Geometric Methods for Spherical Data Analysis

Domenico Marinucci

Department of Mathematics

Università di Roma Tor Vergata

Chicheley Hall, July 14, 2014

Research Supported by ERC Grant 277742 Pascal

Joint Projects with

- **Y. Fantayè (Roma Tor Vergata) + F.K.Hansen (Oslo), D.Maino (Milano)**
- **Sreekar Vadlamani (Tata Institute for Fundamental Research, Bangalore)**
- **Igor Wigman (King's College London)**
- **Valentina Cammarota (Roma Tor Vergata)**

Geometry of Gaussian Fields

- Let M be a general Riemannian manifold. In particular, for CMB we can think of M as a sphere S^2.
- The basic set of random geometrical objects are $\mathcal{R}$ valued random field $f(x)$ defined on M and its excursion sets A

$$A_u(f, M) = \{x \in M : f(x) \geq u\}$$

Lipschitz-Killing Curvatures

- Lipschitz-Killing Curvatures (LKCs) (Minkowski Functionals (MFs)), can be defined using a tube formula:

$$\mu(Tube(M,\rho)) = \sum_{j=0}^{n=dim(M)} \omega_j \mathcal{L}_{n-j}(M)\rho^j$$

where $Tube(M,\rho) = \{t \in \mathcal{R}^N : dist(M,x) \leq \rho\}$ is a tube of radius ρ bounding M; μ is Lebesgue measure; and w_j is the volume of a unit ball in $\mathcal{R}^j$.

- LKCs depend on the Riemannian metric, and are a measure of the k-dimensional size of the Riemannian manifold M.

Lipchitz-Killing Curvatures II

In particular, in two dimensions

- $\mathcal{L}_0(A_u(f))$ is the genus or the Euler-Poincarè characteristic (minima+maxima-saddles) of the excursion regions, i.e. the third Minkowski functional (2 for the sphere).
- $\mathcal{L}_1(A_u(f))$ is half the boundary length of the excursion regions, e.g. the second Minkowski functional (0 for the sphere).
- $\mathcal{L}_2(A_u(f))$ is the area of the excursion regions, e.g. the first Minkowski functional (4π for the sphere).

Gaussian Kinematic Formula (GKF)

- Due to Adler and Taylor, it allows to evaluate expected values of Lipshitz-Killing curvatures (LKCs)/Minkowski Functionals (MFs) for excursion regions under very general circumstances.

$$\mathbb{E}\mathcal{L}_i^f(A_u(f,M)) = \sum_{k=0}^{dim M - i} \begin{bmatrix} i+k \\ k \end{bmatrix} \mathcal{L}_{i+k}^f(M)\mathcal{M}_k([u,\infty))$$

$$\begin{bmatrix} i+k \\ k \end{bmatrix} = \binom{i+k}{k} \frac{\omega_{i+k}}{\omega_k \omega_i}$$

Gaussian MF

- $\mathcal{M}_k$ is given by

$$\mathcal{M}_j^{\gamma_k}([u,\infty)) = (2\pi)^{-1/2} H_{j-1}(u) e^{-u^2/2}.$$

where H_j denotes the Hermite polynomials: $H_0(u) = 1$, $H_1(u) = 2u$, $H_2(u) = 4u^2 - 1$, $H_3(u) = 8u^3 - 12u$

Beware: Gaussian MF are not the MF of a Gaussian field!

Advantages of GKF

- Splits the role of the correlation structure from the threshold level.
- The $\mathcal{L}_k^f(M)$ part depends only metric properties, and hence on correlation; if the metric is scaled by λ, $\mathcal{L}_k(M)$ scales by λ^k.
- Allows to cover easily masked data
- Allows to cover important forms of nonGaussianity

Spherical Gaussian fields

Recall that

$$T_\ell(x) = \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(x) \text{ and } \beta_j(x) = \sum_\ell b(\frac{\ell}{B^j}) T_\ell(x) \ ,$$

and normalizing

$$\widetilde{T}_\ell(x) = \frac{T_\ell(x)}{\sqrt{\frac{2\ell+1}{4\pi} C_\ell}} \text{ , and } \widetilde{\beta}_j(x) = \frac{\beta_j(x)}{\sqrt{\sum_\ell b^2(\frac{\ell}{B^j}) \frac{(2\ell+1)}{4\pi} C_\ell}} \ .$$

Of course

$$T(x) = \sum_{\ell=1}^{\infty} T_\ell(x) = \sum_{j=1}^{\infty} \beta_j(x)$$

Needlets Fields

Needlet component fields are defined by

$$\beta_j(x) = \sum_\ell b(\frac{\ell}{B^j}) T_\ell(x) \text{ , } j = 1, 2, 3...$$

where the needlet kernel is given by

$$\Psi_j(\langle x, y \rangle) \quad : \quad = \sum_\ell b(\frac{\ell}{B^j}) \frac{2\ell + 1}{4\pi} P_\ell(\langle x, y \rangle)$$

The function $b(.)$

1. $b^2(.)$ has support in $[\frac{1}{B}, B]$, and hence $b(\frac{\ell}{B^j})$ has support in $\ell \in [B^{j-1}, B^{j+1}]$
2. the function $b(.)$ is infinitely differentiable in $(0, \infty)$.
3. we have
$$\sum_{j=1}^{\infty} b^2(\frac{\ell}{B^j}) \equiv 1 \text{ for all } \ell > B. \qquad (1)$$

(partitions of unity)
We need $B > 1$, for instance $B = 2$

THE SHAPE OF $b(\frac{\cdot}{B^j})$

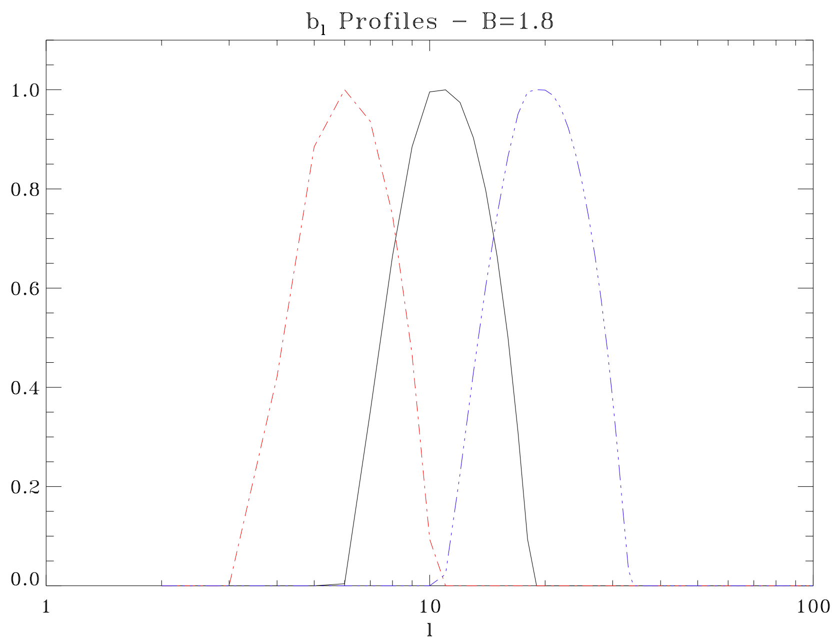

Figure 1: Partition of unity

Localization property

Localization property
For any M there exists a constant c_M s.t.,for every $\xi \in \mathbb{S}^2$:

$$|\Psi_j(x,y)| \leq \frac{c_M B^j}{(1 + B^j \arccos\langle x, y\rangle)^M} \; .$$

(Quasi-Exponential localization) Recall that
$\arccos\langle x, y\rangle \to d(x, y)$, geodesic distance on the sphere.

THE ROLE OF j

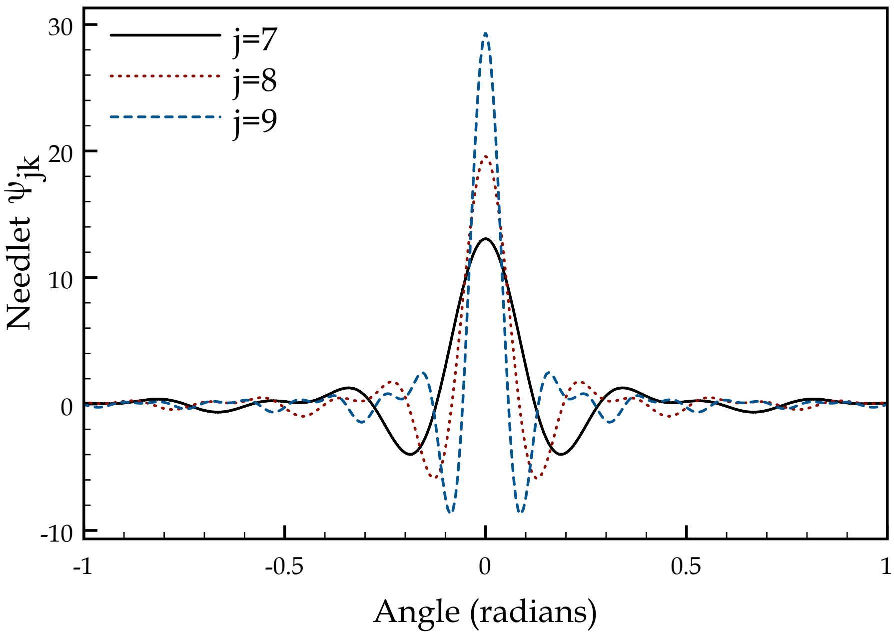

Figure 2: Needlets

Needlets Fields

The component can hence be viewed as projections:

$$\beta_j(x) = \int_{S^2} \Psi_j(\langle x, y \rangle) T(y) dy = \sum_\ell b(\frac{\ell}{B^j}) T_\ell(x)$$

- Other approaches to spherical wavelets have been developed by many other people in this meeting; similar applications of GKF are possible.

Asymptotic Uncorrelation

Under some regularity conditions on C_l, uncorrelation inequality:

$$|Corr(\beta_j(x), \beta_j(y))| \leq \frac{C_M}{(1 + B^j d(x,y))^M} \tag{2}$$

where $d(x,y) = \arccos(\langle x, y \rangle)$.
The needlet fields at any finite distance are asymptotically uncorrelated.
IMPORTANT NOTICE: this is NOT due to localization.

GKF on the sphere

The scaling λ equals the derivative of the covariance function at the origin; in the case of random spherical harmonics and needlet fields it is given by:

$$\lambda = \begin{cases} \sqrt{\frac{\ell(\ell+1)}{2}}, & \text{if } f(x) = T_\ell(x) \\ \sqrt{\frac{\sum_\ell b^2(\frac{\ell}{2^s})C_\ell \frac{2\ell+1}{4\pi}\frac{\ell(\ell+1)}{2}}{\sum_\ell b^2(\frac{\ell}{2^s})C_\ell \frac{2\ell+1}{4\pi}}}, & \text{if } f(x) = \beta_j(x) \end{cases}$$

Applications of GKF in cosmology

- Our interest here is to compute the expected values of the LKCs (MFs) in harmonic and needlet space.
- The advantages of implementing LKCs on needlet space are:
  - Needlets enjoy very good localization in pixel space - are minimally affected by masked regions, especially at high-frequency j.
  - The double-localization properties of needlets (in real and harmonic space) allow a precise interpretation of any possible anomalies - offer a scale-by-scale probe of asymmetries and relevant features e.g. *Cold Spot*.

LKCs for a Gaussian field

- First LKC (e.g. Euler-Poincarè characteristic)

$$\mathbb{E}\mathcal{L}_0(A_u(f(x), S^2)) = 2\{1 - \Phi(u)\} + \lambda^2 \frac{u e^{-u^2/2}}{\sqrt{(2\pi)^3}} 4\pi \ ;$$

- Second LKC (e.g., half the boundary length)

$$\mathbb{E}\mathcal{L}_1(A_u(f(x), S^2)) = \pi \times \lambda e^{-u^2/2} \ ;$$

- Third LKC (e.g., area)

$$\mathbb{E}\mathcal{L}_2(A_u(f(x), S^2)) = 4\pi \times \{1 - \Phi(u)\} \ .$$

Quadratic case $\beta_j^2(x)$

- Goal: anisotropic fluctuations in the power spectrum.

- First LKC (e.g. Euler-Poincarè characteristic)

$$\mathbb{E}\mathcal{L}_0(A_u(H_{2s}(x), S^2))$$

$$= 4(1 - \Phi(\sqrt{u+1})) + 4\lambda^2 \frac{e^{-(u+1)/2}}{\sqrt{2\pi}} \sqrt{u+1} \; ;$$

- Second LKC (half the boundary length)

$$\mathbb{E}\mathcal{L}_1(A_u(H_{2s}(x), S^2)) = 2\pi\lambda e^{-(u+1)/2} \; ;$$

- Third LKC (area)

$$\mathbb{E}\mathcal{L}_2(A_u(H_2(x), S^2)) = 4\pi \times 2(1 - \Phi(\sqrt{u+1})) \; .$$

Cubic case $\beta_j^3(x)$

- Useful to study local fluctuations in non-Gaussianity.
- First LKC: Euler characteristic

$$\mathbb{E}\mathcal{L}_0(A_u(H_{3s}(x), S^2)) = 2(1 - \Phi(\sqrt[3]{u})) + 2\lambda^2 \frac{e^{-(\sqrt[3]{u})^2/2}}{\sqrt{2\pi}} \sqrt[3]{u} \text{ ;}$$

- Second LKC: (half) boundary length

$$\mathbb{E}\mathcal{L}_1(A_u(H_{3s}(x), S^2)) = \pi\lambda e^{-(\sqrt[3]{u})^2/2} \text{ ;}$$

- Third LKC: Area

$$\mathbb{E}\mathcal{L}_2(A_u(H_{3s}(x), S^2)) = 4\pi(1 - \Phi(\sqrt[3]{u})) \text{ .}$$

Masked case, $M := S^2 \backslash G$

- Euler characteristic

$$\mathbb{E}\mathcal{L}_0(A_u(f(x), M)) = \{1 - \Phi(u)\}\, \mathcal{L}_0(M)$$

$$+\frac{\pi}{2}\lambda_s \frac{1}{2\pi} e^{-u^2/2} \mathcal{L}_1(M) + \lambda^2 \frac{u e^{-u^2/2}}{\sqrt{(2\pi)^3}} \mathcal{L}_2(M)\ ;$$

- half the boundary length

$$\mathbb{E}\mathcal{L}_1(A_u(f(x), M)) = 2\{1 - \Phi(u)\}\, \mathcal{L}_1(M) + \frac{\pi}{2}\lambda \rho_1(u) \mathcal{L}_2(M)\ ;$$

- the area

$$\mathbb{E}\mathcal{L}_2(A_u(f(x), S^2)) = \{1 - \Phi(u)\}\, \mathcal{L}_2(S^2).$$

Computing LKCs from a (CMB) map

- Harmonic space - obtain $T_\ell(x)$ maps; normalize each map by the expected RMS; power transform normalized T_ℓ maps to obtain NG maps.
- Needlet space - apply the standard needlet filter to the spherical harmonic coefficients; we used B=1.5.
- The area functional is computed by finding the ratio of Healpix pixels above a certain temperature threshold.
- The length and genus functionals are computed by using the method described in Eriksen et. al. 2004 paper.

Masked algorithm

- Fix C_ℓ, $L_{\max} = 10$, and generate Gaussian maps
- Fix some threshold values u_i, evaluate LKC by Monte Carlo
- Use least square regression to estimate $\mathcal{L}_i(S^2 \backslash G)$, $i = 0, 1, 2$
- Use these estimates obtained in point 3 as an input for GKF

Asymmetries in the angular power spectrum

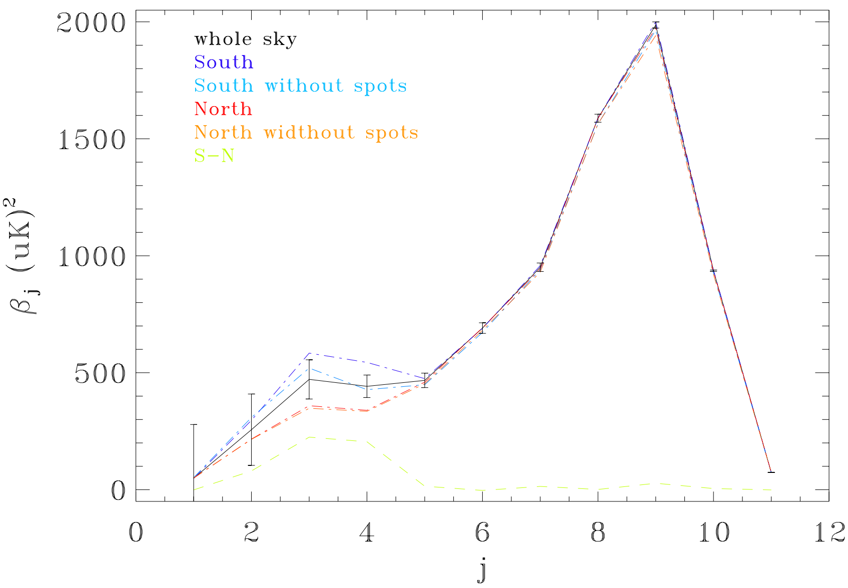

Figure 3: Angular power spectrum estimator

Nonlocal transform

As argued earlier

$$\mathbb{E}\left\{\beta_j^2(x)\right\} = \sum_\ell b(\frac{\ell}{B^j})\frac{2\ell+1}{4\pi}C_\ell \ ,$$

providing a natural local estimator for the angular power spectrum. Let us now introduce the smoothed sequences

$$g_{j;2}(z) := \int_{S^2} K(\langle z,x\rangle)\beta_j^2(x)dx \ .$$

Nonlinear transform

For instance, (hemispherical asymmetry)

$$g_{j;2}(N) := \int_{S^2} K(\langle N, x \rangle)\beta_j^2(x)dx \text{ , } g_{j;2}(S) := \int_{S^2} K(\langle S, x \rangle)\beta_j^2(x)dx \text{ ,}$$

where $K(\langle a, . \rangle) := \mathbb{I}_{[0, \frac{\pi}{2}]}(\langle a, . \rangle)$, and N, S denote the North and South Poles (Hansen et al. (2009), Pietrobon et al. (2009), Bennett (2012), Planck anisotropy papers). More generally

$$g_{j;q}(z) := \int_{S^2} K(\langle z, x \rangle)H_q(\beta_j(x))dx \text{ ,} \quad (3)$$

Nonlocal transforms of Gaussian fields

We introduce, for every $x \in S^2$

$$g_{j;q}(x) := \int_{S^2} K(\langle x, y\rangle) H_q(\widetilde{\beta}_j(y))dy \; ; \tag{4}$$

where we assume:

$$K(\langle x, y\rangle) = \sum_{\ell}^{L_K} \frac{2\ell+1}{4\pi} \kappa(\ell) P_\ell(\langle x, y\rangle) \text{ , some } L_K \in \mathbb{N} \text{ .}$$

Motivations: local estimates of angular power spectrum, bispectrum.

LKCs at high j

- Euler-Poincarè characteristic

$$\mathbb{E}\mathcal{L}_0(A_u(f(x),S^2)) = 2\left\{1-\Phi(u)\right\} + \lambda_{j;2}^2 \frac{ue^{-u^2/2}}{\sqrt{(2\pi)^3}} 4\pi \;;$$

- Boundary length

$$\mathbb{E}\mathcal{L}_1(A_u(f(x),S^2)) = \pi \times \lambda_{j;2} e^{-u^2/2} \;;$$

- Area of the excursion region

$$\mathbb{E}\mathcal{L}_2(A_u(f(x),S^2)) = 4\pi \times \left\{1-\Phi(u)\right\} \;.$$

Nonlinear parameters

In the previous slides, we have used the constants:

$$\lambda_{j;q} = \frac{\sum_{\ell=1}^{L} \frac{2\ell+1}{4\pi} C_{\ell;j,q} P'_{\ell}(1)}{\sum_{\ell=1}^{L} \frac{2\ell+1}{4\pi} C_{\ell;j,q}} \,. \tag{5}$$

and

$$C_{\ell;j,2} = 2\kappa^2(\ell) \sum_{\ell_1 \ell_2} b^2(\frac{\ell_1}{B^j}) b^2(\frac{\ell_2}{B^j}) \frac{(2\ell_1+1)(2\ell_2+1)}{4\pi} C_{\ell_1} C_{\ell_2} \begin{pmatrix} \ell & \ell_1 & \ell_2 \\ 0 & 0 & 0 \end{pmatrix}$$

with obvious generalizations to $q > 2$

Euler-Poincaré Heuristics

- EP characteristic = connected components - holes
- For u large, only one connected component
- Expected value = probability to go above u

Excursion probabilities

We have that

$$\left| P\left\{ \sup_{x \in M} f(x) \geq u \right\} - \mathbb{E}\left\{ \mathcal{L}_0(A_u(f;M)) \right\} \right| < O\left(\exp\left(-\frac{\alpha u^2}{2\sigma^2} \right) \right), \tag{6}$$

where $\mathcal{L}_0(A_u(f;M))$ is, as defined earlier, the Euler-Poincaré characteristic of the excursion set $A_u(f;M) = \{x \in M : f(x) \geq u\}$, and $\alpha > 1$ is a constant, which depends on the field f and can be determined (see Theorem 14.3.3 of RFG).

Excursion probabilities

(M and Vadlamani, 2013) For u large enough

$$\limsup_{j\to\infty}\left|\Pr\left\{\sup_{x\in S^2}\tilde{g}_{j;q}(x)>u\right\}-\{2(1-\Phi(u))+2u\phi(u)\lambda_{j;q}\}\right| \tag{7}$$

$$\leq \exp\left(-\frac{\alpha u^2}{2}\right), \tag{8}$$

where $\tilde{g}_{j;q}(x)$ has been normalized to have unit variance, $\phi(.), \Phi(.)$ denote standard Gaussian density and distribution function, $\alpha > 1$ is some constant and the parameters $\lambda_{j;q}$ has been defined above.

Some generalizations - Area

The GKF only refers to expected values - it is of interest to have some results on variances and CLT as well. For the third LKC, these results are simple (MW,2011,2014):

$$\frac{\mathcal{L}_2(A_u(T_\ell(x), S^2)) - E\mathcal{L}_2(A_u(T_\ell(x), S^2))}{\sqrt{\ell^{-1} u\phi(u)}} \to N(0,1)$$

Two remarkable features:

- For the needlets, the variance decays faster ($= O(\ell^{-2})$)
- "Berry cancellation" at $u = 0$.

Some generalizations

For the EP characteristic, Cammarota, M and Wigman (2014) have recently shown that

$$Var\mathcal{L}_0(A_u(T_\ell(x), S^2)) = \frac{\ell^3}{4}\frac{e^{-u^2}}{2\pi}[H_3(u) + H_1(u)]^2 + O(\ell^5/2) \text{ .}$$

Some generalizations

This expression looks "Gaussian kinematic" and shows that after normalization the variance is $O(\ell^{-1})$; for the needlet case, convergence to zero is faster, $Var = O(\ell^{-2})$. The result is a corollary of a more general statement concerning the asymptotic convergence of maxima, minima and saddles, in the high-frequency limit. It is also possible to combine the statistics at different frequencies into a single value - applications under way.

Figure 4: Variance EP - analytic prediction