u_2 \left((\mathbf{x}_i - \widehat{\boldsymbol{\mu}})^t \, \mathbf{V}_*^{-1} \left(\mathbf{x}_i - \widehat{\boldsymbol{\mu}} \right) \right) - \ln\left(\det\left(\mathbf{V}_*\right)\right)$$
(4.2)

where \mathcal{P}_d is the vector space of $d \times d$ symmetric positive definite matrices, $\mathbf{V}_* = \mathbf{V} + \lambda Trace(\mathbf{V}) \mathbf{I}_d$, $\lambda > 0$ and $u_1(s)$ and $u_2(s)$ satisfy assumptions similar to those defined in Section 2 of [23]. A similar estimator of scatter to Equation 4.2 has been studied in [6] where Tyler's matrix was used instead of the identity matrix. This estimator can be viewed as a shrinkage estimator of scatter, where the shrinkage is towards the identity matrix. The parameter λ is a tuning constant that depends on both the sample size and the underlying scatter structure of the data and controls how much the estimator is shrunk to the identity. The form of the matrix Σ_* dictates that it is nonsingular, hence $\hat{\mu}$ will be defined and unique from the sample mean vector, even when n < d. This fact is an an obvious advantage over affinely equivariant estimators. However, because these estimators incorporate an estimate of the underlying scatter structure into them, it is speculated that they will have better efficiencies than orthogonally equivariant estimators. The existence and uniqueness of these estimators still must be established as well as their robustness properties and asymptotic behavior. In addition, these estimators will not be as computationally intensive as high breakdown point affinely equivariant estimators.

The motivation of this research direction is in the field of sparse high-dimensional data analysis where $n \ll d$. As mentioned in section 2, any affinely equivariant estimate of location and scatter reduce to the sample mean vector and sample covariance matrix respectively [38]. Consequently, orthogonally equivariant estimators of the location vector are often implemented in this setting because they are defined despite the fact n < d; in addition they are unique from the former two estimators. However, using orthogonally equivariant estimators could be disadvantageous considering their tendency to favor certain scatter structures. The hybrid estimators defined in Equations 4.1 and 4.2 proposed would serve as alternatives to orthogonally equivariant estimators since they do not favor any particular scatter structure.

References

- M. ABRAMOWITZ AND I. A. STEGUN, eds., Handbook of mathematical functions with formulas, graphs, and mathematical tables, Dover Publications Inc., New York, 1992. Reprint of the 1972 edition.
- [2] S. N. AFRIAT, Orthogonal and oblique projectors and the characteristics of pairs of vector spaces, Proc. Cambridge Philos. Soc., 53 (1957), pp. 800–816.
- [3] H. BENSMAIL AND G. CELEUX, Regularized Gaussian discriminant analysis through eigenvalue decomposition, J. Amer. Statist. Assoc., 91 (1996), pp. 1743– 1748.
- [4] A. BJÖRCK AND G. GOLUB, Sketch of a proof that an odd perfect number relatively prime to 3 has at least eleven prime factors, Math. Comp., 27 (1973), pp. 579–694.
- [5] B. M. BROWN, Statistical uses of the spatial median, J. Roy. Statist. Soc. Ser. B, 45 (1983), pp. 25–30.
- [6] Y. CHEN, A. WIESEL, AND A. O. HERO, III, Robust shrinkage estimation of highdimensional covariance matrices, IEEE Trans. Signal Process., 59 (2011), pp. 4097– 4107.
- [7] C. CROUX, E. OLLILA, AND H. OJA, Sign and rank covariance matrices: statistical properties and application to principal components analysis, in Statistical data analysis based on the L_1 -norm and related methods (Neuchâtel, 2002), Stat. Ind. Technol., Birkhäuser, Basel, 2002, pp. 257–269.
- [8] C. CROUX AND A. RUIZ-GAZEN, High breakdown estimators for principal components: the projection-pursuit approach revisited, J. Multivariate Anal., 95 (2005), pp. 206–226.
- [9] P. L. DAVIES, Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices, Ann. Statist., 15 (1987), pp. 1269–1292.
- [10] D. L. DONOHO AND M. GASKO, Breakdown properties of location estimates based on halfspace depth and projected outlyingness, Ann. Statist., 20 (1992), pp. 1803– 1827.
- [11] L. DÜMBGEN, On Tyler's M-functional of scatter in high dimension, Ann. Inst. Statist. Math., 50 (1998), pp. 471–491.
- [12] D. GERVINI, Robust functional estimation using the median and spherical principal components, Biometrika, 95 (2008), pp. 587–600.
- [13] J. GOWER, The mediancentre, J. Roy. Statist. Soc. Ser. C, 23 (1974), pp. 466–470.

- [14] P. J. HUBER, Robust statistical procedures, Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1977. CBMS-NSF Regional Conference Series in Applied Mathematics, No. 27.
- [15] J. H. B. KEMPERMAN, The median of a finite measure on a Banach space, in Statistical data analysis based on the L_1 -norm and related methods (Neuchâtel, 1987), North-Holland, Amsterdam, 1987, pp. 217–230.
- [16] J. T. KENT AND D. E. TYLER, Constrained M-estimation for multivariate location and scatter, Ann. Statist., 24 (1996), pp. 1346–1370.
- [17] N. LOCANTORE, J. S. MARRON, D. G. SIMPSON, N. TRIPOLI, J. T. ZHANG, AND K. L. COHEN, *Robust principal component analysis for functional data*, Test, 8 (1999), pp. 1–73. With discussion and a rejoinder by the authors.
- [18] H. P. LOPUHAÄ, On the relation between S-estimators and M-estimators of multivariate location and covariance, Ann. Statist., 17 (1989), pp. 1662–1683.
- [19] H. P. LOPUHAÄ, Multivariate τ -estimators for location and scatter, Canad. J. Statist., 19 (1991), pp. 307–321.
- [20] H. P. LOPUHAÄ, Asymptotics of reweighted estimators of multivariate location and scatter, Ann. Statist., 27 (1999), pp. 1638–1665.
- [21] A. MAGYAR AND D. TYLER, The asympotitic efficiency of the spatial median for elliptically symmetric distributions, Sankhyā Ser. B, 73 (2012).
- [22] J. I. MARDEN, Some robust estimates of principal components, Statist. Probab. Lett., 43 (1999), pp. 349–359.
- [23] R. A. MARONNA, Robust M-estimators of multivariate location and scatter, Ann. Statist., 4 (1976), pp. 51–67.
- [24] R. A. MARONNA, R. D. MARTIN, AND V. J. YOHAI, *Robust statistics*, Wiley Series in Probability and Statistics, John Wiley & Sons Ltd., Chichester, 2006. Theory and methods.
- [25] R. A. MARONNA, W. A. STAHEL, AND V. J. YOHAI, Bias-robust estimators of multivariate scatter based on projections, J. Multivariate Anal., 42 (1992), pp. 141– 161.
- [26] J. M. MIAO AND A. BEN-ISRAEL, On principal angles between subspaces in Rⁿ, Linear Algebra Appl., 171 (1992), pp. 81–98.
- [27] P. MILASEVIC AND G. R. DUCHARME, Uniqueness of the spatial median, Ann. Statist., 15 (1987), pp. 1332–1333.
- [28] J. MÖTTÖNEN, K. NORDHAUSEN, AND H. OJA, Asymptotic theory of the spatial median, in Nonparametrics and robustness in modern statistical inference and time series analysis: a Festschrift in honor of Professor Jana Jurečková, vol. 7 of Inst. Math. Stat. Collect., Inst. Math. Statist., Beachwood, OH, 2010, pp. 182–193.
- [29] J. NEVALAINEN, D. LAROCQUE, AND H. OJA, On the multivariate spatial median for clustered data, Canad. J. Statist., 35 (2007), pp. 215–231.

- [30] C. R. RAO, Criteria of estimation in large samples, Sankhyā Ser. A, 25 (1963), pp. 189–206.
- [31] P. ROUSSEEUW, Multivariate estimation with high breakdown point, in Mathematical statistics and applications, Vol. B (Bad Tatzmannsdorf, 1983), Reidel, Dordrecht, 1985, pp. 283–297.
- [32] K. S. TATSUOKA AND D. E. TYLER, On the uniqueness of S-functionals and Mfunctionals under nonelliptical distributions, Ann. Statist., 28 (2000), pp. 1219– 1243.
- [33] D. E. TYLER, REDUNDANCY ANALYSIS AND ASSOCIATED ASYMPTOTIC DISTRIBUTION THEORY, ProQuest LLC, Ann Arbor, MI, 1979. Thesis (Ph.D.)-Princeton University.
- [34] D. E. TYLER, Asymptotic inference for eigenvectors, Ann. Statist., 9 (1981), pp. 725–736.
- [35] D. E. TYLER, A distribution-free M-estimator of multivariate scatter, Ann. Statist., 15 (1987), pp. 234–251.
- [36] D. E. TYLER, Statistical analysis for the angular central Gaussian distribution on the sphere, Biometrika, 74 (1987), pp. 579–589.
- [37] D. E. TYLER, Finite sample breakdown points of projection based multivariate location and scatter statistics, Ann. Statist., 22 (1994), pp. 1024–1044.
- [38] D. E. TYLER, A note on multivariate location and scatter statistics for sparse data sets, Statist. Probab. Lett., 80 (2010), pp. 1409–1413.
- [39] Y. VARDI AND C.-H. ZHANG, A modified Weiszfeld algorithm for the Fermat-Weber location problem, Math. Program., 90 (2001), pp. 559–566.
- [40] S. VISURI, V. KOIVUNEN, AND H. OJA, Sign and rank covariance matrices, J. Statist. Plann. Inference, 91 (2000), pp. 557–575. Prague Workshop on Perspectives in Modern Statistical Inference: Parametrics, Semi-parametrics, Nonparametrics (1998).
- [41] J. WANG AND M. G. GENTON, The multivariate skew-slash distribution, J. Statist. Plann. Inference, 136 (2006), pp. 209–220.
- [42] A. WEBER, Über der Standort der Industrien, University of Chidago Press, Chicago, Il., 1909. English Translation by C Frederich (1929): Alfred Webers theory of location of industries.
- [43] E. WEISZFELD, On the point for which the sum of the distances to n given points is minimum, Ann. Oper. Res., 167 (2009), pp. 7–41. Translated from the French original [Tohoku Math. J. 43 (1937), 355–386] and annotated by Frank Plastria.

Vita

Andrew Magyar

2006-12	Ph. D. in Statistics, Rutgers University
2001-05	B. Sc. in Mathematics, The University of Pennsylvania

2010-12 Lecturer - Department of Statistics, Rutgers University