Variance-Covariance from a Metropolis Chain on a Curved, Singular Manifold

A. Ronald Gallant
University of North Carolina, Chapel Hill

Paper: http://www.aronaldg.org/papers/sdev.pdf
Slides: http://www.aronaldg.org/papers/sdevclr.pdf
Code: http://www.aronaldg.org/webfiles/npb

Preamble

- There are many uses for scale measures: Adhering to the convention of reporting both location and scale in the presentation of statistical results. Tuning an MCMC chain. Etc.

- Variance and covariance are Euclidean concepts. A curved, singular manifold is not typically a Euclidean space. We explore some suggestions on how to adapt a Euclidean concept to a non-Euclidean space.

Motivating Problem: Bayes Subject to Moment Conditions

The parameters $(\rho, \theta) \in \mathbb{R}^{d_a}$ of the likelihood

$$f(y \,|\, x, \rho) = \prod_{t=1}^{n} f(y_t \,|\, x_{t-1}, \rho) \tag{1}$$

are to be estimated subject to the moment conditions

$$0 = q(\rho, \theta) = \frac{1}{n} \sum_{t=1}^{n} \int m(y, x_{t-1}, \rho, \theta) f(y \,|\, x_{t-1}, \rho) \, dy \quad m \in \mathbb{R}^m \tag{2}$$

the support conditions

$$h(\rho, \theta) > 0, \qquad h \in \mathbb{R}^l \tag{3}$$

and the prior

$$\pi(\rho, \theta). \tag{4}$$

Nonparametric Bayes

- Bayesian estimation can be regarded as nonparametric when

$$f(y_t \,|\, x_{t-1}, \rho)$$

  is a sieve.

- A sieve is a density with a variable number K of parameters

$$\rho = (\rho_1, \rho_2, \ldots, \rho_K)$$

  that is dense for some norm, e.g. Sobolev norm, as $K \to \infty$.

  – Code uses the SNP sieve (Gallant and Tauchen, 1989, ECTA).

Clash of Notation

To adhere to the notational conventions of both the econometric and numerical analysis literature:

- Italic represents data: x_t, y_t, x, y
- Sans serif represents parameters: x, y
  - i.e., $\mathsf{x} = (\rho, \theta)$ and $\mathsf{y} = (\rho, \theta)$

Overidentification

- The support of the posterior is the manifold

$$M = \left\{ \mathsf{x} \in \mathbb{R}^{d_a} : q_i(\mathsf{x}) = 0, i = 1,..,m, h_j(\mathsf{x}) > 0, j = 1,..,l \right\} \quad (5)$$

- The problem is interesting when θ is overidentified, i.e., when the dimension m of q is larger than the dimension of θ because then M is singular with respect to Lebesgue measure on R^{d_a}.
  - Specialized algorithms are required: Gallant (2022, JoE)
  - Otherwise the problem is boring.

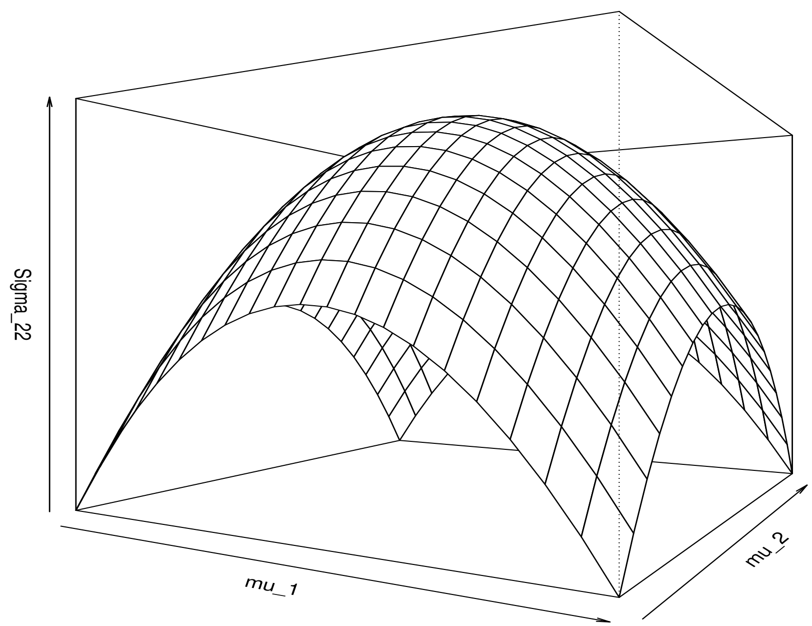


Figure 1. A Curved, Singular Manifold

Geodesics – 1

- On a manifold $M \subset \mathbb{R}^{d_a}$ of dimension $d < d_a$, distance is computed along geodesics.

- One computes distance by traversing a geodesic from a starting point s to an end point p and accumulating (infinitesimal increments of) a Hausdorff weight function $\delta_M(s,p)$ defined on M (Morgan, 2016).

- For a point cloud on M one can compute approximate geodesics from a d_a-dimensional set M_ϵ that is the union of ϵ-balls centered at the points using Euclidean distance $\delta(s,p)$ provided ϵ is large enough that M_ϵ is a connected set (Memoli and Sapiro, 2001).

Geodesics – 2

- Because the contours of the density that the chain targets are not spheres, our ϵ-balls for determining $\mathcal{G}_\epsilon$ are rectangles with sides k equal to $\Delta \max\{|\mathsf{x}_{k,i} - \mathsf{x}_{k,i-1}| \,:\, \mathsf{x}_i \in \mathcal{D}\}$ where $\mathcal{D} = \{\mathsf{x}_i\}_{i=1}^N$ denotes the MCMC chain and $\mathsf{x}_{k,i}$ denotes the kth element of x_i.

- If M_ϵ is a connected set, then the MCMC draws may be viewed as nodes p_j of a graph $\mathcal{G}_\epsilon$ connected by edges $e_{j,j'}$ that have Euclidean length $\delta(p_j, p_{j'})$ and that stay within M_ϵ.

- From a start s, Dijkstra's algorithm finds the shortest path that traverses edges to every node p_j (Dijkstra, 1959).
  - This is the same algorithm that Google maps uses for routing.

Choose Delta at Inflection

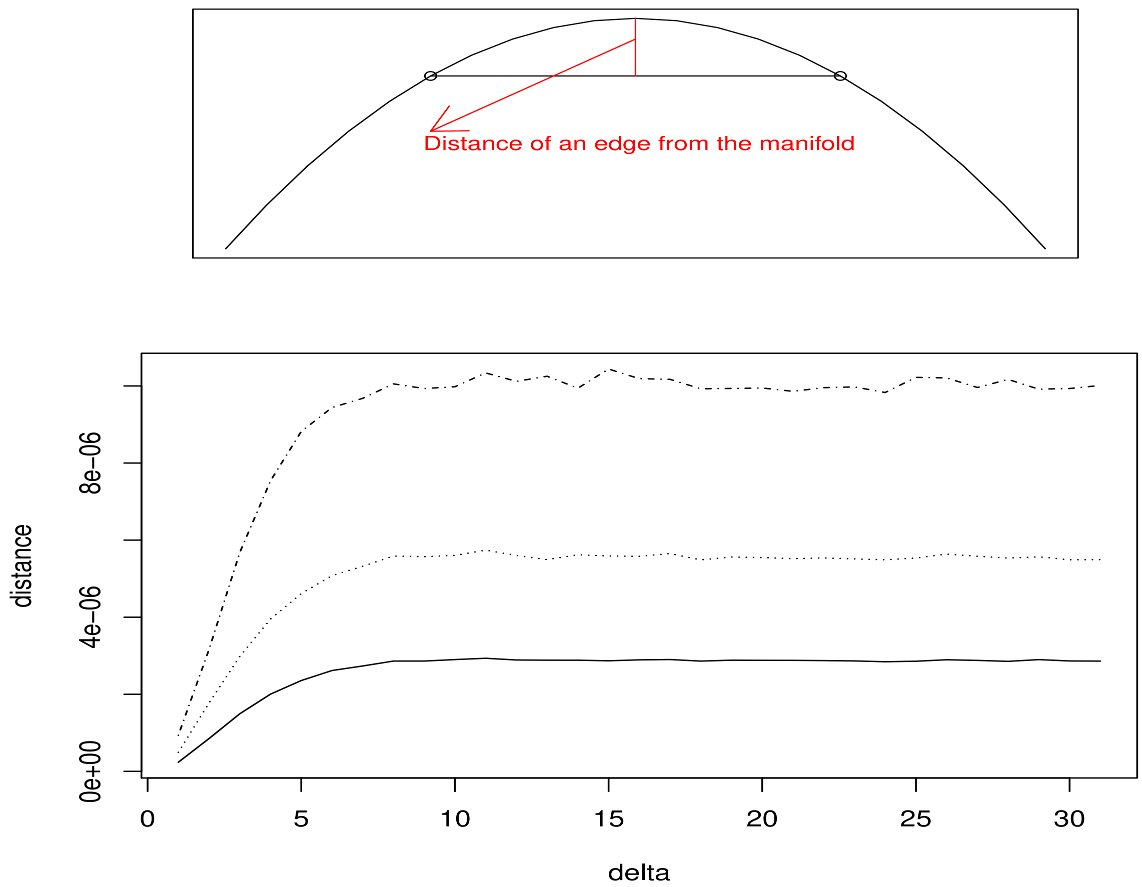


Figure 2. Lower panel: The dotdash line is the 99th percentile of all edges, dotted the 90th percentile, and solid the mean.

Intrinsic Mean

- There seems to be general agreement on how to define a mean over a curved, nonlinear manifold and estimate it from a sample.

- It is the intrinsic mean, $\bar{\text{x}}$, that is the start s that minimizes $\frac{1}{N}\sum_{i=1}^{N} \delta^2(s, p_{j(i)})$.
  - The MCMC chain has duplicate draws due to rejections.
  - $j(i)$ is the mapping from the draws i to the distinct elements of the chain.
  - The distinct elements are the nodes p_j of the edges.

- The extrinsic mean, $\tilde{\text{x}}$, is the ordinary sample average.

Scale

- The notions of variance and covariance are flat space concepts, i.e., Euclidean space concepts, and it is not obvious how to extend them to a curved, nonlinear manifold.

- We shall consider four possible definitions
  - Extrinsic variance-covariance centered at $\tilde{x}$: V_{EC}
  - Extrinsic variance-covariance centered at $\bar{x}$: V_{IC}
  - Modified extrinsic variance-covariance: V_{ME}
  - Modified Riemann variance-covariance: V_{MR}

Extrinsic centered at $\tilde{\mathsf{x}}$: V_{EC}

- $V_{EC} = \frac{1}{N}\sum_{i=1}^{N}(\mathsf{x}_i - \tilde{\mathsf{x}})(\mathsf{x}_i - \tilde{\mathsf{x}})^\top$

- Disregard the geometry of M and view $\{\mathsf{x}_i\}_{i=1}^{N}$ as a sample from a probability space like any other.

- A credibility interval such as

$$R_\tau = \times_{i=1}^{d_a}[\bar{\mathsf{x}}_i - \tau\,\mathsf{sdev}(\mathsf{x}_i),\ \bar{\mathsf{x}}_i + \tau\,\mathsf{sdev}(\mathsf{x}_i)]$$

constructed from V_{EC} need not intersect M.

Extrinsic centered at $\bar{\mathsf{x}}$: V_{IC}

- $V_{IC} = \frac{1}{N}\sum_{i=1}^{N}(\mathsf{x}_i - \bar{\mathsf{x}})(\mathsf{x}_i - \bar{\mathsf{x}})^\top$

- A credibility interval such as

$$R_\tau = \times_{i=1}^{d_a}[\bar{\mathsf{x}}_i - \tau\,\mathsf{sdev}(\mathsf{x}_i),\ \bar{\mathsf{x}}_i + \tau\,\mathsf{sdev}(\mathsf{x}_i)]$$

constructed from V_{IC} does intersect M.

Modified extrinsic variance-covariance: V_{ME}

- Same as an extrinsic computation but one increases each coordinate of a point p_j by its geodesic distance in that direction

- Specifically, for the path $(j_1^p, j_2^p, ..., j_k^p)$ that connects $\bar{x}$ to p_j, where j_1^p indexes node $\bar{x}$ and indexes j_k^p node p_j,

$$D_j = \mathrm{diag}[\mathrm{sgn}(p_j - \bar{x})] \sum_{\ell=2}^{k} |p_{j_\ell^p} - p_{j_{\ell-1}^p}| \quad D_j \in \mathbb{R}^{d_a}$$

- The estimated variance-covariance matrix is

$$V_{ME} = \frac{1}{N} \sum_{i=1}^{N} D_{j(i)} D_{j(i)}^\top$$

- See figure on next slide

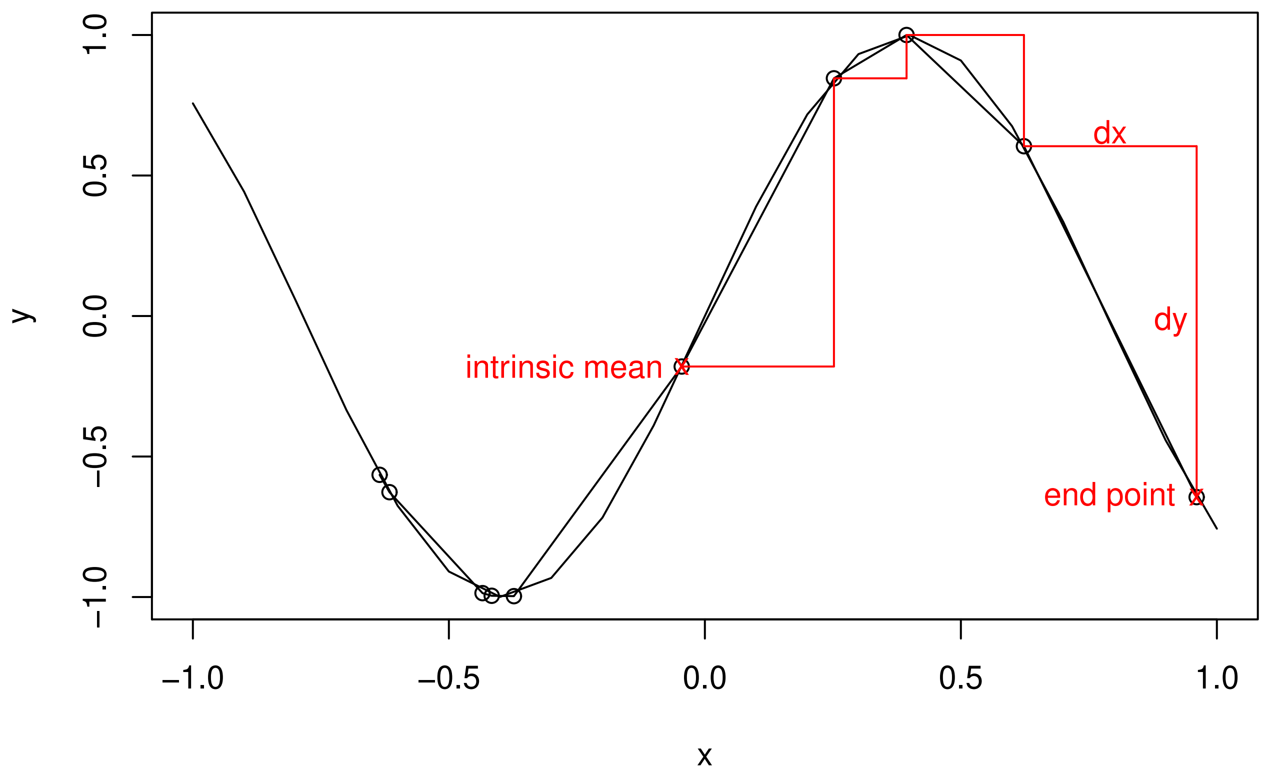


Figure 3. For V_{ME} the contribution to D_j of the end point is the sum absolute values of the increments, $\begin{pmatrix} |dx| \\ |dy| \end{pmatrix}$, whereas the contribution to $(\mathrm{x}_i - \bar{\mathrm{x}})(\mathrm{x}_i - \bar{\mathrm{x}})^\top$ of V_{EC} is the absolute value of the sum.

Riemannian Geometry

- Represent the manifold as a flat space called a chart and then compute variances and covariances in the usual way on the chart
  - Think of a Mercator projection of the globe centered at Greenwich, England.

- The flat space is the plane $T_{\bar{x}}M$ tangent to the manifold M at the mean $\bar{x}$. Note $T_{\bar{x}}M \subset \mathbb{R}^d$, $d < d_a$

- Requires a differentiable, analytic expression for geodesics $\gamma(t)$ with $\gamma(0) = \bar{x}$

- A point $x \in M$ is plotted on $T_{\bar{x}}M$ in the direction $(d/dt)\gamma(0)$ at the distance $\delta(x, \bar{x})$

Modified Riemann variance-covariance: V_{MR}

- The Riemannian approach is not possible if all we have is a point cloud on a manifold because we do not have an analytic expression for geodesics but we can borrow the basic ideas:

- Orthogonally project x_i onto the chart $T_{\bar{\mathsf{x}}}M$

$$v_i = T_{\bar{\mathsf{x}}} T_{\bar{\mathsf{x}}}^{\top} (\mathsf{x}_i - \bar{\mathsf{x}})$$

- Plot the marker for x_i at $z_i = \delta_i \frac{v_i}{\|v_i\|}$

- Modified Riemann variance is

$$V_{MR} = \frac{1}{N} \sum_{i=1}^{N} z_i z_i^{\top}$$

  – Note that V_{MR} is $d_a \times d_a$ and singular of rank d

Examples

- In the paper (`www.aronaldg.org/papers/sdev.pdf`)
  - A Simple Demand and Supply Example (simulation)
  - Extraction of the Stochastic Discount Factor (data)
  - A Curved Manifold Example (simulation).
- We'll look at the curved manifold example

Curved Manifold Example – 1

Likelihood:

$$
\begin{aligned}
y_t &\sim n_2(y_t \,|\, \mu, \Sigma) \\
\Sigma &= RR' \\
\rho &= (\mu_1, \mu_2, R_{1,1}, R_{1,2}, R_{2,2}) \in \mathbb{R}^5
\end{aligned}
$$

Moment conditions:

$$
\begin{aligned}
m_{c,1}(y_t, y_{t-1}, \rho, \theta) &= y_{1,t}^2 + y_{2,t}^2 - 4\theta \\
m_{c,2}(y_t, y_{t-1}, \rho, \theta) &= (y_{1,t} - y_{1,t-1})^2 - 2\theta \\
\theta \in \mathbb{R}^1 & \\
\rho & \quad \text{not used}
\end{aligned}
$$

Curved Manifold Example – 2

- Data, $n = 500$, simulated with $\mu_1 = 0$, $\mu_2 = 0$, $\Sigma_{1,1} = 5$, $\Sigma_{1,2} = \Sigma_{2,1} = 6.12372$, $\Sigma_{2,2} = 15$, and $\theta = 5$.

- Prior for ρ is independent normal with location the unconstrained maximum likelihood estimates and standard deviation 5.0.

- Prior for θ is normal with mean 5.0 and standard deviation 5.0.

- The support conditions are that diagonals of R must be positive and θ must be positive.

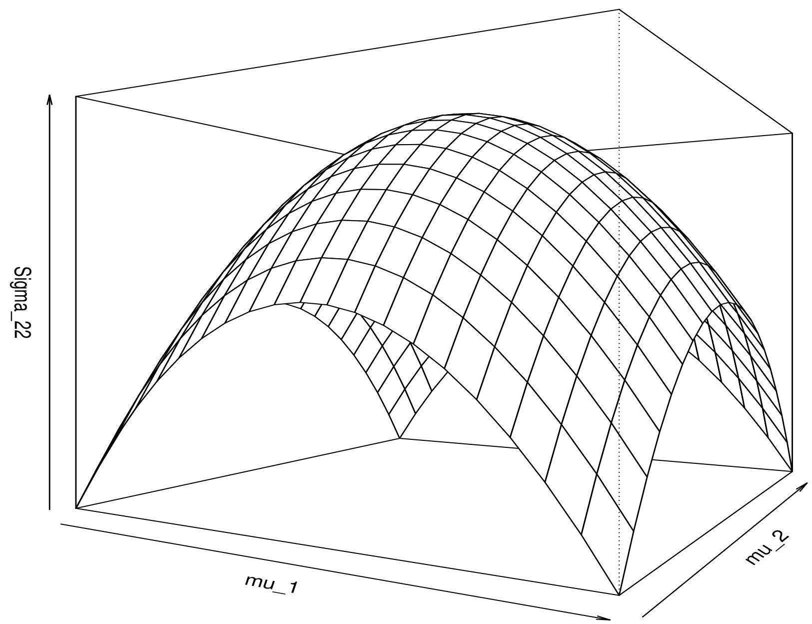
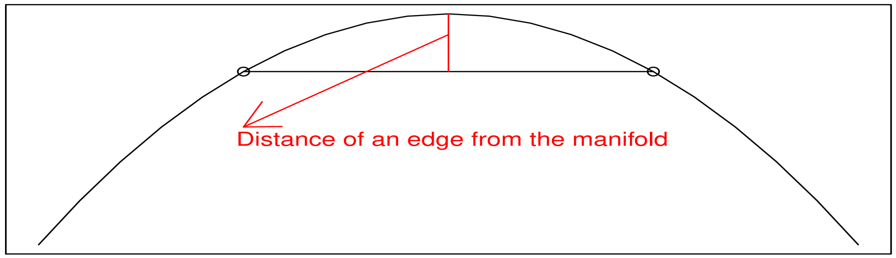


Figure 4. Curved, Singular Manifold. The missing dimensions, $\Sigma_{1,1}$, $\Sigma_{1,2}$, and θ, are held constant at 5, 6.12372, and 5, respectively.

Curved Manifold: Distance vs. Delta



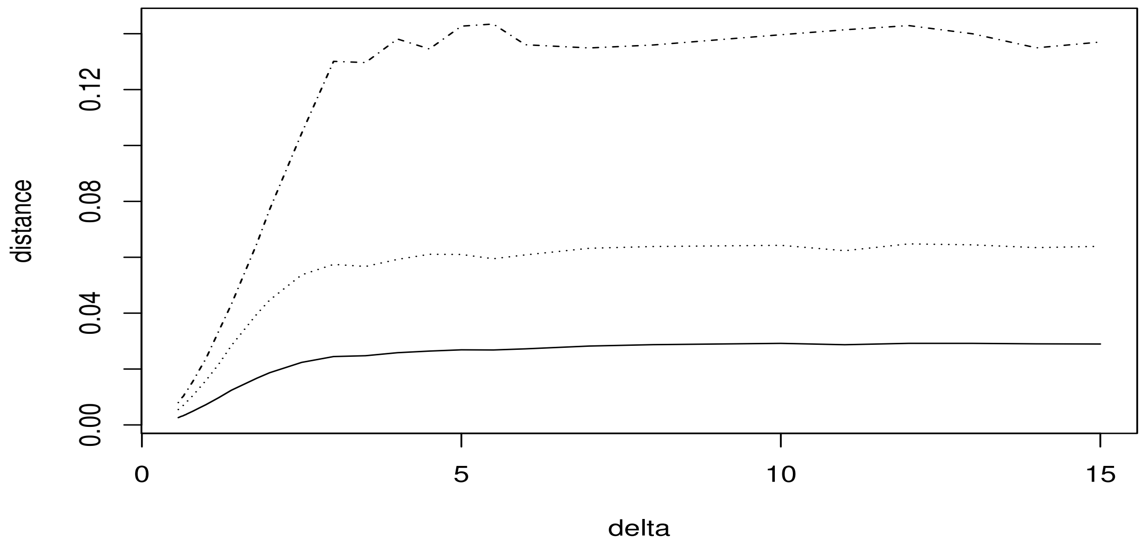


Figure 5. Lower panel: The dotdash line is the 99th percentile of all edges, dotted the 90th percentile, and solid the mean.

Table 1. Curved Manifold Example, $\Delta = 0.57$

	Mean		Standard Deviation or Correlation			
			Extrinsic		Modified	
Parameter	Extrinsic	Intrinsic	Extr Ctr	Intr Ctr	Extrinsic	Riemann
μ_1	0.003030	0.001782	0.044938	0.044956	0.256304	0.045930
μ_2	0.010777	0.008102	0.046710	0.046787	0.282870	0.047894
$R_{1,1}$	0.997487	0.992473	0.030209	0.030622	0.199155	0.031385
$R_{1,2}$	-0.011216	-0.008383	0.021103	0.021293	0.133476	0.021763
$R_{2,2}$	1.029374	1.030792	0.010518	0.010614	0.066102	0.010639
θ	5.379109	5.377738	0.155378	0.155384	0.975752	0.159227
$\rho(\mu_1, \mu_2)$			-0.078107	-0.076362	-0.043754	-0.075747
$\rho(\mu_1, R_{1,1})$			-0.038925	-0.033837	-0.031125	-0.030951
$\rho(\mu_1, R_{1,2})$			-0.014263	-0.017826	-0.008660	-0.010185
$\rho(\mu_1, R_{2,2})$			-0.049502	-0.052750	-0.032077	-0.050698
$\rho(\mu_1, \theta)$			-0.034030	-0.033771	-0.025612	-0.031390
$\rho(\mu_2, R_{1,1})$			-0.000003	0.009360	-0.028539	0.010071
$\rho(\mu_2, R_{1,2})$			0.061739	0.053481	-0.018607	0.054120
$\rho(\mu_2, R_{2,2})$			-0.230838	-0.236033	-0.068083	-0.223426
$\rho(\mu_2, \theta)$			0.003121	0.003620	-0.025818	0.003372
$\rho(R_{1,1}, R_{1,2})$			-0.149040	-0.167514	0.333217	-0.161162
$\rho(R_{1,1}, R_{2,2})$			0.439249	0.407558	0.463633	0.440073
$\rho(R_{1,1}, \theta)$			0.467925	0.463035	0.475693	0.462643
$\rho(R_{1,2}, R_{2,2})$			0.762028	0.766259	0.793649	0.771083
$\rho(R_{1,2}, \theta)$			0.801051	0.792721	0.819296	0.798896
$\rho(R_{2,2}, \theta)$			0.960119	0.950296	0.936706	0.973192

Table 2. Curved Manifold Example, $\Delta = 3.0$

	Mean		Standard Deviation or Correlation			
			Extrinsic		Modified	
Parameter	Extrinsic	Intrinsic	Extr Ctr	Intr Ctr	Extrinsic	Riemann
μ_1	0.003030	0.001782	0.044938	0.044956	0.064382	0.044967
μ_2	0.010777	0.008102	0.046710	0.046787	0.069694	0.046818
$R_{1,1}$	0.997487	0.992473	0.030209	0.030622	0.049044	0.030692
$R_{1,2}$	-0.011216	-0.008383	0.021103	0.021293	0.034104	0.021243
$R_{2,2}$	1.029374	1.030792	0.010518	0.010614	0.015268	0.010386
θ	5.379109	5.377738	0.155378	0.155384	0.228527	0.155419
$\rho(\mu_1, \mu_2)$			-0.078107	-0.076362	-0.039940	-0.076271
$\rho(\mu_1, R_{1,1})$			-0.038925	-0.033837	-0.013112	-0.031901
$\rho(\mu_1, R_{1,2})$			-0.014263	-0.017826	0.006544	-0.012268
$\rho(\mu_1, R_{2,2})$			-0.049502	-0.052750	-0.003076	-0.053070
$\rho(\mu_1, \theta)$			-0.034030	-0.033771	-0.004220	-0.033868
$\rho(\mu_2, R_{1,1})$			-0.000003	0.009360	0.000939	0.009883
$\rho(\mu_2, R_{1,2})$			0.061739	0.053481	0.020551	0.054497
$\rho(\mu_2, R_{2,2})$			-0.230838	-0.236033	-0.046810	-0.223549
$\rho(\mu_2, \theta)$			0.003121	0.003620	0.013745	0.003537
$\rho(R_{1,1}, R_{1,2})$			-0.149040	-0.167514	0.163611	-0.162194
$\rho(R_{1,1}, R_{2,2})$			0.439249	0.407558	0.320706	0.440313
$\rho(R_{1,1}, \theta)$			0.467925	0.463035	0.329882	0.462884
$\rho(R_{1,2}, R_{2,2})$			0.762028	0.766259	0.653799	0.770204
$\rho(R_{1,2}, \theta)$			0.801051	0.792721	0.698690	0.798101
$\rho(R_{2,2}, \theta)$			0.960119	0.950296	0.886439	0.973122

Table 3. Curved Manifold Example, $\Delta = 15.0$

	Mean		Standard Deviation or Correlation			
			Extrinsic		Modified	
Parameter	Extrinsic	Intrinsic	Extr Ctr	Intr Ctr	Extrinsic	Riemann
μ_1	0.003030	0.001782	0.044938	0.044956	0.045979	0.044963
μ_2	0.010777	0.008102	0.046710	0.046787	0.049665	0.046814
$R_{1,1}$	0.997487	0.992473	0.030209	0.030622	0.036357	0.030688
$R_{1,2}$	-0.011216	-0.008383	0.021103	0.021293	0.024844	0.021239
$R_{2,2}$	1.029374	1.030792	0.010518	0.010614	0.010586	0.010383
θ	5.379109	5.377738	0.155378	0.155384	0.158211	0.155382
$\rho(\mu_1, \mu_2)$			-0.078107	-0.076362	-0.017182	-0.076274
$\rho(\mu_1, R_{1,1})$			-0.038925	-0.033837	-0.003856	-0.031906
$\rho(\mu_1, R_{1,2})$			-0.014263	-0.017826	-0.022850	-0.012280
$\rho(\mu_1, R_{2,2})$			-0.049502	-0.052750	-0.035418	-0.053089
$\rho(\mu_1, \theta)$			-0.034030	-0.033771	-0.033630	-0.033885
$\rho(\mu_2, R_{1,1})$			-0.000003	0.009360	0.011497	0.009885
$\rho(\mu_2, R_{1,2})$			0.061739	0.053481	0.028671	0.054493
$\rho(\mu_2, R_{2,2})$			-0.230838	-0.236033	-0.083846	-0.223585
$\rho(\mu_2, \theta)$			0.003121	0.003620	0.016311	0.003531
$\rho(R_{1,1}, R_{1,2})$			-0.149040	-0.167514	-0.047995	-0.162304
$\rho(R_{1,1}, R_{2,2})$			0.439249	0.407558	0.186387	0.440271
$\rho(R_{1,1}, \theta)$			0.467925	0.463035	0.209996	0.462847
$\rho(R_{1,2}, R_{2,2})$			0.762028	0.766259	0.448719	0.770156
$\rho(R_{1,2}, \theta)$			0.801051	0.792721	0.519493	0.798059
$\rho(R_{2,2}, \theta)$			0.960119	0.950296	0.797676	0.973115