Diffusion Tensor Imaging, and Deconvolution Density Estimation on Spaces of Positive Definite Symmetric Matrices

Peter Kim (University of Guelph)

D.R. (Penn State University and SAMSI)

A model for diffusion tensor imaging (DTI)

DTI: An imaging method to determine the orientation, structure, pathology of biological tissue

In brain imaging, each DTI image is represented as a 3×3 positive definite (symmetric) matrix

A nice introductory tutorial is available here

Hasan and Narayana

Determine the structure and orientation of white matter brain fibers, track the diffusion of water along those fibers

Pathology in organ and tissue types such as the human breast, kidney, lingual, cardiac, skeletal muscles, spinal cord

Highly promising: Comparing the human brain in normal states with abnormal states caused by strokes, epileptic seizures, tumors, white matter abnormalities

Multiple sclerosis lesions, traumatic brain injuries, aging (Alzheimer's disease), alcoholism, developmental disorders

On-going investigations: Potential applications to psychiatric conditions, e.g., schizophrenia, autism, cognitive and learning disabilities

DTI may be the only non-invasive *in vivo* procedure which enables the study of deep brain white matter fibers

Zhu's abstract: "... [MRI] data, from which diffusion tensors are estimated, inherently contain noise [leading to] uncertainty in estimated diffusion tensors ... eigenvalues and eigenvectors, ..."

MRI measurements are taken in a multivariate frequency domain space

The MR magnitude images are corrupted with Gaussian noise

Basu-Fletcher-Whitaker, "Rician noise removal in diffusion tensor MRI"

DTI measurements contain *Rician noise*, arising from the corrupted MRI measurements

Due to noise, DTI returns imperfect image representations

X: A true image

Each DT image is a distorted version of X

Y: The recorded DT image, a distorted version of X

X and Y are represented as positive definite matrices

$\mathcal{P}_m$: The space of $m \times m$ positive definite matrices

For $X, Y \in \mathcal{P}_m$, there is a unique geodesic joining X and Y

$\mathcal{P}_m$ is also a homogeneous space: $\mathcal{P}_m = \mathrm{GL}(m, \mathbb{R})/\mathrm{O}(m)$

There exists a unique $V \in \mathrm{GL}(m, \mathbb{R})$ such that $Y = VXV'$

Analogy with linear regression: Observed = True + Error

Think of this in terms of the log-map on $\mathcal{P}_m$

We observe Y and record y, where $\exp(y) = Y$

Similarly, represent the true X by x where $\exp(x) = X$

y and x are in the tangent space of $\mathcal{P}_m$

There exists a “small” v in the tangent space such that

$$y = x + v$$

Exponentiate this relationship to obtain

$$Y = V'XV$$

Technical details are needed to show how $y = x + v$ leads to an equation of the form $Y = V'XV$ for some $V \in \mathrm{GL}(m, \mathbb{R})$

Terras (1988), p. 16, Theorem 1

Terras (1988). *Harmonic Analysis on Symmetric Spaces and Applications, II.* Springer, NY

Note that V is a random matrix

Assumption: The errors are isotropic, i.e., have no preferred orientation, so they are biinvariant under $\mathrm{O}(m)$

The distribution of V is biinvariant under $\mathrm{O}(m)$:

$$V \stackrel{\mathcal{L}}{=} k_1 V k_2 \quad \text{for all } k_1, k_2 \in \mathrm{O}(m)$$

Model: $Y \stackrel{\mathcal{L}}{=} V' X V$ where the error V is $\mathrm{O}(m)$-biinvariant

Since V is biinvariant, we can replace V by $B^{1/2}$ where $B \in \mathcal{P}_m$

Polar coordinates representation: $V = kB^{1/2}$, $k \in \mathrm{O}(m)$

Therefore $V \stackrel{\mathcal{L}}{=} k'V = k'kB^{1/2} = B^{1/2}$

Conclude: $Y \stackrel{\mathcal{L}}{=} B^{1/2} X B^{1/2}$

The distribution of B is invariant under $\mathrm{O}(m)$:

$$k' B k = k' V' V k \stackrel{\mathcal{L}}{=} V' V, \quad k \in \mathrm{O}(m)$$

Notation: $G = \mathrm{GL}(m, \mathbb{R})$, $K = \mathrm{O}(m)$

G acts transitively on $\mathcal{P}_m$ by the action

$$G \times \mathcal{P}_m \to \mathcal{P}_m, \quad Y \mapsto g' Y g, \qquad \text{where } g \in G, Y \in \mathcal{P}_m$$

Under this action, the isotropy group of the identity in G is K

The homogeneous space $K \backslash G$ can be identified with $\mathcal{P}_m$ by

$$K \backslash G \to \mathcal{P}_m, \quad K g \mapsto g' g.$$

A random matrix $Y \in \mathcal{P}_m$ with p.d.f. f is K-*invariant* if

$$f(k'yk) = f(y) \quad \text{for all } y \in \mathcal{P}_m, k \in K$$

A similar definition holds for K-invariant functions on $\mathcal{P}_m$

We will write $f \in L^1(\mathcal{P}_m/K)$ whenever f is K-invariant

We identify K-invariant random matrices on $\mathcal{P}_m$ with $K-$biinvariant random matrices on the group G

Consider random $Y, Z \in \mathcal{P}_m$

Let $\tilde{Y}, \tilde{Z}$ be the corresponding group elements in G

When ordinary matrix multiplication in G is translated into a "composition" on $\mathcal{P}_m$, we are led naturally to the:

Definition: Let $Y, Z \in \mathcal{P}_m$ be random matrices where Z is $K-$invariant. The *composition* of Y and Z is

$$Y \circ Z = Z^{1/2} Y Z^{1/2}$$

where $Z^{1/2} \in \mathcal{P}_m$ is the square root of Z.

How do we find the distribution of $Y \circ Z$?

The Helgason-Fourier transform

$Y = (y_{ij}) \in \mathcal{P}_m$

$|Y|$: The determinant of Y

The G-invariant measure on $\mathcal{P}_m$ is

$$d_*Y = |Y|^{-(m+1)/2} \prod_{1 \le i \le j \le m} dy_{ij}$$

$|Y_j|$: The principal minor of order j of Y, $1 \le j \le m$

For $s = (s_1, \ldots, s_m) \in \mathbb{C}^m$, the *power function* p_s on $\mathcal{P}_m$ is

$$p_s(Y) = \prod_{j=1}^{m} |Y_j|^{s_j}, \quad Y \in \mathcal{P}_m$$

$C_c^\infty(\mathcal{P}_m)$: The space of infinitely differentiable, compactly supported $f : \mathcal{P}_m \to \mathbb{C}$

The *Helgason-Fourier transform* of $f \in C_c^\infty(\mathcal{P}_m)$ is

$$\mathcal{H}f(s,k) = \int_{\mathcal{P}_m} f(Y)\,\overline{p_s(k'Yk)}\,d_*Y, \qquad s \in \mathbb{C}^m, k \in K$$

Harish-Chandra; Helgason; Terras (1988), p. 87

$m = 1$: The H-F transform reduces to the Mellin transform

$$\hat{f}(s) = \int_0^\infty f(Y)\,Y^{\overline{s}}\,\frac{dY}{Y}$$

dk: The normalized Haar measure on K

The *zonal spherical function*:

$$h_s(Y) = \int_K p_s(k'Yk)\, dk, \quad Y \in \mathcal{P}_m,$$

If the s_j are nonnegative integers then h_s is a zonal polynomial

If f is $K-$invariant, then we can replace p_s by h_s to get the *zonal spherical transform*:

$$\hat{f}(s) = \int_{\mathcal{P}_m} f(Y)\, \overline{h_s(Y)} dY,$$

Ding, Gross and D.R. (1996), Pacific J. Math.: Applications to hypergeometric functions of matrix argument

How do we invert the Helgason-Fourier transform?

$A = \{\mathrm{diag}\,(a_1, \ldots, a_m) : a_j > 0, j = 1, \ldots, m\}$: The group of diagonal positive definite matrices

$N = \{n = (n_{ij}) \in G : n_{ij} = 0, 1 \leq j < i \leq p; n_{jj} = 1, 1 \leq j \leq m\}$: The group of upper triangular matrices with 1's on the diagonal

The Iwasawa decomposition: Each $g \in G$ can be written as

$$g = kan, \qquad (k, a, n) \in (K, A, N)$$

(k, a, n) are called the *Iwasawa coordinates* of g

The classical beta function: For $\mathrm{Re}(a), \mathrm{Re}(b) > 0$,

$$B(a,b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}$$

The *Harish-Chandra c-function*

$$c_m(s) = \prod_{1 \le i < j \le m-1} \frac{B\big(\frac{1}{2}, s_i + \cdots + s_j + \frac{1}{2}(j-i+1)\big)}{B\big(\frac{1}{2}, \frac{1}{2}(j-i+1)\big)}$$

Let $\rho = (\frac{1}{2}, \ldots, \frac{1}{2}, \frac{1}{4}(1-m))$ and

$$\omega_m = \frac{\prod_{j=1}^{m} \Gamma(j/2)}{(2\pi i)^m \, \pi^{m(m+1)/4} \, m!}$$

Notice that $\omega_1 = 1/2\pi i$

The space of diagonal matrices with non-zero entries ± 1:

$$M = \left\{ \begin{pmatrix} \pm 1 & & \\ & \ddots & \\ & & \pm 1 \end{pmatrix} \right\}$$

M is a group of order 2^m and is a subgroup of K

There exists an invariant measure $d\bar{k}$ on K/M such that

$$\int_{\bar{k} \in K/M} d\bar{k} = 1.$$

Helgason-Fourier inversion

Notation

$$d_*s := |c_m(s)|^{-2}\, ds$$
$$\mathbb{C}^m(\rho) := \{s \in \mathbb{C}^m : \mathrm{Re}(s) = -\rho\}$$

Helgason-Fourier Inversion: For $f \in C_c^\infty(\mathcal{P}_m)$,

$$f(Y) = \omega_m \int_{\mathbb{C}^m(\rho)} \int_{\bar{k}\in K/M} \mathcal{H}f(s,k)\, p_s(k'Yk)\, d\bar{k}\, d_*s$$

$m = 1$: H-F inversion reduces to Mellin inversion

$$f(Y) = \frac{1}{2\pi i} \oint_{\mathrm{Re}(s)=0} Y^{-s}\, \hat{f}(s)\, ds, \quad Y > 0$$

The Plancherel Formula for the H-F transform

For $f \in C_c^\infty(\mathcal{P}_m)$,

$$\int_{\mathcal{P}_m} |f(Y)|^2 \, dY = \int_{\mathbb{C}^m(\rho)} \int_{\bar{k} \in K/M} |\mathcal{H}(s, \bar{k})|^2 \, d\bar{k} \, d_* s$$

If f is also K-invariant then the Plancherel formula reduces to

$$\int_{\mathcal{P}_m} |f(Y)|^2 \, dY = \int_{\mathbb{C}^m(\rho)} |\hat{f}(s)|^2 \, d_* s$$

Terras, p. 88, Theorem 1

The Plancherel formula implies

$$\|f\|_{L^2(\mathcal{P}_m, dY)} = \|\hat{f}\|_{L^2(\mathbb{C}^m, d_* s)}$$

The *convolution* of $f, h \in L^1(\mathcal{P}_m)$:

$$(f * h)(X) = \int_{\mathcal{P}_m} f(Y)\, h(Y^{-1/2} X Y^{-1/2})\, dY$$

$f * h$ is the density function of $Y^{-1/2} X Y^{-1/2}$

The convolution property: If $f \in C_c^\infty(\mathcal{P}_m)$, $h \in C_c^\infty(\mathcal{P}_m/K)$ then

$$\mathcal{H}(f * h)(s, k) = \mathcal{H}f(s, k)\, \hat{h}(s), \quad s \in \mathbb{C}^m, k \in K$$

The convolution property means that $\mathcal{H}(f * h) = \mathcal{H}f \cdot \mathcal{H}h$

$\mathcal{H}h = \hat{h}$ because h is K-invariant

Terras (1988), Theorem 1, p. 88

The Laplace-Beltrami operator on $\mathcal{P}_m$:

$$\Delta = -\text{tr}\left(\left(Y\frac{\partial}{\partial Y}\right)^2\right), \quad \frac{\partial}{\partial Y} = \left(\tfrac{1}{2}(1+\delta_{ij})\frac{\partial}{\partial y_{ij}}\right)$$

Siegel/Selberg/Maass: p_s is an eigenfunction of Δ (also of all G-invariant differential operators)

$$\Delta p_s(Y) = \lambda_s\, p_s(Y)$$

where

$$\lambda_s = \|s\|^2 + \tfrac{1}{48}m(1-m^2)$$

The H-F transform changes the effect of invariant differential operators on functions to pointwise multiplication:

$$\mathcal{H}(\Delta f) = \lambda_s \mathcal{H} f, \qquad f \in C_c^\infty(\mathcal{P}_m)$$

Two Sobolev spaces

$\|\cdot\|$: The $L^2(\mathcal{P}_m)$-norm w.r.t. dY

For $\sigma > 0$, define the Sobolev space

$$H_\sigma(\mathcal{P}_m) := \left\{ f \in C^\infty(\mathcal{P}_m) : \left\|\Delta^{\sigma/2} f\right\| < \infty \right\}$$

For $\sigma, Q > 0$, define the bounded Sobolev class

$$H_\sigma(\mathcal{P}_m, Q) := \left\{ f \in C^\infty(\mathcal{P}_m) : \left\|\Delta^{\sigma/2} f\right\| < Q \right\}$$

Deconvolution for positive defi nite matrix space

A statistical model on $\mathcal{P}_m$: We observe $Z = \varepsilon^{1/2} X \varepsilon^{1/2}$

$X \in \mathcal{P}_m$ is the true, unobservable, measurement

$\varepsilon \in \mathcal{P}_m$ is an independent random noise

f_ε, the p.d.f. of ε, is assumed known and K-invariant

f_X, f_Z, the p.d.f.'s of X, Z, respectively, are unknown and assumed $K-$invariant

$$f_Z = f_X * f_\varepsilon$$

Given an i.i.d. sample $Z_1, \ldots, Z_n$ from Z, we are to estimate f_X

Apply the convolution property of the Helgason-Fourier transform

$$\mathcal{H}f_Z(s,k) = \mathcal{H}f_X(s,k)\,\hat{f}_\varepsilon(s),$$

Form the *empirical Helgason-Fourier transform*

$$\mathcal{H}^n f_Z(s,k) = \frac{1}{n}\sum_{\ell=1}^{n} \overline{p_s(k'Z_\ell k)}.$$

Crucial assumption: $\hat{f}_\varepsilon(s) \neq 0$ for all s

An example of such a function? (Hint: Gaussian distributions)

We obtain

$$\mathcal{H}^n f_X(s,k) = \frac{\mathcal{H}^n f_Z(s,k)}{\hat{f}_\varepsilon(s)}$$

Smoothing parameter $T = T(n) \to \infty$ as $n \to \infty$

Notation: $\mathbb{C}^m(\rho, T) := \{s \in \mathbb{C}^m(\rho) : \lambda_s < T\}$

To define an estimator of f_X, apply “truncated” H-F inversion

$$f_X^n(Y) = \omega_m \int_{\mathbb{C}^m(\rho,T)} \int_{\bar{k} \in K/M} \mathcal{H}^n f_X(s,k)\, p_s(k'Yk)\, d\bar{k}\, d_* s$$

f_X^n is our nonparametric deconvolution density estimator for f_X

Rates of convergence

Theorem: Suppose that

1. f_ε satisfies $\|\hat{f}_\varepsilon(s)\|^{-2} \ll T^\beta$ as $T \to \infty$,
2. $\beta \geq 0$,
3. $s \in \mathbb{C}^m(\rho, T)$,
4. $\sigma > \dim \mathcal{P}_m/2$, and
5. $f_X \in H_\sigma(\mathcal{P}_m, Q)$.

Then, as $n \to \infty$,

$$\mathbb{E}\|f_X^n - f_X\|^2 \ll n^{-2\sigma/(2\sigma+2\beta+\dim \mathcal{P}_m)}.$$

Example: Suppose $\hat{f}(s) = (1 + \gamma\lambda_s)^{-\beta}$ where $\gamma > 0$

Apply H-F inversion

$$f(Y) = \omega_m \int_{\mathbb{C}^m(\rho)} (1 + \gamma\lambda_s)^{-\beta}\, h_s(Y)\, d_* s$$

As $\beta \to 0$, $\hat{f}(s)$ approaches the Dirac measure at I_m

Interpretation: The observations are made without error

Corollary: If the distribution of ε is concentrated at I_m, $f_X \in H_\sigma(\mathcal{P}_m, Q)$, and $\sigma > \dim \mathcal{P}_m/2$ then, as $n \to \infty$,

$$\mathbb{E}\|f_X^n - f_X\|^2 \ll n^{-2\sigma/(2\sigma + \dim \mathcal{P}_m)}$$

Theorem: Suppose that f_ε satisfies $\|\hat{f}_\varepsilon(s)\|^{-2} \ll \exp(T^\beta/\gamma)$ as $T \to \infty$ where $\beta, \gamma > 0$, $s \in \mathbb{C}^m(\rho, T)$; $f_X \in H_\sigma(\mathcal{P}_m, Q)$, and $\sigma > \dim \mathcal{P}_m/2$. Then, as $n \to \infty$,

$$\mathbb{E}\|f_X^n - f_X\|^2 \ll (\log n)^{-\sigma/\beta}$$

Example: $\hat{f}(s) = \exp(-\gamma \lambda_s^\beta)$, $\gamma > 0$, $s \in \mathbb{C}^m(\rho, T)$.

Again by H-F inversion

$$f(Y) = \omega_m \int_{\mathbb{C}^m(\rho)} \exp(-\gamma \lambda_s^\beta)\, h_s(Y)\, d_* s$$

$\beta = 1$ is an important case (the heat or Gaussian kernel)

Example: The Wishart distribution

The multivariate gamma function

$$\Gamma_m(s_1,\ldots,s_m) = \pi^{m(m-1)/4} \prod_{j=1}^{m} \Gamma\big(s_j + \cdots + s_m - \tfrac{1}{2}(j-1)\big),$$

where $\mathrm{Re}(s_j + \cdots + s_m) > (j-1)/2$, $j = 1,\ldots,m$.

Relative to the invariant measure d_*Y, the Wishart p.d.f. is

$$f(Y) = \frac{1}{2^{mN/2}\Gamma_m(0,\ldots,0,N/2)} (\det Y)^{N/2} \exp\Big(-\tfrac{1}{2}\,\mathrm{tr}\,Y\Big),$$

$Y \in \mathcal{P}_m$

The Helgason-Fourier transform of f is

$$\int_{\mathcal{P}_m} f(y)\, h_s(y)\, d_* y = 2^{s_1+\cdots+s_m} \frac{\Gamma_m((0,\ldots,0,N/2)+s^*)}{\Gamma_m(0,\ldots,0,N/2)}$$

where $s^* = (s_{m-1}, s_{m-2}, \ldots, s_2, s_1, -(s_1+\cdots+s_m))$

Apply Stirling's formula for the gamma function

Proposition: For $N > (3m+1)/2$ and some $\gamma > 0$, the Wishart distribution satisfies

$$|\widehat{f}(s)|^{-2} \ll \exp(T^{1/2}/\gamma)$$

as $T \to \infty$, where $s \in \mathbb{C}^m(\rho, T)$

Corollary: Suppose that f_ε is Gaussian, $f_X \in H_\sigma(\mathcal{P}_m, Q)$ and $\sigma > \dim \mathcal{P}_m/2$. Then, as $n \to \infty$,

$$\mathbb{E}\|f_X^n - f_X\|^2 \ll (\log n)^{-\sigma}$$

The case of the Wishart distribution: If $N > (3m+1)/2$, $f_X \in H_\sigma(\mathcal{P}_m, Q)$, $\sigma > \dim \mathcal{P}_m/2$, then

$$\mathbb{E}\|f_X^n - f_{\|}^2 \ll (\log n)^{-2\sigma},$$

as $n \to \infty$

The Wishart distribution has faster convergence than the Gaussian distribution in its Helgason-Fourier transform. This leads to slower recovery in the deconvolution problem.

The Proofs

Decompose the integrated mean-squared error into its variance and bias components and bound each part individually.

$$\mathbb{E}\|f_X^n - f_X\|^2 = \mathbb{E}\|f_X^n - \mathbb{E}f_X^n\|^2 + \|\mathbb{E}f_X^n - f_X\|^2.$$

To bound the IMSE, apply the Plancherel formula:

$$\begin{aligned}
\|\mathbb{E}f_X^n - f_X\|^2 &= \int_{s\in\mathbb{C}^m(\rho,T),\bar{k}} |\mathcal{H}f_X(s,\bar{k})|^2\,|c_m(s)|^{-2}\,d\bar{k}\,ds \\
&\leq T^{-\sigma}\int_{s\in\mathbb{C}^m(\rho,T),\bar{k}} \lambda_s^\sigma\,|\mathcal{H}f_X(s,\bar{k})|^2\,|c_m(s)|^{-2}\,d\bar{k}\,ds \\
&\leq T^{-\sigma}\int_{\bar{k},\mathrm{Re}(s)=-\rho} \lambda_s^\sigma\,|\mathcal{H}f_X(s,\bar{k})|^2\,|c_m(s)|^{-2}\,d\bar{k}\,ds \\
&\leq Q\,T^{-\sigma},
\end{aligned}$$

This uses the fact that $f_X \in H_\sigma(\mathcal{P}_m, Q)$ and $\sigma > \dim \mathcal{P}_m/2$.

Conclude: If $f_X \in H_\sigma(\mathcal{P}_m, Q)$ and $\sigma > \dim \mathcal{P}_m/2$ then

$$\|\mathbb{E} f_X^n - f_X\|^2 \ll T^{-\sigma}.$$

The integrated variance: The details are involved

We have to bound the variance of the empirical Helgason-Fourier transform.

We keep a close eye on the detail in the classical case

Result: For $s \in \mathbb{C}^m \cap \{\mathrm{Re}(s) = -\rho\}$ and $k \in K/M$,

$$\mathbb{E}|\mathcal{H}f_Z^n(s,k) - \mathbb{E}\mathcal{H}f_Z^n(s,k)|^2 = n^{-1}(1 - |\mathcal{H}f_Z(s,k)|^2)$$

Proof: Direct calculation using some clever properties of the spherical functions.

Lemma: As $T \to \infty$,

$$\mathbb{E}\|f_X^n - \mathbb{E}f_X^n\|^2 \ll \sup_{s \in \mathbb{C}^m(\rho,T)} |\widehat{f_\varepsilon}(s)|^{-2} \frac{T^{\dim \mathcal{P}_m/2}}{n}.$$

Proof. Begin with the Plancherel formula.

Putting together these two bounds concludes the proof of the first theorem.